STATISTISK TIDSKRIFT

TREDJE FÖLJDEN
ÅRGÅNG 21
1983

STATISTISKA CENTRALBYRÅN

STOCKHOLM

STATISTICAL REVIEW

THIRD SERIES
VOL. 21
1983

STATISTICS

STOCKHOLM, SWEDEN

INLEDNING

TILL

Statistisk tidskrift / utgiven av Kungl. Statistiska centralbyrån. – Stockholm : Norstedt, 1860-1913, 1952-1984.
Efter ett uppehåll 1914-1951 uppstod Statistiska tidskrift igen åren 1952-1984.
H. 165, utg. 1913 utgör register till 1860-1913. – I Statistisk tidskrift årgångarna 1871-1913 publicerades *Sveriges officiella statistik i sammandrag* som första nummer.

Efterföljare:

Journal of official statistics : JOS : an international quarterly / published by SCB, Statistics Sweden. – Stockholm : Statistics Sweden [Statistiska centralbyrån], 1985-.
Tidskriften finns från 1985 fritt tillgänglig i fulltext på internet.

Anmärkning: Vid digitalisering av Statistisk tidskrift har häftena med *Sveriges officiella statistik i sammandrag* och tre bilagor brutits ut ur sin årgång och behandlats som separata häften. Register för 1952-1962 utbrutet ur årgång 1962.

Statistisk tidskrift, 1983
Digitaliserad av Statistiska centralbyrån 2012.

URN:NBN:SE:SCB-st-1983

STATISTISK TIDSKRIFT

UTGIVEN AV STATISTISKA CENTRALBYRÅN
Ansvarig utgivare: Sten Johansson. Redaktör: Nils Welander
Adr.: Karlavägen 100. Postadr.: 11581 Stockholm
Telefoner: 08/14 05 60, ankn. 4260 (redaktören personligen) och 4265 (Helen Lidén)

Prenumeration genom bokhandel eller hos Liber distribution, Prenumerationsorder 162 89 Stockholm
Utkommer med 5 häften om året
Pris kr. 105:– per år eller kr. 20:– per häfte (supplementet gratis)

Tredje följden Årg. 21 Supplement 1983

Innehåll

Register för Statistisk tidskrift 1983 3
Författarregister 3
Sakregister 4
Litteratur 7

Statistiska publikationer utgivna 1983
Allmänna publikationer 13
Sveriges officiella statistik (SOS) 14
Statistiska meddelanden (SM) 16
Am – Arbetsmarknad 16
Be – Befolkning och levnadsförhållanden 19
Bo – Bostäder och byggnader 19
E – Energi 21
F – Företag 23
H – Handel och servicenäringar 23
HS – Hälso- och sjukvård 24
I – Industri, branschuppgifter 25
Iv – Industri, varuuppgifter 26
J – Jordbruk, skogsbruk och fiske 28
K – Kreditmarknad 29
N – Nationalräkenskaper och offentliga finanser 30
P – Priser och konsumtion 31
R – Rättsväsen 32
S – Socialtjänst och socialförsäkring 33
T – Transport och kommunikationer 34
U – Utbildning, forskning och kultur 35
Information i prognosfrågor (IPF) 37

Contents

Index to the Statistical Review, 1983 8
Author register 8
Subject register 9
Literature 12

Statistical Publications issued in 1983
General Publications 13
Official Statistics of Sweden (SOS) 14
Statistical Reports (SM) 16
Am – Labour market 16
Be – Population and living conditions 19
Bo – Housing and construction 21
E – Energy 21
F – Enterprises 23
H – Trade and service trades 23
HS – Public health and medical care 24
I – Manufacturing. Branch data 25
Iv – Manufacturing. Data on commodities 26
J – Agriculture, forestry and fishery 28
K – Credit market, banking and insurance 29
N – National accounts and public finance 30
P – Prices and consumption 31
R – Legal statistics 32
S – Social welfare 33
T – Transport and communication 34
U – Education, research and culture 35
Forecasting Information (IPF) 37

Register till Statistisk tidskrift 1983

Siffran inom parentes anger häftets nummer.

Författarregister

ANDERSSON, RONNIE, Kodning av arbetsställe vid folk- och bostadsräkningen 1980 .....102 (2)

BERG, SVEN, Factorial Series Distributions and Estimation Based on the Distinct Units of a Sample 41 (5)

BJÖRK, LEIF och HAGBERG, JAN, Probability – Knowledge – Action: Comments on Some Recent Controversies on Applications of Probabilistic Methods 207 (5)

BOURKE, PATRICK D., Randomized Response Designs With Attribute-Based Randomization 125 (5)

CASSEL, CLAES M., Statistical Matching – Statistical Prediction – What is the Difference? 55 (5)

CHRISTIANSON, ANDERS, Balancing Recall Bias Against Sampling Error in the Swedish TV Audience Surveys 133 (5)

CHRISTOFFERSEN, NYGAARD, MOGENS, Simpel sammenligning af indkomstoplysninger 191 (3)

LA COUR, AAGE, Statistik over landegrænserne 7 (1)

COX, LAWRENCE H., Some Mathematical Problems Arising from Confidentiality Concerns 179 (5)

CRICENTI, PASQUALE, Klasser och partival i Sverige – en longitudinell studie om svenska väljare 19 (1)

DUNCAN, JOSEPH W., Confidentiality in a Decentralized Statistical System 203 (5)

FRANK, OVE, Statistical Disclosure Control 173 (5)

FRITZELL, YNGVE, Yrkesfördelningen 1753–1805 enligt Tabellverket: de särskilda städerna 279 (4)

GONZALES, MARIA ELENA, Organization of the U.S. Federal Committee on Statistical Methodology 167 (5)

HAGBERG, JAN. Se BJÖRK, LEIF

HOEM, JAN M., Balancing Bias in Vital Rates Due to an Informative Sampling Plan ... 81 (5)

HÄGGLUND, BJÖRN och RANNEBY, BO, Riksskogstaxeringen 36 (1)

JABINE, THOMAS B., Teaching Questionnaire Design 157 (5)

JUSTUSSON, BO. Se POLFELDT, THOMAS

KATZOFF, MYRON J., Two Sequential Procedures for Obtaining Subdomain Estimates 49 (5)

KLEVMARKEN, ANDERS N., Pooling Incomplete Data Sets 67 (5)

LANKE, JAN, Hot Deck Imputation Techniques that Permit Standard Methods for Assessing Precision of Estimates 105 (5)

LEINIÖ, TARJA-LIISA, Invandrare som brottsoffer 209 (3)

LETTENSTRÖM, GERD, Befolkningsstatistikens utveckling i Norge 188 (3)

LICHTIG, LEO K., Confidentiality Issues in Health Statistics: An American Perspective 191 (5)

LYBERG, INGRID, On the Practical Use of Subsampling Among Nonrespondents .. 89 (5)

LYBERG, LARS, The Development of Procedures for Industry and Occupation Coding at Statistics Sweden 139 (5)

MATÉRN, BERTIL, On the Efficiency of Sampling with Unequal Probabilities 27 (5)

MEDIN, KNUT, The Introduction of Probability Sampling in Swedish Acreage and Livestock Statistics in 1950. A Note on the Background and Later Developments 19 (5)

NIEMI, IIRIS, Systematic Bias in Hours Worked? 326 (4)

NORDSTRÖM, LARS, Statistikbehov i framtiden 5 (1)

POLFELDT, THOMAS och JUSTUSSON, BO, On Applications of Random Techniques in Environmental Statistics 217 (5)

RANNEBY, BO. Se HÄGGLUND, BJÖRN

RENNERMALM, BO, The Advantage of good Methods. – Considerations on the Application of Markov Models 201 (3)

RÅBÄCK, GÖRAN. Se SÄRNDAL, CARL-ERIK

SJÖSTRÖM, OLLE, Vad är statistisk metod? 109 (2)

SKOTTE, LENE, Befolkningsstatistikkens udvikling i Danmark indtil 1849 ..... 182 (3)

STRENGELL, GÖRAN, Nordiska och finländska tankar med anledning av 200-årsminnet av Pehr Wargentins död 184 (3)

STÅHLBERG, ANN-CHARLOTTE, Avgiftsprocenten i ATP 197 (3)

SWENSSON, BENGT, On Double Sampling for Stratification with Non-Response in the Second Phase 111 (5)

SÄRNDAL, CARL-ERIK och RÅBÄCK, GÖRAN, Variance Reduction and Unbiasedness for Small Domain Estimators 33 (5)

TEGLE, STIG, Deltidssysselsatta i SCBs arbetskraftsundersökningar 1965–1982. Synpunkter och förslag 121 (2)

THORBURN, DANIEL, On Methods for Disclosure Control in Longitudinal Studies 93 (2)

Wargentin P. W. 1717–1783. "Den svenska statistikens fader" 181 (3)

WRETMAN, JAN HÅKAN, On Variance Estimation in Survey Sampling with Measurement Errors 117 (5)

YLANDER, HANS, Ett par kartografiska metoder i statistikproduktionen vilka bygger på SCBs tätortsavgränsningar 269 (4)

Sakregister

Arbetsmarknad

ILOs 13e arbetsmarknadsstatistikerkonferens 141 (2)
NIEMI, IIRIS, Systematic Bias in Hours Worked? 326 (4)
OECD-möten:
Utbildnings- och sysselsättningsåtgärder för ungdomar i 16–19 årsåldern 227 (3)
SCBs yttrande över betänkandet "Information om arbetsmiljörisker" (SOU 1982: 30) 242 (3)
TEGLE, STIG, Deltidssysselsatta i SCBs arbetskraftsundersökningar 1965–1982. Synpunkter och förslag 121 (2)

Brottslighet och Rättsväsen

LEINIÖ, TARJA-LIISA, Invandrare som brottsoffer 209 (3)
Seminarium om kriminalstatistik 60 (1)

Byggnader och bostäder

Nya metoder för folk- och bostadsräkningar 158 (2)
SCBs yttrande över underhållsfondsutredningens slutbetänkande "Återinvestera i bostäder" (SOU 1982: 65) 339 (4)

Databehandling

COM-tekniken vid Norges Sentralbyrå 157 (2)
Databehandling i statistiken (COMPSTAT 82) 64 (1)
IIASAs seminarium om informationssystem för integrerad regional utveckling 237 (3)
Internationell kartografisk konferens . 63 (1)
Konferensen "NordDATA 82" 155 (2)
PL/1-GUIDE-möte 59 (1)
Programmeringsmöte i Östberlin 156 (2)
SCBs yttrande över rapport 2 från UHÄs datareferensgrupp "Förstärkning av den administrativt inriktade utbildningen inom dataområdet" 239 (3)
Statistical Computing Project 238 (3)
Statistical Computing Project: seminarium i RAPID-frågor 157 (2)
Statistical Computing Project: Table Generation 59 (1)
Världskonferens i programmeringsspråket APL 57 (1)

Demografi

ANDERSSON, RONNIE, Kodning av arbetsställe vid folk- och bostadsräkningen 1980 (FoB 80) 102 (2)
Europarådets befolkningskonferens .. 55 (1)
FRITZELL, YNGVE, Yrkesfördelningen 1753–1805 enligt Tabellverket: de särskilda städerna 279 (4)
Konferens om den aktuella befolkningsutvecklingen i Europa 155 (2)
LEINIÖ, TARJA-LIISA, Invandrare som brottsoffer 209 (3)
LETTENSTRÖM, GERD, Befolkningsstatistikens utveckling i Norge 188 (3)
Nordiska statistiska utskott:
Utskottet för befolkningsstatistik ... 331 (4)
Nya metoder för folk- och bostadsräkningar 158 (2)
SCB hos Census Research Group 157 (2)
SKOTTE, LENE, Befolkningsstatistikkens udvikling i Danmark indtil 1849 ..... 182 (3)
STRENGELL, GÖRAN, Nordiska och finländska tankar med anledning av 200-årsminnet av Pehr Wargentins död 184 (3)
Studiebesök vid Statistics Canada och Bureau of the Census samt ASAs årsmöte 1982 72 (1)

Handel och Industri

Nordiska statistiska utskott:
Utskottet för näringsstatistik (NUN) 135 (2)

Hushåll

Europeiska statistiska möten:
Studiegrupp för livsmedels- och jordbruksstatistik 48 (1)

Hälso- och sjukvård

LICHTIG, LEO K., Confidentiality Issues in Health Statistics:
An American Perspective 191 (5)
Nordiskt olycksfallsforskningsseminarium 1982 58 (1)
SCBs yttrande över betänkandet "Information om arbetsmiljörisker" (SOU 1982: 30) 242 (3)

Inkomst och Förmögenhet

CHRISTOFFERSEN, NYGAARD, MOGENS, Simpel sammenligning af indkomstoplysninger 191 (3)
Konferens om dynamiska analysmöjligheter inom inkomststatistiken 57 (1)

Internationella organisationer, möten m. m.

Arkivmöten 155 (2)
LA COUR, AAGE, Statistik over landegrænserne 7 (1)
Databehandling i statistiken (COMPSTAT 82) 64 (1)
Europarådets befolkningskonferens ... 55 (1)
Europeiska statistiska möten:
Arbetsgruppen för nationalräkenskaper 47 (1)
Informellt möte om kvalitetsredovisning m. m. (ECE) 50 (1)
Metodfrågor i miljöstatistiken (ECE) 331 (4)
Miljöstatistiska indikatorer (ECE) .. 226 (3)
Möte om miljöstatistik 49 (1)

Möte om utbildningsstatistik och utbildningsindikatorer (UNESCO/ECE) 51 (1)
Studiegrupp för livsmedels- och jordbruksstatistik 48 (1)
Expertmöte om miljöstatistik (FN) 57 (1)
IIASAs seminarium om informationssystem för integrerad regional utveckling 237 (3)
ILOs 13e arbetsmarknadsstatistikerkonferens 141 (2)
Internationell kartografisk konferens . 63 (1)
Internationella havsforskningsrådet 154 (2)
Kommitté för yrkeskod för statistiker (ISI) 154 (2)
Konferens om den aktuella befolkningsutvecklingen i Europa 155 (2)
Konferens om dynamiska analysmöjligheter inom inkomststatistiken 57 (1)
Konferensen "NordDATA 82" 155 (2)
Möte om kulturstatistik och seminarium om kulturindustri 153 (2)
Möten om kulturstatistik 151 (2)
NORDFORSKs kontaktforum för FoU-statistik 62 (1)
Nordiska chefsstatistikermötet 1982 ... 228 (3)
Nordiska statistiska utskott:
Seminarium om vattendata 335 (4)
Utskottet för befolkningsstatistik ... 331 (4)
Utskottet för miljöstatistik 334 (4)
Utskottet för nationalräkenskaper (NUNA) 53 (1)
Utskottet för näringsstatistik (NUN) 135 (2)
Utskottet för utbildnings- och kulturstatistik 336 (4)
Nordiskt olycksfallsforskningsseminarium 1982 58 (1)
OECD-möten:
Expertmöte om FoU-statistik 226 (3)
Möte om miljöstatistik 338 (4)
Möten om sociala indikatorer 134 (2)
Utbildnings- och sysselsättningsåtgärder för ungdomar i 16–19 årsåldern 227 (3)
PL/1-GUIDE-möte 59 (1)
Programmeringsmöte i Östberlin 156 (2)
Seminarium om frågekonstruktion 56 (1)
Seminarium om kriminalstatistik 60 (1)
Sextonde nordiska statistikermötet 139 (2)
Statistical Computing Project 238 (3)
Statistical Computing Project: seminarium i RAPID-frågor 157 (2)
Statistical Computing Project: Table Generation 59 (1)
Världskonferens i programmeringsspråket APL 57 (1)

Jordbruk och binäringar

Europeiska statistiska möten:
Studiegrupp för livsmedels- och jordbruksstatistik 48 (1)
HÄGGLUND, BJÖRN och RANNEBY, BO, Riksskogstaxeringen 36 (1)
Internationella havsforskningsrådet 154 (2)
MEDIN, KNUT, The Introduction of Probability Sampling in Swedish Acreage and Livestock Statistics in 1950. A Note on the Background and Later Developments 19 (5)
Skördestatistiska nämnden 340 (4)
YLANDER, HANS, Ett par kartografiska metoder i statistikproduktionen vilka bygger på SCBs tätortsavgränsningar 269 (4)

Kommunala, regionala förhållanden

IIASAs seminarium om informationssystem för integrerad regional utveckling 237 (3)
SCBs yttrande över länsdemokratikommitténs betänkande "Vidgad länsdemokrati" (SOU 1982: 24) 241 (3)

Miljö

Det finsk-ungersk-svenska samarbetet kring miljöstatistiska metodfrågor 248 (3)
Europeiska statistiska möten:
Metodfrågor i miljöstatistiken (ECE) 331 (4)
Miljöstatistiska indikatorer (ECE) .. 226 (3)
Möte om miljöstatistik 49 (1)
Expertmöte om miljöstatistik (FN) 57 (1)
Nordiska statistiska utskott:
Seminarium om vattendata 335 (4)
Utskottet för miljöstatistik 334 (4)
OECD-möten:
Möte om miljöstatistik 338 (4)
POLFELDT, THOMAS och JUSTUSSON, BO, On Applications of Random Techniques in Environmental Statistics 217 (5)
Studieresa till USA för miljöstatistik .. 70 (1)

Nationalräkenskaper

Europeiska statistiska möten:
Arbetsgruppen för nationalräkenskaper 47 (1)
Nordiska statistiska utskott:
Utskottet för nationalräkenskaper (NUNA) 53 (1)

Sociala förhållanden

CRICENTI, PASQUALE, Klasser och partival i Sverige – en longitudinell studie om svenska väljare 19 (1)
LEINIÖ, TARJA-LIISA, Invandrare som brottsoffer 209 (3)
OECD-möten:
Möten om sociala indikatorer 134 (2)
Seminarium om frågekonstruktion 56 (1)
STÅHLBERG, ANN-CHARLOTTE, Avgiftsprocenten i ATP 197 (3)
TEGLE, STIG, Deltidssysselsatta i SCBs arbetskraftsundersökningar 1965–1982. Synpunkter och förslag 121 (2)

Statistikens organisation och historia

COM-tekniken vid Norges Sentralbyrå 157 (2)
Contact 83 ..... 339 (4)
LA COUR, AAGE, Statistik over landegrænserne 7 (1)

Det finsk-ungersk-svenska samarbetet kring miljöstatistiska metodfrågor 248 (3)
GONZALES, MARIA ELENA, Organization of the U.S. Federal Committee on Statistical Methodology 167 (5)
Kommitté för yrkeskod för statistiker (ISI) 154 (2)
Konsultation vid Ungerns statistiska centralbyrå samt besök i Berlin och Bratislava .. 71 (1)
LETTENSTRÖM, GERD, Befolkningsstatistikens utveckling i Norge 188 (3)
NORDSTRÖM, LARS, Statistikbehov i framtiden 5 (1)
Nya metoder för folk- och bostadsräkningar 158 (2)
SCB hos Census Research Group 157 (2)
Statistiska centralbyrån:
Personalnotiser 163 (2), 340 (4)
SCB får ny chef 225 (3)
SCBs styrelse 73 (1)
Yttrande över betänkandet "Information om arbetsmiljörisker" (SOU 1982: 30) 242 (3)
Yttrande över förslag till ämnesbeskrivning vid prövning av ledigblivande professur i statistik, särskilt offentlig statistikproduktion 66 (1)
Yttrande över gymnasieutredningens betänkande "En reformerad gymnasieskola" (SOU 1981: 96) 67 (1)
Yttrande över länsdemokratikommitténs betänkande "Vidgad länsdemokrati" (SOU 1982: 24) 241 (3)
Yttrande över rapport 2 från UHÄs datareferensgrupp "Förstärkning av den administrativt inriktade utbildningen inom dataområdet" 239 (3)
Yttrande över underhållsfondsutredningens slutbetänkande "Återinvestera i bostäder" (SOU 1982: 65) 339 (4)
Sextonde nordiska statistikermötet 139 (2)
SKOTTE, LENE, Befolkningsstatistikkens udvikling i Danmark indtil 1849 ..... 182 (3)
Skördestatistiska nämnden 340 (4)
Statistiska föreningens verksamhet år 1982 160 (2)
STRENGELL, GÖRAN, Nordiska och finländska tankar med anledning av 200-årsminnet av Pehr Wargentins död 184 (3)
Studiebesök i Danmark och Norge med anledning av SCBs fördelningsmodellprojekt 246 (3)
Studieresa till USA för miljöstatistik ... 70 (1)
Studiebesök vid Statistics Canada och Bureau of the Census samt ASAs årsmöte 1982 72 (1)
Wargentin P. W. 1717–1783. "Den svenska statistikens fader" 181 (3)

Statistikens teori och metod

BERG, SVEN, Factorial Series Distributions and Estimation Based on the Distinct Units of a Sample 41 (5)
BJÖRK, LEIF och HAGBERG, JAN, Probability – Knowledge – Action: Comments on Some Recent Controversies on Applications of Probabilistic Methods 207 (5)
BOURKE, PATRICK D., Randomized Response Designs With Attribute-Based Randomization 125 (5)
CASSEL, CLAES M., Statistical Matching – Statistical Prediction – What is the Difference? 55 (5)
CHRISTIANSON, ANDERS, Balancing Recall Bias Against Sampling Error in the Swedish TV Audience Surveys 133 (5)
COX, LAWRENCE H., Some Mathematical Problems Arising from Confidentiality Concerns 179 (5)
DUNCAN, JOSEPH W., Confidentiality in a Decentralized Statistical System 203 (5)
Europeiska statistiska möten:
Informellt möte om kvalitetsredovisning m. m. (ECE) 50 (1)
Metodfrågor i miljöstatistiken (ECE) 331 (4)
FRANK, OVE, Statistical Disclosure Control 173 (5)
HOEM, JAN M., Balancing Bias in Vital Rates Due to an Informative Sampling Plan ... 81 (5)
HÄGGLUND, BJÖRN och RANNEBY, BO, Riksskogstaxeringen 36 (1)
JABINE, THOMAS B., Teaching Questionnaire Design 157 (5)
KATZOFF, MYRON J., Two Sequential Procedures for Obtaining Subdomain Estimates 49 (5)
KLEVMARKEN, ANDERS N., Pooling Incomplete Data Sets 67 (5)
Konferens om dynamiska analysmöjligheter inom inkomststatistiken 57 (1)
LANKE, JAN, Hot Deck Imputation Techniques that Permit Standard Methods for Assessing Precision of Estimates 105 (5)
LICHTIG, LEO K., Confidentiality Issues in Health Statistics: An American Perspective 191 (5)
LYBERG, INGRID, On the Practical Use of Subsampling Among Nonrespondents .. 89 (5)
LYBERG, LARS, The Development of Procedures for Industry and Occupation Coding at Statistics Sweden 139 (5)
MATÉRN, BERTIL, On the Efficiency of Sampling with Unequal Probabilities 27 (5)
MEDIN, KNUT, The Introduction of Probability Sampling in Swedish Acreage and Livestock Statistics in 1950. A Note on the Background and Later Developments 19 (5)
NIEMI, IIRIS, Systematic Bias in Hours Worked? 326 (4)
POLFELDT, THOMAS och JUSTUSSON, BO, On Applications of Random Techniques in Environmental Statistics 217 (5)
RENNERMALM, BO, The Advantage of good Methods. – Considerations on the Application of Markov Models 201 (3)
Seminarium om frågekonstruktion 56 (1)
SJÖSTRÖM, OLLE, Vad är statistisk metod? 109 (2)
Studiebesök i Danmark och Norge med anled-

ning av SCBs fördelningsmodellprojekt 246 (3)
SWENSSON, BENGT, On Double Sampling for Stratification with Non-Response in the Second Phase 111 (5)
SÄRNDAL, CARL-ERIK och RÅBÄCK, GÖRAN, Variance Reduction and Unbiasedness for Small Domain Estimators 33 (5)
THORBURN, DANIEL, On Methods for Disclosure Control in Longitudinal Studies 93 (2)
WRETMAN, JAN HÅKAN, On Variance Estimation in Survey Sampling with Measurement Errors 117 (5)

Undervisning och forskning
Europeiska statistiska möten:
Möte om utbildningsstatistik och utbildningsindikatorer (UNESCO/ECE) 51 (1)
Möten om kulturstatistik 151 (2)
Möte om kulturstatistik och seminarium om kulturindustri 153 (2)
NORDFORSKs kontaktforum för FoU-statistik 62 (1)
Nordiska statistiska utskott:
Utskottet för utbildnings- och kulturstatistik 336 (4)
OECD-möten:
Expertmöte om FoU-statistik 226 (3)
Utbildnings- och sysselsättningsåtgärder för ungdomar i 16–19 årsåldern 227 (3)
Statistiska centralbyrån:
Yttrande över gymnasieutredningens betänkande "En reformerad gymnasieskola" (SOU 1981:96) 67 (1)
Yttrande över förslag till ämnesbeskrivning vid prövning av ledigblivande professur i statistik, särskilt offentlig statistikproduktion 66 (1)
Yttrande över rapport 2 från UHÄs datareferensgrupp "Förstärkning av den administrativt inriktade utbildningen inom dataområdet" 239 (3)

Val, allmänna
CRICENTI, PASQUALE, Klasser och partival i Sverige – en longitudinell studie om svenska väljare 19 (1)

Litteratur

Bokanmälningar
BERGMAN, L. R. och WÄRNERYD, B., Om datainsamling i surveyundersökningar (Per Kempe) 255 (3)
BRÄNDSTRÖM, A. & SUNDIN, J. (ed), Tradition and Transition (Rune Tryggveson) .. 258 (3)
GHISELLI, E., CAMPBELL, J. & ZEDECK, S., Measurement Theory for the Behavioral Sciences (Lars R. Bergman) 80 (1)
HÁJEK, J., Sampling from a finite population (Claes Cassel och Ingrid Lyberg) ... 170 (2)
HOFSTEN, E., Befolkningslära – demografi för lekmän (Ingvar Holmberg) 347 (4)
– Demografins grunder (Ingvar Holmberg) 263 (3)
– Indien. Om samhället och befolkningsproblemen (Ingvar Holmberg) 261 (3)
Hur jämställda är vi? (Tarja-Liisa Leiniö) 85 (1)
Lärare för livet? (Björn Steneberg) 254 (3)
Pehr Wargentin – den svenska statistikens fader (Ingvar Holmberg) 253 (3)
Replik till Joakim Palme (Joachim Vogel) 83 (1)
Replik till Tarja-Liisa Leiniö (Lars-Göran Andersson) 87 (1)
ROLSKI, T., Stationary Random Processes Associated with Point Processes (Martin Ribe) 263 (3)
SACHS, L., Applied Statistics. A handbook of techniques (Jörgen Dalén) 347 (4)
SHAPIRO, S. S. & GROSS, A. J., Statistical Modeling Techniques (Tomas Garås) ..... 88(1)
Social rapport om ojämlikheten i Sverige (Joakim Palme) 80 (1)
Social Trends II. (Joakim Palme) 83 (1)
VINOD, HRISHIKESH D. & ULLAH, AMAN, Recent Advances in Regression Methods (Martin Ribe) 167 (2)
ÖHRVIK, J., Robusteffektiv lägesestimation [Akad. avh.] (Marianne Frisén) 173 (2)

Övrigt
Nyare svensk och utländsk litteratur .. 78 (1), 344 (4)
Arkivmöten 155 (2)

Index to the Statistical Review, 1983

The figures in brackets indicate the issue numbers.

Author register

ANDERSSON, RONNIE, Coding of Establishments in the 1980 Population and Housing Census 102, 164 (2)

BERG, SVEN, Factorial Series Distributions and Estimation Based on the Distinct Units of a Sample 41 (5)

BJÖRK, LEIF and HAGBERG, JAN, Probability – Knowledge – Action: Comments on Some Recent Controversies on Applications of Probabilistic Methods 207 (5)

BOURKE, PATRICK D., Randomized Response Designs With Attribute-Based Randomization 125 (5)

CASSEL, CLAES M., Statistical Matching – Statistical Prediction – What is the Difference? 55 (5)

CHRISTIANSON, ANDERS, Balancing Recall Bias Against Sampling Error in the Swedish TV Audience Surveys 133 (5)

CHRISTOFFERSEN, NYGAARD, MOGENS, Information on income obtained from an interview survey compared with information given to taxation authorities 191, 249 (3)

LA COUR, AAGE, International Statistics ... 7, 75 (1)

COX, LAWRENCE H., Some Mathematical Problems Arising from Confidentiality Concerns 179 (5)

CRICENTI, PASQUALE, Classes and electoral behaviour in Sweden 19, 75 (1)

DUNCAN, JOSEPH W., Confidentiality in a Decentralized Statistical System 203 (5)

FRANK, OVE, Statistical Disclosure Control 173 (5)

FRITZELL, YNGVE, The Swedish Town Population in the 18th Century: 68 Towns by Occupation and Social Status 279, 341 (4)

GONZALES, MARIA ELENA, Organization of the U.S. Federal Committee on Statistical Methodology 167 (5)

HAGBERG, JAN. See BJÖRK, LEIF

HÄGGLUND, BJÖRN and RANNEBY, BO, The Swedish National Forest Survey . 36, 76 (1)

HOEM, JAN M., Balancing Bias in Vital Rates Due to an Informative Sampling Plan ... 81 (5)

JABINE, THOMAS B., Teaching Questionnaire Design 157 (5)

JUSTUSSSON, BO. See POLFELDT, THOMAS

KATZOFF, MYRON J., Two Sequential Procedures for Obtaining Subdomain Estimates 49 (5)

KLEVMARKEN, ANDERS N., Pooling Incomplete Data Sets 67 (5)

LANKE, JAN, Hot Deck Imputation Techniques that Permit Standard Methods for Assessing Precision of Estimates 105 (5)

LEINIÖ, TARJA-LIISA, Immigrants as Victims of Crime 209, 252 (3)

LETTENSTRÖM, GERD, Befolkningsstatistikens utveckling i Norge (Without English summary) 188 (3)

LICHTIG, LEO K., Confidentiality Issues in Health Statistics: An American Perspective 191 (5)

LYBERG, INGRID, On the Practical Use of Subsampling Among Nonrespondents .. 89 (5)

LYBERG, LARS, The Development of Procedures for Industry and Occupation Coding at Statistics Sweden 139 (5)

MATERN, BERTIL, On the Efficiency of Sampling with Unequal Probabilities 27 (5)

MEDIN, KNUT, The Introduction of Probability Sampling in Swedish Acreage and Livestock Statistics in 1950. A Note on the Background and Later Developments 19 (5)

NIEMI, IIRIS, Systematic Bias in Hours Worked? 326 (4)

NORDSTRÖM, LARS, Statistics demands in the future 5, 74 (1)

POLFELDT, THOMAS and JUSTUSSON, BO, On Applications of Random Techniques in Environmental Statistics 217 (5)

RABACK, GÖRAN. See SARNDAL, CARL-ERIK

RANNEBY, BO. See HÄGGLUND, BJÖRN

RENNERMALM, BO, The Advantage of good Methods. – Considerations on the Application of Markov Models 201 (3)

SARNDAL, CARL-ERIK and RABACK, GÖRAN, Variance Reduction and Unbiasedness for Small Domain Estimators 33 (5)

SJÖSTRÖM, OLLE, What is statistical method? 109, 164 (2)

SKOTTE, LENE, Befolkningsstatistikkens udvikling i Danmark indtil 1849 (Without English summary) 182 (3)

STÅHLBERG, ANN-CHARLOTTE, ATP-Contributions 197, 251 (3)

STRENGELL, GÖRAN, Nordiska och finländska tankar med anledning av 200-årsminnet av Pehr Wargentins död (Without English summary) 184 (3)

SWENSSON, BENGT, On Double Sampling for Stratification with Non-Response in the Second Phase 111 (5)

TEGLE, STIG, Part-time employment in the Swedish Labour Force Surveys 121, 165 (2)

THORBURN, DANIEL, On Methods for Disclosure Control in Longitudinal Studies 93 (2)

Wargentin, P. W. 1717–1783. "The Father of Swedish Statistics" 181, 249 (3)

WRETMAN, JAN HÅKAN, On Variance Estimation in Survey Sampling with Measurement Errors 117 (5)

YLANDER, HANS, Two cartographic methods based on the delimitation of urban areas at Statistics Sweden 269, 341 (4)

Subject register

Agriculture and adjacent industries
European statistical meetings:
Study group on statistics of food consumption and agriculture 48 (1)
HAGGLUND, BJORN and RANNEBY, BO. The Swedish National Forest Survey . 36, 76 (1)
The Harvest Statistics Board 340 (4)
The International Council for the Exploration of the Sea (ICES) 154 (2)
MEDIN, KNUT. The Introduction of Probability Sampling in Swedish Acreage and Livestock Statistics in 1950. A Note on the Background and Later Developments 19 (5)
YLANDER, HANS. Two cartographic methods based on the delimitation of urban areas at Statistics Sweden 269, 341 (4)

Commerce and manufacturing
Inter-nordic statistical committees:
Committee on food statistics 135 (2)

Communal, regional conditions
IIASAs seminar on information system for integrated regional development 237 (3)
Statistics Sweden:
Comments on the report "Extended County Democracy" by the Committee on County Democracy (SOU 1982: 24) 241 (3)

Criminology and Justice
LEINIÖ, TARJA-LIISA. Immigrants as Victims of Crime 209, 252 (3)
Seminar on criminal statistics 60 (1)

Data processing
APL-world conference 57 (1)
Conference of the International Cartographic Association 63 (1)
Conference "Nord-DATA 82" 155 (2)
Data processing in statistics (COMPSTAT 82) 64 (1)
IIASAs seminar on information systems for integrated regional development 237 (3)
Meeting on programming in the German Democratic Republic 156 (2)
PL/1-GUIDE-meeting 59 (1)
Statistical Computing Project 238 (3)
Statistical Computing Project: Seminar on RAPID-questions 157 (2)
Statistical Computing Project: Table Generation 59 (1)
Statistics Sweden:
Comments on the report no. 2 by the National Swedish Board of Universities and Colleges "Intensification of the administrative aimed education within the data processing sector" 239 (3)
Visit at the Central Bureau of Statistics in Norway on COM-technique 157 (2)

Demography
ANDERSSON, RONNIE. Coding of Establishments in the 1980 Population and Housing Census 102, 164 (2)
Conference on Current Population Problems in Europe 155 (2)
The Council of Europe: Conference on demography 55 (1)
FRITZELL, YNGVE. The Swedish Town Population in the 18th Century: 68 Towns by Occupation and Social Status 279, 341 (4)
Future Population and Housing Census: New methods 158 (2)
Inter-nordic statistical committees:
Committee on demographic statistics 331 (4)
LEINIÖ, TARJA-LIISA. Immigrants as Victims of Crime 209, 252 (3)
LETTENSTROM, GERD. Befolkningsstatistikens utveckling i Norge (Without English summary) 188 (3)
SKOTTE, LENE. Befolkningsstatistikkens udvikling i Danmark indtil 1849 (Without English summary) 182 (3)
STRENGELL, GÖRAN. Nordiska och finländska tankar med anledning av 200-årsminnet av Pehr Wargentins död (Without English summary) 184 (3)
Study visit at Statistics Canada, and US Bureau of the Census, and the annual meeting of the American Statistics Association in 1982 72 (1)
Visit at Census Research Group in London 157 (2)

Education, Research, etc.
European statistical meetings:
Meeting on educational statistics and indicators (UNESCO/ECE) 51 (1)
Inter-nordic statistical committees:
Committee on educational and cultural statistics 336 (4)
Meeting on cultural statistics 151 (2)
Meeting on cultural statistics and seminar on industrialized culture 153 (2)
NORDFORSK: Contact forum on FoU-statistics 62 (1)
OECD-meetings:
Educational and employment measures for youth 16–19 years old 227 (3)
Expert meeting on FoU-statistics ... 226 (3)
Statistics Sweden:
Comments on the draft subject description in the consideration of the chair of statistics, in particular public statistics production 66 (1)
Comments on the report no. 2 by the National Swedish Board of Universities and Colleges "Intensification of the administrative aimed education within the data processing sector" 239 (3)
Comments on the Secondary School Commit-

tee report "A revised secondary school" (SOU 1981:96) 67 (1)

Elections
CRICENTI, PASQUALE. Classes and electoral behaviour in Sweden 19, 75 (1)

Environment protection
Co-operation agreement between Finland, Hungary and Sweden in the field of methods for environment statistics 248 (3)
European statistical meetings:
Meeting on environment statistics .. 49 (1)
Methodological problems in environmental statistics (ECE) 226 (3), 331 (4)
Expert meeting on environment statistics (UN) 57 (1)
Inter-nordic statistical committees:
Committee on environmental statistics 334 (4)
Seminar on water information and water statistics 335 (4)
OECD-meetings:
Meeting on environmental statistics 338 (4)
POLFELDT, THOMAS and JUSTUSSON, BO, On Applications of Random Techniques in Environmental Statistics 217 (5)
Study visit to USA for investigation of environmental trends 70 (1)

Household
European statistical meetings:
Study group on statistics of food consumption and agriculture 48 (1)

Housing
Future Population and Housing Census: New methods 158 (2)
Statistics Sweden:
Comments on the final report "To Reinvest in Housing" (SOU 1982:65) 339 (4)

International organizations, conferences, etc.
APL-world conference 57 (1)
Conference about the possibility of dynamic analysis of income statistics 57 (1)
Conference "Nord-DATA-82" 155 (2)
Conference of the International Cartographic Association 63 (1)
Conference on Current Population Problems in Europe 155 (2)
Committee on professional code for statisticians (ISI) 154 (2)
The Council of Europe: Conference on demography 55 (1)
LA COUR, AAGE, International Statistics ... 7, 75 (1)
Data processing in statistics (COMPSTAT 82) 64 (1)
European statistical meetings:
Informal meeting on timeliness, cost and quality attributes of statistics 50 (1)
Meeting on educational statistics and indicators (UNESCO/ECE) 51 (1)
Meeting on environment statistics .. 49 (1)
Methodological problems in environmental statistics (ECE) 226 (3), 331 (4)
Study group on statistics of food consumption and agriculture 48 (1)
Working party on national accounts . 47 (1)
Expert meeting on environment statistics (UN) 57 (1)
IIASAs seminar on information systems for integrated regional development 237 (3)
The International Council for the Exploration of the Sea (ICES) 154 (2)
International Council on Archives: Meeting 155 (2)
Inter-nordic statistical committees:
Committee on demographic statistics 331 (4)
Committee on educational and cultural statistics 336 (4)
Committee on environmental statistics 334 (4)
Committee on food statistics 135 (2)
Committee on national accounts (NUNA) 53 (1)
Seminar on water information and water statistics 335 (4)
Meeting on cultural statistics 151 (2)
Meeting on cultural statistics and seminar on industrialized culture 153 (2)
Meeting on programming in the German Democratic Republic 156 (2)
NORDFORSK: Contact forum on FoU-statistics 62 (1)
Nordic research seminar on accidents in 1982 58 (1)
OECD-meetings:
Educational and employment measures for youth 16–19 years old 227 (3)
Expert meeting on FoU-statistics ... 226 (3)
Meeting on environmental statistics . 338 (4)
Meeting on social indicators 134 (2)
PL/1-GUIDE-meeting 59 (1)
Seminar on criminal statistics 60 (1)
Seminar on problems to frame questions for investigations 56 (1)
Sixteenth meeting of the nordic statisticians 139 (2)
Statistical Computing Project 238 (3)
Statistical Computing Project: Seminar on RAPID-questions 157 (2)
Statistical Computing Project: Table Generation 59 (1)
13th ILO-conference of the labour market statisticians 141 (2)
The 1982 meeting of the heads of the national statistical offices of the Nordic countries 228 (3)

Justice
See: Criminology

Labour market
NIEMI, IIRIS, Systematic Bias in Hours Worked? 326 (4)
OECD-meetings:
Educational and employment measures for youth 16–19 years old 227 (3)
Statistics Sweden:
Comments on the report "Information on risks in the working environment" (SOU 1982: 30) 242 (3)
TEGLE, STIG, Part-time employment in the Swedish Labour Force Surveys 121, 165 (2)
13th ILO-conference of the labour market statisticians 141 (2)

National accounts
European statistical meetings:
Working party on national accounts 47 (1)
Inter-nordic statistical committees:
Committee on national accounts (NUNA) 53 (1)

Organization and history of statistics
Committee on professional code for statisticians (ISI) 154 (2)
Consultation at the Central Statistical Office in Hungary and visit in Berlin and Bratislava 71 (1)
Contact -83 ..... 339 (4)
Co-operation agreement between Finland, Hungary and Sweden in the field of methods for environment statistics 248 (3)
LA COUR, AAGE, International Statistics ... 7, 75 (1)
Future Population and Housing Census: New methods 158 (2)
GONZALES, MARIA ELENA, Organization of the U.S. Federal Committee on Statistical Methodology 167 (5)
The Harvest Statistics Board 340 (4)
LETTENSTRÖM, GERD, Befolkningsstatistikens utveckling i Norge (Without English summary) 188 (3)
NORDSTRÖM, LARS, Statistics demands in the future 5, 74 (1)
Sixteenth meeting of the nordic statisticians 139 (2)
SKOTTE, LENE, Befolkningsstatistikkens udvikling i Danmark indtil 1849 (Without English summary) 182 (3)
Statistics Sweden:
The Board of Statistics Sweden 73 (1)
Comments on the draft subject description in the consideration of the chair of statistics, in particular public statistics production 66 (1)
Comments on the final report "To Reinvest in Housing" (SOU 1982: 65) 339 (4)
Comments on the report "Extended County Democracy" by the Committee on County Democracy (SOU 1982: 24) 241 (3)
Comments on the report "Information on risks in the working environment" (SOU 1982: 30) 242 (3)
Comments on the report no. 2 by the National Swedish Board of Universities and Colleges "Intensification of the administrative aimed education within the data processing sector" 239 (3)
Comments on the Secondary School Committee report "A revised secondary school" (SOU 1981: 96) 67 (1)
New head of Statistics Sweden appointed 225 (3)
Personnel notes 163 (2), 340 (4)
STRENGELL, GÖRAN, Nordiska och finländska tankar med anledning av 200-årsminnet av Pehr Wargentins död (Without English summary) 184 (3)
Study visit at Statistics Canada, and US Bureau of the Census, and the annual meeting of the American Statistical Association in 1982 72 (1)
Study visit in Denmark and Norway concerning SCBs project of a distribution model . 246 (3)
Study visit to USA for investigation of environmental trends 70 (1)
Swedish statistical society: Activity in 1982 160 (2)
Visit at the Central Bureau of Statistics in Norway on COM-technique 157 (2)
Visit at Census Research Group in London 157 (2)
Wargentin, P. W. 1717–1783. "The Father of Swedish Statistics" 181, 249 (3)

Personal income
CHRISTOFFERSEN, NYGAARD, MOGENS, Information on income obtained from an interview survey compared with information given to taxation authorities 191, 249 (3)
Conference about the possibility of dynamic analysis of income statistics 57 (1)

Public health
LICHTIG, LEO K., Confidentiality Issues in Health Statistics: An American Perspective 191 (5)
Nordic research seminar on accidents in 1982 58 (1)
Statistics Sweden:
Comments on the report "Information on risks in the working environment" (SOU 1982: 30) 242 (3)

Social conditions
CRICENTI, PASQUALE, Classes and electoral behaviour in Sweden 19, 75 (1)
LEINIÖ, TARJA-LIISA, Immigrants as Victims of Crime 209, 252 (3)

OECD-meetings:
Meetings on social indicators 134 (2)
Seminar on problems to frame questions for investigations 56 (1)
STÅHLBERG, ANN-CHARLOTTE, ATP-Contributions 197, 251 (3)
TEGLE, STIG, Part-time employment in the Swedish Labour Force Surveys ... 121, 165 (2)

Theory and method of statistics

BERG, SVEN, Factorial Series Distributions and Estimation Based on the Distinct Units of a Sample 41 (5)
BJÖRK, LEIF and HAGBERG, JAN, Probability – Knowledge – Action: Comments on Some Recent Controversies on Applications of Probabilistic Methods 207 (5)
BOURKE, PATRICK D., Randomized Response Design With Attribute-Based Randomization 125 (5)
CASSEL, CLAES M., Statistical Matching – Statistical Prediction – What is the Difference? 55 (5)
CHRISTIANSON, ANDERS, Balancing Recall Bias Against Sampling Error in the Swedish TV Audience Surveys 133 (5)
Conference about the possibility of dynamic analysis of income statistics 57 (1)
COX, LAWRENCE H., Some Mathematical Problems Arising from Confidentiality Concerns 179 (5)
DUNCAN, JOSEPH W., Confidentiality in a Decentralized Statistical System 203 (5)
European statistical meetings:
Informal meeting on timeliness, cost and quality attributes of statistics 50 (1)
Methodological problems in environmental statistics 331 (4)
FRANK, OVE, Statistical Disclosure Control 173 (5)
HAGGLUND, BJÖRN and RANNEBY, BO, The Swedish National Forest Survey . 36, 76 (1)
HOEM, JAN M., Balancing Bias in Vital Rates Due to an Informative Sampling Plan ... 81 (5)
JABINE, THOMAS B., Teaching Questionnaire Design 157 (5)
KATZOFF, MYRON J., Two Sequential Procedures for Obtaining Subdomain Estimates 49 (5)
KLEVMARKEN, ANDERS N., Pooling Incomplete Data Sets 67 (5)
LANKE, JAN, Hot Deck Imputation Techniques that Permit Standard Methods for Assessing Precision of Estimates 105 (5)
LICHTIG, LEO K., Confidentiality Issues in Health Statistics:
An American Perspective 191 (5)
LYBERG, INGRID, On the Practical Use of Subsampling Among Nonrespondents .. 89 (5)
LYBERG, LARS, The Development of Procedures for Industry and Occupation Coding at Statistics Sweden 139 (5)
MATÉRN, BERTIL, On the Efficiency of Sampling with Unequal Probabilities 27 (5)
MEDIN, KNUT, The Introduction of Probability Sampling in Swedish Acreage and Livestock Statistics in 1950. A Note on the Background and Later Developments 19 (5)
NIEMI, IIRIS, Systematic Bias in Hours Worked? 326 (4)
POLFELDT, THOMAS and JUSTUSSON, BO, On Applications of Random Techniques in Environmental Statistics 217 (5)
RENNERMALM, BO, The Advantage of good Methods. – Considerations on the Application of Markov Models 201 (3)
Seminar on problems to frame questions for investigations 56 (1)
SJÖSTRÖM, OLLE, What is statistical method? 109, 164 (2)
Study visit in Denmark and Norway concerning SCBs project of a distribution model . 246 (3)
SWENSSON, BENGT, On Double Sampling for Stratification with Non-Response in the Second Phase 111 (5)
SÄRNDAL, CARL-ERIK and RÅBÄCK, GÖRAN, Variance Reduction and Unbiasedness for Small Domain Estimators 33 (5)
THORBURN, DANIEL, On Methods for Disclosure Control in Longitudinal Studies 93 (2)
WRETMAN, JAN HÅKAN, On Variance Estimation in Survey Sampling with Measurement Errors 117 (5)

Literature

Book reviews (see p. 7, Bokanmälningar)
Recent Swedish and foreign literature .. 78 (1), 344 (4)
International Council on Archives: Meeting 155 (2)

Statistiska publikationer utgivna 1983

Statistical Publications issued in 1983

Publikationerna är producerade och utgivna av statistiska centralbyrån (SCB), där ej annat anges. *Where not otherwise indicated, the publications are produced and issued by Statistics Sweden (earlier the National Central Bureau of Statistics, Sweden).*

ALLMÄNNA PUBLIKATIONER

General Publications

Arbetskraftsundersökningen. Råtabeller. *Labour force survey. Tables*
- december 1982
- 4e kv 1982
- Årsmedeltal 1982
- januari–november 1983
- 1a kv 1983
- 2a kv 1983
- 3e kv 1983

Arbetskraftsundersökningen. Standardavvikelser
- Årsmedeltal 1982
- månad 1983
- kvartal 1983

Bakgrundsmaterial från prognosinstitutet.
- 1983:1. Fruktsamhet inom olika befolkningsgrupper

Belysning av den svenska industrins konkurrenskraft 1970–1983

BNP/Kvartal. Bruttonationalprodukten i 1980 års priser
- utkommer kvartalsvis

Du och SCBs register (broschyr)

Folk- och bostadsräkningen 1980 (FoB 80)
- Diverse broschyrer
- Alfabetiskt yrkesregister (Ykod, SEI)

HINK – Hushållens inköpsplaner. SCBs årliga inkomstfördelningsundersökning

Indelningsändringar. Läns- och grupplista
- 1982:12+tillägg
- 1982:13+tillägg
- 1983:1–11

Kelp. Kommunal ekonomisk långtidsplanering 1983–87 (utgiven av bostadsstyrelsen och SCB)

Koncernregistret

Leveranser och order inom industrin
- 1982 oktober–december
- 1983 januari–september

Levnadsnivå och ojämlikhet i Norden. En komparativ analys av de nordiska levnadsnivåundersökningarna utgiven av Nordiska rådet och Nordiska Statistiska sekretariatet

Meddelanden i samordningsfrågor (MIS)
Reports on statistical co-ordination (MIS)
- 1983:1 Policy för användarorienterad kvalitetsredovisning av statistik
- 1983:2 Områdesindelningar i lantbruksstatistiken 1983
- 1983:3 Regionala koder för län, kommuner och församlingar + rättelseblad Ändringstryck 1983 till MIS 77:4 (onumrerat)

Metodinformation
- Databehandling 1983:1–3
- Mätteknik 1983:1
- Rapportförteckningar 1982:2, 1983:1
- Statistiska metoder 1983:1–7

Nyckeltal
- Branschvisa
- Vad utmärker lönsamma företag
- Nyckeltal för kommuner

Nyförvärv. Nytillskott i SCBs bibliotek
- 1983:1–24

Näringsgrens- och verksamhetsförteckning

PB-nytt
- 1983:1–3

Promemorior från P/STM
- nr 8–10

Promemorior från SCB *Memoranda from the SCB*
1983:1 Skördeskadeskyddet. Principer och metoder. 1983 års upplaga
1983:2 Beskrivning av de objektiva skördeuppskattningarna
1983:3 Export och import av tjänster 1982

Publikationer från Statistiska centralbyrån 1983 (broschyr)

Rapporter från Prognosinstitutet
- 1983:1–4

REKO (Information om statistik för regional och kommunal planering)
- 1983:4

Rikets indelningar 1983

Rätts- och socialvårdsstatistik, RS-promemorior
- 1983:1–14

SCB-Indikatorer. Ekonomisk månadsöversikt för Sverige
- 1983:1–12

Siffror om Sverige 1983 (broschyr)

Statistisk tidskrift. Tredje följden. Årg. 21 (1983):1–4. *Statistical Review, series 3: vol. 21 (1983):1–4*

Statistisk tidskrift 1982, supplement. Register till Statistisk tidskrift 1982. *Statistical Review 1982 supplement. Index to the Statistical Review 1982. Statistical Review*/Statistisk tidskrift

– 1983: 5 *Essays in Honour of Tore E. Dalenius*
Sveriges Riksbank 1982. Statistisk årsbok
Sweden 1983 (broschyr)
Tabeller ur Företagsregistret
Urval nummer 16. H. Lindström. Non-Response Errors in sample Surveys. A discussion of causes and effects of non-response in sample surveys of living conditions
Verksamhetsberättelse för SCB 1982/83
Välfärdsbulletinen. Fakta och analys om välfärden och dess fördelning (broschyr)
– 1983: 1–4
Årssysselsättningen 1982 och utbildningsnivå februari 1983. Tilläggsfrågor till Arbetskraftsundersökningen (AKU) i februari 1983
Rättelse till Årssysselsättningen 1981. Tabellerna 5E, 6, 7, 8, 9 och 24

SVERIGES OFFICIELLA STATISTIK (SOS)

Alkoholstatistik 1981 (utgiven av socialstyrelsen)

Allmän månadsstatistik
– 1983: 1–12
– Supplement
Allmän försäkring 1981 (utgiven av Riksförsäkringsverket)
Allmänna valen 1982
– Del 1. Riksdagsvalet den 19 september 1982
– Del 2. Kommunala valen den 19 september 1982
Arbetsmarknadsstatistisk årsbok 1982–1983
Arbetsskador 1980 (utgiven av Arbetarskyddsstyrelsen och SCB)

Bankerna (månadsstatistik utgiven av Bankinspektionen)
– 1982: november och december
– 1983: januari–november
Bankaktiebolagen, fondkommissionärerna, fondbörsen och VPC 1982 (utgiven av Bankinspektionen)
Befolkningsförändringar 1981
– Del 3. Hela riket och länen m. m.

Befolkningsförändringar 1982
– Del 1. Församlingar, kommuner och A-regioner
– Del 2. Flyttning mellan kommuner
Bostads- och byggnadsstatistisk årsbok 1983

Den uträknade skolan
Dödsorsaker 1981
Enskilda försäkringsanstalter 1981 (utgiven av Försäkringsinspektionen)
Folkmängd 31 december 1982 enligt indelningen 1 januari 1983.
– Del 1–2. Kommuner och församlingar
– Del 3. Fördelning efter kön, ålder, civilstånd och medborgarskap m. m.
Företagen 1981. Ekonomisk redovisning
– Appendix nr 1. Tillverkningsindustri (SNI 3; 2-siffernivå)
– Appendix nr 2. Tillverkningsindustri (SNI 3; 3–5 siffernivå)
– Appendix nr 3. Byggnadsindustri (SNI 5; 3–5 siffernivå)

OFFICIAL STATISTICS OF SWEDEN

Alcohol Statistics 1981 (published by the National Board of Health and Welfare)
Monthly Digest of Swedish Statistics
– 1983: 1–12
– Supplement
National Insurance 1981 (published by the National Social Insurance Board)
General Elections 1982
– Vol. 1. The election to the Riksdag in 1982
– Vol. 2. The Communal Elections 1982

Yearbook of Labour Statistics 1982–1983
Occupational Injuries 1980 (published by the National Board of Occupational Safety and Health and the Statistics Sweden)
Banking Statistics (published by the Bank Inspection Board)
– 1982: November and December
– 1983: January–November
Banking and Stock Exchange Statistics 1982 (published by the Bank Inspection Board)

Population Changes 1981
– Part 3. The whole country and the counties etc.
Population Changes 1982
– Part 1. Parishes, communes etc.

– Part 2. Migration between communes
Yearbook of Housing and Building Statistics 1983
The calculated school
Causes of death 1981
Private Insurance Companies 1981 (published by the Swedish Private Insurance Supervisory)
Population December 31, 1982, according to subdivisions January 1, 1983.
– Part 1–2. Communes and Parishes
– Part 3. Distribution by sex, age, marital status and citizenship by commune etc.
Enterprises 1981. Financial accounts

- Appendix nr 4. Varuhandel, restaurang- och hotellrörelse (SNI 6; 2–4 siffernivå)
- Appendix nr 5. (SNI 7; 3–5 siffernivå)
- Appendix nr 6. Uppdragsverksamhet, tjänster m. m. (SNI 1, 4, 8, 9; 2–5 siffernivå)
- Appendix nr 7. Bolagssektorn med minst 50 anställda uppdelade efter ägarkategori och bransch (SNI 1–9, 1-siffernivå)

Industri 1981. Del 1. Data fördelade enligt Svensk standard för näringsgrensindelning (SNI)

Industri 1981. Del 2. Produktion av varor och tjänster fördelade enligt CCC-nomenklaturen och enligt SITC-Revision 2

Jordbruksstatistisk årsbok 1983

Kommunal personal 1982. Landstingspersonal, primärkommunal och kyrkokommunal personal

Kommunernas finanser 1982
- Appendix

Konsumentpriser och indexberäkningar 1982
Lantbruksräkningen 1981
- Del 1. Företag, företagare. Ägoslag. Åkerbruk

- Del 2. Husdjur. Skogsbruk. Övrigt

Levnadsförhållanden
- Rapport 35. Utbildning och utbildningseffekter
- Rapport 36. Arbetslöshetens offer
- Rapport 37. Oregelbundna och obekväma arbetstider

Löner 1981. Del 2. Lantarbetare, industriarbetare, butiks- och lagerpersonal inom varuhandel m. fl. (utkom i november 1982)

Löner 1982
- Del 1. Tjänstemän inom industri, varuhandel m. fl.
- Del 2. Lantarbetare, industriarbetare, butiks- och lagerpersonal inom varuhandel, restauranganställda m. fl.

Post Office 1980 (utgiven av Postverket)

Postverket 1981 (utgiven av Postverket)

Rättsstatistisk årsbok 1983
Siffor om högskolan
- 1. Rekrytering till högskolan 1977/78–1980/81

- 2. Enstaka kurser 1977/78–1981/82

Sjöfart 1981
Skogsstatistisk årsbok 1981–1983
Socialvården 1981
Sparbankerna 1982
Statens Vattenfallsverk 1 juli 1981–30 juni 1982 (utgiven av Statens Vattenfallsverk)

Statistik 84. Publiceringsplan för Sveriges officiella statistik
Statistisk årsbok för Sverige 1984

Manufacturing 1981. Part 1. Data grouped according to Swedish Standard Industrial Classification of all Economic Activities
Manufacturing 1981. Part 2. The production of commodities and services grouped according to the CCC-Nomenclature and according to the SITC, Revision 2
Yearbook of Agricultural Statistics 1983

Employees in municipalities 1982. Employees in county councils, employees in primary local authorities and employees in the parishes of the Church of Sweden
Local Government Finance 1982
- Appendix

Consumer Prices and Index Computations 1982
Census of Agriculture 1981
- Part 1. Holdings, holders. Type of land. Arable land
- Part 2. Livestock. Forestry. Other tables

Living Conditions
- Report no. 35. Education and effects of education
- Report no. 36. The victims of unemployment
- Report no. 37. Irregular and inconvenient hours of work

Wages 1981. Part 2. Agricultural and industrial workers, salesmen, shop assistants and storemen in wholesale and retail trade etc. (was published in November 1982)
Wages 1982
- Part 1. Salaried employees in manufacturing, employees in wholesale and retail trade, etc.
- Part 2. Agricultural and industrial workers, salesmen, shop assistants and storemen in wholesale and retail trade etc.

L'Administration des postes en 1980 (published by Direction Générale des Postes)
Post Office 1981 (published by The Swedish General Directorate of Posts)
Yearbook of Judicial Statistics 1983
Numbers on higher education
- 1. Recruitment to higher education 1977/78–1980/81
- 2. Separate single-subject courses 1977/78–1981/82

Navigation 1981
Statistical Yearbook of Forestry 1981–1983
Social Welfare Statistics 1981
Savings Banks 1982
Swedish State Power Board 1 July 1981–30 June 1982 (published by the Swedish State Power Board)
Statistics 84. A plan for the publications of the official statistics of Sweden
Statistical Abstract of Sweden 1984

Statsanställda 1982. Del 1. Lönegrad, tjänstetidsklass. Myndigheter, kommuner

Televerket 1 juli 1981–30 juni 1982 (utgiven av Televerkets centralförvaltning)

Understödsföreningar 1981 (utgiven av Försäkringsinspektionen)
Utrikeshandel 1982. Årsstatistik. Import och export enligt CCCN med varutexter
Utrikeshandel 1982. Årsstatistik. Import och export. Fördelning land/vara enligt SITC

Vägtrafikolyckor med personskada 1982
Årets tryck
– 1982
Årsbok för Sveriges kommuner 1983

Government employees 1982. Part 1. Salary grade, age, increment. Authorities, municipalities
Telecommunications Administration 1st July, 1981–30th June, 1982 (published by the Central Administration of Swedish Telecommunication)
Friendly Societies 1981 (published by the Swedish Private Insurance Supervisory Service)
Foreign Trade 1982. Imports and exports. Commodities according to the CCCN
Foreign Trade 1982. Imports and exports. Distribution by country/commodity according to the SITC
Road traffic accidents with personal injury 1982
Publications of the Year
– 1982
Statistical yearbook of administrative districts of Sweden 1983

Statistiska meddelanden

Serie Am – Arbetsmarknad

Am 1983

6* Appendix 1
Arbetskraftsundersökningarna 1982.
Länsredovisning för Stockholms län

6* Appendix 2
Arbetskraftsundersökningarna 1982
Länsredovisningar för Östra Mellansverige
Uppsala län
Södermanlands län
Östergötlands län
Örebro län
Västmanlands län

6* Appendix 3
Arbetskraftsundersökningarna 1982
Länsredovisningar för Småland med öarna
Jönköpings län
Kronobergs län
Kalmar län
Gotlands län

6* Appendix 4
Arbetskraftsundersökningarna 1982
Länsredovisningar för Södra Götaland
Blekinge län
Kristianstads län
Malmöhus län

6* Appendix 5
Arbetskraftsundersökningarna 1982
Länsredovisningar för Västra Götaland
Hallands län
Göteborgs och Bohus län
Älvsborgs län
Skaraborgs län

Statistical Reports

Series Am – Labour market

Am 1983

6* Appendix 1
The labour force surveys 1982

6* Appendix 2
The labour force surveys 1982

Uppsala län
Södermanlands län
Östergötlands län
Örebro län
Västmanlands län

6* Appendix 3
The labour force surveys 1982

Jönköpings län
Kronobergs län
Kalmar län
Gotlands län

6* Appendix 4
The labour force surveys 1982

Blekinge län
Kristianstads län
Malmöhus län

6* Appendix 5
The labour force surveys 1982

Hallands län
Göteborgs och Bohus län
Älvsborgs län
Skaraborgs län

Serie Am 1983 (forts.)

6*	Appendix 6 Arbetskraftsundersökningarna 1982 Länsredovisningar för Norra Mellansverige Värmlands län Kopparbergs län Gävleborgs län
6*	Appendix 7 Arbetskraftsundersökningarna 1982 Länsredovisningar för Mellersta Norrland Västernorrlands län Jämtlands län
6*	Appendix 8 Arbetskraftsundersökningarna 1982 Länsredovisningar för Övre Norrland Västerbottens län Norrbottens län

Am 1983

1	Publiceringsplan för Statistiska meddelanden 1983, serie Am
2	Sysselsättning och löner för arbetare inom industri
2.1	– oktober 1982
2.2	– november 1982
2.3	– december 1982
2.4	– januari 1983. Preliminära uppgifter
2.5	– februari 1983. Preliminära uppgifter
2.6	– januari, februari och mars 1983. Definitiva uppgifter
2.7	– april 1983. Definitiva uppgifter
2.8	– maj 1983. Definitiva uppgifter
2.9	– juni 1983
2.10	– juli 1983
2.11	– augusti 1983
2.12	– september 1983
2.13	– oktober 1983
3	Arbetskraftsundersökningen
3.1	– december 1982
3.2	– januari 1983
3.3	– februari 1983
3.4	– mars 1983
3.5	– april 1983
3.6	– maj 1983
3.7	– juni 1983
3.8	– juli 1983
3.9	– augusti 1983
3.10	– september 1983
3.11	– oktober 1983
3.12	– november 1983
4	Arbetskraftsundersökningarna
4.1	– 4e kv 1982
4.2	– 1a kv 1983
4.3	– 2a kv 1983
4.4	– 3e kv 1983
5	Arbetskostnadsindex för industriarbetare
5.1	– november 1982
5.2	– februari 1983
5.3	– maj 1983
5.4	– augusti 1983
6	Arbetskraftsundersökningarna 1982

Series Am 1983 (cont.)

6*	Appendix 6 The labour force surveys 1982 Värmlands län Kopparbergs län Gävleborgs län
6*	Appendix 7 The labour force surveys 1982 Västernorrlands län Jämtlands län
6*	Appendix 8 The labour force surveys 1982 Västerbottens län Norrbottens län

Am 1983

1	Publishing Plan for the Statistical Reports 1983, Series Am
2	Employment and wages in mining, quarrying and manufacturing
2.1	– October 1982
2.2	– November 1982
2.3	– December 1982
2.4	– January 1983. Preliminary data
2.5	– February 1983. Preliminary data
2.6	– January, February and March 1983. Final data
2.7	– April 1983. Final data
2.8	– May 1983. Final data
2.9	– June 1983
2.10	– July 1983
2.11	– August 1983
2.12	– September 1983
2.13	– October 1983
3	The labour force survey
3.1	– December 1982
3.2	– January 1983
3.3	– February 1983
3.4	– March 1983
3.5	– April 1983
3.6	– May 1983
3.7	– June 1983
3.8	– July 1983
3.9	– August 1983
3.10	– September 1983
3.11	– October 1983
3.12	– November 1983
4	The labour force surveys
4.1	– 4th quarter 1982
4.2	– 1st quarter 1983
4.3	– 2nd quarter 1983
4.4	– 3rd quarter 1983
5	Labour cost index for workers in mining, quarrying and manufacturing
5.1	– November 1982
5.2	– February 1983
5.3	– May 1983
5.4	– August 1983
6	The labour force surveys 1982

Serie Am 1983 (forts.)

6*	Appendix: se föreg. sida
7	Sysselsättning och lönesummor för tjänstemän inom industri
7.1	– november 1982
7.2	– februari 1983 (1982 års urval)
7.3	– maj 1983
7.4	– augusti 1983
8	Löner för anställda vid apotek 1982
9	Sysselsättning och lönesummor för anställda inom varuhandel
9.1	– 4e kv 1982
9.2	– 1a kv 1983
9.3	– 2a kv 1983
9.4	– 3e kv 1983
10	Kyrkokommunal personal 1982
11	Landstingspersonal 1982
12	Primärkommunal personal 1982
13	Sysselsättning och lönesummor för personal i offentlig tjänst, statliga sektorn
13.1	– 4e kv 1982
13.2	– 1a kv 1983
13.3	– 2a kv 1983
13.4	– 3e kv 1983
14	Sysselsättning och lönesummor för arbetare inom byggnadsverksamhet
14.1	– 4e kv 1982
14.2	– 1a kv 1983
14.3	– 2a kv 1983
14.4	– 3e kv 1983
15	Sysselsättningen i storskogsbruket. Definitiva uppgifter
15.1	– 15 februari 1983
15.2	– 17 maj 1983
15.3	– 16 augusti 1983
15.4	– 15 november 1983
16	Löner för lantarbetare och skogsarbetare år 1982
17	Löner för befälspersonal på lastfartyg i utrikes sjöfart och stor kustsjöfart september 1982
18	Löner för tjänstemän inom industri, byggnadsverksamhet, varuhandel m. m. 1982
19	Jordbrukets arbetskraft 1982
20	Löner för arbetare vid buss- och åkeriföretag samt oljedistributionsföretag 1982
21	Löner för anställda inom varuhandel, 2a kvartalet 1982. Butikspersonal samt lagerpersonal och chaufförer m. fl. Definitiva resultat
22	Löner för arbetare inom industri m. m. 1982
23	Löner för arbetare inom byggnadsverksamhet 1982
24	Arbetskraften i skogsbruket under 1982. Definitiva uppgifter
25	Löner för anställda vid banker 1983

Series Am 1983 (cont.)

6*	Appendices: see previous page
7	Employment and salaries for salaried employees in mining, quarrying and manufacturing
7.1	– November 1982
7.2	– February 1983
7.3	– May 1983
7.4	– August 1983
8	Salaries of employees in pharmacies in 1982
9	Employment and total wages for employees within the whole sale and retail trade
9.1	– 4th quarter 1982
9.2	– 1st quarter 1983
9.3	– 2nd quarter 1983
9.4	– 3rd quarter 1983
10	Employees in parishes of the Church of Sweden 1982
11	Employees in county councils 1982
12	Employees in primary local authorities 1982
13	Employment and total wages for public employees, the governmental sector
13.1	– 4th quarter 1982
13.2	– 1st quarter 1983
13.3	– 2nd quarter 1983
13.4	– 3rd quarter 1983
14	Employment and amount of wages for workers in construction
14.1	– 4th quarter 1982
14.2	– 1st quarter 1983
14.3	– 2nd quarter 1983
14.4	– 3rd quarter 1983
15	Employment in the large scale forestry. Definitive data
15.1	– 15 February 1983
15.2	– 17 May 1983
15.3	– 16 August 1983
15.4	– 15 November 1983
16	Wages for agricultural and forestry workers 1982
17	Salaries in cargo ships, September 1982
18	Salaries in manufacturing, in construction and in wholesale and retail trade, in 1982
19	The agricultural labour force in 1982
20	Wages in omnibus and motor truck transport and in petroleum marketing 1982
21	Wages of employees in wholesale and retail trade, 2nd quarter 1982. Salesmen and shop assistants, storemen and drivers. Definitive data
22	Hourly earnings of wage-earners in mining, quarrying, manufacturing, power plants and stevedoring 1982
23	Hourly earnings of wage-earners in construction 1982
24	The labour force in forestry in 1982. Definitive data
25	Salaries of employees in banks in 1983

Serie Am 1983 (forts.)

26	Löner för anställda vid försäkringsbolag 1983
27	Retroaktiva lönetillägg för arbetare inom industri under januari–juli 1983
28	Löner för anställda inom hotell och restauranger 1982
29	Arbetskostnadsindex för industritjänstemän 1982
30	Förteckning över Statistiska meddelanden i serie Am 1983

Series Am 1983 (cont.)

26	Salaries of employees in insurance companies in 1983
27	Retroactive wage supplements in mining, quarrying and manufacturing January–July 1983
28	Wages of hotels and restaurant personnel 1982
29	Labour cost index for salaried employees in mining, quarrying and manufacturing, in 1982
30	List of Statistical Reports, Series Am 1983

Serie Be 1983 – Befolkning och levnadsförhållanden

1	Publiceringsplan för Statistiska meddelanden 1983, serie Be
2	Inrikes och utrikes flyttningar 1982. Definitiva uppgifter
3	Kommunernas folkmängd 31 december 1982 och befolkningsförändringar 1982. Definitiva resultat
4	Inkomstfördelningsundersökningen
4.1	– 1981:1
4.2	– 1981:2. Inkomstutveckling för anställda åren 1973 till 1981
4.3	– 1981:3. Inkomstfördelning för hushåll
5	Valdeltagande vid riksdagsvalet 1982
6.1	Inkomst- och förmögenhetsfördelningen i riket 1981. Taxeringsmässiga begrepp
6.1*	Appendix Inkomst- och förmögenhetsfördelningen 1981. Taxeringsmässiga begrepp. Riket – län – kommuner
7	Valdeltagande bland utländska medborgare vid kommunala valen 1982
8	Folkmängd i hela riket, länen och kommunerna 31 december 1983 enligt indelningen 1 januari 1984, snabbstatistik. Preliminära uppgifter
9	Förteckning över Statistiska meddelanden i serie Be 1983

Series Be 1983 – Population and living conditions

1	Publishing Plan for the Statistical Reports 1983, Series Be
2	Internal and external migration 1982. Definitive data
3	Population by commune, Dec. 31, 1982 and population changes 1982. Final data
4	Income distribution survey
4.1	– 1981:1
4.2	– 1981:2. Income development for employees in 1973 to 1981
4.3	– 1981:3. Income distribution for households
5	Participation at the election to the riksdag 1982
6.1	Distribution of income and assets in the whole country in 1981. Taxable concepts
6.1*	Appendix Distribution of income and assets in 1981. Taxable concepts. The whole country – Counties – Municipalities
7	Participation among aliens at the communal elections 1982
8	Population in the whole country, the counties and the communes Dec. 31, 1983. Preliminary data
9	List of Statistical Reports, Series Be 1983

Serie Bo 1983 – Bostäder och byggnader

1	Publiceringsplan för Statistiska meddelanden 1983, serie Bo
2	Byggnadsinventeringen. Definitiva uppgifter
2.1	– 10 november 1982
2.2	– 10 februari 1983
2.3	– maj 1983
2.4	– 10 augusti 1983
3	Byggnadsprojekt påbörjade. Preliminära uppgifter
3.1	– november 1982

Series Bo 1983 – Housing and construction

1	Publishing Plan for the Statistical Reports 1983, Series Bo
2	Building projects under construction. Final data
2.1	– 10 November 1982
2.2	– 10 February 1983
2.3	– May 1983
2.4	– 10 August 1983
3	Building projects begun. Preliminary data
3.1	– November 1982

Serie Bo 1983 (forts.)

3.2 – december 1982
3.3 – januari 1983
3.4 – februari 1983
3.5 – mars 1983
3.6 – år 1982
3.7 – april 1983
3.8 – maj 1983
3.9 – juni 1983
3.10 – juli 1983
3.11 – augusti 1983
3.12 – september 1983
3.13 – oktober 1983
4 Bostadsbyggandet 1982, påbörjade bostadslägenheter. Preliminära uppgifter
4.1 – januari–november 1982
4.2 – januari–december 1982
4.3 – januari 1983
4.4 – januari–februari 1983
4.5 – januari–mars 1983
4.6 – januari–april 1983
4.7 – januari–maj 1983
4.8 – januari–juni 1983
4.9 – januari–juli 1983
4.10 – januari–augusti 1983
4.11 – januari–september 1983
4.12 – januari–oktober 1983
5 Bostadsbyggandet, påbörjade och inflyttningsfärdiga bostadslägenheter. Preliminära uppgifter
5.1 – 4e kv 1982
5.2 – 1a kv 1983
5.3 – 2a kv 1983
5.4 – 3e kv 1983
6 Påbörjade och avslutade moderniseringar av flerbostadshus. Preliminära uppgifter
6.1 – 4e kv 1982
6.2 – 1a kv 1983
6.3 – 1a halvåret 1983
6.4 – 3e kv 1983
7 Bostadsbyggandet 1982. Nybyggnad
7.1 – Del 1. Kommuner och län
7.2 – Del 2. Översiktliga uppgifter för 1977–1982. Detaljerade uppgifter för 1982
8 Bostads- och hyresundersökningen 1982. Lägenhetsdata för hyres- och bostadsrättslägenheter. Preliminära uppgifter
9 Modernisering och rivning av flerbostadshus 1982
10 Outhyrda lägenheter i statsbelånade flerbostadshus färdigställda under
10.1 – 2a halvåret 1982
10.2 – 1a halvåret 1983
11 Outhyrda bostadslägenheter i flerbostadshus den 1 mars 1983
12 Bostadshyror m. m. i statsbelånade flerbostadshus färdigställda under år 1982
13 Låneobjektsstatistik 1982

Series Bo 1983 (cont.)

3.2 – December 1982
3.3 – January 1983
3.4 – February 1983
3.5 – March 1983
3.6 – 1982
3.7 – April 1983
3.8 – May 1983
3.9 – June 1983
3.10 – July 1983
3.11 – August 1983
3.12 – September 1983
3.13 – October 1983
4 Housing construction 1982, dwellings begun. Preliminary data
4.1 – January–November 1982
4.2 – January–December 1982
4.3 – January 1983
4.4 – January–February 1983
4.5 – January–March 1983
4.6 – January–April 1983
4.7 – January–May 1983
4.8 – January–June 1983
4.9 – January–July 1983
4.10 – January–August 1983
4.11 – January–September 1983
4.12 – January–October 1983
5 Housing construction, dwellings begun and dwellings completed. Preliminary data
5.1 – 4th quarter 1982
5.2 – 1st quarter 1983
5.3 – 2nd quarter 1983
5.4 – 3rd quarter 1983
6 Dwellings begun and dwellings completed in restored multi-dwelling buildings. Preliminary data
6.1 – 4th quarter 1982
6.2 – 1st quarter 1983
6.3 – 1st half of 1983
6.4 – 3rd quarter 1983
7 Housing construction 1982. New construction
7.1 – Part 1. Municipalities and counties
7.2 – Part 2. Surveyable data for 1977–1982. Detailed data for 1982
8 Survey of housing and rents 1982. Data concerning rented and co-operative building society dwellings. Preliminary results
9 Restoration and demolition of multi-dwelling buildings 1982
10 Unlet dwellings in multi-dwelling buildings with government housing loans completed during
10.1 – 2nd half of 1982
10.2 – 1st half of 1983
11 Unlet dwellings in multi-dwelling buildings the 1st of March 1983
12 Rents/fees etc. in multi-dwelling buildings with government housing loans completed during 1982
13 Statistics of government housing loan projects 1982

Serie Bo 1983 (forts.)

13.1 – Flerbostadshus med preliminärt beslut om statligt bostadslån 1982

13.2 – Småhus med beslut om statligt bostadslån 1982

13.3 – Pantvärden och produktionskostnader i flerbostadshus och gruppbyggda småhus med preliminärt beslut om statligt bostadslån år 1982. Definitiva uppgifter

13.4 – Ombyggnadskostnader för flerbostadshus med slutligt beslut om statligt bostadslån 1982

14 Bostads- och hyresundersökningen 1982
– Del 1. Lägenhetsdata för hyres- och bostadsrättslägenheter. Definitiva uppgifter

15 Byggnads- och byggkonsultverksamhet 1981

16 Outhyrda bostadslägenheter i flerbostadshus ägda av allmännyttiga bostadsföretag den 1 september 1983

17 Intäkts- och kostnadsundersökningen (IKU) för flerbostadshus 1982

18 Förteckning över Statistiska meddelanden i serie Bo 1983

Series Bo 1983 (cont.)

13.1 – Multi-dwelling buildings with government housing loans preliminary granted in 1982

13.2 – One- or two-dwelling buildings with government housing loans granted in 1982

13.3 – Mortgage values and production costs of multi-dwelling buildings and collectively built one- or two-dwelling buildings with government housing loans preliminary granted in 1982. Final data

13.4 – Conversion costs of multi-dwelling buildings with government housing loans finally granted in 1982

14 Survey of housing and rents 1982
– Part 1. Data concerning rented dwellings and co-operative buildings society dwellings. Final data

15 Construction 1981

16 Unlet dwellings in multi-dwelling buildings owned by semi-public bodies the 1st September 1983

17 The survey of receipts and costs of multi-dwelling buildings in 1982

18 List of Statistical Reports, Series Bo 1983

Serie E 1983 – Energi

1 Publiceringsplan för Statistiska meddelanden 1983, serie E

2.1 Elförsörjningen. Elproduktion och teknisk utrustning i kraftstationerna 1981. Preliminära uppgifter

2.2 Elförsörjningen och fjärrvärmeförsörjningen, användning av elenergi och fjärrvärme 1981. Preliminära uppgifter

2.3 Rättelse

3 Prisindexar för olika energivaror

3.1 – november 1982
3.2 – december 1982
3.3 – januari–februari 1983
3.4 – mars 1983
3.5 – april 1983
3.6 – maj 1983
3.7 – juni 1983
3.8 – juli 1983
3.9 – augusti 1983
3.10 – september 1983
3.11 – oktober 1983

4 Elförsörjningen. Preliminära uppgifter

4.1 – november 1982
4.2 – december 1982 samt år 1982
4.3 – januari 1983
4.4 – februari 1983
4.5 – mars 1983
4.6 – april 1983

Series E 1983 – Energy

1 Publishing Plan for the Statistical Reports 1983, Series E

2.1 Electric energy supply. Electric energy production and technical equipment in power stations in 1981. Preliminary data

2.2 Electric energy supply and district heating, consumption of electric energy and steam and hot water from district heating plants in 1981. Preliminary data

2.3 Correction

3 Price indices for different energy commodities

3.1 – November 1982
3.2 – December 1982
3.3 – January–February 1983
3.4 – March 1983
3.5 – April 1983
3.6 – May 1983
3.7 – June 1983
3.8 – July 1983
3.9 – August 1983
3.10 – September 1983
3.11 – October 1983

4 Production and consumption of electric energy. Preliminary data

4.1 – November 1982
4.2 – December 1982 and during 1982
4.3 – January 1983
4.4 – February 1983
4.5 – March 1983
4.6 – April 1983

Serie E 1983 (forts.)

4.7	– maj 1983
4.8	– juni 1983
4.9	– juli 1983
4.10	– augusti 1983
4.11	– september 1983
4.12	– oktober 1983
5	Elförsörjningen. Snabbstatistik
5.1	– december 1982
5.2	– januari 1983
5.3	– februari 1983
5.4	– mars 1983
5.5	– april 1983
5.6	– maj 1983
5.7	– juni 1983
5.8	– juli 1983
5.9	– augusti 1983
5.10	– september 1983
5.11	– oktober 1983
5.12	– november 1983
6	Bränslen. Tillförsel och leveranser av petroleumprodukter. Preliminära uppgifter
6.1	– november 1982
6.2	– december 1982
6.3	– januari 1983
6.4	– februari 1983
6.5	– mars 1983
6.6	– april 1983
6.7	– maj 1983
6.8	– juni 1983
6.9	– juli 1983
6.10	– augusti 1983
6.11	– september 1983
6.12	– oktober 1983
7	Energiförsörjningen. Preliminära uppgifter
7.1	– 3e kv 1981 och 1982
7.2	– 4e kv 1981 och 1982 samt åren 1981 och 1982
7.3	– 1a kv 1982 och 1983
7.4	– 2a kv 1982 och 1983
8	Bränslen. Produktion och förbrukning av stadsgas 1980 och 1981. Preliminära uppgifter
9	Industristatistik. Användning av inköpt energi 1981. Regional och branschvis fördelning
10	Bränslen. Leveranser och förbrukning av bränslen och smörjmedel
10.1	– 4e kv 1982 samt år 1982
10.2	– 1a kv 1983
10.3	– 2a kv 1983
11	Elförsörjningen. Produktion och förbrukning av elenergi månadsvis 1981. Definitiva uppgifter
12	Elförsörjningen och fjärrvärmeförsörjningen 1981
13	Leveranser av drivmedel och eldningsoljor samt förbrukning av eldningsoljor inom industri, el-, gas- och värmeverk. Fördelning på kommuner 1982
14	Energistatistik

Series E 1983 (cont.)

4.7	– May 1983
4.8	– June 1983
4.9	– July 1983
4.10	– August 1983
4.11	– September 1983
4.12	– October 1983
5	Production and consumption of electric energy. GWh (preliminary data)
5.1	– December 1982
5.2	– January 1983
5.3	– February 1983
5.4	– March 1983
5.5	– April 1983
5.6	– May 1983
5.7	– June 1983
5.8	– July 1983
5.9	– August 1983
5.10	– September 1983
5.11	– October 1983
5.12	– November 1983
6	Fuels. Supply and deliveries of petroleum products. Preliminary data
6.1	– November 1982
6.2	– December 1982
6.3	– January 1983
6.4	– February 1983
6.5	– March 1983
6.6	– April 1983
6.7	– May 1983
6.8	– June 1983
6.9	– July 1983
6.10	– August 1983
6.11	– September 1983
6.12	– October 1983
7	Energy supply. Preliminary data
7.1	– 3rd quarter 1981 and 1982
7.2	– 4th quarter 1981 and 1982 and the whole years 1981 and 1982
7.3	– 1st quarter 1982 and 1983
7.4	– 2nd quarter 1982 and 1983
8	Fuels. Production and consumption of gaswork gas during 1980 and 1981. Preliminary data
9	Industrial statistics. Consumption of purchased energy in 1981. Data by regions and branch
10	Fuels. Deliveries and consumption of fuels and lubricants during
10.1	– 4th quarter 1982 and during 1982
10.2	– 1st quarter 1983
10.3	– 2nd quarter 1983
11	Electric energy. Production and consumption of electric energy monthly in 1981. Definitive figures
12	Electric energy supply and district heating 1981
13	Deliveries and consumption of gasoline and fuel oil divided into municipalities in 1982
14	Energy statistics

Serie E 1983 (forts.)

14.1 – för småhus 1982
14.2 – för lokaler 1982
14.3 – för flerbostadshus 1982
15.1 Elförsörjningen. Elproduktion och teknisk utrustning i kraftstationerna 1982. Preliminära uppgifter
16 Bränslen. Produktion och förbrukning av stadsgas 1981 och 1982. Preliminära uppgifter
17 Förteckning över Statistiska meddelanden i serie E 1983

Series E 1983 (cont.)

14.1 – for one- or two-dwelling houses in 1982
14.2 – for buildings with premises in 1982
14.3 – for multi-dwelling buildings in 1982
15.1 Electric energy supply. Electric energy production and technical equipment in power stations in 1982. Preliminary data
16 Fuels. Production and consumption of gaswork gas during 1981 and 1982. Preliminary data
17 List of Statistical Reports, Series E 1983

Serie F 1983 – Företag

1 Publiceringsplan för Statistiska meddelanden 1983, serie F
2 Företagens utländska tillgångar och skulder samt skattning av bytesbalansutfallet
2.1 – 3e kv 1982
2.2 – 4e kv 1982
3 Ej utgiven
4 Likvida medels utveckling inom större industri- och varuhandelsföretag fjärde kvartalet 1982
5 Investeringar 1982–1983 enligt
5.1 – februarienkäten 1983
5.2 – majenkäten 1983
5.2 – Rättelseblad
5.3 – augustienkäten 1983
6 Företagens taxerade inkomster m. m. under verksamhetsåren 1981 och 1982
7 Större industri- och varuhandelsföretag. Resultat 1982
8 Industrins investeringar för yttre miljöskydd 1981
9 Industriföretagen 1981–1982. Ekonomisk redovisning
10 Större industriföretag. Delårsresultat för 1983
11 Förteckning över Statistiska meddelanden i serie F 1983

Series F 1983 – Enterprises

1 Publishing Plan for the Statistical Reports 1983, Series F
2 Foreign assets and liabilities of enterprises and estimate of the balance of current account
2.1 – 3rd quarter 1982
2.2 – 4th quarter 1982
3 Has not been published
4 The development of liquid assets in large manufacturing wholesale and retail trade enterprises in the fourth quarter of 1982
5 Investments 1982–1983 according to the survey in
5.1 – February 1983
5.1 – May 1983
5.2 – Correction
5.3 – August 1983
6 Assessed incomes and deductions for limited companies the income years 1981 and 1982
7 Gross operating income and result of larger enterprises in manufacturing, wholesale and retail trade in 1982
8 Investments in 1981 in environmental protection in industry
9 Manufacturing enterprises and mining and quarrying 1981–1982. Financial accounts
10 Larger manufacturing enterprises. Result in the first part of 1983
11 List of Statistical Reports, Series F 1983

Serie H 1983 – Handel och servicenäringar

1 Publiceringsplan för Statistiska meddelanden 1983, serie H
2 Utrikeshandel. Månadsstatistik
2.1 – oktober 1982
2.2 – november 1982
2.3 – december 1982. Årsstatistik
2.4 – januari 1983
2.5 – februari 1983
2.6 – mars 1983
2.7 – april 1983
2.8 – maj 1983

Series H 1983 – Trade and service trades

1 Publishing Plan for the Statistical Reports 1983, Series H
2 Foreign Trade. Monthly Bulletin
2.1 – October 1982
2.2 – November 1982
2.3 – December 1982. Annual data
2.4 – January 1983
2.5 – February 1983
2.6 – March 1983
2.7 – April 1983
2.8 – May 1983

Serie H 1983 (forts.)

2.9	– juni 1983
2.10	– juli 1983
2.11	– augusti 1983
2.12	– september 1983
3	Inkvarteringsstatistik
3.1	– november 1982
3.2	– december 1982
3.3	– januari 1983
3.4	– februari 1983
3.5	– mars 1983
3.6	– 1982
3.7	– april 1983
3.8	– maj 1983
3.9	– juni 1983
3.10	– juli 1983
3.11	– augusti 1983
3.12	– september 1983
3.13	– oktober 1983
4	Konjunkturbarometer för varuhandeln
4.1	– december 1982
4.2	– mars 1983
4.3	– juni 1983
4.4	– september 1983
5	Utrikeshandel. Kvartalsstatistik. Import
5.1	– januari–december 1982. Årsstatistik
5.2	– januari–mars 1983
5.3	– januari–juni 1983
5.4	– januari–september 1983
6	Utrikeshandel. Kvartalsstatistik. Export
6.1	– januari–december 1982. Årsstatistik
6.2	– januari–mars 1983
6.3	– januari–juni 1983
6.4	– januari–september 1983
7.1	Lager inom varuhandeln 4e kv 1982. Preliminära uppgifter
8	Omsättning inom detaljhandeln och vissa tjänstenäringar. Preliminära uppgifter
8.1	– 4e kv 1982
8.2	– 1982
8.3	– 1a och 2a kv 1983
9	Förteckning över Statistiska meddelanden i serie H 1983

Series H 1983 (cont.)

2.9	– June 1983
2.10	– July 1983
2.11	– August 1983
2.12	– September 1983
3	Accomodation statistics
3.1	– November 1982
3.2	– December 1982
3.3	– January 1983
3.4	– February 1983
3.5	– March 1983
3.6	– 1982
3.7	– April 1983
3.8	– May 1983
3.9	– June 1983
3.10	– July 1983
3.11	– August 1983
3.12	– September 1983
3.13	– October 1983
4	A business tendency survey in wholesale and retail trade
4.1	– December 1982
4.2	– March 1983
4.3	– June 1983
4.4	– September 1983
5	Foreign Trade. Quarterly Bulletin. Imports
5.1	– January–December 1982. Annual data
5.2	– January–March 1983
5.3	– January–June 1983
5.4	– January–September 1983
6	Foreign Trade. Quarterly Bulletin. Exports
6.1	– January–December 1982. Annual data
6.2	– January–March 1983
6.3	– January–June 1983
6.4	– January–September 1983
7.1	Stocks in wholesale and retail trade 4th quarter 1982. Preliminary data
8	Turnover in retail trade and certain service trades. Preliminary data
8.1	– 4th quarter 1982
8.2	– 1982
8.3	– 1st and 2nd quarter 1983
9	List of Statistical Reports. Series H 1983

Serie HS 1983 – Hälso- och sjukvård (producent Socialstyrelsen[1])

1	Publiceringsplan för Statistiska meddelanden 1983, serie HS
2	Personalsituationen inom folktandvården 1982-10-01 avseende tandläkare
3	Aborter 1981

[1] Med undantag av Arbetsskador, nr 6, som är producerad av Arbetarskyddsstyrelsen och SCB.

Series HS 1983 – Public health and medical care (produced by the National Board of Health and Welfare[2])

1	Publishing Plan for the Statistical Reports 1983, Series HS
2	Dentists in public dental care 1982-10-01
3	Abortions 1981

[2] Except Occupational Injuries, No. 6, which is produced by the National Board of Occupational Safety and Health and the SCB.

Serie HS 1983 (forts.)

4	Verkställda aborter. Preliminär sammanställning
4.1	– januari–december 1982
4.2	– januar–juni 1983
5	Medicinsk födelseregistrering 1979 och 1980
6	Arbetsskador (producerad av Arbetarskyddsstyrelsen och SCB gemensamt)
6.1	– 1a halvåret 1982
6.2	– 2a halvåret 1982
6.3	– 1982. Preliminär sammanställning
7	Verkställda steriliseringar januari–december 1982. Preliminär sammanställning
8	Läkarstatistik 1982
9	Aborter 1982
10	Hälso- och sjukvårdspersonal, sjukgymnaster 1982
11	Hälso- och sjukvårdspersonal, sjuksköterskor 1982
12	Personalsituationen inom folktandvården 1983-10-03 avseende tandläkare
13	Förteckning över Statistiska meddelanden i serie HS 1983

Series HS 1983 (cont.)

4	Induced abortions. Preliminary report
4.1	– January–December 1982
4.2	– January–June 1983
5	Medical birth registration in 1979 and 1980
6	Occupational injuries
6.1	– 1st half of 1982
6.2	– 2nd half of 1982
6.3	– 1982. Preliminary report
7	Sterilizations January–December 1982. Preliminary report
8	Physicians 1982
9	Abortions 1982
10	Registered physiotherapists 1982
11	Registered nurses 1982
12	Dentists in public dental care 1983-10-03
13	List of Statistical Reports, series HS 1983

Serie I 1983 – Industri, branschuppgifter

1	Publiceringsplan för Statistiska meddelanden 1983, serie I
2	Industrins lager. Preliminära uppgifter
2.1	– 3e kv 1982
2.2	– 4e kv 1982
2.3	– 1a kv 1983
2.4	– 2a kv 1983
2.5	– 3e kv 1983
3	Industriproduktionsindex
3.1	– oktober 1982
3.2	– november 1982
3.3	– december 1982
3.4	– januari 1983
3.5	– februari 1983
3.6	– mars 1983
3.7	– april 1983
3.8	– maj 1983
3.9	– juni 1983
3.10	– juli 1983
3.11	– augusti 1983
3.12	– september 1983
4	Import
4.1	– år 1982 samt 1a halvåret 1983 enligt importenkäten i november 1982
4.2	– år 1982 samt första halvåret och tredje kvartalet 1983 enligt importenkäten i februari 1983
4.3	– år 1983 samt första kvartalet 1984 enligt importenkäten i augusti 1983

Series I 1983 – Manufacturing, branch data

1	Publishing Plan for the Statistical Reports 1983, Series I
2	Stocks industry. Preliminary data
2.1	– 3rd quarter 1982
2.2	– 4th quarter 1982
2.3	– 1st quarter 1983
2.4	– 2nd quarter 1983
2.5	– 3rd quarter 1983
3	Monthly index numbers of mining and manufacturing
3.1	– October 1982
3.2	– November 1982
3.3	– December 1982
3.4	– January 1983
3.5	– February 1983
3.6	– March 1983
3.7	– April 1983
3.8	– May 1983
3.9	– June 1983
3.10	– July 1983
3.11	– August 1983
3.12	– September 1983
4	Imports
4.1	– 1982 including the 1st half year 1983 according to the survey in November 1982
4.2	– 1982 including the first half year and the third quarter 1983 according to the survey in February 1983
4.3	– 1983 including the first quarter 1984 according to the survey in August 1983

Serie I 1983 (forts.)

5	Industrins leverans- och orderläge. Preliminära uppgifter
5.1	– november 1982
5.2	– december 1982
5.3	– januari 1983
5.4	– februari 1983 samt reviderade värden för 1980, 1981 och 1982
5.5	– mars 1983
5.6	– april 1983
5.7	– maj 1983
5.8	– juni 1983
5.9	– juli–augusti 1983
5.10	– september 1983
6	Export
6.1	– 1983 och första halvåret 1984 enligt exportenkäten i februari 1983
6.2	– 1983 och första halvåret 1984 enligt exportenkäten i maj 1983
6.3	– 1983 och 1984 enligt exportenkäten i augusti 1983
6.4	– 1983 och 1984 enligt exportenkäten i oktober 1983
7	Industristatistik. Regional redovisning av data för åren 1980 och 1981
8	Industristatistik. Strukturstatistik för industrin för åren 1980 och 1981
9	Industri 1982. Preliminära branschdata
10	Industrins kapacitetsutnyttjande. Andra kvartalet 1983
11	Förteckning över Statistiska meddelanden i serie I 1983

Series I 1983 (cont.)

5	The delivery and orderbook situation in industry. Preliminary data
5.1	– November 1982
5.2	– December 1982
5.3	– January 1983
5.4	– February 1983 and revised figures for 1980, 1981 and 1982
5.5	– March 1983
5.6	– April 1983
5.7	– May 1983
5.8	– June 1983
5.9	– July–August 1983
5.10	– September 1983
6	Exports
6.1	– 1983 and the first half of 1984 according to the survey in February 1983
6.2	– 1983 and the first half of 1984 according to the survey in May 1983
6.3	– 1983 and 1984 according to the survey in August 1983
6.4	– 1983 and 1984 according to the survey in October 1983
7	Industrial statistics. Regional statistics 1980 and 1981
8	Industrial Statistics. Structural Statistics 1980 and 1981
9	Manufacturing 1982. Preliminary data
10	Capacity utilization in industry. Second quarter 1983
11	List of Statistical Reports, Series I 1983

Serie Iv 1983 – Industri. Varuuppgifter

1	Publiceringsplan för Statistiska meddelanden 1983, serie Iv
2	Gummi och gummivaror. Produktion, tillförsel och förbrukning
2.1	– januari–september 1982
2.2	– januari–december 1982
2.3	– januari–mars 1983
2.4	– januari–juni 1983
3	Verkstadsprodukter: Marknaden samt verkstadsindustrins utveckling
3.1	– januari 1981–september 1982
3.2	– januari 1981–december 1982
3.3	– januari 1982–mars 1983
3.4	– januari 1982–juni 1983
4	Sågat virke av barrträd. Produktion
4.1	– november 1982
4.2	– december 1982
4.3	– januari 1983
4.4	– februari 1983
4.5	– mars 1983
4.6	– april 1983
4.7	– maj 1983

Series Iv 1983–Manufacturing. Data on commodities

1	Publishing Plan for the Statistical Reports 1983, Series Iv
2	Rubber and rubber products. Production, domestic supply and consumption
2.1	– January–September 1982
2.2	– January–December 1982
2.3	– January–March 1983
2.4	– January–June 1983
3	The market for commodities from and the situation for the engineering industry
3.1	– January 1981–September 1982
3.2	– January 1981–December 1982
3.3	– January 1982–March 1983
3.4	– January 1982–June 1983
4	Sawn wood of conifer. Production
4.1	– November 1982
4.2	– December 1982
4.3	– January 1983
4.4	– February 1983
4.5	– March 1983
4.6	– April 1983
4.7	– May 1983

Serie Iv 1983 (forts.)

4.8	– juni 1983
4.9	– juli 1983
4.10	– augusti 1983
4.11	– september 1983
4.12	– oktober 1983
5	Handelsfärdigt stål. Marknaden samt järn- och stålverkens utveckling
5.1	– juli–september 1982
5.2	– 1982
5.3	– januari–mars 1983
5.4	– april–juni 1983
6	Kemikalier, gödselmedel och plaster. Produktion och tillförsel
6.1	– januari–september 1982
6.2	– januari–december 1982
6.3	– januari–mars 1983
6.4	– januari–juni 1983
7	Garn och vävnader. Marknaden samt textilindustrins utveckling
7.1	– januari–september 1982
7.2	– januari–december 1982
7.3	– januari–mars 1983
7.4	– januari–juni 1983
8	Byggvaror. Produktion och leveranser
8.1	– januari–september 1982
8.2	– november 1982
8.3	– december 1982
8.4	– januari 1983
8.5	– januari–december 1982
8.6	– februari 1983
8.7	– mars 1983
8.8	– april 1983
8.9	– januari–mars 1983
8.10	– maj 1983
8.11	– juni 1983
8.12	– januari–juni 1983
8.13	– juli–augusti 1983
8.14	– september 1983
8.15	– oktober 1983
9	Skodon, läder, hudar och skinn
9.1	– Produktion och tillförsel januari–september 1982
9.2	– Produktion och tillförsel januari–december 1982
9.3	– Marknaden och skoindustrins utveckling januari–mars 1983
9.4	– Marknaden och skoindustrins utveckling januari–juni 1983
10	Kläder (trikå och konfektion). Marknaden samt beklädnadsindustrins utveckling
10.1	– januari–september 1982
10.2	– januari–december 1982
10.3	– januari–mars 1983
10.4	– januari–juni 1983
11	Metaller och legeringsmalmer. Marknaden samt ickejärnmetallverkens utveckling under 1982

Series Iv 1983 (cont.)

4.8	– June 1983
4.9	– July 1983
4.10	– August 1983
4.11	– September 1983
4.12	– October 1983
5	The market for finished iron and steel and the situation for the iron and steelworks
5.1	– July–September 1982
5.2	– 1982
5.3	– January–March 1983
5.4	– April–June 1983
6	Chemicals, fertilizers and plastic materials. Production and domestic supply
6.1	– January–September 1982
6.2	– January–December 1982
6.3	– January–March 1983
6.4	– January–June 1983
7	The market for yarn, woven and knitted fabrics and the situation for the manufacture of textiles
7.1	– January–September 1982
7.2	– January–December 1982
7.3	– January–March 1983
7.4	– January–June 1983
8	Building materials. Production and deliveries
8.1	– January–September 1982
8.2	– November 1982
8.3	– December 1982
8.4	– January 1983
8.5	– January–December 1982
8.6	– February 1983
8.7	– March 1983
8.8	– April 1983
8.9	– January–March 1983
8.10	– May 1983
8.11	– June 1983
8.12	– January–June 1983
8.13	– July–August 1983
8.14	– September 1983
8.15	– October 1983
9	Shoes, leather, hides and skins
9.1	– Production and supply January–September 1982
9.2	– Production and supply January–December 1982
9.3	– The market for shoes, leather and hides January–March 1983
9.4	– The market for shoes, leather and hides, January–June 1983
10	The market for garments and the situation for the manufacture of wearing apparel
10.1	– January–September 1982
10.2	– January–December 1982
10.3	– January–March 1983
10.4	– January–June 1983
11	The market for metals and metalic ores and the situation for non-ferrous metal basic industries 1982

Serie Iv 1983 (forts.)

12	Livsmedel, tvål och tvättmedel. Lager 31 december 1982
13	Förteckning över Statistiska meddelanden i serie Iv 1983

Series Iv 1983 (cont.)

12	Food, soap and washing powder. Stocks 31 of December 1982
13	List of Statistical Reports, Series Iv 1983

Serie J 1983 – Jordbruk, skogsbruk och fiske

1	Publiceringsplan för Statistiska meddelanden 1983, serie J
2	Saltsjöfiskets fångster. Preliminära uppgifter
2.1	– november 1982
2.2	– december 1982 och hela år 1982
2.3	– januari 1983
2.4	– februari 1983
2.5	– mars 1983
2.6	– april 1983
2.7	– maj 1983
2.8	– juni 1983
2.9	– juli 1983
2.10	– augusti 1983
2.11	– september 1983
3	Galt- och betäckningsstatistik
3.1	– november och december 1982
3.2	– december 1982 och januari 1983
3.3	– januari och februari 1983
3.4	– februari och mars 1983
3.5	– mars och april 1983
3.6	– april och maj 1983
3.7	– maj och juni 1983
3.8	– juni och juli 1983
3.9	– juli och augusti 1983
3.10	– augusti och september 1983
3.11	– september och oktober 1983
3.12	– oktober och november 1983
4	Fiskeflottans storlek och sammansättning den 31 december 1982
5	Yrkesfiskarnas inkomster och utgifter 1981
6	Rapporter från lantbrukets företagsregister 1982
6.1	– Jordbruksföretag och företagare den 10 juni 1982
6.2	– Ägoslag och åkerarealens användning den 10 juni 1982
6.3	– Husdjur den 10 juni 1982
6.4	– Förändringar i jordbruksföretagen 1981–1982. Längdsnittsdata
7.1	Skördevärde-, skördeskade- och ersättningsstatistik för län och riket 1982
7.2	Normskördar för skördeområden, län och riket 1983
8	Rapporter från de objektiva skördeuppskattningarna 1982
8.1	– Hektarskördar. Definitiva resultat, metoder m. m.

Series J 1983 – Agriculture, forestry and fishery

1	Publishing Plan for the Statistical Reports 1983, Series J
2	Swedish sea-fisheries. Preliminary data
2.1	– November 1982
2.2	– December 1982 and the year 1982
2.3	– January 1983
2.4	– February 1983
2.5	– March 1983
2.6	– April 1983
2.7	– May 1983
2.8	– June 1983
2.9	– July 1983
2.10	– August 1983
2.11	– September 1983
3	Statistics on boars and covered sows
3.1	– November and December 1982
3.2	– December 1982 and January 1983
3.3	– January and February 1983
3.4	– February and March 1983
3.5	– March and April 1983
3.6	– April and May 1983
3.7	– May and June 1983
3.8	– June and July 1983
3.9	– July and August 1983
3.10	– August and September 1983
3.11	– September and October 1983
3.12	– October and November 1983
4	The size and composition of fishing fleet as per 31st December 1982
5	The assessed income and expenditure, etc., of professional fishermen in 1981
6	Reports from the farm register
6.1	– Holdings and holders the 10th June 1982
6.2	– Type of land and use of arable land the 10th June 1982
6.3	– Livestock the 10th June 1982
6.4	– Structural changes in agriculture 1981–1982. Longitudinal data
7.1	Statistics about yield values, yield damages and compensations for counties and the whole country in 1982
7.2	Standard yields for yield survey districts, counties and the whole country 1983
8	Objective crop yield estimation 1982
8.1	– Objective crop yield estimation

Serie J 1983 (forts.)

8.2	– Spannmålsskördens kvalitet
8.3	– Undersökningar rörande växtodlingen
9	Deklarationsundersökningen för jordbrukare 1981
10	Skördens storlek 1982. Definitiva uppgifter
11	Jordbrukets arbetskraft 1982
12	Årsväxtrapport för den
12.1	– 18 maj 1983
12.2	– 27 juli 1983
12.3	– 12 oktober 1983
13	Byggnadsverksamheten i lantbruket 1981–1982
14	Rapporter från de objektiva skördeuppskattningarna 1983
14.1	– Åkerarealens användning. Preliminära uppgifter
14.2	– Första skörd på slåttervall. Preliminära uppgifter
14.3	– Vallskördens kvalitet
14.4	– Skörd av höstsäd
14.5	– Skörd av vårsäd
14.6	– Skörd av potatis
14.7	– Höstsådda arealer
15	Rapporter från lantbrukets företagsregister 1983. Husdjur, antal den 9 juni 1983. Preliminära uppgifter för företag med mer än 2,0 hektar åkermark
16	Jordbruksekonomiska undersökningen 1982
17	Förteckning över Statistiska meddelanden i serie J 1983

Series J 1983 (cont.)

8.2	– Quality of grain
8.3	– Surveys about the plants husbandry
9	The 1981 survey of farmers' assessed incomes. Expenditures, net earnings etc.
10	Size of yield in 1982. Definitive data
11	The agricultural labour force in 1982
12	Crop out-look report on
12.1	– May 18th, 1983
12.2	– July 27th, 1983
12.3	– October 12th, 1983
13	Building in the agricultural sector 1981–1982
14	Objective crop yield estimation 1983
14.1	– The use of arable land. Preliminary data
14.2	– Yield of hay. Preliminary data
14.3	– The quality of hay
14.4	– Yield of winter wheat and winter rye
14.5	– Yield of spring wheat and coarse grain
14.6	– Yield of potatoes
14.7	– Areas of wheat, rye, rape and turnip rape sown in the autumn
15	Number of livestock the 9th June 1983. Preliminary data
16	The 1982 farm economics survey
17	List of Statistical Reports, Series J 1983

Serie K 1983 – Kreditmarknad

1	Publiceringsplan för Statistiska meddelanden 1983, serie K
2	Försäkringsbolagen. Preliminära uppgifter
2.1	– 4e kv 1982
2.2	– 1a kv 1983
2.3	– 1982 och första halvåret 1983
2.4	– 3e kv 1983
3	Kreditmarknaden. Kvartalsvis redovisning av kreditmarknadsstatistik och finansräkenskaper
3.1	– 1. 1983
3.2	– 2. 1983
3.3	– 3. 1983
4	In- och utlåning på kreditmarknaden – fördelning efter kundkategorier. Definitiva uppgifter för 1982

Series K 1983 – Credit market, banking and insurance

1	Publishing Plan for the Statistical Reports 1983, Series K
2	Swedish insurance companies. Preliminary data
2.1	– 4th quarter of 1982
2.2	– 1st quarter of 1983
2.3	– 1982 and 1st quarter of 1983
2.4	– 3rd quarter of 1983
3	Credit market statistics. Quarterly report on credit market statistics and financial accounts
3.1	– 1. 1983
3.2	– 2. 1983
3.3	– 3. 1983
4	Deposits and advances on the credit market – by groups of depositors and borrowers. Conclusive data for 1982

Serie K 1983 (forts.)

5	Finansbolag och andra finansföretag 1982
6	Investmentbolag och aktiefonder 1982
7	Förteckning över Statistiska meddelanden i serie K 1983

Series K 1983 (cont.)

5	Swedish finance houses and other finance companies 1982
6	Swedish investment companies and funds 1982
7	List of Statistical Reports, Series K 1983

Serie N 1983 – Nationalräkenskaper och offentliga finanser

1	Publiceringsplan för Statistiska meddelanden 1983, serie N
2	Nationalräkenskaperna kvartalsvis
2.1	– 1a kv 1981–3e kv 1982
2.2	– 1a kv 1981–4e kv 1982
2.3	– 1a kv 1981–1a kv 1983
2.4	– 2a kv 1983
2.5	– 1970–1982
2.5*	Appendix 1–5 se nedan
2.6	– 3e kv 1983
2.5*	Appendix 1 Konsumtion 1970–1982
2.5*	Appendix 2 Realkapitalinvestering samt realkapitalstockar 1970–1982
2.5*	Appendix 3 Försörjningsbalanser Utlandstransaktioner Förädlingsvärden och sparande för företagsbranscher
2.5*	Appendix 4 Produktion och faktorinsats 1970–1982
2.5*	Appendix 5 Sysselsättning och lönesummor 1970–1982
3	Taxeringsutfallet 1982
4	Kreditmarknadsstatistiska uppgifter för kommuner och landsting den 31 december 1982
5	Kommunernas och landstingens utgifter och inkomster enligt februarienkäten 1983
6	Utfallet av statsbudgeten för budgetåret 1981/82
7.1	Riksrevisionsverkets taxeringsstatistiska undersökning taxeringsåret 1981 (inkomståret 1980)
7.2	Riksrevisionsverkets taxeringsstatistiska undersökning taxeringsåret 1982 (inkomståret 1981)
8	Kommunernas finanser 1982. Preliminära uppgifter
9	Skattekronor och medelskattekraft enligt 1983 års taxering. Preliminära uppgifter
10	Särskild fastighetstaxering 1983
11	Finansräkenskaper 1978–1982
12	Kommunernas och landstingens investe-

Series N 1983 – National accounts and public finance

1	Publishing Plan for the Statistical Reports 1983, Series N
2	Quarterly national accounts
2.1	– 1st quarter 1981–3rd quarter 1982
2.2	– 1st quarter 1981–4th quarter 1982
2.3	– 1st quarter 1981–1st quarter 1983
2.4	– 2nd quarter 1983
2.5	– 1970–1982
2.5*	Appendix 1–5 see below
2.6	– 3rd quarter 1983
2.5*	Appendix 1 Consumption
2.5*	Appendix 2 Capital formation and stocks of fixed capital
2.5*	Appendix 3 GDP by expenditure External transactions Value added and savings for enterprises
2.5*	Appendix 4 Production and factor income
2.5*	Appendix 5 Employment and compensation of employees
3	Tax assessment 1982
4	Credit market statistics for the municipalities and county councils 31st of December 1982
5	Estimated expenditure and revenue of the municipalities and the county councils according to the survey in February 1983
6	The outcome of the central government budget for the fiscal year 1981/82
7.1	Sample survey of income tax returns for individuals at the assessment in 1981
7.2	Sample survey of income tax returns for individuals at the assessment in 1982
8	Finances municipalities 1982. Preliminary data
9	Skattekronor and tax capacity according to the 1983 assessment. Preliminary data
10	Special assessment of real estate in 1983
11	Financial accounts 1978–1982
12	Investments, wages and salaries of the

Serie N 1983 (forts.)

ringar och löner enligt augustienkäten 1983
13 Kommunernas ekonomi inför 1984 – kommunernas och landstingens budgeter, skatteunderlag och skattesatser. Preliminära uppgifter
14 Förteckning över Statistiska meddelanden i serie N 1983

Series N 1983 (cont.)

municipalities and the county councils according to the survey in August 1983
13 Tax basis, tax rates and budget statistics for the municipalities and the county councils. Preliminary data
14 List of Statistical Reports, Series N 1983

Serie P 1983 – Priser och konsumtion

1 Publiceringsplan för Statistiska meddelanden 1983, serie P
2 Producent-, export- och importprisindex samt prisindex för inhemsk tillgång
2.1 – november 1982
2.2 – december 1982
2 Prisindex i producent- och importled
2.3 – januari 1983
2.4 – februari 1983
2.5 – mars 1983
2.6 – april 1983
2.7 – maj 1983
2.8 – juni 1983
2.9 – juli 1983
2.10 – augusti 1983
2.11 – september 1983
2.12 – oktober 1983
3 Konsumentprisindex
3.1 – december 1982
3.2 – januari 1983
3.3 – februari 1983
3.4 – mars 1983
3.5 – april 1983
3.6 – maj 1983
3.7 – juni 1983
3.8 – juli 1983
3.9 – augusti 1983
3.10 – september 1983
3.11 – oktober 1983
3.12 – november 1983
4 Byggnadsindex. Byggnadsprisindex för flerbostadshus och gruppbyggda småhus
4.1 – 1a och 2a kv 1982
4.2 – 3e kv 1982
4.3 – 1982. Funktionsprisindex och kvalitetsindex för flerbostadshus och gruppbyggda småhus 1968–1982
4.4 – 1a kv 1983
5 Faktorprisindex för bostäder
5.1 – oktober–december och 4e kv 1982
5.2 – januari–mars och 1a kv 1983

Series P 1983 – Prices and consumption

1 Publishing Plan for the Statistical Reports 1983, Series P
2 Price indices in producer, export and import stages
2.1 – November 1982
2.2 – December 1982
2 Price indices in producer and import stages
2.3 – January 1983
2.4 – February 1983
2.5 – March 1983
2.6 – April 1983
2.7 – May 1983
2.8 – June 1983
2.9 – July 1983
2.10 – August 1983
2.11 – September 1983
2.12 – October 1983
3 Consumer price index
3.1 – December 1982
3.2 – January 1983
3.3 – February 1983
3.4 – March 1983
3.5 – April 1983
3.6 – May 1983
3.7 – June 1983
3.8 – July 1983
3.9 – August 1983
3.10 – September 1983
3.11 – October 1983
3.12 – November 1983
4 Price indices for multiple-unit dwellings and single-unit dwellings
4.1 – 1st and 2nd quarter 1982
4.2 – 3rd quarter 1982
4.3 – 1982. Indices for the price per flat (functional unit) and quality indices per flat for multiple-unit dwellings and single-unit dwellings 1968–1982
4.4 – 1st quarter 1983
5 Indices of construction costs for housing
5.1 – October–December and 4th quarter 1982
5.2 – January–March and 1st quarter 1983

Serie P 1983 (forts.)

5.3 – april–juni och 2a kv 1983
5.4 – juli–september och 3e kv 1983
6 Hushållens inköpsplaner
6.1 – januari 1983
6.2 – april 1983
6.3 – juli 1983
6.4 – oktober 1983
7 Fastighetsprisstatistik 1981
7.1 – Jord- och skogsbruk
7.2 – Småhus, hyreshus och industrifastigheter
7.3 – Obebyggd tomtmark
8 Fastighetsprisstatistik 1980. Småhus och hyreshus. Definitiva uppgifter
9 Konsumentprisindextal 1914–1982. Konsumentprisindex, levnadskostnadsindex, nettoprisindex, basbeloppet, pensionspristalet, procenttal för ändring av vissa underhållsbidrag
10 Fastighetsprisstatistik 1982
10.1 – Småhus, hyreshus och industrifastigheter
10.2 – Obebyggd tomtmark
10.3 – Jord- och skogsbruk
11 Fastighetsprisstatistik för 1a och 2a kvartalen 1983. Småhus, hyreshus och industrifastigheter. Preliminära uppgifter
12 Förteckning över Statistiska meddelanden i serie P 1983

Series P 1983 (cont.)

5.3 – April–June and 2nd quarter 1983
5.4 – July–September and 3rd quarter 1983
6 Consumer buying expectations
6.1 – January 1983
6.2 – April 1983
6.3 – July 1983
6.4 – October 1983
7 Prices of real estates 1981
7.1 – Agriculture and forestry
7.2 – One- or two-dwelling, multi-dwelling and commercial buildings and industrial real estates
7.3 – Sites
8 Prices of real estate 1980. One- or two-dwelling, multi-dwelling and commercial buildings. Final data
9 Consumer Price Index Numbers 1914–1982
10 Prices of real estates 1982
10.1 – One- or two-dwelling, multi-dwelling and commercial buildings and industrial real estates
10.2 – Sites
10.3 – Agriculture and forestry
11 Prices of real estate for 1st and 2nd quarter of 1983. One- and two-dwelling houses, multi-dwelling and commercial buildings, industrial buildings. Preliminary data
12 List of Statistical Reports, series P 1983

Serie R 1983 – Rättsväsen

1 Publiceringsplan för Statistiska meddelanden 1983, serie R
2 Polisrapporterade vägtrafikolyckor med personskada. Preliminära uppgifter
2.1 – december 1982 och år 1982
2.2 – januari 1983
2.3 – februari 1983
2.4 – 1982. Definitiva uppgifter
2.5 – mars 1983 och första kvartalet 1983
2.6 – april 1983
2.7 – maj 1983
2.8 – juni 1983 och första halvåret 1983
2.9 – juli 1983
2.10 – augusti 1983 och perioden maj–augusti
2.11 – september 1983 och de tre första kvartalen 1983
2.12 – oktober 1983
2.13 – november 1983
3 Kriminalstatistik. Brott som kommit till polisens kännedom. Preliminära uppgifter
3.1 – november 1982
3.2 – december 1982 och år 1982

Series R 1983 – Legal statistics

1 Publishing Plan for the Statistical Reports 1983, Series R
2 Road traffic accidents reported to the police with personal injury. Preliminary data
2.1 – December 1982 and year 1982
2.2 – January 1983
2.3 – February 1983
2.4 – 1982. Definitive data
2.5 – March 1983 and the first quarter 1983
2.6 – April 1983
2.7 – May 1983
2.8 – June 1983 and the first half of 1983
2.9 – July 1983
2.10 – August 1983 and May–August
2.11 – September 1983 and the three first quarters 1983
2.12 – October 1983
2.13 – November 1983
3 Criminal statistics. Crimes known to the police. Preliminary data
3.1 – November 1982
3.2 – December 1982 and the whole year 1982

Serie R 1983 (forts.)

3.3 – januari–mars 1983
3.4 – april–juni 1983
3.5 – juli–september 1983
4 Kriminalstatistik, brott m.m. 1982. Anmälda brott, uppklarade brott och för brott misstänkta personer. Definitiva uppgifter
5 Inteckningar 1982
6 Kriminalstatistik. För brott lagförda personer, 1982
6.1 – Huvudbrott och huvudpåföljd
6.2 – Den grövre brottsligheten
6.3 – Brottslagrum
7 Kriminalstatistik. För narkotikabrott lagförda personer, 1982
7* Rättelse
8 Förteckning över Statistiska meddelanden i serie R 1983

Series R 1983 (cont.)

3.3 – January–March 1983
3.4 – April–June 1983
3.5 – July–September 1983
4 Criminal statistics, offences 1982. Reported offences, cleared offences and persons suspected of offences. Definitive data
5 Registrations 1982
6 Criminal statistics. Persons found guilty of criminal offences, 1982
6.1 – Principal offences and principal sanctions
6.2 – Serious offences
6.3 – Sections of law concerning criminal offences
7 Criminal statistics. Persons found quilty of drug offences, 1982
7* Correction
8 List of Statistical Reports, Series R 1983

Serie S 1983 – Socialtjänst och socialförsäkring

1 Publiceringsplan för Statistiska meddelanden 1983, serie S
2 Nykterhetsvård 1981
3 Utgifter för sociala ändamål inom socialdepartementets, arbetsmarknadsdepartementets och bostadsdepartementets verksamhetsområde 1983/84. Förslag enligt budgetpropositionen 1983
4 Beslut om atgärder enligt barnavårdslagen, 1981
5 Beslut om atgärder enligt barnavårdslagen, 1981. Kommunvis fördelning
6 Socialvårdsstatistik på kommunniva 1981
7 Socialutgifterna och dessas finansiering 1981
8 Vägledande socialbidragsnormer 1 oktober 1982, med diagram
9 Adoptivbarn 1982. Preliminära uppgifter
10 Ungdomsvårdsskoleelever 1982
11 Barnomsorgsundersökningen 1982/83. Del 1. Yngre skolbarn (7–10 år)
12 Barnomsorgsundersökningen 1982/83. Del 2. Förskolebarn (0–6 år)
13 Social hemhjälpsverksamhet under 1982 samt verksamhet i den öppna vården för äldre och handikappade den 31 december 1982. Kommunvis fördelning
14 Social hemhjälpsverksamhet under 1982 samt verksamhet i den öppna vården för äldre och handikappade den 31 december 1982 och vecka 43, 1982
15 Uppföljningsstatistik inom nykterhetsvården med avseende på nykterhetsvårdslagen m.m. 1976–1981

Series S 1983 – Social welfare

1 Publishing Plan for the Statistical Reports 1983, Series S
2 Care of inebriates 1981
3 The social welfare expenditure under the Ministry for Social Affairs, the Ministry of the Labour Market and the Ministry of Housing 1983/84. From the Government Budget Proposals 1983
4 Decisions on measures according to the Child Welfare Act, 1981
5 Decisions on measures according to the Child Welfare Act, 1981
6 Social welfare statistics for municipalities 1981
7 The cost and financing of the social welfare services in 1981
8 Social assistance-principles on October 1st, 1982, with charts
9 Adopted children 1982. Preliminary data
10 Pupils in reformatory schools 1982
11 Survey of childcareneeds 1982/83. Part 1. Junior school-children (7–10 years old)
12 Survey of childcareneeds 1982/83. Part 2. Pre-school children (0–6 years old)
13 The home-help service during 1982 and assistance given in the home or other facilities for the aged and the handicapped on 31st December 1982. Distribution by municipality
14 The home-help service during 1982 and assistance given in the home or other facilities for the aged and handicapped on 31st December 1982 and week 43, 1982
15 Follow-up statistics for temperance care under the Temperance Act etc., 1976–1981

Serie S 1983 (forts.)

16 Socialbidrag under 1982. Kommunvis fördelning. Preliminära uppgifter
17 Färdtjänst 31 december 1982
18 Förskolor, fritidshem och familjedaghem den 31 december 1982
19 Faderskap och underhåll under 1982
20 Förteckning över Statistiska meddelanden i serie S 1983

Series S 1983 (cont.)

16 Social assistance 1982. Distribution by municipality. A preliminary report
17 Transport service on 31st December 1982
18 Children's day homes and family day-homes on December 31, 1982
19 Establishment of paternity and maintenance allowances during 1982
20 List of Statistical Reports, Series S 1983

Serie T 1983 – Transport och kommunikation

1 Publiceringsplan för Statistiska meddelanden 1983, serie T
2 Varutransporter med lastbil
2.1 – 3e kv 1982
2.2 – 4e kv 1982
2.3 – 1a kv 1983
2.4 – 1982. Definitiva uppgifter
2.5 – 2a kv 1983
3 Bosättningsmönster och långväga personbilsresor
4 Fordon enligt bilregistret
4.1 – december 1982 och hela 1982
4.2 – januari 1983
4.3 – februari 1983
4.4 – mars 1983
4.5 – april 1983
4.6 – maj 1983
4.7 – juni 1983 och första halvåret 1983
4.8 – juli 1983
4.9 – augusti 1983
4.10 – september 1983
4.11 – oktober 1983
4.12 – november 1983
5 Utrikessjöfarten
5.1 – oktober–december 1982 och hela år 1982
5.2 – januari–mars 1983
5.3 – april–juni 1983
5.4 – juli–september 1983
6 Utrikes- och kustsjöfart (SNI grupp 7121) och hamn- och lotsväsen, stuverirörelse (SNI grupp 7123) produktions- och importkonton för åren 1972–1981, löpande priser. Preliminära uppgifter 1981
7 Handelsflottans förändringar. Preliminära uppgifter
7.1 – 1982
7.2 – januari–juni 1983
8 Specialfartyg vid 1982 års slut
9 Svenska rederiers bruttointäkter från utrikes sjöfart och kostnader erlagda i utlandet
9.1 – 1982
9.2 – 1a kv 1983
9.3 – 2a kv 1983

Series T 1983 – Transport and communication

1 Publishing Plan for the Statistical Reports 1983, Series T
2 Road goods transport by lorries
2.1 – 3rd quarter 1982
2.2 – 4th quarter 1982
2.3 – 1st quarter 1983
2.4 – 1982. Final data
2.5 – 2nd quarter 1983
3 The relationship between residential patterns and demand for long-distance car travel
4 Registered vehicles
4.1 – December 1982 and 1982
4.2 – January 1983
4.3 – February 1983
4.4 – March 1983
4.5 – April 1983
4.6 – May 1983
4.7 – June 1983 and the first half of 1983
4.8 – July 1983
4.9 – August 1983
4.10 – September 1983
4.11 – October 1983
4.12 – November 1983
5 Sweden's overseas shipping
5.1 – October–December 1982 and in the year 1982
5.2 – January–March 1983
5.3 – April–June 1983
5.4 – July–September 1983
6 Ocean and coastal water transport (SNI 7121) and supporting services to water transport (SNI 7123), production and import accounts 1972–1981, current prices. Preliminary figures 1981
7 Increase and decrease of the Swedish merchant fleet. Preliminary data
7.1 – 1982
7.2 – January–June 1983
8 Special vessels at the end of 1982
9 Gross receipts obtained by Swedish shipping companies in foreign trade and costs paid abroad
9.1 – 1982
9.2 – 1st quarter 1983
9.3 – 2nd quarter 1983

Serie T 1983 (forts.)

10 I utrikes sjöfart transporterade godskvantiteter fördelade på varukategorier och på fartygsstorlek 1982
10* Rättelse
11 Utrikeshandeln fördelad på transportområden och efter varornas transportsätt 1981
12 Utrikes varutrafik i svenska hamnar 1981
12* Bilaga
13 Utrikeshandeln fördelad på transportområden och efter varornas transportsätt 1982
14 Flygplan i civil luftfart 1982 (Bestånd och nyregistreringar)
15 Antal resande till och från Sverige efter använt transportmedel 1977–1982
16 Stora container och storflak lossade från och lastade på fartyg i utrikes fart 1980–1982
17 Pris- och volymindexar för yrkesmässiga lastbilstransporter under 1982. Definitiva uppgifter
18 Fordon i län och kommuner den 31 december 1982 (enl. kommunindelningen 1 januari 1983)
19 Busstrafik 1980–1981
20 Utrikes och inrikes varutrafik på fartyg 1982 inkl. regional statistik
21 Fordon enligt bilregistret, personbilar 1982
22 Utlänningar i svenska handelsfartyg år 1982
23 Rederinäringens bruttointäkter samt handelsflottans användning under år 1982
24 Förteckning över Statistiska meddelanden i serie T 1983

Series T 1983 (cont.)

10 Goods unloaded and loaded in foreign trade at Swedish ports by commodities and by size of the vessels in 1982
10* Correction
11 Foreign trade by transport area and by mode of transport in 1981
12 Goods loaded and unloaded in foreign trade in Swedish ports 1981
12* Appendix
13 Foreign trade by transport area and by mode of transport in 1982
14 Civil aircraft on register in Sweden on December 31st 1982 and civil aircraft entered into the register during 1982
15 Number of passengers arriving and leaving Sweden by modes of transport 1977–1982
16 Large containers and large flats unloaded from and loaded on to ships in foreign trade during 1980–1982
17 Price and volume indices for goods transports by lorries 1982. Transports for hire or reward. Final data
18 Vehicles in counties and municipalities on December 31st, 1982
19 Bus traffic services
20 Shipping goods in 1982
21 Registered passenger cars 1982
22 Foreigners employed by the Swedish mercantile marine in 1982
23 Gross receipts of Swedish shipping and mercantile marine according to different routes
24 List of Statistical Reports, Series T 1983

Serie U 1983 – Utbildning, forskning och kultur

1 Publiceringsplan för Statistiska meddelanden 1983, serie U
2 Folkbiblioteken 1981. Mediebestånd, låneverksamhet, resurser och uppsökande verksamhet m. m.
3 Högskolan 1981/82. Forskarutbildning: Registrerade och nybörjare 1981/82. Definitiva uppgifter avseende höstterminen 1981 samt preliminära uppgifter avseende vårterminen 1982
4 Elevpanel för longitudinella studier. Beskrivning av panelen samt uppgifter avseende skolgången i årskurs 6 (1980) och årskurs 7 (1981) i grundskolan

Series U 1983 – Education, research and culture

1 Publishing Plan for the Statistical Reports 1983, Series U
2 Public libraries 1981. Media stock, lending, resources and extension work etc.
3 Higher education 1981/82. Post-graduate education: Registration and first year students 1981/82. Definite figures for autumn semester 1981 and preliminary figures for spring semester 1982
4 Panel of pupils for longitudinal studies. Description of the panel and data on the school attendance in grade 6 (1980) and grade 7 (1981) in compulsory comprehensive school

Serie U 1983 (forts.)

5 Viss gymnasial och eftergymnasial utbildning 1981/82. Närvaro höstterminen 1981 och examination läsåret 1981/82

6 Statlig vuxenutbildning 1981/82. Elever, kursdeltagare och studieresultat

7 Folkhögskolan 1982/83. Elever i kurser om minst 15 veckors längd höstterminen (ht) 1982

8 Specialskolan 1982/83. Elever i statliga specialskolor samt syn- och hörselskadade elever i vissa andra skolformer den 1 november 1982

9 Högskolan 1978/79–1980/81. Forskarutbildning: Examina och studietider läsåren 1978/79–1980/81

10 Högskolan 1977/78. Elevuppföljningar 1982. Studier, förvärvsarbete etc. 4 år efter examen. Examinerade vårterminen 1978

11 Högskolan 1980/81. Elevuppföljningar 1982. Studier, förvärvsarbete etc. 1 år efter examen. Examinerade vårterminen 1981

12 Högskolan. Nybörjare på enstaka kurs höstterminen 1979-studiesituation och sysselsättning under perioden 1979–1981

13 Studerandes ålder hösten 1981

14 Grundskolan och gymnasieskolan 1982/83. Hemspråk och hemspråksundervisning

15 Grundskolan 1982/83. Elever och klasser den 15 september 1982

16 Grundskolan 1979/80. Elevuppföljning våren 1982 bland 16-åringar 1980 avgångna från grundskolan våren 1980 och ej intagna till gymnasieskolan hösten samma år. Studier, förvärvsarbete etc. två år efter grundskolan

17 Högskolan 1977/78–1980/81. Grundutbildning: Examina läsåret 1977/78 t. o. m. läsåret 1980/81

18 Högskolan 1977/78–1980/81. Grundutbildning: Nybörjare läsåret 1977/78 t. o. m. läsåret 1980/81

19 Gymnasieskolan 1981/1982. Elever som avgått från gymnasieskolan med eller utan avslutad utbildning läsåret 1981/82

20 Gymnasieskolan 1982/83. Sökande och intagna höstterminen 1982

21 Högskolan 1982. Forskarutbildning: Registrerade och nybörjare 1982. Definitiva uppgifter avseende vårterminen 1982 och preliminära uppgifter avseende höstterminen 1982

22 Gymnasieskolan 1982/83. Elever den 15 september 1982

Series U 1983 (cont.)

5 Some second and third level education 1981/82. Registrations autumn 1981 and graduations during the academic year 1981/82

6 The state schools for adult education 1981/82. Students, participants and achieved qualifications

7 People's colleges 1982/83. Students in courses of at least 15 weeks given during the autumn 1982

8 Special schools 1982/83. Pupils in special schools and pupils with visual and hearing handicaps in some other schools on the 1st of November 1982

9 Higher education 1978/79–1980/81. Post graduate education: Degrees and lengths of study 1978/79–1980/81

10 Higher education 1977/78. Student surveys 1982. Studies, occupation etc. 4 years after graduation. Graduates in spring 1978

11 Higher education 1980/81. Student surveys 1982. Studies, occupation etc. 1 year after graduation. Graduates in spring 1981

12 Higher education

13 Age distribution of students, autumn 1981

14 Comprehensive school and integrated upper secondary school 1982/83. Home language and home language instruction

15 Comprehensive school 1982/83. Pupils (students) and classes on the 15th of September 1982

16 Comprehensive school

17 Higher education 1977/78–1980/81. Basic education: Graduation during the academic years 1977/78–1980/81

18 Higher education 1977/78–1980/81. Basic education: First year students academic years 1977/78–1980/81

19 Integrated upper secondary school 1981/1982. Students with a leaving certificate and drop-outs during the school year 1981/82

20 Integrated upper secondary school 1982/83. Applicants and admitted autumn 1982

21 Higher education 1982. Postgraduate education: Registration and first year students 1982. Definite figures for spring term 1982 and preliminary figures for autumn term 1982

22 Integrated upper secondary school 1982/83. Students on the 15th of September 1982

Serie U 1983 (forts.)

23 Särskolan 1982/83. Psykiskt utvecklingsstörda barn och ungdomar i särskola m. m. den 1 november 1982
24 Högskolan 1980/81 och 1981/82. Anmälan och tillströmning till högskolans utbildningslinjer höstterminen 1980–vårterminen 1982
25 Gymnasieskolan 1980/81. Elever i kurser som börjat under perioden 16 september 1980–30 juni 1981 eller som både börjat och slutat under perioden 1 juli–14 september 1980
26 Gymnasieskolan 1981/82. Elever i kurser som börjat under perioden 16 september 1981–30 juni 1982 eller som både börjat och slutat under perioden 1 juli–14 september 1981
27 Forskningsstatistik. Teknisk och naturvetenskaplig forskning och utveckling inom företagssektorn 1981
28 Högskolan 1978/79. Elevuppföljningar 1982. Studier, förvärvsarbete etc. 4 år efter examen. Examinerade höstterminen 1978
29 Elevpanel för longitudinella studier. Uppgifter avseende skolgången i årskurs 7 (1981) och årskurs 8 (1982) i grundskolan
30 Högskolan 1981/82. Elevuppföljningar 1982. Studier, förvärvsarbete etc. 1 år efter examen. Examinerade höstterminen 1981
31 Forskningsstatistik 1981/82. Teknisk och naturvetenskaplig forskning och utveckling i offentliga sektorn 1981/82
32 Lärare 1983
33 Forskningsstatistik. Statliga anslag till FoU-verksamhet 1983/84
34 Förteckning över Statistiska meddelanden i serie U 1983

Series U 1983 (cont.)

23 Special schools for mentally retarded 1982/83. Pupils at Special Schools for mentally retarded, 1st of November 1982
24 Higher education 1980/81 and 1981/82. Applicants for higher education and new students autumn 1980–spring 1982
25 Integrated upper secondary school 1980/81. Students in courses which started during the period 16 September 1980–30 June 1981 or started and ended during the period 1 July–14 September 1980
26 Integrated upper secondary school 1981/82. Students in courses which started during the period 16 September 1981–30 June 1982 or started and ended during the period 1 July–14 September 1981
27 Research statistics. Research and development in technology and natural sciences in business enterprice sector 1981
28 Higher education 1978/79. Student surveys 1982. Studies, occupation etc. 4 years after graduation. Graduates in autumn 1978
29 Panel of pupils for longitudinal studies. Data on the school attendance in grade 7 (1981) and grade 8 (1982) in compulsory comprehensive school.
30 Higher education 1981/82. Student surveys 1982. Studies, occupation etc. 1 year after graduation. Graduates in autumn 1981
31 Research and development in technology and natural sciences in general government sector 1981/82
32 Teachers 1983
33 Government R & D funds
34 List of Statistical Reports, Series U 1983

Information i prognosfrågor (IPF) 1983

1 Länens arbetsmarknad 1976–1982
2 Sveriges framtida befolkning. Prognos för åren 1983–2025
3 Arbetskraftsbarometern 1983. Arbetsmarknadsutsikter för vissa utbildningsgrupper hösten 1983
4 Arbete och barn. Kvinnors sysselsättning i de barnafödande åldrarna

Forecasting Information (IPF) 1983

1 The Labour Market by Counties 1976–1982
2 Population projection for Sweden 1983–2025
3 Labour Market Tendency Survey for some Training Categories in the Autumn 1983
4 Work and children. Employment patterns among women in the childbearing ages

Liber/ALLF 41083013
Norstedts Tryckeri, Stockholm 1984

Oi-C

EJ HEMLÅN

STATISTISK TIDSKRIFT

UTGIVEN AV STATISTISKA CENTRALBYRÅN

Ansvarig utgivare: Lennart Nilsson. Redaktör: Thore Grönqvist

Adr: Karlavägen 100. Postadress: 115 81 Stockholm

Telefoner: 08/14 05 60, ankn. 4260 (redaktören personligen)

och 4265 (Helen Lidén)

Prenumeration genom bokhandel eller hos Liber distribution, Prenumerationsorder, 162 89 Stockholm

Utkommer med 5 häften om året

Pris ca 105: – per år eller ca 20: – per häfte (inkl. moms)

Tredje följden Årg. 21 Nr 1 1983

Innehåll

Artiklar

Lars Nordström:
Statistikbehov i framtiden 5

Aage la Cour:
Statistik over landegrænserne 7

Pasquale Cricenti:
Klasser och partival i Sverige – en longitudinell studie om svenska väljare 19

Björn Hägglund och Bo Ranneby:
Riksskogstaxeringen 36

Meddelanden

Europeiska statistiska möten:
Arbetsgruppen för nationalräkenskaper 47
Studiegrupp för livsmedels- och jordbruksstatistik 48
Möte om miljöstatistik 49
Informellt möte om kvalitetsredovisning m. m. (ECE) 50
Möte om utbildningsstatistik och utbildningsindikatorer (UNESCO/ECE) 51
Nordiska statistiska utskott:
Utskottet för nationalräkenskaper (NUNA) 53
Europarådets befolkningskonferens 55

Seminarium om frågekonstruktion ... 56
Expertmöte om miljöstatistik (FN) ... 57
Konferens om dynamiska analysmöjligheter inom inkomststatistiken ... 57
Världskonferens i programmeringsspråket APL ... 57
Nordiskt olycksfallsforskningsseminarium 1982 ... 58
Statistical Computing Project: Table Generation ... 59
PL/1-GUIDE-möte ... 59
Seminarium om kriminalstatistik ... 60
NORDFORSKs kontaktforum för FoU-statistik ... 62
Internationell kartografisk konferens ... 63
Databehandling i statistiken (COMPSTAT 82) ... 64
SCBs yttrande över förslag till ämnesbeskrivning vid prövning av ledigblivande professur i statistik, särskilt offentlig statistikproduktion ... 66
SCBs yttrande över gymnasieutredningens betänkande "En reformerad gymnasieskola" (SOU 1981:96) ... 67
Studieresa till USA för miljöstatistik ... 70
Konsultation vid Ungerns statistiska centralbyrå samt besök i Berlin och Bratislava 71
Studiebesök vid Statistics Canada och Bureau of the Census samt ASAs årsmöte 1982 72
Statistiska centralbyråns styrelse ... 73

Summary in English ... 74

Litteratur

Nyare svensk och utländsk litteratur ... 78
Bokanmälningar:
Ghiselli, E., Campbell, J. and Zedeck, S., Measurement Theory for the Behavioral Sciences (Lars R. Bergman) ... 80
Social rapport om ojämlikheten i Sverige (Joakim Palme) ... 80
Replik till Joakim Palme (Joachim Vogel) ... 83
Social Trends II. (Joakim Palme) ... 83
Hur jämställda är vi? (Tarja-Liisa Leiniö) ... 85
Replik till Tarja-Liisa Leiniö (Lars-Gunnar Andersson) ... 87
Shapiro, S. S. and Gross, A. J., Statistical Modeling Techniques (Tomas Garås) ... 88

STATISTICAL REVIEW

PUBLISHED BY STATISTICS SWEDEN, STOCKHOLM, SWEDEN

Five issues a year

Third series. Vol. 21 No. 1 1983

Contents

Articles

Lars Nordström:
Statistics demands in the future 5, 74

Aage la Cour:
International Statistics 7, 75

Pasquale Cricenti:
Classes and electoral behaviour in Sweden 19, 75

Björn Hägglund and Bo Ranneby:
The Swedish National Forest Survey 36, 76

Notes

European statistical meetings:
- Working party on national accounts 47
- Study group on statistics of food consumption and agriculture 48
- Meeting on environment statistics 49
- Informal meeting on timeliness, cost and quality attributes of statistics 50
- Meeting on educational statistics and indicators (UNESCO/ECE) 51

Inter-nordic statistical committees:
- Committee on national accounts (NUNA) 53

The Council of Europe: Conference on demography 55
Seminar on problems to frame questions for investigations 56
Expert meeting on environment statistics (UN) 57
Conference about the possibility of dynamic analysis of income statistics 57
APL-world conference 57
Nordic research seminar on accidents in 1982 58
Statistical Computing Project: Table Generation 59
PL/1-GUIDE-meeting 59
Seminar on criminal statistics 60

NORDFORSK: Contact forum on FoU-statistics 62
Conference of the International Cartographic Association 63
Data processing in statistics (COMPSTAT 82) 64
Statistics Sweden: Comments on the draft subject description in the consideration of the chair of statistics, in particular public statistics production 66
Statistics Sweden: Comments on the Secondary School Committee report "A revised secondary school" (SOU 1981:96) 67
Study visit to USA for investigation of environmental trends 70
Consultation at the Central Statistical Office in Hungary and visit in Berlin and Bratislava 71
Study visit at Statistics Canada, and US Bureau of the Census, and the annual meeting of the American Statistical Association in 1982 72
The Board of Statistics Sweden 73

Summary in English 74

Literature

Recent Swedish and foreign literature 78
Book reviews:
Ghiselli, E., Campbell, J. and Zedeck, S., Measurement Theory for the Behavioral Sciences (Lars R Bergman) 80
Social rapport om ojämlikheten i Sverige (Joakim Palme) 80
Replik till Joakim Palme (Joachim Vogel) 83
Social Trends II. (Joakim Palme) 83
Hur jämställda är vi? (Tarja-Liisa Leiniö) 85
Replik till Tarja-Liisa Leiniö (Lars-Gunnar Andersson) 87
Shapiro, S. S. and Gross, A. J., Statistical Modeling Techniques (Tomas Garås) 88

Statistikbehov i framtiden

av docent Lars Nordström[1]

Om framtiden vet vi intet påstås en lärd person ha sagt. Sannolikt har han rätt även om många anser sig duktiga på att göra prognoser och förutsägelser.

Vilken statistik behöver vi i ett längre perspektiv? I frågan ligger ett antagande om att man i någon mån måste prioritera mellan olika typer av statistikproduktion. I ett samhälle med obegränsade resurser skulle annars svaret varit "så mycket statistik som möjligt inom så många områden som möjligt". En obegränsad statistikproduktion innebär självfallet att det finns beredskap inför framtiden att kunna svara på i dag inte förutsedda frågor från planerare, forskare, politiker eller andra statistikkonsumenter.

Nu finns inte obegränsade resurser och inte heller under 1980-talet kommer det att vara möjligt med en mycket omfattande statistikproduktion. En prioritering är därför nödvändig och denna måste självfallet baseras på en uppfattning om hur den framtida statistikfrågan kommer att se ut.

Den centrala frågan blir därför att söka klarlägga vilka frågeställningar kommer de framtida statistikkonsumenterna att söka besvara.

För att ge en belysning åt denna fråga vill jag dela in samhällsutvecklingen efter kriget i några faser och utifrån denna indelning blicka framåt.

Tidsperiod	Benämning	Statistikbehov
1945–1960	Oredans tidsålder	Litet statistikbehov
1960–1970	Planeringens tidsålder	Stort behov av underlagsmaterial, tidsserier, ekonomiska variabler
1970–1980	Ruelsens tidsålder	Sociala variabler, krisstatistik
1980–1990	Förändringens tidsålder	?

Uppställningen syftar till att åskådliggöra att varje tidsålder och dess karaktär kräver sin speciella statistik.

Under den stora planeringens tidsålder på 1960-talet behövdes stora mängder tidsseriedata eftersom alla förutsatte en fortsatt trendmässig tillväxt. Tidsserier kunde användas för att motivera den utbyggnad av samhället som ansågs säkerlagd.

Under 1970-talet bröts de flesta kurvor och en mängd problem – inte minst sociala – som tidigare sopats under mattan dök upp till ytan. För att lösa de då aktuella frågeställningarna gav tidsserier inte stor hjälp utan helt nya statistikbehov uppstod. Vi började bl. a. få en snabb ökning av enkäter och SIFO-undersökningar.

Nu under 1980-talet befinner vi oss i ett läge där vi med säkerhet vet att

- den ekonomiska tillväxten blir svag
- stora strukturförändringar kommer att aktualiseras

[1] Kulturgeografiska institutionen, Göteborgs universitet.

- Sverige blir allt mer styrt utifrån i en allt mer komplex värld
- de sociala problemen kommer att öka.

Denna nya framtidsbild kommer att ställa delvis helt nya krav på statistikproduktionen.

Jag skall i korthet beröra några sådana krav. Fler torde finnas.

1. 80-talet kommer att bl. a. kännetecknas av snabba inte alltid förutsedda händelser typ uppkomsten av nya krisbranscher. För att åtgärda dessa kriser krävs snabbt tillgängligt statistiskt material vars karaktär inte går att i förväg förutse. Det krävs följaktligen en beredskap för snabb datainsamling, produktion och sammanställning av data.
2. Ett stort problemområde under hela 80-talet kommer sysselsättningen att vara. Statistik som snabbt och rättvisande visar var nya sysselsättningskriser uppstår, vilka åtgärder som leder till positiva resp. negativa effekter och över huvud taget hur den svenska arbetsmarknaden fungerar kommer att vara av stor betydelse.
3. Sveriges allt mer omfattande inlämmande i en internationell företagsvärld har hittills skett utan att vi kunnat följa processen på ett tillräckligt säkert sätt. Inför framtiden kommer kraven att öka på statistik över handelsströmmar, ägarkonstellationer över nationsgränser, internleveranser inom multinationella företag etc. För att kunna påverka den svenska utvecklingen krävs således kunskaper om vad som är påverkbart i Sverige och vad som är påverkan utifrån.
4. Trots ett av världens högsta materiella välstånd kommer stora grupper i det svenska samhället att ha betydande sociala problem. Statistik som visar det svenska folkets hälsoläge i ett brett perspektiv kommer därför att visa sig värdefull. Detta gäller såväl utifrån planerares som forskares synpunkt.

De nu nämnda områdena innebär i sig en prioritering gentemot dagsläget. Ökar inte resurserna till statistikproduktion måste rimligtvis andra delar av bl. a. SCBs verksamhet skäras eller läggas ner.

En sådan omprioritering är ju på intet sätt något nytt utan ligger inom ramen för de aktiviteter som SCB regelmässigt gör. Det nya skulle möjligen vara att de hårda ekonomiska villkoren kommer att leda till mer drastiska förändringar än vad man tidigare varit van vid.

Statistik over landegrænserne

af Aage la Cour[1]

1. Indledning

Det statistiske samarbejde på tværs af landegrænserne går langt tilbage i tiden. Det gælder globalt, og det gælder regionalt, hvor navnlig samarbejdet indenfor Europa og specielt mellem de nordiske lande har været intenst gennem mange år.

Om tre år er det hundrede år siden, det Internationale statistiske institut (ISI) blev stiftet i Haag, og i 1985 fejres hundredåret for stiftelsen netop i Haag, som et led i den almindelige regel i ISI om at have et internationalt statistisk møde hvert andet år. Det er ganske imponerende, at dette møde i 1985 bliver møde nr. 45 i løbet af de 100 år, og det på trods af afbrydelser på grund af krige eller andre hindringer i perioder med begrænsninger i internationalt samarbejde. Interessen for disse møder er betydelig, da der normalt er flere hundrede deltagere.

Det nordiske statistiske samarbejde har som bekendt været et integreret led i de nordiske institutters virksomhed gennem mange år. Det første nordiske chefstatistikermøde afholdtes allerede i 1889[2], og det netop afviklede nordiske chefmøde er det 30te i rækken siden 1889. Som noget formentlig helt specielt for Norden, er der endog regelmæssigt samarbejde mellem de statistiske kontorer, der er knyttet til hovedstædernes eller de større byers administrationer og tillige mellem de nordiske arbejdsgiverorganisationers statistiske afdelinger. Hertil kommer så selvfølgelig samarbejdet mellem de statistiske foreninger i de nordiske lande, der står som arrangører af disse nordiske statistiske møder, og som nu gennem en længere årrække har været gennemført med stor deltagelse hvert 3die år i tilknytning til chefmøderne. Det har altid været berigende at deltage i disse møder, og jeg er meget glad for, at arrangørerne har bedt mig sige noget her om statistik over landegrænserne.

2. Strukturen i det internationale statistiske samarbejde

2.1. Globalt

Det overordnede og samlende organ i det internationale samarbejde på det praktiske statistiske område er selvsagt FN's Statistical Commission, der mødes i New York hvert andet år. Statistical Commission sorterer under FN's økonomiske og sociale råd, og kommissionens ca 25 medlemmer vælges efter særlige regler, der skulle sikre repræsentation fra de forskellige grupper geografisk, udviklingsmæssigt, øst–vest osv. (Se fig. 1.)

I den sidst valgte Statistical Commission, repræsenterer Finland de nordiske lande, og herudover er Vesteuropa for tiden repræsenteret af Frankrig, England, Irland, Østrig og Spanien. Bemærkelsesværdigt er det, at USA ikke er valgt medlem i denne periode. Det er første gang USA ikke er medlem, og det forekommer temmelig uheldigt af mange grunde, specielt på baggrund

[1] Generaldirektør i De europæiske Fællesskabers statistiske kontor "EUROSTAT", Luxembourg.
[2] Se Nordisk statistisk skriftserie 15. Det nordiske Sjefsstatistikermöte i Oslo 1967. Oslo 1968.

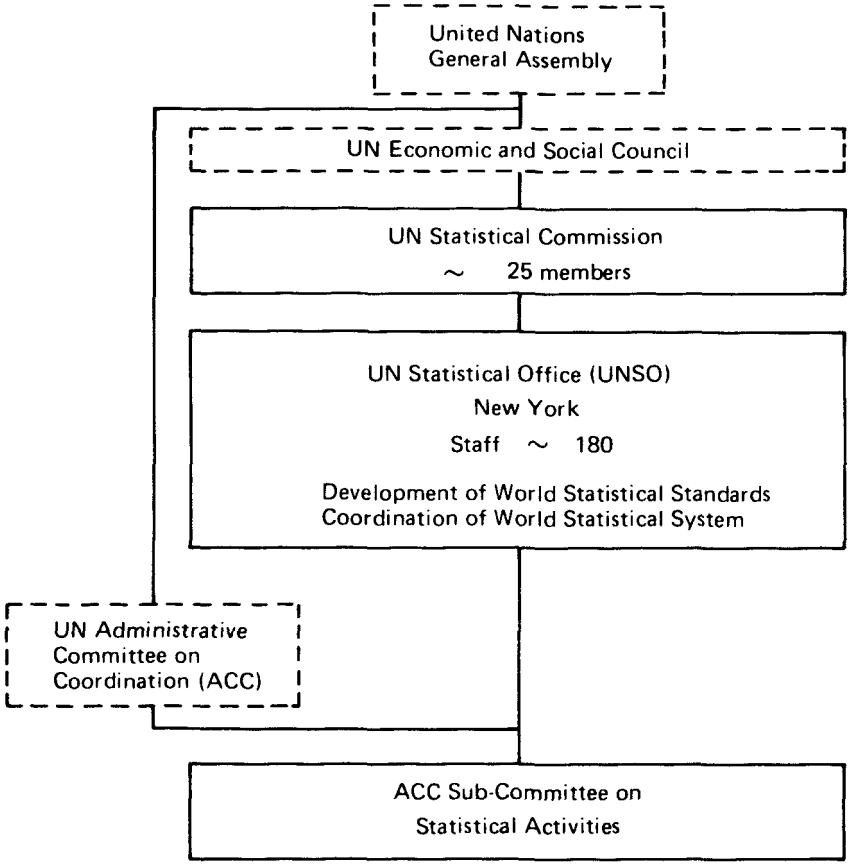


Fig. 1. World Statistical System

af den betydelige indsats, der gennem år er ydet af navnlig chefen for det organ, der indtil for nylig har stået for den overordnede koordinering af USA's centrale statistiske arbejdsopgaver og ansvarlig for USA's samarbejde med andre lande på det statistiske område. Han var endog formand for Statistical Commission indtil det nye valg, der fandt sted i 1981 gældende for perioden 1982–83. Statistical Commission's næste møde er fastlagt til marts 1983, og det bliver Kommissionens 22. møde.

Udover de valgte medlemmer deltager også repræsentanter fra de mange statistiske kontorer i FN's forskellige organer, herunder fra de regionale kontorer – eller Economic Commissions; ECE (Europa) ECA (Afrika), ECLA (Latinamerika), ECWA (Vestasien), ESCAP (Asien og Stillehavets lande). Endvidere fra FN-familien i øvrigt, såsom FAO, ILO, UNESCO, WHO, World Bank, IMF, UNICEF, UNCTAD, UNIDO, WTO, osv. (se fig. 2). Hertil kommer repræsentanter fra andre statslige

UN Statistical Office (UNSO)

Regional offices

ECE-Europe
ECA-Africa
ECLA-Latin America
ECWA-West Asia
ESCAP-Asia and Pacific

Agencies etc.

FAO (Food and Agriculture Organization)
ILO (International Labour Organization)
UNCTAD (Conference on Trade and Development)
World Bank
IMF (International Monetary Fund)
UNESCO (Educational, Scientific and Cultural Organization)
UNICEF (Children's Emergency Fund)
WHO (World Health Organization)
UNIDO (Industrial Development Organization)
WTO (World Tourism Organization)
etc.

Fig. 2. The UN-Family (within Statistics)

organisationer bl. a. GATT, og Det internationale toldsamarbejde, samt EF repræsenteret ved Eurostat, endvidere OECD og CMEA, den østeuropeiske samarbejdsorganisation. Yderligere repræsentanter fra ikke statslige organer såsom ISI og andre. Endelig deltager et antal observatører fra forskellige ikke direkte valgte lande, således deltog Kina i 1981 for første gang i Statistical Commission's møde som observatør, og redegjorde der for de enorme problemer de stod overfor ved forberedelse til en folketælling i dette verdens folkerigeste land. Dette så meget mere som Kinas statistiske kontorer blev nedlagt under kulturrevolutionen i slutningen af 1960'erne, og nu fra slutningen af 1970erne forsøges genetableret. Nedlægningen fulgtes endog op med nedlægning af statistikundervisningen på læreanstalterne, hvorved genetableringen er blevet ekstra kompliceret. Som et beskedent bidrag til denne genetablering har vi i Eurostat i Luxembourg fornylig haft tre kinesere fra Kinas toldvæsen, der gennem nogle måneder studerede det europæiske udenrigshandelsstatistiske system.

FN's Statistical Commission har til opgave at koordinere medlemslandenes statistik. Statistical Commission har ingen særlige beføjelser til at pålægge landene harmonisering, men i samarbejdet med medlemslandene udfører FN gennem Statistical Commission og FN's statistiske kontor et omfattende arbejde på adskillige områder, specielt indenfor det internationale nomenklatursamarbejde, og i det hele taget i udviklingen af fælles internationale statistiske regler.

Det centrale statistiske FN organ, der også virker som sekretariat for Statistical Commission er FN's statistiske kontor i New York (UNSO), der har et samlet personale på ca 180 personer. Det er her trådene samles, specielt samarbejdet indenfor FN-familiens organer og med andre internationale organisationer, altsammen efter retningslinier fastlagt af Statistical Commission. (Se fig. 2.) Dette kontor (UNSO) har som bekendt haft nordmanden Svein Nordbotten som chef gennem de sidste tre år indtil 1. august 1982, og før ham, gennem en længere årrække canadieren Simon Goldberg.

Som et led i koordineringsarbejdet indenfor FN's mange institutioner og agencies virker United Nations Administrative Committee on Coordination (ACC). Under denne komité er der The ACC Sub-Committee on Statistical Activities (se fig. 1). Medlemmer er repræsentanter fra de forskellige statistiske kontorer indenfor FN-familien – inklusive de regionale kontorer. I de senere år har også Eurostat og OECD deltaget i disse møder, der afholdes en gang om året skiftende mellem FN-organernes forskellige hovedkontorer. Denne ACC Sub-Commit-

tee forestår koordineringen mellem Statistical Commission's møder og medvirker i øvrigt ved forberedelsen af Kommissionens møder. Til dette sidste formål virker endvidere en mindre samarbejdsgruppe på en halv snes personer, hvoriblandt også repræsentanter fra OECD og Eurostat.

Som det vil ses, er dette samarbejde mellem de internationale organisationer og over landegrænserne lagt i ganske faste rammer, med mange kontakter mellem de forskellige organer. Det har udviklet og stabiliseret sig betydeligt i de senere år, bl.a. på grund af et stærkt stigende behov for koordinering og samarbejde i en tid, hvor det er blevet stadigt vanskeligere at imødekomme det stigende behov for statistiske oplysninger med ofte ganske betydelige begrænsninger i de midler, der er til rådighed. Dobbeltarbejde m.v. må derfor søges undgået, hvilket samarbejdet mellem landene kan medvirke til. For de mere teknisk udviklede lande er det da også væsentligt, at OECD og EF's Eurostat er blevet mere og mere inddraget i dette samarbejde specielt på det, der måske kan kaldes sekretariatsplan.

Ved siden af dette globale samarbejde på det praktiske statistiske plan, er der, som allerede nævnt, et omfattende samarbejde gennem International Statistical Institute (ISI), der også omfatter Bernoulli Society for Mathematical Statistics and Probability, International Association of Survey Statisticians, International Association for Regional and Urban Statistics and International Association for Statistical Computing. ISI har sine større møder eller sessions hvert andet år, næste gang i 1983 i Madrid og derefter i 1985 i Haag, hvor hundredåret for stiftelsen, som anført, vil blive markeret. En meget væsentlig del af aktiviteterne indenfor ISI med tilsluttede organisationer er koncentreret om hovedsageligt teoretiske problemer og stærkt knyttet til universiteterne og research-arbejdet rundt om i landene. Medlemsskab af ISI er i stor udstrækning personligt, idet der kræves en særlig anbefalelses- og godkendelsesprocedure forinden optagelsen. Cheferne for de nationale statistiske institutioner m.fl. er dog "fødte" medlemmer.

Blandt det meget store antal spørgsmål, der behandles på ISI's sessions har også været spørgsmålet om, hvorledes afstanden mellem universitetsstatistikere og praktiske statistikere kunne formindskes. Udviklingen har mange steder været den modsatte, uagtet der er et stort behov for at slå bro mellem de to altfor adskilte verdener. Men selv om dette emne har været stærkt fremme på de sidste sessioner, synes der ikke at være kommet så meget ud af det, hvilket i høj grad må beklages.

Et andet internationalt forum, hvor samarbejdet om statistiske problemer er dominerende, er International Association for Research into Income and Wealth. Denne organisation IARIW har møde hvert andet år, næste gang i 1983 i Luxembourg med Eurostat som vært.

2.2. Regionalt samarbejde

På det regionale plan er samarbejdet indenfor Europa det mest udviklede. (Se fig. 3.) Her er det igen FN, der gennem sit hovedkontor i Genève er det centrale, og under ECE (Economic Commission for Europe) har Den europæiske chefstatistikerkonference netop i år kunnet fejre sin beståen gennem 30 år (se fig. 3). Peter Jakob Bjerve, der i sin tid var med blandt stifterne, har præsenteret dette samarbejdes betydning i den netop i anledning af 30-året af ECE udsendte nye kvartals-publikation[1]. Jeg

[1] Statistical Journal, Vol. 1. Nr. 1, June 1982, ECE Genève.

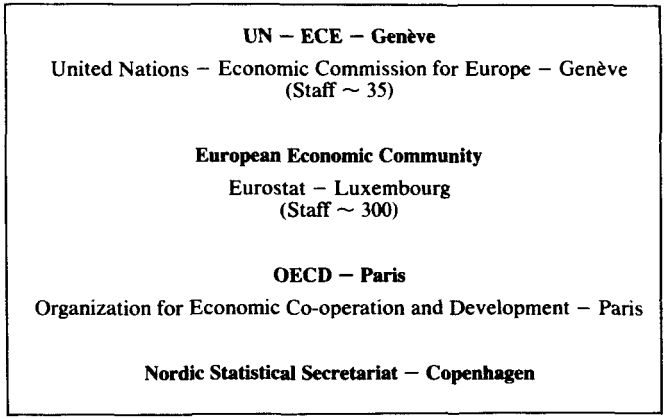


Fig. 3. The European Statistical System

skal ikke her gentage, hvad Bjerve så klart har beskrevet, men anbefale Bjerves artikel til interesserede.

Den europæiske chefstatistikerkonference holdes hvert år i den sidste fulde uge i juni i Genève, og der drøftes de aktuelle samarbejdsopgaver og arbejdsprogram for de kommende år. I disse chefmøder er der normalt særdeles aktiv deltagelse fra repræsentanter fra USA og Canada, samt selvsagt fra ILO, FAO, UNESCO og andre FN-agencies. Endvidere deltager repræsentanter fra Eurostat, OECD og CMEA. Under dette ECE-samarbejde virker en række arbejdsgrupper indenfor de forskellige specielle statistikområder. Som sekretariat virker ECE's statistiske kontor i Genève, der har ca 35 medarbejdere i alt. Denne stab forestår såvel sekretariatsarbejdet, som udgivelsen af kontorets statistiske publikationer m. m.

Medens FN-samarbejdet under ECE i Genève selvsagt dækker både Øst- og Vesteuropa, med som nævnt deltagelse også fra USA og Canada, er det statistiske samarbejde indenfor OECD begrænset til OECD's i alt 24 medlemslande, samt Jugoslavien, som ikke er fuldt medlem. Det statistiske samarbejde indenfor denne medlemskreds kan vel siges at være begrænset til et mere snævert område indenfor den praktiske statistik, hovedsageligt omfattende de mest centrale økonomiske problemer, herunder også energiproblemer, der behandles i nært samarbejde med Det Internationale Energiagentur (IEA).

For såvel FN-samarbejdet som indenfor OECD, er hovedformålet at medvirke til at fastlægge så vidt muligt ensartede retningslinier i det praktiske statistiske arbejde, og søge at formå landene til at følge disse retningslinier. Samarbejdet må derfor i følge sagens natur bygge på frivillighed.

Indenfor det Europæiske Økonomiske Fællesskab (EF), der nu tæller 10 medlemslande, har samarbejdet derimod en langt stærkere struktur. Her bygger den praktiske statistik i vid udstrækning på direktiver og forordninger, der er bindende for landene, og hovedsigtet her har været at harmonisere medlemslandenes statistik, for derved at skabe mulighed for sammenligninger mellem landene på harmoniseret grundlag. Dette må ses i lyset af, at EF-

statistikken i stor udstrækning skal dække EF-administrationens krav, på samme måde som den nationale statistik skal kunne anvendes i den nationale administration, medens kravene til andre internationale organisationers statistik er mindre regoristiske end i EF-samarbejdet.

EF's statistiske kontor Eurostat, eller SOEC, har ialt ca 300 medarbejdere, og samarbejdet med landene er tilrettelagt gennem to årlige chefstatistikermøder, og herudover gennem i alt ca 70 arbejdsgrupper, der tilsammen holder mellem 90 og 100 arbejdsgruppemøder om året, i næsten alle tilfælde med deltagelse fra samtlige 10 medlemslande. Herudover deltager Spanien og Portugal nu i stigende grad i disse møder. I chefmøderne har Spanien og Portugal ligeledes deltaget gennem det sidste par år. Endvidere er ECE og OECD repræsenteret på Eurostats chefmøder[1]. Til chefmøderne udarbejdes alle arbejdspapirer m.v. på fire sprog, ligesom der er simultantolkning til og fra fire sprog på møderne, medens arbejdsgruppemøderne normalt har tolkning til og fra syv sprog og med dokumenter på alle syv fællesmarkedssprog.

Udover dette meget intense samarbejde indenfor EF med deltagelse fra nu 12 af Vesteuropas lande, har Eurostat et omfattende samarbejde med de "tredielande", der i øvrigt har en eller anden form for tilknytning til eller samarbejde med EF. Det gælder i særlig grad de såkaldte AVS-lande, d.v.s. lande i Afrika, Carribien og Pacific. Dette samarbejde går såvel på praktisk statistik specielt på udenrigshandelsområdet, som på uddannelse. Således forberedes for tiden et træningsprogram for statistikere i 37 afrikanske lande, hvortil EF har bevilget betydelige midler med det formål at opbygge den praktiske statistik i disse mange lande. Hertil kommer, at der allerede gennem mange år har været adgang for statistikere fra AVS-lande til efteruddannelse på forskellige træningscentre der nærmest må karakteriseres som højskoler. Eksempelvis findes der et sådant center i München, der normalt har to større kurser hvert år, af hver 4 måneders varighed, alternerende på henholdsvis fransk og engelsk. Dette center samt et center i Paris og andre centre i Afrika, styres af et særlig råd bestående af bl.a. statistikcheferne fra EF's medlemslande, og finansieres i et samarbejde mellem EF, et særligt fond i Tyskland og INSEE, det centrale franske statistiske institut, der tillige er uddannelsesinstitut på universitetsniveau for økonomer og statistikere såvel fra Frankrig som fra de fransktalende AVS-lande.

Om dette vigtige efteruddannelsesarbejde for udviklingslandenes statistikere, hvor mange er udgået fra europæiske læreanstalter, gælder, at der her gives en undervisning, som i vid udstrækning forener det praktiske behov i udviklingslandene med den nødvendige basisteori. Det kan nok vække til eftertanke, at udviklingslandenes statistikere har denne mulighed, medens vi i Europa ikke har været i stand til at opbygge noget tilsvarende. Jeg mener, at der i høj grad er behov herfor ikke mindst som et led i bestræbelserne på at slå bro mellem praktikere og teoretikere.

Institut National de la Statistique et des Etudes Economiques (INSEE) i Paris er jo en af de berømte institutioner i den statistiske verden, og central i det franske ud-

[1] I øvrigt kan henvises til: Statistisk tidskrift 1979: 5. Policy and recent developments in Eurostat—methods of disseminating statistics. Endvidere: Statistisches Amt Mitteilungsblatt, Heft Nr. 44, December 1981 udgivet af Oesterreichische Gesellschaft für Statistik und Information: The Statistical System in the European Economic Community. Begge artikler af Aage la Cour.

dannelsessystem for såvel økonomer som statistikere, med normalt omkring 600 studerende. Her er etableret et frugtbart samvirke mellem teori og praksis. Den nuværende chef er Edmund Malinvaud, hvis lærebøger er kendt stof blandt statistikere og matematikere, ligesom hans forgænger Jean Ripert er en kendt international personlighed, nu vicegeneralsekretær i FN.

Det regionale samarbejde på nordisk plan sådan som det nu efterhånden indgår også som et led i aktiviteterne omkring Nordisk Råd er selvfølgelig af særlig interesse for nordiske statistikere. Det er på den anden side så velkendt for dette mødes deltagere, at jeg vil begrænse mig til at henvise til det omfattende materiale der foreligger herom bl.a. i Nordisk statistisk skriftserie, og i publikationerne og notaterne, der udgår fra Nordisk Statistisk Sekretariat. Dette sekretariat er jo ikke blot sekretariat for de nordiske chefstatistikermøder, men også aktivt i tilrettelæggelsen af møderne i de 15 nordiske statistikerudvalg, hvis mødeprogram fastlægges af cheferne på chefmøderne, og normalt omfatter 10–15 møder hvert år. Herudover forestår sekretariatet i samarbejde med landene tilrettelæggelsen af seminarer og andre aktiviteter i det nordiske statistiske samarbejde.

Yderligere må nævnes det statistiske samarbejde indenfor de østlige lande. Dette samarbejde er en del af Council for Mutual Economic Assistance (CMEA).

Indenfor det videregående regionale statistiske samarbejde kan også nævnes Coordinating Working Party on Atlantic Fishery Statistics, som på initiativ af FAO i 1959 blev etableret med henblik på at koordinere og videreudvikle den fiskeristatistik, der er nødvendig for de internationale organer, som arbejder med de atlantiske fiskeriproblemer. Foruden FAO og andre organer med FN status deltager også OECD og Eurostat, der siden 1976 har været fuldt medlem af arbejdsgruppen. Den sidst afholdte session, dvs. den 11te session, fandt sted i Luxembourg, hvor Eurostat var vært.

Om regionalt statistisk samarbejde kan endelig nævnes, at der også indenfor Europarådet's regi finder en vis koordinerende virksomhed sted, hovedsageligt koncentreret om befolkningsstatistiske problemer. Et egentligt struktureret samarbejde på organisationsplan mellem Europarådet, der nu har 29 medlemslande, og de øvrige nævnte internationale organisationer er ikke etableret. Europarådet blev som bekendt etableret i Strassbourg i 1949, eller 3 år før den europæiske kul- og stålunion, som i 1958 blev til det europæiske økonomiske Fællesskab EF, oprindelig med seks medlemmer i 1973 udvidet til ni og i 1981 til ti.[1]

3. Indholdet i det internationale statistiske samarbejde

Der skal ikke her gøres noget forsøg på at give en udtømmende beskrivelse af indholdet i det internationale statistiske samarbejde. En del er fremgået af det allerede sagte, men nogle af de mere væsentlige områder indenfor samarbejdet på det praktiske plan skal nævnes. Det gælder nomenklatur- og klassifikationsområdet, nationalregnskab, verdenshandlen, befolkning- og boligstatistik og de igangværende projekter vedrørende sammenlignelig pris- og levestandardsundersøgelser. Indenfor de givne rammer kan disse hver især meget store problemkomplekser kun blive overfladisk berørt.

[1] Oprindelig Frankrig, Tyskland, Italien, Holland, Belgien og Luxembourg, derefter England, Irland og Danmark og endelig også Grækenland.

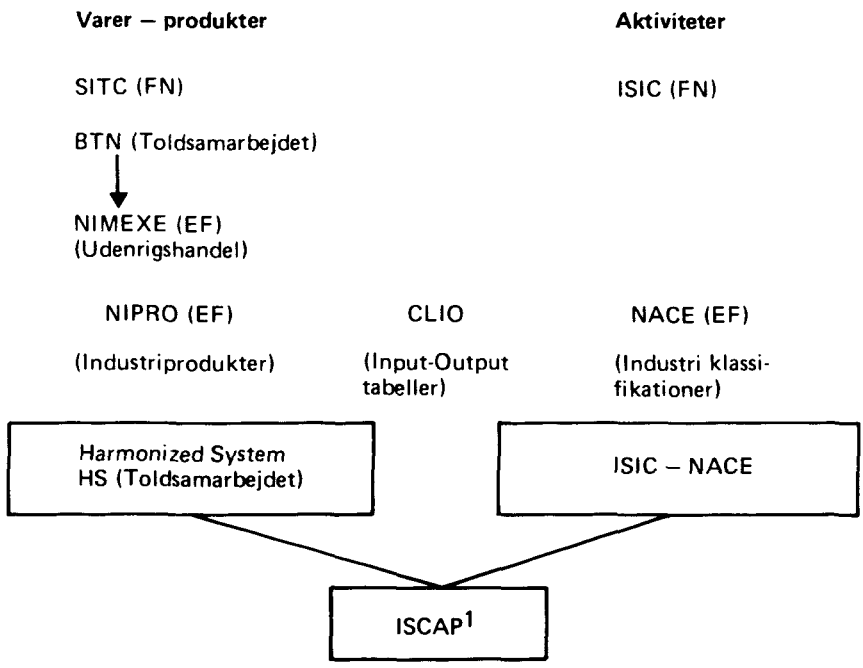


Fig. 4. Klassifikationer

Ensartede statistiske nomenklaturer er selvsagt helt centralt i det internationale samarbejde. Allerede på det første nordiske chefmøde i 1889 var varenomenklaturen i udenrigshandelsstatistikken og ensartede stillingsklassifikationer i folketællingsopgørelserne på dagsordenen. På globalt plan er FN's klassifikationer i dag de centrale. Det gælder varenomenklaturen, Standard International Trade Classification (SITC) og på aktivitetsområdet International Standard Industrial Classification of all Economic Activities (ISIC). Endvidere stillingsklassifikationen International Standard Classification of Occupations (ISCO) samt nationalregnskabssystemet, System of National Accounts (SNA), for at nævne de mest centrale.

Så enkelt er det imidlertid ikke i virkligheden. Dels er disse centrale nomenklaturer under stadig revision, dels er der utallige andre mere eller mindre afvigende nomenklaturer indenfor de samme områder.[2] Det skaber selvsagt betydelige vanskeligheder.

Af særlig betydning har her været det vanskelige ved at udarbejde nomenklaturer, der er anvendelige i lande på vidt forskelligt udviklingstrin. Dette var da også baggrunden for, at EF i de første pionerår i

[1] ISCAP: Koordineret vare- og aktivitetsklassifikation

[2] Der kan henvises til en af Statistisches Bundesamt, Wiesbaden i 1981 udgivet: Verzeichnis wichtiger internationaler Abkürzungen, som indeholder et afsnit på 15 sider om klassifikationssystemer.

1950erne og 1960erne kastede sig ud i at skabe specielle nomenklaturer med henblik på at dække de særlige behov, som EF-strukturen syntes at nødvendiggøre. Dermed har vi fået et EF-nationalregnskabssystem, European System of Integrated Economic Accounts (ESA) ved siden af FN's SNA og østlandenes særlige system Market Product System (MPS). På vareområdet tilkom oprindelig Bruxellesnomenklaturen (BTN) udarbejdet indenfor det internationale toldsamarbejde, men ikke fuldt ud harmoniseret med FN. BTN er siden suppleret med Nomenclature of Goods for the External Trade Statistics of the Community and Statistics of Trade between Member States (NIMEXE). På aktivitetsområdet tilkom Nomenclature générale des activités économiques dans les Communautés européennes (NACE) (se fig. 4).

Dette er en meget summarisk skitse, der dog skulle give indtryk af hovedproblemerne, som snart skulle vise sig større end det formentlig var forudset. På denne baggrund nedsattes i sidste halvdel av 1970erne en uformel samarbejdsgruppe efter aftale mellem UNSO i New York og Eurostat. Gruppens opgave er at udarbejde forslag til harmonisering indenfor vare- og aktivitetsklassifikationerne. Endvidere at søge opstillet et system, således at der på et mere aggregeret niveau kan sikres korrespondance mellem vare- og aktivitetsklassifikationernes hovedgrupperinger. Arbejdet er ret vidt fremskredet, men ventes dog ikke at kunne fremlægges for Statistical Commission tidligere end ved Kommissionens møde i 1985.

Yderligere udfoldes der indenfor varenomenklaturområdet en betydelig aktivitet på at tilvejebringe en klassifikation, som skulle kunne dække hele det internationale behov på området. Dette arbejde foregår under ledelse af det internationale toldsamarbejde og vil resultere i et såkaldt Harmonized Commodity Description and Coding System (HS), der forudses iværksat fra 1985. Den endelige behandling af dette system kan dog ikke nå frem til drøftelse og vedtagelse før Statistical Commission mødes i 1985, og det må forudses at systemet ikke bliver benyttet i FN's medlemslande førend i løbet af sidste halvdel av 1980erne.

Indenfor nationalregnskabsstatistikken er der ligeledes bestræbelser i gang på at nærme systemerne mere og mere til hinanden, gennem samarbejde mellem UNSO, OECD og Eurostat, samt ECE. Hovedsigtet er at bringe overensstemmelse mellem SNA og ESA, men der stiles tillige mod at opnå en koordinering til østlandenes system, MPS.

Det kan næppe undre, at disse klassifikationsproblemer er særdeles komplicerede og tidkrævende, samtidig med, at det er meget vanskeligt at skaffe tilstrækkelig expertise til dette arbejde, der selvsagt må foregå over lange tidshorisonter. Også stillingsklassifikationer og uddannelsesgrupperinger indebærer betydelige problemer, som trænger sig på, bl. a. på grund af de hastigt skiftende forhold på disse områder.

For de internationale varebevægelser er varenomenklaturen selvsagt essentiel. Men også harmonisering af udenrigshandelens begreber er vigtig. Det drejer sig foruden om en ensartet landenomenklatur, også om ensartede regler m. h. t. værdiangivelser, f. o. b. eller c. i. f., vægt med eller uden emballage eller mængdeangivelse, oprindelses- og bestemmelsessted, eller produktions- og forbrugsland m. m.

For befolknings- og boligområdet er det klassiske internationale problem knyttet til såvel tidspunkt for tællingernes afholdelse, som vedrørende ensartede opdelinger i de endelige tabeller. Her er planlægningen

også vigtig, hvis tællingerne skal forventes afholdt i flest mulige lande indenfor en begrænset interval omkring de internationalt anbefalede tællingsår.

Ved siden af de her nævnte mere traditionelle problemer må nævnes det meget ambitiøse program, FN's Statistical Commission nu gennem flere år har tumlet med, således som det er styret af UNSO, med støtte af specielt engagerede experter. Nemlig National Household Survey Capability Programme (NHSCP), der i korthed går ud på at inddrage flest mulige udviklingslande i Household Surveys, hvorved det tillige skulle være muligt at engagere udviklingslande i en statistisk koordineret opgave til gavn for landene selv og i mange tilfælde som løftestang for udvikling af det statistiske apparatur i lande med en mere eller mindre svag national statistik. Adskillige af FN's øvrige organer såsom FAO, ILO og World Bank medvirker aktivt i dette arbejde. EF's statistiske kontor og EF-Kommissionen i øvrigt er også gået ind i dette arbejde med finansiel støtte til flere af de implicerede udviklingslande.

Et andet betydeligt, men mindre projekt, er International Price Comparison Programme (ICP). Det var vel egentlig dette program, der på et tidligt stade gav anledning til det mere vidtfavnende Household Survey Programme, der dermed kom mere i forgrunden, da dette sidste nok må siges at dække langt flere generelle behov end ICP.

Prissammenligningsprogrammet er dog efter min opfattelse af ganske betydelig interesse, idet det går ud på at nå frem til sammenligninger mellem de forskellige landes leveniveauer på realøkonomisk grundlag. Det er velkendt, at sammenligninger mellem lande er foretaget ved at omregne nationalproduktet pr. indbygger i den nationale valuta til $ efter gældende vekselkurs. Det er evident, at disse resultater oftest er mere vild- end vejledende, specielt i perioder med stærkt varierende valutakurser. Eksempelvis kan nævnes at USA, der normalt betragtes som verdens rigeste land med hensyn til gennemsnitsindtægt, pludselig fra et år til et andet røg langt ned på rangstigen, kun fordi dollaren stod svagt i dette år. For at afbøde på dette har UNSO og Eurostat gennem adskellige år samarbejdet på dette område, specielt med hensyn til de mere teoretiske aspekter. Jeg skal her begranse mig til at nævne at Eurostat siden 1979 årligt har publiceret resultaterne af sådanne realøkonomiske sammenligninger dækkende EF's medlemslande tilbage til 1975.[1] Der forestår imidlertid nu offentliggørelse af lignende sammenligninger dækkende flere lande, hvilket sker i nært samarbejde mellem Eurostat, ECE og OECD samt UNSO. I dette samarbejde optræder Østrig som bindeled ved sammenligninger mellem Øst- og Vesteuropa.

Problemerne på dette område er betydelige såvel praktisk som teoretisk, og lykkeligvis er der stor interesse for dette projekt i begge lejre. Der synes dog at være nogen skepsis blandt nogle af udviklingslandenes repræsentanter i Statistical Commission, formentlig hovedsagelig af mere realøkonomiske årsager. De traditionelle sammenligninger på grundlag af $ vil i almindelighed vise langt større forskelle mellem udviklede og mindre udviklede lande, end sammenligninger på realøkonomisk grundlag.

4. Fremtidsperspektiver og konklusion

Set ud fra udviklingen i det praktiske statistiske samarbejde på internationalt plan er der stærkt øget forståelse for nødvendighe-

[1] Se bl. a. National Accounts ESA-Aggregates, 1960–1980, publiceret af Eurostat i 1982.

den af fælles retningslinier og fælles eller koordinerede klassifikationer. Det stadig øgede internationale økonomiske samkvem øger kravene til den bedst mulige sammenlignelighed i den tilgængelige statistiske information. Der er nået ret meget med henblik på at samordne de internationale organers krav til landene, så behovene i de internationale organer mere enkelt end tidligere kan dækkes gennem samordnede aktiviteter. Den stigende computerisering med øget brug af databaser eller andre computeriserede informationssystemer har generelt lettet dataformidlingen mellem landene indbyrdes og i kommunikationen med eller mellem de internationale organisationer.

De videre perspektiver i så henseende er lovende, så meget mere som den gensidige forståelse i det internationale statistiske samarbejde synes at være blevet væsentligt stimuleret i de senere år. Chefstatistikere og andre ledende folk i dette arbejde mødes hyppigt; i Europa adskillige gange om året, således at udviklingen kan styres og tilrettelægges for længere eller kortere perioder frem i tiden. Dette er i nogen grad under pres af nødvendigheden, idet begrænsede ressourcer og stadige nedskæringer gør en koordineret prioritering mere og mere påkrævet. At chefstatistikerne lægger vægt på betydningen af det internationale samarbejde, ses da også af, at cheferne kun sjældent undlader at deltage i de forskellige arrangementer på chefplan.

Det må imidlertid forudses, at det kan blive stadig mere kompliceret at få tilvejebragt det nødvendige basismateriale for den statistiske informationstjeneste. Dels må der på flere områder noteres en stigende uvilje og modstand mod at besvare statistiske spørgeskemaer, specielt for så vidt angår erhvervsstatistikken. Her er det navnlig de mange mindre virksomheder, der føler dette arbejde byrdefuldt og uden større værdi for dem selv. På individstatistikkens område er vanskelighederne vel nok væsentligt mindre, men store totale folke- og boligtællinger har dog i nogle af de lande, hvor de fortsat forekommer helt nødvendige, stødt på modstand. I Vesttyskland således navnlig på grund af de store omkostninger, der er forbundet med sådanne tællinger, men i nogen grad også på grund af uvilje i dele af befolkningen. I Holland er det hovedsagelig modvilje i befolkningen, der resulterede i, at en ellers planlagt folke- og boligtælling i 1980 er helt opgivet. På denne baggrund er interessen steget for at finde andre udveje til at tilvejebringe det statistiske materiale, der i mange henseender er nødvendigt for de lovgivende og administrative organer.

Den stigende anvendelse af administrative registre giver nogen mulighed for at tilvejebringe den påkrævede statistiske information uden at ulejlige de enkelte borgere eller virksomheder. På den anden side rejser administrative registre mange problemer, dels nationalt, dels internationalt. Nationalt skal sådanne registre være indrettet, så de kan anvendes statistisk, det vil bl. a. sige, at grundidentifikationerne må foreligge i en form, der gør det muligt at fastholde de statistisk set relevante definitioner og inddelinger m. v. Da de nationale administrative registre nødvendigvis må følge de enkelte landes lovgivninger betyder dette ofte, at den internationale sammenlignelighed vanskeliggøres. Det er utopisk at forestille sig, at der indenfor en kortere årrække, om overhovedet, kan sikres en koordinering af de nationale administrative registre, således at den statistik, der kan uddrages af disse registre, uden videre vil kunne bringes i overensstemmelse med de for tiden anerkendte og realiserede koordi-

neringer på internationalt niveau. Det må håbes, at de statistiske nationale og internationale behov med hensyn til den statistiske anvendelse af administrative registre gives en høj prioritet, når sådanne registre planlægges og opbygges. De nordiske lande er adskilligt længere fremme på dette område end andre lande.

Det her behandlede drejer sig næsten udelukkende om praktisk statistik, men mange områder ville kunna drage nytte af et mere indgående samarbejde mellem praktikere og teoretikere. Det vil kræve en større gensidighed såvel med hensyn til interesse og indsigt i de to altfor adskilte verdener. Jeg vil slutte med at udtrykke håb om, at det i fremtiden må blive muligt at finde frem til en kommunikationsform mellem praktikere og teoretikere, det i højere grad kan forstås og accepteres af begge parter.

Klasser och partival i Sverige – en longitudinell studie om svenska väljare

av fil. kand. Pasquale Cricenti[1]

1. Inledning

Det svenska partisystemet har sedan införandet av allmän rösträtt varit mycket stabilt. Det kan kortfattat karakteriseras som ett trepartisystem fram till 1917, bestående av de konservativa, liberala och socialdemokratiska partierna. Efter 1917 ledde en utbrytning från det socialdemokratiska partiet till bildandet av det kommunistiska partiet. Vid denna tidpunkt bildades också det nuvarande centerpartiet.

Efter första världskriget har det svenska partisystemets form inte förändrats, det har blivit ett typiskt fempartisystem. I tab. 1 redovisas valutgångarna vid de olika riksdagsvalen sedan 1921. Av tabellen framgår att dessa fem partier tillsammans fick mellan 96 procent (1968 års val) och 100 procent (1921 års val) av de giltiga valsedlarna. Försök att bryta detta system har gjorts, främst under 60-talet och bl.a. av KDS[2], men utan framgång.

Inom statsvetenskaplig forskning är brukligt att indela de fem riksdagspartierna i två block. Socialdemokraterna och kommunisterna bildar det socialistiska blocket. Moderaterna, centerpartiet och folkpartiet bildar det borgerliga blocket. Valutgångarna visar, med undantag av några riksdagsval, att fördelningen mellan blocken är ganska jämn, särskilt under 1950- och 70-talet.

Även om valutgångarna är ganska jämna mellan blocken får man inte glömma den viktiga betydelsen som den socialdemokratiska frammarschen inom väljarkåren under 1930- och 40-talet har haft för den svenska politiken. Tack vare det växande stödet från väljarkåren kunde det socialdemokratiska partiet skjuta den politiska stridslinjen åt vänster. Denna förskjutning märks mest inom socialpolitiken. Det har lett till att även borgerliga partier, främst det liberala partiet, har accepterat denna politik (för en djupare diskussion om denna utveckling se Korpi, 1981: 39–43).

Moderna sociologer och statsvetare har mer och mer intresserat sig för varför folk röstar som de gör. Vilka faktorer kommer att påverka partival, osv.

Valforskarna är eniga om att valbeteende kan förklaras av olika bakgrundsvariabler. Vad gäller de skandinaviska länderna visar det sig, mer än i övriga västländer, att partival har ett starkt samband med klasstillhörighet. En av de viktiga orsakerna till det starka sambandet mellan yrkestillhörighet och partival, som funnits särskilt bland svenska väljare, beror främst på att religiösa och regionala motsättningar spelar en mycket mindre roll än i övriga västländer. I de länder där dessa konflikter existerar spelar klasstillhörigheten inte samma roll för partivalet.

Syftet med denna uppsats är att analysera klassröstning i Sverige utifrån väljarnas objektiva klasstillhörighet, samt illustrera var de olika socio-ekonomiska skikten ligger i

[1] Sociologiska institutionen, Stockholms universitet.
[2] Kristen Demokratisk Samling.

Tab. 1. De fem riksdagspartiernas procentuella fördelning av de giltiga valsedlarna vid några riksdagsval

Parti[1]	Valår[2]							
	1921	1932	1944	1956	1968	1973	1976	1979
Vpk	4,6	8,3	10,3	5,0	3,0	5,3	4,8	5,6
S	39,4	41,7	46,7	44,6	50,1	43,6	42,7	43,2
C	11,1	14,1	13,6	9,4	15,7	25,1	24,1	18,1
Fp	19,1	11,7	12,9	23,8	14,3	9,4	11,1	10,6
M	25,8	23,5	15,9	17,1	12,9	14,3	15,6	20,3
Summa	100	99,3	99,4	99,9	96,0	97,7	98,3	97,8
Valdeltagande	54,2	68,6	71,9	79,8	89,3	90,8	91,8	90,7

Källa: Sveriges officiella statistik: Allmänna valen 1979, del 1; tab. 1.

[1] Partiernas beteckning: Vpk = Vänsterpartiet kommunisterna, tidigare Kommunisterna (inräknat Socialister); S = Arbetarepartiet–Socialdemokraterna (inräknat Vänstersocialister); C = Centerpartiet, tidigare Bondeförbundet (inräknat Jordbrukarnas Riksförbund); Fp = Folkpartiet, tidigare Liberala och Frisinnade; M = Moderata Samlingspartiet, tidigare Högerpartiet (inräknat Liberala Försvarsvänner).
[2] T. o. m. 1968 andrakammarval.

den politiska konfliktmodellen.

Uppsatsen behandlar först klassröstning, där en rad bakgrundsvariablers (t. ex. klasstillhörighet, ålder och social rörlighet) inverkan på röstningssättet studeras. Vidare analyseras partival efter yrkesgrupp. Denna del bygger huvudsakligen på samma yrkesgrupper som den första delen, men syftet är att studera om det finns, utöver de ekonomiska gränserna, en politisk modell som bättre kan förklara de politiska stridslinjerna än den vanliga vänster-högerdimensionen.

Uppsatsens statistiska redovisning kommer från publicerat material från statistiska centralbyråns Allmänna valen 1970, 1973, 1976 och 1979 del 3 samt från Olof Petersson, "Väljarna och valet 1976" (1977) och Sören Holmberg, "Svenska väljare" (1981).

2. Klassröstning

För att uttrycka klassröstningen inom valbeteendeforskningen har man ställt upp en modell som utvecklades av Robert Alford. Han utvecklade en klassröstningsindex, dvs. procent av väljarna med manuell sysselsättning som röstar för vänsterpartierna minus procent av väljarna med icke-manuell sysselsättning som röstar på samma partier (Alford, 1963: 79–80).

Inom nordisk valforskning har denna modell antagits utan större kritik. Henry Valen kritiserar var gränsen för manuellt/icke-manuellt arbete går. Han skriver:

"Det er ikke alltid lett å sette grensen mellom manuelle og ikke-manuelle yrker. Det kan for eksempel være många tvilstilfelle blant selvstendig næringsdrivende og lavere funksjonærer. Enda mer usikkerhet knyter det seg til gruppen gårdbrukere og fiskere, som vel i hovedsak må regnes som manuell yrkesgruppe, men politisk gir de overvekt i borgerlig retning. De er under tvil blitt plassert sammen med funksjonærer og selvstendig næringsdrivende" (Valen, 1981: 129).

Uusitalo har gått ett steg längre än bara påpekande om tvivel. Han har helt enkelt uteslutit bönderna från klassröstningsindex. Han motiverar sitt val på detta sätt:

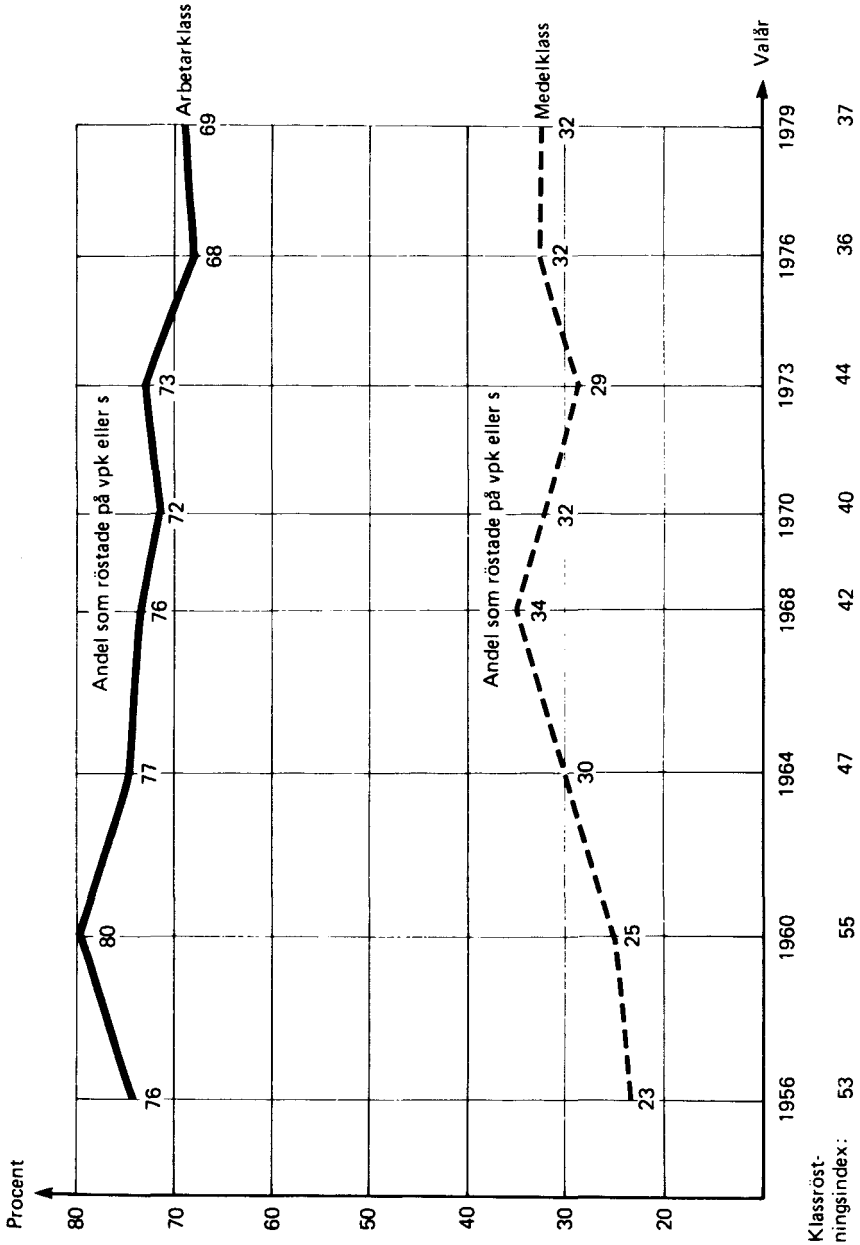


Fig. 1. Klassröstning vid valen 1956–1979 (procent) (Holmberg, 1981:311)

"... to place them on a manual/non-manual dimension is more confusing than illuminating" (Uusitalo, 1975:27–28).

Petersson (1977) och Holmberg (1981) har ställt upp samma dimension som Alford (1963). Gemensamt för dessa författare är att den manuella/icke-manuella dimensionen överförs till svenska förhållanden under begreppet arbetarklass/medelklass dimension. Där består medelklassen av olika yrkesgrupper med objektiva skilda intressen.

Walter Korpi hör till dem som ställer sig skeptiska till detta mått på klassröstning. Han skriver:

"Att använda det här måttet som uttryck för klassröstning kan leda till feltolkningar. Utifrån en annan definition av klassbegreppet kan man hävda, att tjänstemän och arbetare som enligt det här måttet placeras i olika klasser, i sin egenskap av löntagare borde placeras i samma klass. Beteckningen antyder också, att det på något sätt skulle vara naturligt för tjänstemän och arbetare att rösta på olika block" (Korpi, 1981:83).

Alfords klassröstningsindex var ursprungligen avsett att sammanfatta graden av klassröstning i det amerikanska samhället. Detta samhälle skiljer sig både politiskt och socioekonomiskt från de nordiska länderna[1]. Följaktligen är, enligt min mening, en mekanisk överföring av den amerikanska klassröstningens analys till den skandinaviska analysen felaktig.

[1] Jag tänker inte gå djupare in i analysen av dessa samhällen. Men här kan påminnas om den skillnad som finns när det gäller den politiska vänsterdimensionen. Inom amerikanska forskningen avses med vänster vad som i Europa betecknas som mitten eller liberal. Vidare att den dimension som studerar klassröstning i Amerika rör sig inom det borgerliga fältet, där det är naturligt att ett borgerligt parti (demokraterna) representerar arbetarna (blue collar workers). Så är inte fallet i Europa.

Jag kommer i det följande att gå igenom klassröstning i Sverige från 1956 till 1979 genom att analysera publicerat material från de intervjuundersökningar som har gjorts av statistiska centralbyrån (SCB) i samarbete med de statsvetenskapliga institutionerna vid Göteborgs och Uppsala universitet sedan 1956 års riksdagsval.

Detta material ger möjligheten att följa hur partival i olika yrkesgrupper har utvecklats sedan 1956. Jag kommer att använda Alfords klassröstningsindex mellan de olika yrkesgrupperna. De yrkesgrupper som jag kommer att analysera är: arbetare, tjänstemän i mellanställning, högre tjänstemän/storföretagare samt jordbrukare. Både Petersson (1977) och Holmberg (1981) har jämfört den första yrkesgruppen (arbetare) med de övriga tillsammans för att studera klassröstning i Sverige. Resultatet av denna analys tyder på att klassröstningen har minskat markant mellan 1956 och 1979, se fig. 1 på sid. 21.

Utifrån det tidigare antagandet, att dessa grupper har olika objektiva intressen kan vi sätta upp följande frågeställningar: har klassröstningen mellan arbetarna och jordbrukarna och mellan arbetarna och högre tjänstemän/storföretagare minskat? Har tjänstemän i mellanställning den perfekta klassröstningen i jämförelse med högre tjänstemän/storföretagare, dvs. röstar tjänstemän i mellanställning på icke-socialistiska partier i samma utsträckning som högre tjänstemän/storföretagare?

Klassröstningen mellan industriarbetarna och jordbrukarna, tab. 2 visar att det har inträffat en förändring under 60-talet för att sedan under 70-talet stabiliseras till samma nivå som vid 1956 års val. Andelen industriarbetare som röstade med de socialistiska partierna sjönk från 87 procent 1960 till 76 procent 1970 och var oförändrad mellan

Tab. 2. Andel socialistiska röster (S+Vpk) bland industriarbetare och jordbrukare 1956–1979 (procent)

Yrkesgrupp	Valår							
	1956	1960	1964	1968	1970	1973	1976	1979
Industriarbetare	80	87	84	81	76	75	73	76
Jordbrukare	14	8	7	6	5	9	10	7
Klassröstningsindex	66	79	77	75	71	66	63	69

Källa: SOS Allmänna valen 1973, del 3: tab. 32. Petersson, O., Väljarna och valet 1976: tab. 2.1. Holmberg, S., Svenska väljare: tab. 13.1.

Tab. 3. Andel socialistiska röster (S+Vpk) bland industriarbetare och högre tjänstemän/storföretagare 1956–1979 (procent)

Yrkesgrupp	Valår							
	1956	1960	1964	1968	1970	1973	1976	1979
Industriarbetare	80	87	84	81	76	75	73	76
Högre tjänstemän/storföretagare	4	10	8	15	16	18	24	23
Klassröstningsindex	76	77	76	66	60	57	49	53

Källa: Samma som tab. 2.

1970 och 1979, då andelen också var 76 procent. Klassröstningen mellan industriarbetare och jordbrukare ökade kraftigt mellan 1956 och 1960 för att sedan minska under hela 60-talet och fram till 1976. Vid 1979 års val ökade den åter kraftigt. Sammanfattningsvis har klassröstning mellan dessa två grupper ökat mellan det första och det sista undersökningstillfället.

En intressant utveckling får vi fram i tab. 3 där andelen industriarbetare som röstat socialistiskt jämförs med högre tjänstemän/storföretagare som röstat på samma sätt. Tendensen till klassröstning har också här minskat från 1960 till 1976 för att återigen öka 1979. Andelen högre tjänstemän/storföretagare som röstat med socialdemokraterna eller vänsterpartiet kommunisterna har ökat stabilt från årtionde till årtionde. Det var 4 procent som röstade socialistiskt 1956 och 10 procent 1960. Tio år senare var andelen uppe i 16 procent och 1979 stödde 23 procent av högre tjänstemän/storföretagare de socialistiska partierna. Klassröstningen mellan dessa två grupper har minskat sedan 1956, men den befinner sig på en annan nivå än den som visas i fig. 1 på sid. 21.

Enligt den klassindelning som redovisas av Petersson, Holmberg och andra författare borde samtliga tjänstemän oberoende av sin ställning rösta på icke-socialistiska partier, eller åtminstone är det naturligt att de skulle göra så.

I det följande kommer jag att analysera klassröstning mellan två olika tjänstemannagrupper. Grupperna är högre tjänstemän/storföretagare[1] och tjänstemän i mellanställning. Klassröstningsindex är skillnaden i procent icke-socialistiröstande mellan högre tjänstemän/storföretagare och tjänste-

[1] Den perfekta jämförelsen störs på grund av att även storföretagare ingår i denna grupp. Men tyvärr har jag ingen möjlighet att skilja dem åt.

Tab. 4. Andel icke-socialistiska röster (C+Fp+M) bland tjänstemän i mellanställning och högre tjänstemän/storföretagare 1956–1979 (procent)

Yrkesgrupp	Valår							
	1956	1960	1964	1968	1970	1973	1976	1979
Högre tjänstemän/storföretagare	96	90	83	83	81	78	76	76
Tjänstemän i mellanställning	62	57	49	49	54	57	57	54
Klassröstningsindex	34	33	34	34	27	24	19	22

Källa: SOS, Allmänna valen 1973, tab. 32. Petersson, O., Väljarna och valet 1976, tab. 2.1. Holmberg, S., Svenska väljare, tab. 13.1 och fig. 13.2.

Tab. 5. Andel socialistiska röster (S+Vpk) bland arbetare och tjänstemän i mellanställning 1956–1979 (procent)

Yrkesgrupp	Valår							
	1956	1960	1964	1968	1970	1973	1976	1979
Arbetare	76	80	77	76	72	73	68	69
Tjänstemän i mellanställning	38	43	47	49	45	40	41	44
Klassröstningsindex	38	37	30	27	27	33	27	25

Källa: Samma som tab. 4.

män i mellanställning (tab. 4). Jag kommer även att se på skillnaden i procent socialiströstande mellan arbetare och tjänstemän i mellanställning (tab. 5).

Av tab. 4 framgår att majoriteten av tjänstemän i mellanställning har röstat vid samtliga val (undantag 1968 års val) på de icke-socialistiska partierna. Men skillnaden mellan denna yrkesgrupp och högre tjänstemän är relativt hög. Klassröstningsindex mellan dessa två grupper har minskat. Minskningen beror inte på att gruppernas stöd till de icke-socialistiska partierna ökat, snarare tvärtom.

Skillnaden å ena sidan i socialiströstande mellan tjänstemän i mellanställning och arbetare (tab. 5) och å andra sidan i icke-socialiströstande mellan tjänstemän i mellanställning och högre tjänstemän är ungefär lika. Av detta kan vi konstatera att tjänstemän i mellanställning intar en mellanställning mellan de socialistiska och de icke-socialistiska partierna, eller om man så vill mellan arbetar- och borgarklassen, och konsekvent vore det kanske mer befogat att skilja tjänstemän i mellanställning från högre tjänstemän/storföretagare vid en analys av klassröstning.

Sammanfattningsvis kan vi säga att klassröstningen bland svenska väljare är stark. Den har ökat mellan arbetarna och jordbrukarna och minskat mellan arbetarna och högre tjänstemän/storföretagare. I det sista fallet ligger det nära till hands att anta att minskningen inte beror på att storföretagare har blivit mer socialistiska men snarare på grund av att de högre tjänstemännens rekryteringsbas har förändrats under tiden. Detta kan betyda att tidigare rekryterades högre tjänstemän uteslutande bland socialgrupp I, medan under de senaste årtiondena möjligheten för barn i arbetarhem att uppnå högre yrkesstatus har ökat. Denna hypotes kan testas genom att studera klassröstning bland de högre tjänstemännen med socialbakgrund i arbetarklassen.

3. Social rörlighet, ålder och klassröstning

Av det material som jag har tillgång till får jag inte någon möjlighet att analysera hur högre tjänstemän med arbetarklassbakgrund röstar. Social rörlighet och partival redovisas av både Petersson och Holmberg efter yrkesgruppindelning: arbetare, tjänstemän, småföretagare och jordbrukare, eller genom dikotomisering i arbetarklass och medelklass. Ändå finns det bland det publicerade materialet en intressant redovisning av social rörlighet och partival efter olika åldersgrupper, som med en viss generalisering kan testa min hypotes, dvs. högre tjänstemän (medelklass i tab. 6 och 7) med fadern som arbetare och/eller med uppväxtfamiljen från arbetarklassen röstar socialistiskt i större utsträckning än högre tjänstemän med fadern i samma yrke och/eller med uppväxtfamiljen i medelklassen.

Av tab. 6 och 7 framgår att faderns yrke och/eller familjens klass spelar en stor roll vid partival, särskilt bland de icke-socialt rörliga väljarna.

Tab. 6. Andel socialistiskt (S+Vpk) röstande 1976 efter faderns yrke, eget yrke och ålder (procent)

Faderns yrke	Eget yrke	Ålder						
		18–30	31–40	41–50	51–60	61–70	71–80	Samtliga
Arbetare	Arbetare	79	80	83	80	85	80	81
Tjänstemän/ företagare[1]	Arbetare	39	40	54	55	60	64	51
Arbetare	Tjänstemän/ företagare	44	49	54	42	44	32	46
Tjänstemän/ företagare[1]	Tjänstemän/ företagare	38	23	23	21	21	13	24

[1] I gruppen "tjänstemän/företagare" ingår även jordbrukare.
Källa: Petersson, O., Väljarna och valet 1976, tab. 2.14, sid. 31.

Tab. 7. Andel socialistiskt (S+Vpk) röstande 1979 efter egen klasstillhörighet, familjens klasstillhörighet och ålder (procent)

Familjens klass[1]	Egen klass	Ålder			
		18–30	31–50	51–80	Samtliga[2]
Arbetarklass	Arbetarklass	80	83	81	80
Medelklass	Arbetarklass	47	51	69	59
Arbetarklass	Medelklass	67	43	47	48
Medelklass	Medelklass	31	25	19	24

Källa: Holmberg, S., Svenska väljare, tab. 13.19, sid. 327.

[1] Holmberg har följande kommentar till tabell 13.19: "Uppväxtfamiljens klass har definierats med hjälp av uppgifter om både moderns och faderns yrke. Endast intervjupersoner som vuxit upp i homogena arbetarhem eller homogena medelklasshem är medtagna i analysen. Nio procent av intervjupersonerna hade föräldrar som yrkesarbetade i olika klasser. De är inte medtagna i tabell 13.19. Arton procent av intervjupersonerna hade vuxit upp i hem där båda föräldrarna arbetade och tillhörde samma yrkesklass. I 73 procent av fallen uppgav intervjupersonerna att modern var hemmafru. I sådana fall utnyttjade vi uppgifterna om faderns yrke för att avgöra familjens klasstillhörighet. Andelen Vpk- eller S-röstande har beräknats med totala antalet röstande på de fem riksdagspartierna som procentbas. Antal intervjupersoner som ingår i analysen är för 18–30-åringarna 404, för 31–50-åringarna 769 och för 51–80-åringarna 927".
[2] Denna kolumn bygger på tabell 13.18 där analysen görs efter faderns klasstillhörighet.

Andelen socialiströsstande bland stabila arbetare är mycket hög i samtliga åldersgrupper och vid båda mättillfällena. Skillnaden mellan åldersgrupperna är mycket liten. 81 procent av den stabila arbetarklassen stödde socialdemokraterna eller kommunisterna vid 1976 års val och 80 procent tre år senare.

Andelen icke-socialiströsstande bland den stabila medelklassen är också mycket hög men visar stora skillnader mellan olika åldersgrupper. Vid valet 1976 stödde 38 procent av väljarna bland den stabila medelklassen i åldern 18–30 år socialdemokraterna eller kommunisterna, vid 1979 års val har denna andel minskat till 31 procent. Sammanlagt är det 24 procent av den stabila medelklassens väljare som stödde socialdemokraterna eller vänsterpartiet kommunisterna vid båda mättillfällena.

Den uppåtrörliga medelklassens väljare, dvs. uppåtrörlig från arbetarklass till medelklass, röstade på de socialistiska partierna i större utsträckning än den stabila medelklassens väljare. Utifrån dessa två mättillfällen kan vi inte dra slutsatsen att den uppåtrörliga medelklassens väljare med tiden kommer att assimilera samma valbeteende som råder inom den nya gruppen. Av tab. 6 framgår inte något som tyder på en assimileringsprocess. De uppåtrörliga väljarna visar inte så stora skillnader mellan de olika åldersgrupperna. Undantag finns i åldersgruppen 41–50 som är extremt hög, i denna åldersgrupp stödde 54 procent socialdemokraterna eller kommunisterna, och i åldersgruppen 71–80 år som är extremt låg, 32 procent. Av tab. 7 får vi en dramatisk förändring bland de uppåtrörliga väljarna i åldersgruppen 18–30 år där andelen socialiströsstande ökat från 44 procent 1976 till 67 procent 1979.

Medelklassens valbeteende 1976 och 1979 stöder min hypotes att den ökade socialiströsstningen bland medelklassen (högre tjänstemän) sedan 1956 har ett starkt samband med sitt eget sociala ursprung. Medelklassens väljare med arbetarklassbakgrund tenderar att stödja de socialistiska partierna i större utsträckning än den stabila medelklassens väljare.

De nedåtrörliga väljarna, dvs. nedåtrörlig från medelklass till arbetarklass, röstar socialistiskt i mindre utsträckning än den stabila arbetarklassens väljare. Bland arbetarna med medelklassbakgrund finner vi ett samband mellan röstningssätt och ålder, andelen socialiströsstande i denna grupp visar en stigande tendens från de yngsta till de äldsta åldersgrupperna. Samma mönster gäller för båda mättillfällena. Vid valet 1976 stödde 51 procent av arbetarna med me-

Tab. 8. Andel väljare inom arbetarklassen som röstade på S eller Vpk 1960, 1970 och 1979 efter åldersgrupp (procent)

Valår	Åldersgrupp					
	20–30	31–40	41–50	51–60	61–70	71–80
1960	78	86	84	76	77	57
1970	67	73	73	73	74	72
1979	67	69	65	73	71	86

Källa: Holmberg, S., Svenska väljare, tab. 13.10, sid. 311.

Tab. 9. Andel väljare inom medelklassen som röstade på S eller Vpk 1960, 1970 och 1979 efter åldersgrupp (procent)

Valår	Åldersgrupp					
	20–30	31–40	41–50	51–60	61–70	71–80
1960	33	29	24	20	22	18
1970	45	32	29	32	23	26
1979	47	35	32	30	25	25

Källa: Samma som tab. 8.

delklassbakgrund socialdemokraterna eller vänsterpartiet kommunisterna, andelen ökade till 59 procent vid 1979 års val.

Sammanfattningsvis kan vi säga att det finns ett samband mellan partival och social bakgrund. De socialt rörliga väljarna avviker från de socialt stabila väljarna. Denna avvikelse visar sig bäst när vi kontrollerar åldern.

Åldern bland de socialt rörliga väljarna kan, enligt min mening, också förklara utvecklingen i klassröstning bland svenska väljare. I tab. 8 och 9 redovisas hur stor andel av arbetarväljare och medelklassväljare inom olika åldersgrupper som röstade på de socialistiska partierna vid 1960, 1970 och 1979 års val.

I tab. 8 och 9 kan vi se en dramatisk förändring mellan 1960 och 1970 i både arbetarklassen och medelklassen, även om den inom arbetarklassen är något mer markant.

De unga väljarna inom både arbetarklassen och medelklassen visar ett annat röstningsmönster än de äldre väljarna. De unga inom medelklassen tenderar att rösta på de socialistiska partierna i större utsträckning än de äldre. De unga inom arbetarklassen röstar något mindre på de socialistiska partierna än de äldre men skillnaden inom denna grupp är inte så markant som bland medelklassens väljare.

Holmberg skriver vad det gäller de unga väljarna och klassröstning:

"Unga väljare har kanske alltid röstat mindre klassmässigt än äldre väljare. Det tar tid innan människor får sådana erfarenheter att de upptäcker vilka partier som bäst tillvaratar deras ekonomiska och sociala intressen. Medelklassväljare kommer i så fall att rösta mer borgerligt ju äldre de blir medan arbetarväljare i högre grad kommer att stödja socialistiska partier" (Holmberg 1981: 310).[1]

Enligt mitt sätt att se kan Holmbergs sammanfattning stämma för den stabila medelklassens och arbetarklassens väljare, men inte för de socialt rörliga väljarna.

Vi har redan konstaterat att medelklassens väljare med arbetarklassbakgrund inte visar så stora skillnader i valbeteende mellan olika åldersgrupper. Medan arbetarväljare med medelklassbakgrund stämmer med Holmbergs sammanfattning. Av detta kan vi sammanfatta att röstningsmönstret bland de socialt rörliga väljarna kan förklaras med utgångspunkt i att dessa väljare är en erfarenhet rikare än de socialt stabila väljarna, dvs. utifrån deras erfarenheter inom de olika sociala grupperna kan de socialt rörliga väljarna t. ex. jämföra de olika ekonomiska och sociala intressena som partierna representerar. De kan ställa mot varandra sådana begrepp som "solidaritet" och "individualism" och utifrån sina erfarenheter kan de bättre förstå betydelsen av att vara solidarisk eller individualist. De resultat som vi får från vår dataredovisning tyder på att de socialt rörliga väljarna är mest inriktade på "solidaritet".

4. Partival efter yrkesgrupp

Henry Valen och Stein Rokkan har analyserat det norska motsättningssystemet genom att sätta upp en tre-dimensionell ekonomisk modell (Valen & Rokkan 1974: 332–333), vilken förefaller att gälla även för Sverige.

Enligt denna modell finns de egentliga motsättningarna på den ekonomiska arenan och modellen baseras på konflikter inom de

[1]Enligt Holmberg är det naturligt att samtliga yrkesgrupper som inte tillhör arbetarklassen röstar borgerligt. Här antyder Holmberg explicit att det är de borgerliga partierna som bäst kan tillvarata en stor del av löntagarnas intressen.

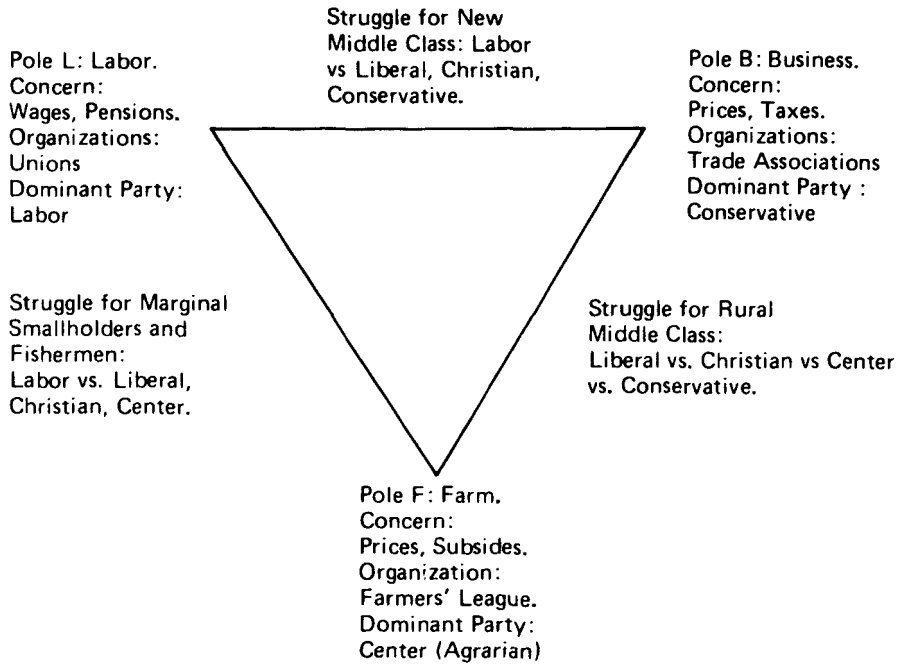


Fig. 2. Valens och Rokkans tre-polsmodell av väljarkåren

fundamentala ekonomiska dimensionerna. Modellen indelas i tre olika ekonomiska poler nämligen labor, business och farming. Varje pol ger uttryck för bestämda ekonomiska intresseorganisationer och politiska partier, se fig. 2. Enligt denna modell befinner väljarkåren sig mellan de tre polerna (a. a.: 332).

I tab. 10 redovisas partival efter olika yrkesgrupper som motsvarar de yrkesgrupper i pol L, B och F samt den mellanliggande kategorin L–B. Tabellen visar att de yrkesgrupper som är placerade vid varje pol i högre grad stöder de resp. partierna vid valet 1979:

Pol L: 76 procent för socialdemokraterna och kommunisterna.

Pol B: 46 procent för moderaterna (konservativa).

Pol F: 65 procent för centerpartiet.

Vad som är av intresse i tab. 10 är den longitudinella utvecklingen på partival efter yrkesgrupp i Sverige.

Vi kan anta (som tidigare) att de fyra yrkesgrupperna som tas upp i tabellen har objektivt skilda intressen och vidare att dessa grupper är kärnan i resp. ekonomisk pol enligt Valens och Rokkans modell, eller i resp. klasser i marxistisk mening[1]. Grupperna är industriarbetare, tjänstemän i mellanställning, högre tjänstemän/storföretagare samt jordbrukare. Utvecklingen av partistöd mellan 1956 och 1979 visar, med avseende på samtliga grupper, dramatiska

[1] Arbetarklassen, medelklassen, småborgerligheten och borgarklassen.

förändringar endast mellan de icke-socialistiska partierna. Högre tjänstemän/storföretagare och jordbrukare visar också stora förändringar mellan blocken.

Jordbrukare: 14 procent bland jordbrukarna stödde socialdemokraterna vid 1956 års val. Det var den högsta andelen som socialdemokraterna uppnådde bland denna yrkesgrupp enligt samtliga valundersökningar. Redan vid 1960 års val halverades jordbrukarnas stöd till socialdemokraterna och sedan dess har det legat i genomsnitt på 7 procent.

Jordbrukarnas stöd till moderaterna är nästan oförändrat mellan 1956 och 1979. Men vissa förändringar har förekommit i mellanvalen. Den lägsta noteringen var vid 1963 års val då endast 13 procent av jordbrukarna stödde moderaterna. Den högsta andelen var vid 1979 års val.

Den största förändringen vad gäller jordbrukarnas partistöd finner vi mellan folkpartiet och centerpartiet. Jordbrukarnas stöd till folkpartiet reducerades kraftigt, från 15 procent 1956 till 5 procent 1979. Partistödet bland jordbrukarna till folkpartiet visar samma trend som socialdemokratin.

Jordbrukarnas stöd till centerpartiet ökade från 49 procent 1956 till 65 procent 1979. 1956 års riksdagsval var det sista valet för bondeförbundet. Året efter bytte bondeförbundet sitt partinamn till centerpartiet. Avsikten var bl. a. att på detta sätt etablera sig bland de övriga samhällsgrupperna. Men det stora lyftet bland övriga yrkesgrupper kom inte förrän 1968–1970, i stället ökade jordbrukarnas stöd till bonde/centerpartiet redan vid det första valet (1960 års val) efter namnbytet. 1960 röstade 70 procent bland jordbrukarna på centerpartiet.

Sammanfattningsvis kan vi konstatera att den longitudinella utvecklingen av partistöd bland jordbrukarna visar, vad gäller de icke-socialistiska partierna, en stor ökning för centerpartiet, en stor minskning för folkpartiet och oförändrat stöd till moderaterna. Vad gäller det socialistiska blocket halverades jordbrukarnas stöd.

Tab. 10. Partival efter yrkesgrupp 1956, 1970 och 1979 (procent)

Parti	Yrkesgrupp			
	Industri-arbetare	Tjänste-män i mellan-ställ-ning	Högre tjänste-män/ storföre-tagare	Jord-bru-kare
1956				
Vpk	3	1	0	0
S	77	37	4	14
C	1	2	0	49
Fp	17	42	40	15
M	2	18	56	22
1970				
Vpk	5	4	1	0
S	71	41	15	5
C	14	20	12	70
Fp	6	22	36	7
M	2	12	33	15
1979				
Vpk	5	8	3	0
S	71	36	20	7
C	12	15	10	65
Fp	5	14	20	5
M	5	25	46	23

Källa: SOS, Allmänna valen 1973, del 3: tab 32, sid. 88–89. Holmberg, S., Svenska väljare, tab. 13.11, sid. 314.

Högre tjänstemän/storföretagare: denna yrkesgrupp visar en revolutionerande longitudinell utveckling sedan 1956 års val, både inom och mellan blocken.

De socialistiska partierna hade nästan inget stöd alls bland denna yrkesgrupp, endast 4 procent röstade för socialdemokraterna vid 1956 års val. Vid 1979 års val röstade 20 procent socialdemokratiskt och 3 procent kommunistiskt.

Också inom det icke-socialistiska blocket visar partistödet bland högre tjänstemän/storföretagare en dramatisk förändring.

Vid 1956 års val fick centerpartiet inget stöd bland högre tjänstemän/storföretagare medan det vid 1979 års val stöds av 10 procent. Det största stödet för centerpartiet inträffade vid 1973 års val, då 21 procent bland högre tjänstemän/storföretagare röstade på centerpartiet.

Folkpartiet visar en dramatisk halvering av sin röstandel bland högre tjänstemän/storföretagare mellan 1956 och 1979. 40 procent bland högre tjänstemän/storföretagare var folkpartister vid 1956 års val och andelen krymptes till 20 procent 1979.

Även stödet till moderaterna minskade bland högre tjänstemän/storföretagare mellan 1956 och 1979. Det minskade från 56 procent 1956 till 46 procent 1979. Valet 1970 markerar den dramatiska utvecklingen av högre tjänstemän/storföretagares stöd till moderaterna. Högre tjänstemän/storföretagares stöd till moderaterna minskade kraftigt under hela 60-talet och nådde 33 procent 1970. Vid 1973 års val började stödet till moderaterna i denna grupp stiga igen.

Den longitudinella utvecklingen av partistödet bland högre tjänstemän/storföretagare visar å ena sidan att andelen för både folkpartiet och moderaterna, som tillsammans hade 96 procent av denna grupps stöd 1956, minskade kraftigt mellan 1956 och 1979. Moderaterna kvarstår som partiet med det största stödet bland högre tjänstemän/storföretagare. Centerpartiet, vänsterpartiet kommunisterna och socialdemokraterna har å andra sidan fått ökat stöd. Förändringen mellan blocken visar att centerpartiets ökning inte är tillräckligt stor för att förhindra ett stort utflöde till de socialistiska partierna.

Tjänstemän i mellanställning: andelen bland tjänstemän i mellanställning som stöder de socialistiska partierna ökade från 38 procent vid 1956 års val till 46 procent 1979. Socialdemokraternas andel är nästan oförändrad, medan vänsterpartiet kommunisternas andel ökade från 1 procent 1956 till 8 procent 1979.

Partistöd bland tjänstemän i mellanställning har förändrats drastiskt mellan de icke-socialistiska partierna.

Folkpartiets ställning bland denna yrkesgrupp visar en dramatisk minskning mellan 1956 och 1979. I valet 1956 var det 42 procent bland tjänstemän i mellanställning som röstade på folkpartiet medan stödet till folkpartiet i denna grupp 1979 reducerades till 14 procent. Den största andelen av väljarna bland tjänstemän i mellanställning som lämnade folkpartiet gick över till centerpartiet eller moderaterna. Endast 2 procent av tjänstemän i mellanställning stödde centerpartiet vid 1956 års val mot 15 procent vid 1979 års val. Moderaterna ökade sin andel från 18 procent vid 1956 års val till 25 procent 1979.

Industriarbetare: industriarbetarnas stöd till de socialistiska partierna är nästan oförändrat mellan 1956 och 1979. Den största omflyttningen, vad gäller industriarbetarna, skedde bland "de borgerliga arbetarna" som flyttade sitt stöd från folkpartiet vid valet 1956 till centern 1979.

Socialdemokraternas andel minskade från 77 procent 1956 till 71 procent 1979. Denna minskning kompenserades inte av vänsterpartiet kommunisternas ökning från 3 procent vid valet 1956 till 5 procent 1979. Sammanlagt gick de socialistiska partierna tillbaka med 4 procent mellan dessa två valår.

Industriarbetarnas stöd till de icke-socialistiska partierna ökade från 20 procent 1956 till 22 procent 1979.

Folkpartiet förlorade sin starka ställning bland industriarbetarna. 1956 var folkpartiet det dominerande partiet bland "de bor-

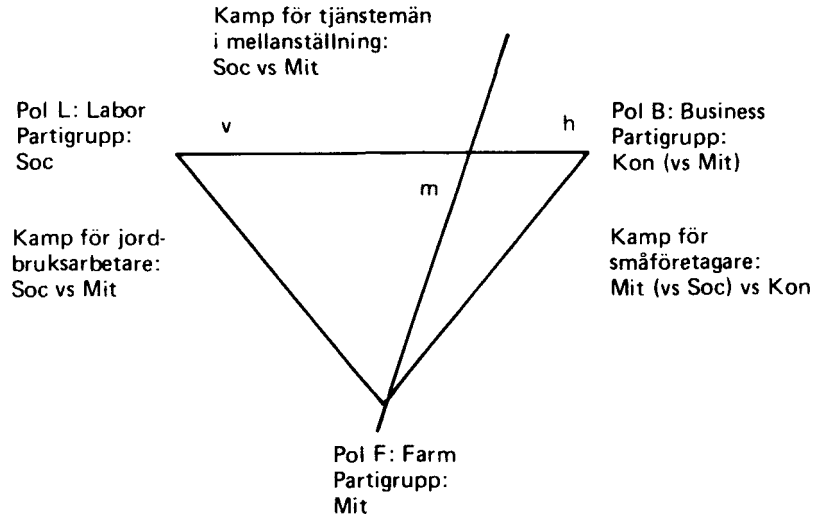


Fig. 3. Modell över den svenska väljarkåren

gerliga arbetarna". Industriarbetarnas stöd till de icke-socialistiska partierna vid valet 1956 resp. 1979 fördelades enligt följande: folkpartiet 17 procent resp. 5 procent, center 1 procent resp. 12 procent och moderaterna 2 procent resp. 5 procent. Den longitudinella utvecklingen mellan 1956 och 1979 visar att av de icke-socialistiska partierna tog centern den dominerande ställningen från folkpartiet bland industriarbetarna.

Utifrån den longitudinella utvecklingen bland svenska väljare ger Valens och Rokkans modell uttryck också för en polarisering av väljarkåren inom bestämda politiska dimensioner i det svenska motsättningssystemet, nämligen Vänster (socialist), Mitten (liberal) och Höger (konservativ). Med denna politiska modell kan vi också kanske komma närmare ett svar på varför valutgången i Sverige brukar vara ganska jämn mellan de socialistiska och de icke-socialistiska partierna.

I fig. 3 har vi en modell över den svenska väljarkåren. Modellen är samma som den i fig. 2 utom de mellanliggande kategorierna på axel L-F och på axel B-F, och med ett tillägg av en skärningslinje som delar den svenska väljarkåren i två politiska dimensioner, Vänster-Mitten dimension (V-M) och Höger-Mitten dimension (H-M).

I tab.11 redovisas samtliga dessa yrkesgruppers partival vid 1970 och 1979 års val. I tabellen grupperas partierna enligt följande: Socialdemokraterna och kommunisterna = Socialist (Soc), centerpartiet och folkpartiet = Mitten (Mit) och moderaterna = Konservativ (Kon). Partigrupperna rangordnas efter den procentuella andelen de får inom varje yrkesgrupp.

Av tab. 11 framgår klart att väljarna inom yrkesgrupp industriarbetare, jordbruksarbetare, jordbrukare och tjänstemän i mellanställning följer ett bestämt röstningsmönster under 70-talet. Samma röstnings-

Tab. 11. Partival efter yrkesgrupp 1970 och 1979 (procent)

Industriarbetare				Jordbruksarbetare				Jordbrukare			
1970		1979		1970		1979		1970		1979	
Soc	(76)	Soc	(76)	Soc	(55)	Soc	(58)	Mit	(77)	Mit	(70)
Mit	(20)	Mit	(17)	Mit	(39)	Mit	(35)	Kon	(15)	Kon	(23)
Kon	(2	Kon	(5)	Kon	(4)	Kon	(7)	Soc	(5)	Soc	(7)
Tjänstemän i mellanställning				Högre tjänstemän/ storföretagare				Småföretagare			
1970		1979		1970		1979		1970		1979	
Soc	(45)	Soc	(44)	Mit	(48)	Kon	(46)	Mit	(55)	Kon	(40)
Mit	(42)	Mit	(29)	Kon	(33)	Mit	(30)	Soc	(29)	Mit	(33)
Kon	(12)	Kon	(25)	Soc	(16)	Soc	(23)	Kon	(13)	Soc	(25)

Källa: SOS, Allmänna valen 1973. del 3: tab. 32, sid. 88–89. Holmberg, S., Svenska väljare, tab. 13.3 och tab. 13.11.

mönster har funnits vid tidigare val (se SOS, Allmänna valen 1973, del 3: tab. 32, sid. 88–89).

De övriga yrkesgrupperna visar en dramatisk utveckling mellan 1970 och 1979. Bland högre tjänstemän hade mittenpartierna en dominerande ställning 1970 följd av de konservativa med ett avstånd på 15 procentenheter. Vid valet 1979 har bilden förändrats och då är det mitten som följer efter de konservativa med ett avstånd på 16 procentenheter. Även bland småföretagare hade mitten en stark ställning 1970. Mer än hälften i denna grupp röstade för de två mittenpartierna 1970, vid valet 1979 minskades deras andel till en tredjedel. De konservativa tog ledningen även i denna grupp. Småföretagare avviker, för övrigt, starkt från den utveckling som övriga yrkesgrupper visar.

Vi har tidigare konstaterat att bland svenska väljare finns det ett samband mellan klasstillhörighet och partival. Av tab. 10 kan vi uppmärksamma att svenska väljare röstar också efter bestämda politiska dimensioner beroende på yrkestillhörighet med undantag av småföretagare som visar sig vara en ganska ombytlig grupp. Fig. 4 visar en grafisk bild av denna tendens.

Figuren representerar en två-dimensionell politisk modell. I modellen kan pol V (vänster), pol M (mitten) och pol H (höger) väl vara de politiska motsvarigheterna till polerna i Valens och Rokkans ekonomiska modell. I denna modell fördelas det politiska motsättningssystemet mellan två politiska dimensioner: Vänster-Mitten dimension och Höger-Mitten dimension.

De svenska väljarna befinner sig inom dessa två dimensioner. På Vänster–Mitten dimensionen gäller den politiska kampen de stora löntagargrupperna. De yrkesgrupper som har huvudrollen i denna dimension är industriarbetare, jordbruksarbetare, tjänstemän i mellanställning; hit kan också räknas övriga arbetare och lägre tjänstemän som inte är medtagna i denna analys. Vänstertpartierna (socialdemokraterna och kommunisterna) och mittenpartierna (centerpartiet och folkpartiet) är de partier som tävlar om deras stöd.

På Höger–Mitten dimensionen har vi egna företagare, privata arbetsgivare och också en hel del löntagare representerade

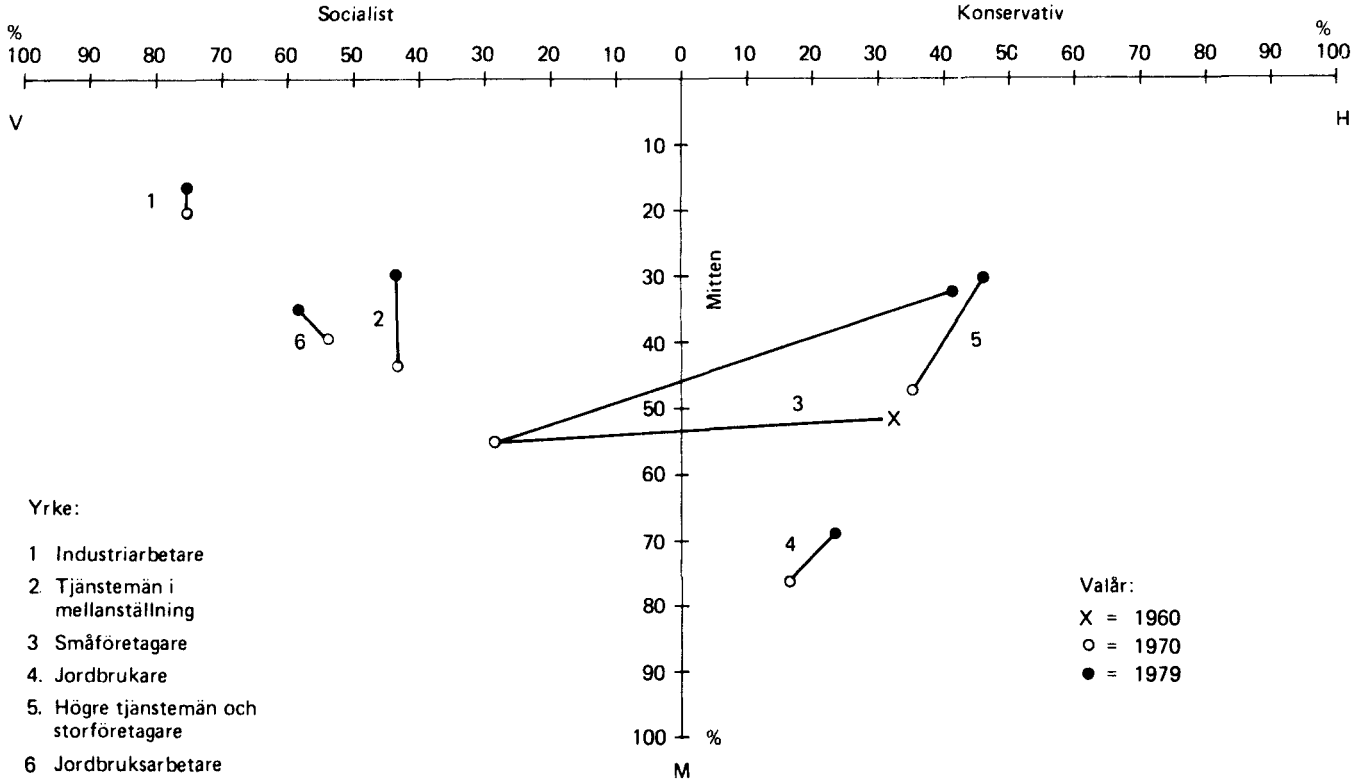


Fig. 4. Väljarnas röstningsbenägenhet på Vänster-Mitten-Höger dimensionen efter socio-ekonomiska grupper vid valen 1960, 1970 och 1979 (procent)

av högre tjänstemän. Här utspelas kampen om deras stöd mellan moderaterna och mittenpartierna.

Linjerna i figuren representerar en grafisk bild av tab. 11 och kan tolkas på följande sätt:

På V–M dimensionen har vi t.ex. industriarbetare (*1*) av vilka 76 procent stödde de socialistiska partierna och 20 procent centerpartiet och folkpartiet vid valet 1970. Tillsammans hade dessa två partigrupper 96 procent av industriarbetarnas stöd. Vid valet 1979 röstade 17 procent på centerpartiet eller folkpartiet, andelen var oförändrad för de socialistiska partierna. Samma sak gäller H–M dimensionen, t.ex. 77 procent bland jordbrukarna röstade vid 1970 års val på centern eller folkpartiet och 15 procent röstade på moderaterna. Kurvan mellan 1970 och 1979 visar att de mittersta partiernas andel bland jordbrukarna minskade medan moderaternas ökade.

Småföretagare visar att de inte har några bestämda rötter i denna modell. I figuren följer vi deras utveckling från 1960 till 1979. Vid valet 1960 befann sig denna grupp inom H–M dimensionen, då mittenpartierna hade den högsta andelen bland småföretagare, tätt följda av moderaterna. Vid 1970 års val hade mittenpartierna fortfarande det största stödet bland denna grupp men de socialistiska partierna gick förbi moderaterna. Vid valet 1979 hade moderaterna den relativa majoriteten i denna grupp. 40 procent av småföretagarna stödde moderaterna, 33 procent mittenpartierna och 25 procent de socialistiska partierna. Moderaternas stora ökning under 70-talet bland småföretagarna kan ha ett samband med de s.k. småföretagaraktionerna i valrörelserna 1976 och 1979, med en stark antisocialistisk propaganda, men ökningen har skett på bekostnad av mittenpartierna.

I inledningen har nämnts, att socialdemokratins framgångar har skjutit politiken mot vänster. De politiska stridslinjerna gällde främst löntagarnas välfärd. Både centerpartiet och folkpartiet har i stort sett anammat de socialreformer och den sysselsättningspolitik som gynnade löntagargrupperna. Detta har gjort det möjligt för dessa två partier att konkurrera med de socialistiska partierna om löntagarnas stöd. Å andra sidan är dessa två partier marknadsekonomins försvarare som gör att de kan tävla med moderaterna om de grupper som gynnas bäst av marknadsekonomin eller, om man så vill, vars existens som yrkesgrupp är beroende av marknadsekonomin.

"Socialreformer utan socialism" är kanske det bästa begreppet som har myntats för att kunna röra sig både till vänster och till höger.

5. Sammanfattning

Klassröstning är en viktig del av det svenska motsättningssystemet. Väljarna är mycket starkt klassbundna och det framkommer bäst när vi kontrollerar klassröstning bland industriarbetarna och jordbrukarna. Bland högre tjänstemän/storföretagare ökade socialiströstandet kraftigt mellan 1956 och 1979, vilket ledde till att klassröstning mellan denna grupp och industriarbetarna minskade. Övriga tjänstemän intar en mellanställning mellan det socialistiska blocket och det borgerliga blocket.

Väljarna befinner sig inom bestämda socio-politiska gränser, inom vilka konkurrensen om deras stöd är begränsad till vissa partier. Bland löntagargrupper tillhörande socialgrupp II och III har socialdemokratin det starkaste stödet, medan moderaterna har det starkaste stödet bland löntagarna i socialgrupp I. Folkpartiets och centerpartiets politiska anpassning har lett till att de

kan konkurrera med socialdemokraterna och moderaterna om stöd från deras resp. kärnväljare.

6. Litteratur

Alford, R. R. (1963): "Party & Society". Chicago: Rand McNally.

Holmberg, S. (1981): "Svenska väljare". Stockholm: LiberFörlag.

Korpi, W. (1981): "Den demokratiska klasskampen". Stockholm: Tidens Förlag.

Petersson, O. (1977): "Väljarna och valet 1976". Stockholm: LiberFörlag.

SOS (1970, 1973, 1976, 1979): "Allmänna valen, del 3". Stockholm: statistiska centralbyrån.

Valen, H. and Rokkan, S. (1974): "Norway: Conflict and Mass Politics in an European Periphery". I Rose, Richard (red.), "Electoral Behavior: A Comparative Handbook". New York: Free Press.

Valen, H. (1981): "Valg og politikk" Oslo: NKS-forlaget.

Uusitalo, H. (1975): "Class Structure and Party Choice: A Scandinavian Comparison". Research Reports 10, 1975. Research Group for Comparative Sociology, University of Helsinki.

Riksskogstaxeringen

av professor Björn Hägglund och fil. dr Bo Ranneby[1]

1. Bakgrund

Skogen har stor betydelse för Sveriges ekonomi och välfärd. Den ger oss välbehövliga exportintäkter, papper och virke för inhemsk konsumtion, sysselsättning på landsbygden, naturupplevelser, skydd mot vädrets växlingar och mycket annat. Karaktäristiskt för utnyttjandet av skogen är bl. a. den långa produktionstiden, skogarnas utbredning över stora arealer och också att skogen själv samtidigt är både produkt och produktionsmedel.

Hushållningen med skogen måste ske under ett samtidigt beaktande av de olika aspekter som nämnts ovan. Den måste baseras på data som korrekt beskriver hela den areal som används för skogsproduktion och som belyser många olika egenskaper hos skogen. Bland dessa egenskaper finns både sådana som avser förutsättningar för skogsproduktion och sådana som beskriver produktionens storlek och sammansättning.

Syftet med riksskogstaxeringen är att insamla och bearbeta de data som behövs för hushållning med skogarna på nationell och regional nivå. I mer formella ordalag är syftet med taxeringen att (Statskontoret, 1977) "ge underlag för planering och kontroll av naturresursen skog på riksnivå och regional nivå samt ge underlag för den skogliga forskningen. Dess centrala arbetsuppgift är därmed att samla in data om tillstånd och förändringar i skog och markanvändning samt bearbeta och redovisa dessa."

[1] Sveriges lantbruksuniversitet, skogsvetenskapliga fakulteten, Umeå.

Nationella skogsinventeringar typ riksskogstaxeringen förekommer i många utvecklade skogsländer. I de nordiska länderna finns stora likheter mellan taxeringarna, medan man t. ex. i USA delvis sökt sig fram längs andra vägar. Ursprunget till den svenska riksskogstaxeringen var en provtaxering av Värmland som utfördes 1911. Denna följdes av den första riksskogstaxeringen, genomförd 1923–29 (SOU 1932: 26), vilken lades upp som en s. k. linjetaxering, där träden mättes in i 10 m breda bälten. Arbetslaget bestod av 9 personer, och dessutom av en eller flera bärare. Taxeringen utfördes län för län, vilket medförde att data från olika delar av landet vid en viss tidpunkt hade olika aktualitet.

Den andra riksskogstaxeringen (SOU 1947: 36) startade 1938 och avslutades 1952. Även denna utfördes län för län, men mätningarna koncentrerades till cirkelprovytor (ca 140 m^2) utefter taxeringslinjerna, vilket (se avsnitt 4) innebar en effektivisering av arbetet. Under den andra taxeringen, år 1944, överfördes ansvaret för riksskogstaxeringen från den särskilda Riksskogstaxeringsnämnden till Skogsforskningsinstitutet sedermera Skogshögskolan sedermera skogsvetenskapliga fakulteten vid Sveriges lantbruksuniversitet.

Den tredje taxeringen, 1953–62 (Hagberg, 1952), innebar stora förändringar jämfört med den andra. Den länsvisa metoden ersattes med ett glest stickprov över hela Sverige varje år. Det främsta motivet för detta var att man skulle kunna skatta den

årliga avverkningen för hela landet. Detta åstadkoms genom mätning av stubbar.

Vid den tredje taxeringen "styckades" och "veks" de långa taxeringslinjerna till kvadratiska kluster, s.k. trakter, utefter vars sidor provytorna var belägna. Traktens storlek anpassades så att den i normalfallet skulle utgöra en dags arbete. I kombination med vägnätets utbyggnad, införandet av tjänstebilar etc. innebar traktmodellen att arbetet underlättades betydligt.

De fjärde (1963–72) och femte (1973–82) taxeringarna (Ohlsson, 1975) innebar vidareutvecklingar av det koncept som den tredje taxeringen byggde på. Systemet med kvadratiska trakter bibehölls, men traktens storlek, traktnätets täthet etc. justerades. Bearbetningen av data överfördes vid mitten av sextiotalet till datorer. Datainsamlingen anpassades till detta genom införandet av s.k. pennstanskort i fält. Vad gäller taxeringens inriktning finns en tendens att man alltmer lagt vikten vid de långsiktiga aspekterna av skogsbruket, främst då bevarandet och utvecklingen av skogarnas produktivitet. Exempelvis infördes 1973 större provytor, vilket underlättar genomförandet av betingade prognoser över skogarnas utveckling, s.k. konsekvensberäkningar.

År 1983 startar den sjätte riksskogstaxeringen, vilket bl.a. innebär att man inför permanenta provytor (Hägglund, 1982). De femte och sjätte taxeringarna kommer att beskrivas mer ingående i det följande.

2. Organisation

Riksskogstaxeringen utförs som ett projekt av institutionen för skogstaxering, skogsvetenskapliga fakulteten, Sveriges lantbruksuniversitet, Umeå. Detta projekt dominerar kraftigt institutionens verksamhet.

Institutionen för skogstaxering omfattar två avdelningar, taxeringsavdelningen och dataavdelningen. Vid taxeringsavdelningen finns året runt ca 40 personer, varav 30 är fast anställda medan 10 avlönas av tillfälliga medel från forsknings- och utvecklingsprojekt samt utredningsuppdrag. Under fältarbetssäsongen tillkommer den mestadels tillfälligt anställda fältarbetspersonalen, vilken organiseras i 19 s.k. ordinarie lag om fem personer och två s.k. kontrollag om fyra personer. Totalt arbetar således ett drygt 100-tal personer i fältarbetet varje år.

3. Insamlade data. Noggrannhetskrav

Riksskogstaxeringen är vad gäller insamlade data starkt orienterad mot hushållningen av skog utifrån skogsindustrins aspekter. Detta kommer också till uttryck i valet av de variabler som registreras. Några viktiga sådana är följande:

Virkesförrådet (mängd stående skog)
Volym, totalt och fördelat på geografiska områden, diameterklasser, trädslag, åldrar etc.
Skador
Tillväxt
Avverkning
m.m.

Uppgifter om växtplatsen
Bonitet (bördighet)
Markvegetation i olika skikt (botten-, fält-, busk- och trädskikt)
Jordart och jordmån (fastställs vid speciella markkarteringar)
m.m.

Arealuppgifter av betydelse för skogsproduktionen
Skogsmarkens areal fördelad på län, ägare, klimatområden etc.
Arealen skog med olika mogenhetsgrad

(huggningsklasser)
Arealen skog med olika ålder (åldersklasser)
Arealen skog med olika utförda åtgärder
Arealen skog i behov av olika åtgärder
m. m.

Speciella (icke rutinmässigt återkommande) uppgifter
Skador på mark och stående träd vid gallring
Tungmetallhalt i mossor
Viltfoder
Förekomst av svamp och bär
m. m.

Den traditionellt viktigaste uppgiften för riksskogstaxeringen är att bestämma virkesförrådets storlek. Ett operativt noggrannhetskrav har knutits till denna skattning, nämligen att medelfelet för en skattning av virkesförrådet baserad på fem års taxeringsmaterial inte bör överstiga 5 % för välrepresenterade åldersklasser inom medelstora och stora skogslän. Detta krav är en viktig hållpunkt vid dimensioneringen av taxeringen.

I samband med 1982 års budgetproposition har regering och riksdag fastställt riksskogstaxeringens framtida ambitionsnivå. Den nya taxeringen är utformad i enlighet med riktlinjerna i propositionen.

4. Teori

Som tidigare nämnts var den första riksskogstaxeringen en bältesinventering. Mätresultaten från punkter, som ligger nära varandra, är positivt korrelerade. Korrelationen blir högre ju närmare varandra punkterna ligger. Den ytterligare information, som en intilliggande punkt ger, blir därför mycket liten. Detta insåg man också inför den andra riksskogstaxeringen, varför mätningarna koncentrerades till provytor med ett inbördes avstånd på några hundra meter, belägna längs taxeringslinjerna.

Vid den tredje taxeringen införde man s. k. trakter – provytorna placerades ut längs periferin på en kvadrat. En sådan hopklumpning av ytorna innebär, på grund av den positiva korrelationen mellan punkter, att alla provytorna inte ger full information. Storleken på trakten anpassades så att den skulle motsvara ett dagsverke. På grund av insparade restider och o. d. ger en inventering av en trakt mer information än en inventering av det fåtal enskilda provytor man skulle hinna med på en dag om ingen hopklumpning av ytorna företagits.

4.1. Traktutformning

För att utnyttja resurserna så effektivt som möjligt måste man göra en lämplig avvägning mellan i första hand antal provytor och traktens sidlängd. Hänsyn måste även tas till andra faktorer som provytestorlek och provträdskvoter. Det optimala utseendet på en trakt varierar beroende på vilken storhet vi vill skatta. Om vi primärt är intresserade av att skatta skogsmarksandelen så skall avståndet mellan provytorna vara större än om vårt primära intresse är att skatta virkesförrådet/ha. Korrelationen mellan närliggande punkter är starkare för "skogsmarksvariabeln" än för variabeln "förråd/ha". Den slutliga utformningen av trakten blir en kompromiss.

Intressant är att notera att för en given sidlängd kan variansen för traktmedelvärdet öka med växande antal provytor om korrelationen mellan ytor blir alltför stor, se t. ex. Ranneby (1979).

Då det optimala traktutseendet påverkas av populationens variation delas Sverige in i olika regioner. För var och en av dessa regioner bestäms traktens storlek så bra som möjligt.

Detta har resulterat i att man fr.o.m. 1983 i södra Sverige, region 5, kommer att använda sig av "halvdagstrakter" i stället för "heldagstrakter". Den minskade hopklumpningen av provytorna som detta medför innebär att arealskattningen kommer att få avsevärt högre precision – medelfelet avtar med faktorn 1,5.

4.2. Provytestorlek

Precisionen hos skattningar av arealer är i stort sett oberoende av provytans storlek. Däremot påverkas precisionen hos skattningar av virkesförrådet/ha av provytestorleken.

Virkesförrådet/ha på en provyta kan ses som en summa av två komponenter där den ena anger ett ungefärligt värde, vilket erhålls via en grovklassificering av provytan, t.ex. angivande av huggningsklass (beståndets mogenhetsgrad), och den andra anger den lokala avvikelsen från det ungefärliga värdet. För en cirkelprovyta med radien r meter och medelpunkten s erhåller vi virkesförrådet/ha $v(s,r)$, som

$$v(s,r) = X(s) + e(s,r) \qquad (1)$$

I formel (1) betecknar $X(s)$ virkesförrådet/ha erhållet från grovklassificeringen och $e(s,r)$ den lokala avvikelsen från den av $X(s)$ givna nivån. Vi antar att variablerna $X(s)$ och $e(s,r)$ är okorrelerade. Då $e(s,r)$ mäter den lokala avvikelsen antar vi att $e(s_1,r)$ och $e(s_2,r)$ är okorrelerade redan för förhållandevis närbelägna punkter s_1 och s_2. Vidare antar vi att processen $\{X(s)\}$ är stationär med en isotropisk korrelationsfunktion, dvs. korrelationen beror endast på avståndet mellan punkterna och ej på riktningen. Variansen för $e(s,r)$ är beroende av provytans storlek medan variansen för $X(s)$ är oberoende av provytestorleken.

Vid riksomfattande inventeringar av typ riksskogstaxeringen innehåller variabeln $X(s)$ mycket information medan $X(s)$ är konstant vid inventering av enskilda bestånd. Detta leder till att den optimala ytstorleken är större vid inventering av enskilda bestånd än vid inventeringar som täcker ett större geografiskt område bestående av många olika bestånd.

Med hjälp av riksskogstaxeringens datamaterial kan erforderliga parametrar skattas så att olika stickprovsutformningar kan jämföras. En sådan jämförelse, se Ranneby (1981), ger vid handen att det vid den svenska riksskogstaxeringen är tillräckligt att inventera cirkelprovytor med radien 7 meter för att uppskatta virkesförrådet.

Genom analys av ett omfattande finskt material har Nyyssönen, Roiko-Jokela & Kilkki (1971) kommit fram till kvalitativt likartade resultat.

4.3. Lokalisering av trakter

I Matérn (1960, kap. 5) har effektiviteten hos några olika metoder att lokalisera samplingenheter jämförts. Vid lokaliseringen av trakterna kan vi använda oss av dessa resultat.

Vi tänker oss att uppskatta medeltalet av alla värden som en funktion $z(x)$ antar i ett visst område. Medeltalet skattas med medelvärdet $\bar{z}$ av de z-värden som observeras i ett antal samplingpunkter inom området. Vid jämförelsen antas att z-värdena kan uppfattas som en realisation av en stationär process med isotropisk och avtagande korrelationsfunktion.

Systematisk sampling i form av ett regelbundet triangelförband visar sig ge högst precision. Det regelbundna kvadratförbandet är dock i det närmaste lika bra och klart bättte än ett obundet slumpmässigt urval.

Det regelbundna triangelförbandet är det mönster som har den största undre gränsen

för avståndet mellan två sampelpunkter (vid en given sampelintensitet). Om korrelationsfunktionen avtar snabbt måste det optimala samplingnätet ha denna egenskap.

Som framgår av Dalenius, Hájek & Zubrzycki (1961) är dock det regelbundna triangelförbandet inte optimalt för samtliga processer som har en isotropisk och avtagande korrelationsfunktion.

4.4. Modeller över variationen i tid och rum

Inför den sjätte taxeringen som startar 1983 har en målsättning varit att undersöka om permanenta provytor och flyg- eller satellitbilder kan användas i den nya taxeringen. Stokastiska modeller som enbart beskriver den rumsliga variationen är inte tillräckliga för att jämföra stickprovsutformningar där permanenta provytor ingår, utan vi behöver modeller som beskriver variationen i såväl tid som rum.

Mosaikmodeller är lämpliga modeller för att bedöma värdet av flyg- och satellitbilder. Utgående från mosaikmodeller i planet har Ranneby (1982) konstruerat modeller som beskriver variationen i tid och rum. Det visar sig då att en stor klass av modeller beskrivs av processer sådana att rums-tidskorrelationen ges av produkten av rums- och tidskorrelationen.

4.5. Tvåfassampling

Riksskogstaxeringen är en systematisk stickprovsundersökning. Ett problem med systematisk sampling är att från observationsmaterialet uppskatta medelfelet. I ett pionjärarbete från 1947 visade Matérn hur man genom teoretiska överväganden kan avgöra vilka medelfelsformler som är användbara.

Vid vissa tillämpningar av dubbelsampling är det särskilt viktigt att man förfogar över korrekta metoder för skattning av precisionen. Som antyds i Matérn (1960, sid. 109) kan en olämplig medelfelsformel medföra att det kombinerade stickprovet utnyttjas mycket dåligt. Skattningen baserad på tvåfassamplingen kan få sämre precision än den skattning där enbart andra fasens mätningar utnyttjas.

5. Riksskogstaxeringens nuvarande uppläggning

Som framgått av det teoretiska avsnittet innebär utformningen av riksskogstaxeringen att man försöker nå en optimal avvägning mellan hög precision för givet antal provytor (ytorna så väl spridda över arealen som möjligt) och stort antal ytor inom givna resursramar (ytorna ligger så nära varandra som möjligt). Denna avvägning har för den pågående taxeringen lett till den utformning som visas i fig. 1. Provytorna är cirkelytor med 10 m radie. Ytorna är samlade till kluster, kvadratiska s. k. trakter med en sidlängd mellan 1 och 1,6 km. Sidorna är längre i norra Sverige än i södra. För de flesta variabler är variationen mellan trakter större i södra Sverige än i norra. Då även länen är mindre i södra Sverige krävs där betydligt tätare stickprov för att uppfylla uppställda noggrannhetskrav. Exempelvis är avståndet mellan två trakter i ett femårsnät 10 km i norra Sverige, 5 km i södra.

Antalet provytor där fullständiga mätningar görs, s. k. förrådsytor, är vanligen 12 per trakt. Dessutom förekommer extra s. k. arealytor, där endast arealbeskrivande faktorer registreras, och stubbytor, där stubbar från det senaste årets avverkningar mäts in.

Provytorna är för närvarande tillfälliga, dvs. de inventeras endast vid ett tillfälle. På varje yta diametermäts – "inklavas" – samtliga träd (vissa undantag finns) på en

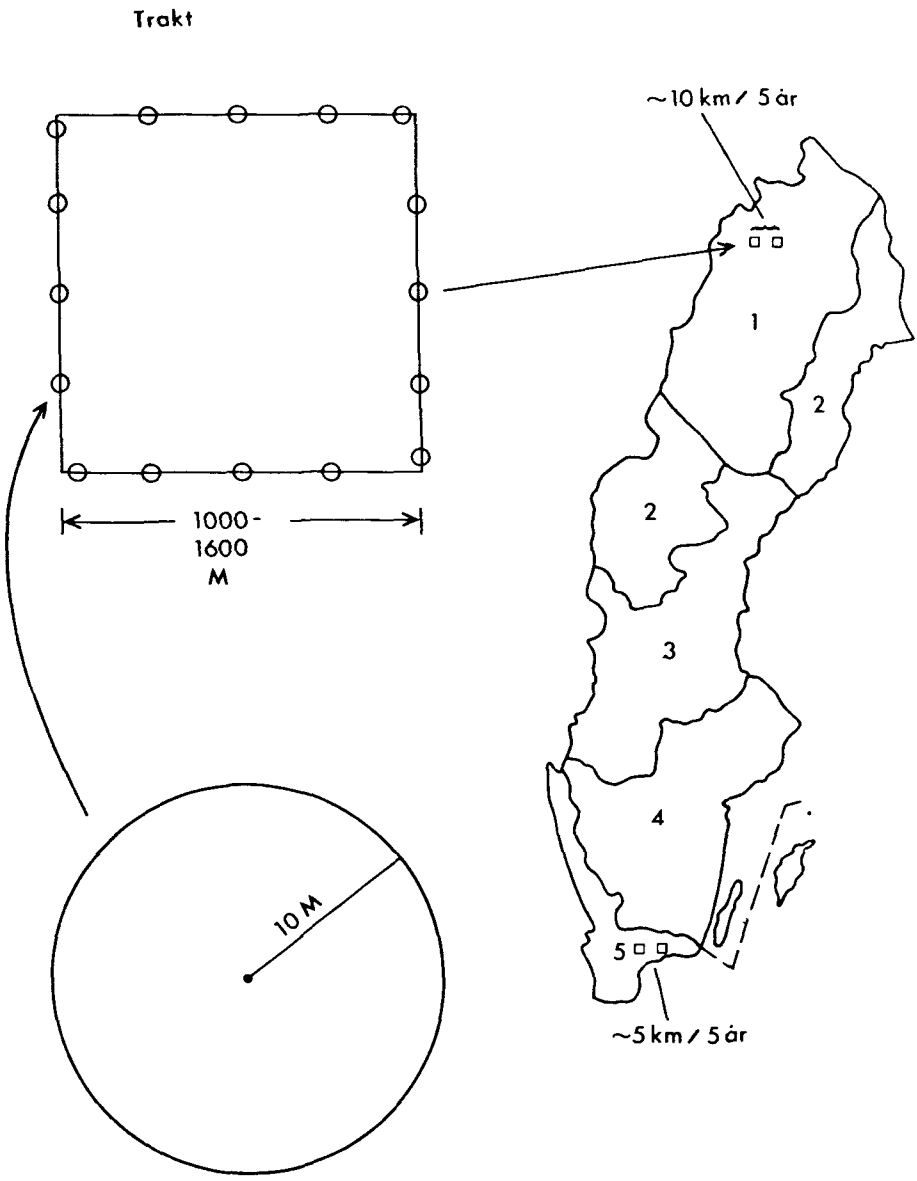


Fig. 1. Riksskogstaxeringen 1973–1982. Kartan till höger visar bl. a. landets indelning i fem taxeringsregioner.

höjd av 1,3 m över mark. Med en sannolikhet, som stiger med trädets diameter, uttas vissa träd – s.k. provträd – för ytterligare mätningar. På dessa träd mäts volymbestämmande faktorer som höjd, kronans höjd över mark, barkens tjocklek, diameter i stubbhöjd m.m. Från varje provträd tas också en borrkärna för bestämning av tillväxt och ålder. Kärnorna sänds till institutionen, där man så småningom mäter bredden på varje årsring.

Riksskogstaxeringens omfattning indikeras av nedanstående sammanställning som avser årligen inventerade kvantiteter.

	Ungefärligt antal 1973–82
Trakter (till någon del på land)	1 500
Förrådsprovytor	19 000
Inklavade träd	220 000
Provträd	20 000

6. Den nya taxeringen

Vid utformningen av den nya taxeringen som startar 1983 har än mer än tidigare betonats taxeringens roll som underlag för långsiktiga bedömningar. Skattning av förändringar ges ökad betydelse. Detta har lett till en uppläggning som innehåller permanenta provytor, vilka kommer att följas lång tid, kanske 20 år. Hela trakten kommer antingen att vara permanent eller tillfällig. Om man bortser från införandet av permanenta provytor påminner den nya uppläggningen om tidigare taxeringar. Sålunda bibehålls modellen med trakter. Antalet trakter, avståndet mellan ytor på traktsidan, provytans storlek, antalet provträd etc. justeras för att uppnå bästa möjliga design.

Förändringar skattas självfallet mycket bättre med permanenta provytor än med tillfälliga. Genom att använda både permanenta och tillfälliga provytor skattas inte enbart förändringar utan även nutillstånd bättre. Teorin för "sampling on successive occasions" har i statistisk litteratur behandlats så tidigt som 1942 av Jessen och av Patterson (1950). I skoglig litteratur har problemet behandlats av bland andra Bickford (1956, 1959) och Ware & Cunia (1962).

I en stickprovsundersökning med både permanenta och tillfälliga stickprovsenheter gäller det att göra en lämplig avvägning mellan antalet enheter av respektive slag. Den optimala andelen permanenta trakter varierar beroende på vilken storhet man vill skatta. De mätningar som registreras vid den svenska riksskogstaxeringen skall även användas som databank för skogsforskningen. Mot denna bakgrund har samma antal permanenta som tillfälliga trakter föreslagits i den nya taxeringen. De permanenta ytorna kommer att ha radien 10 m och de tillfälliga 7,07 m, dvs. de permanenta ytorna är dubbelt så stora som de tillfälliga.

Under antagandet om samma antal permanenta som tillfälliga trakter får vi under den första 10-årsperioden tre lika stora system av trakter nämligen

a) tillfälliga åren 1983–1987
b) permanenta åren 1983–1987, vilka återinventeras åren 1988–1992
c) tillfälliga åren 1988–1992.

Intuitivt skall trakterna i vart och ett av dessa tre system vara jämnt fördelade samt systemen sinsemellan vara sådana att totala antalet trakter bildar ett regelbundet mönster.

Med hjälp av modeller över variationen i tid och rum kan precisionen hos olika traktmönster jämföras. Vid en sådan jämförelse visar det sig att skillnaden mellan olika traktmönster (vilka samtliga är "sunda") är liten.

En permanent trakt är mindre (kortare traktsida och färre provytor) än en tillfällig. Följaktligen är den permanenta trakten före den första återinventeringen inte lika informativ som den tillfälliga. För att uppnå den precision vid skattning av virkesförråd eller skogsmarksandel som erhålls när man lägger ut ett visst antal tillfälliga trakter behövs ca 1,3 gånger så många permanenta trakter.

Utgående från skattade relativa medelfel för virkesförråd/ha och skogsmarksandel för perioden 1973–1977 har relativa medelfel för perioderna 1983–1987 (före den första återinventeringen), 1988–1992 (efter en återinventering) samt 1993–1997 (efter två återinventeringar) beräknats för den nya taxeringen. Resultaten avseende totalt virkesförråd redovisas i tab. 1.

En osäker faktor vid beräkningen av precisionen i den nya taxeringen är korrelationen mellan två på varandra följande observationer på en permanent provyta. För virkesförråd/ha har värdet 0,90 använts och för skogsmarksandel värdet 0,95.

Den valda uppläggningen av taxeringen medför att man redan efter den första återinventeringen skattar virkesförrådet betydligt bättre än om endast tillfälliga trakter hade använts. Den stora precisionsvinsten görs dock vid skattning av förändringar.

Tab. 1. Jämförelse av relativa medelfel för totalt virkesförråd

Län Länsdel	Medelfel i %			
	1973–1977	1983–1987	1988–1992	1993–1997
ACL	3,8	3,8	3,3	3,0
ACK	3,3	3,3	2,7	2,5
Y	2,5	2,7	2,3	2,1
X	3,0	3,0	2,5	2,3
F	3,5	3,5	3,0	2,7
U	5,3	5,2	4,5	4,2
M	10,8	8,1	7,7	7,1
L	5,2	4,2	3,9	3,7

Virkesförrådets förändring från år 1983 till år 1988 skattas med ett relativt medelfel som är $\sqrt{4}$ till $\sqrt{5}$ ggr mindre än om endast tillfälliga trakter används. Förhållandet är detsamma för skattning av skillnaden mellan 5-årsmedeltalen 1983—87 resp. 1988—92.

Avverkat förråd kommer också att skattas med ett lägre medelfel än tidigare. En stor fördel med permanenta trakter vid skattning av avverkad volym är att avverkningen kan skattas utan systematiska fel.

Den nya taxeringen skall rymmas inom samma resursram som den nu pågående och i vissa avseenden ge mer noggranna skattningar. För att klara detta genomförs vissa rationaliseringar. Den viktigaste av dessa är att arbetslagen utrustas med mikrodatorer i fält (Jonsson, 1981), varigenom lagens storlek kan reduceras från 5 till 3–4 (i genomsnitt 3,5) personer.

7. Riksskogstaxeringen som ett informationssystem

Riksskogstaxeringen kan uppfattas som ett informationssystem enligt skissen i fig. 2, vilken hämtats från Svensson (1980). Datainsamlingen innefattar som synes ett antal moment, varav fältarbetet volymmässigt dominerar. Stor vikt läggs vid att upprätthålla en hög kvalitet på insamlade data. Detta åstadkoms bl. a. på följande sätt:

- Lagledare och förmän utbildas varje år i de olika inventeringsmomenten.
- En speciell fältinspektör besöker varje lag 2–3 gånger per säsong. Dessutom besöks lagen ytterligare en gång av en forskare/handläggare.
- Två s. k. kontrollag taxerar om ungefär 10 % av de trakter som taxeras av ordinarie lag. De sistnämnda vet inte vilka trakter som skall omtaxeras. Resultaten jämförs i datorn siffra för siffra.
- Lagen sänder insamlade data till institu-

	FÄLTDATAINSAMLING			PRIMÄRBEARBETNING	RESULTATREDOVISNING
AKTIVITET	Inläggning av taxeringstrakter på kartor	Fältarbete	Årsringsmätning	Granskning	Kubering av provträd, beräkning av volymer för alla träd m. m.	Bearbetning till färdiga resultat samt redovisning av dessa
ger						
DATA	Trakternas läge, gränser mellan ägargrupper	Fältdata, årsringar på borrkärnor	Årsringsbredder	Rättade data, årsringsbredder	Bearbetade grunddata	Arealer, trädantal, plantantal, virkesförråd, provträdsegenskaper, tillväxt, avverkning m. m.
som lagras på						
MEDIUM	Kartor	Hålkort, borrkärnor	Hålremsa	Skivminne	Skivminne	Datalistor, rapporter m. m.

Fig. 2. Översiktlig skiss av riksskogstaxeringen sedd som ett informationssystem

tionen minst en gång per vecka. Data genomgår där omfattande kontroller i ett testprogram. Listor med upptäckta fel returneras till lagen. Felen skall åtgärdas omedelbart.

De olika kontrollerna pågår sålunda kontinuerligt under säsongen, vilket gör det möjligt att exempelvis snabbt korrigera en lagledare som missuppfattat någon punkt i instruktionen.

Data lagras efter primärbearbetningen som en databas på skiva. Basen är tämligen stor – för närvarande 100–150 milj. tecken – och alltid direkt tillgänglig. Varje år utför institutionen ett femtiotal uppdrag som innebär att man direkt går in i basen och gör sammanställningar för olika ändamål. Uppdragsgivare är t. ex. myndigheter, företag och skogsägarnas organisationer. Dessa "mindre uppdrag" ger oss en ytterst värdefull återkoppling till användarna. Härigenom får vi bl. a. impulser till förbättringar av datainsamlingsrutiner och inriktning av publiceringsverksamheten. Den tunga presentationen av data sker dock främst genom egna rapporter, artiklar i tidskrifter, föredrag och – icke minst – genom publicering i anslutning till stora uppdrag. Exempel på stora uppdrag är beräkningar för 1973 års skogsutredning (SOU, 1978:6, 7), inventering av mark för energiskogsodling (Lindroth, 1981) och inventering av virkesspill (Walldén, 1982).

Riksskogstaxeringens data används också i mer obearbetat skick som en "databank". Exempelvis finns i vår databas en stor mängd tillväxtuppgifter från enskilda träd, uppgifter som ofta används i samband med företags inventering av sina marker.

8. Framtiden

Riksskogstaxeringen tar just nu ett stort och viktigt steg framåt i sin utveckling genom att införa de nya taxeringsmetoder som tidigare beskrivits. Dessa är tänkta att bestå ett 20-tal år framåt. Denna långa period krävs för att vi fullt ut skall kunna dra nytta av de permanenta provytorna. Givetvis måste alla möjligheter till förbättringar tas till vara under perioden. Inom 5–10 år kommer förmodligen användandet av satellitbilder som en första samplingfas att aktualiseras. Detta för att erhålla förbättrade arealskattningar. Det bör betonas att satellitbilder knappast kan ersätta fältarbete, eftersom vissa variabler som dimensionerar fältarbetet – t. ex. tillväxten per arealenhet – är svåra att skatta med fjärranalysmetoder. Den moderna satellitbildstekniken torde – i motsats till traditionella flygbilder – utgöra ett relativt billigt komplement till fältarbetet och medverka till att precisionen i areal- och totalsiffror kan höjas.

Den explosiva utvecklingen på dator- och datakommunikationsområdet kommer säkert också att starkt påverka riksskogstaxeringens rutiner. Den dag är väl inte så långt borta när taxeringslagen i skogen via någon typ av radio/telekommunikation kan arbeta "on line" med bearbetningsprogram och databas i centraldatorn. Alternativt kommer fält- och skrivbordsdatorerna att bli så kraftfulla att hela konceptet med en centraldator kan försvinna.

Till syvende och sist kommer riksskogstaxeringens framtid att i hög grad bero på intentionerna i skogspolitiken. Så länge som Sverige vill driva en medveten skogspolitik, inriktad mot uthållig och hög produktion av nyttigheter från skogen, så länge måste vi också driva en riksskogstaxering på hög ambitionsnivå.

9. Referenser

Bickford, C. A. (1956): Proposed design for continuous inventory: a system of

perpetual forest survey for the Northeast. U.S. Forest Serv. Eastern Techniques Meeting, Forest Survey, Cumberland Falls, Ky. Oct. 8–13, 1956.

Bickford, C. A. (1959): A test of continuous inventory for national forest management based upon aerial photographs, double sampling, and remeasured plots. Soc. Amer. Forester Proc., sid. 143–148.

Dalenius, T., Hájek, J. & Zubrzycki, S. (1961): On plane sampling and related geometrical problems. Proc. 4th Berkeley Symp. Math. Statist. Vol. I, 125–150.

Hagberg, E. (1975): Riksskogstaxering efter ny modell. Skogen nr 2.

Hägglund, B. (1982): En ny svensk riksskogstaxering. Manuskript.

Jessen, R. J. (1942): Statistical investigations of a sample survey for obtaining farm facts. Iowa Agr. Exp. Sta. Res. Bull., 304.

Jonsson, B. (1981): An electronic caliper with automatic data storage. Forest Sci. 27, sid. 765–770.

Lindroth, S. (1981): Mark för energiskog. Nämnden för energiproduktionsforskning, projektresultat, NE 1981:26.

Matérn, B. (1947): Metoder att uppskatta noggrannheten vid linje- och provytetaxering. Medd. från statens skogsforskningsinstitut 36, nr 1.

Matérn, B. (1960): Spatial variation. Medd. från Statens skogsforskningsinstitut 49, nr 5.

Nyyssönen, A., Roiko-Jokela, P. & Kilkki, P. (1971): Studies on improvement of the efficiency of systematic sampling in forest inventory. Acta Forestalia Fennica, vol. 116.

Ohlsson, S. (1975): Riksskogstaxeringen – ändamål, utformning och användningsområden. Skogshögskolan, inst. för skogstaxering, rapporter och uppsatser nr 20.

Patterson, H. D. (1950): Sampling on successive occasions with partial replacement of units. Jour. Roy. Stat. Soc., Ser. B, 12, sid. 241–255.

Ranneby, B. (1979): Model studies of tract sizes in forest survey. Forest Resource Inventories, vol. 3, sid. 289–297.

Ranneby, B. (1981): Provytestorlekens betydelse vid skogsinventeringar. Sveriges lantbruksuniversitet. Inst. för skogstaxering, NYTAX 83, rapport 4.

Ranneby, B. (1982): Stochastic models of variation in time and space. I Statistics in Theory and Practice. Essays in Honour of Bertil Matérn, sid. 227–245. Sveriges lantbruksuniversitet, Avd. för skoglig biometri.

SOU, 1932:26. Uppskattning av Sveriges skogstillgångar verkställd åren 1923–29. Riksskogstaxeringsnämnden, Stockholm.

SOU, 1947:36. Vid andra riksskogstaxeringen av Norrland åren 1938–42 använd metodik och härom vunna erfarenheter. 1937 års Riksskogstaxeringsnämnd, Stockholm.

SOU, 1978:6, 7. Skog för framtid. Betänkande av 1973 års skogsutredning.

Statskontoret (1977): Riksskogstaxeringen, en organisationsöversyn. Statskontoret, rapport 1977:8.

Svensson, S. A. (1980): Riksskogstaxeringen 1973–77. Sveriges lantbruksuniv., inst. för skogstaxering, rapport 30.

Walldén, J. (1982): Riksskogstaxeringens inventering av virkesspill 1980. Sveriges lantbruksuniversitet, inst. för skogstaxering, rapport 35.

Ware, K. D. & Cunia, T. (1962): Continuous forest inventory with partial replacement of samples. Forest Science – Monograph 3.

Meddelanden

Europeiska statistiska möten

Arbetsgruppen för nationalräkenskaper

Arbetsgruppen för nationalräkenskaper höll sitt 11e möte i Genève den 1–4 juni 1982. I mötet deltog ett 25-tal länder. Sverige representerades av statistikchef Åke Tengblad, SCB. Till ordförande valdes herr Adler från Kanada.

Huvuddelen av mötet ägnades åt frågorna kring översyn av det internationella nationalräkenskapssystemet System of National Accounts (SNA). Bland andra intressanta frågor som avhandlades var samordning av ekonomisk statistik med hjälp av nationalräkenskaper.

SNA fastställdes för ca 15 år sedan och det finns nu betydande erfarenhet ifråga om dess tillämpbarhet. Systemet skapades mot bakgrunden av de ekonomiska och sociala förhållanden som rådde under 1960-talet. Sedan dess har den höga inflationstakten, den offentliga sektorns ökade betydelse och andra institutionella förändringar klargjort behovet av preciseringar, uppdatering och utbyggnad av SNA. Det finns vidare behov av att få bättre samstämmighet "harmonisering" mellan SNA och relaterade statistiksystem, som utarbetats av andra internationella organisationer, som IMF, ILO och EG.

Det hade utarbetats en rapport rörande fortsatt arbete och ändringar på kort sikt av SNA, och denna hade diskuterats av en expertgrupp. Rapporten och expertgruppens utlåtande låg till grund för arbetsgruppens diskussioner av SNA.

Bland de frågor som diskuterades särskilt var behandlingen av räntesubventioner, tillägg rörande icke-marknadsmässiga aktiviteter, samt några frågor som berör inkomstbegreppet bl. a. behandlingen av arrenden. Vidare behandlades kontoplanen och sektorsavgränsningar, harmonisering och SNAs revidering på kort sikt. Diskussionen om räntesubventionerna blev närmast en probleminventering och behandlingen måste utredas närmare. Beträffande icke-marknadsmässiga aktiviteter t. ex. hushållsarbete framhölls de betydande svårigheterna med värderingen. Skäl som talade för att icke-marknadsmässiga aktiviteter togs med i större utsträckning än för närvarande var att man på så sätt underlättade välfärdsanalys. Skäl som talade mot var att man inte vill försämra nationalräkenskapernas användbarhet för finanspolitisk och monetär analys. Under alla förhållanden borde SNA kompletteras så att man särredovisar marknadsmässiga och icke-marknadsmässiga aktiviteter.

Ifråga om arrenden var man överens om de statistiska svårigheterna att skilja dessa från andra typer av hyresbetalningar. Några ifrågasatte också de teoretiska grunderna för en distinktion.

Diskussionen om kontoplanen och sektorsavgränsningar handlade mest om sektorstillhörighet för personliga företag, offentlig affärsmässig verksamhet och ideella organisationer som betjänar hushållssektorn.

Harmoniseringsfrågan berör främst IMFs riktlinjer för finansstatistik för offentliga sektorn, betalningsbalansstatistik, och internationell finansstatistik, samt EGs riktlinjer för det europeiska nationalräkenskapssystemet.

I diskussionen om harmonisering framhölls bl. a. att ett första steg kunde vara att utarbeta tabeller eller matriser, i vilka man klart kunde se begreppsmässiga eller andra skiljaktigheter mellan olika statistiksystem och SNA. Denna aspekt kommer att tillvaratas i de olika handböcker om nationalräkenskaperna som nu håller på att utarbetas av FN.

Det fanns i expertrapporten en lång lista med förslag till ändringar av SNA som borde genomföras på kort sikt. Av dessa diskuterades främst behandlingen av direktinvesteringar i utlandet – detta är också en av de mest besvärande olikheterna mellan

SNAs och IMFs betalningsbalansstatistik. I Australien hade man arbetat mycket med detta problem och delegaten därifrån skisserade en lösning, som diskuterades och ansågs värd att närmare penetrera.

I diskussionen angående samordning av ekonomisk statistik med hjälp av SNA noterades att det finns stora likheter länderna emellan både ifråga om hur långt man nått och vad gäller de problem man mött. Samordningsarbetet är en kontinuerlig process, som fortlöpande måste bevakas av statistikmyndigheterna. Nationalräkenskaperna har här en framstående roll, både för att förbättra statistikens användning för ekonomisk analys, för att undvika statistiskt dubbelarbete och överbrygga konflikter mellan parallella statistikserier. Samordningsinstrumenten är också rätt likartade mellan länderna, t.ex. nationella, ibland obligatoriska, standards för definitioner, begrepp, varunomenklaturer och andra klassificeringar, bildandet av styr- och samordningsgrupper för berörda ämnesområden, centrala företagsregister/arbetsställeregister, utarbetandet av input-output-tabeller etc. Behovet av samordning mellan mikrodata och makrodata ansågs också betydelsefullt.

Vid diskussionerna om det fortsatta arbetet på europeisk nivå föreslog arbetsgruppen dessa områden som särskilt angelägna:

- Behandlingen av naturresurser och mått på dess utnyttjande.
- Behandlingen av subventioner.
- Sektorsavgränsningarna.
- Samordning av statistik för offentliga sektorn resp. betalningsbalansstatistik med nationalräkenskaperna.
- Behandlingen av illegal verksamhet.
- Behandlingen av direktinvesteringar i utlandet.
- Befolkningens totalkonsumtion.
- Integrering av input-output-tabeller med nationalräkenskaperna.

Det fortsatta arbetet inom statistikerkonferensen ansågs böra ägnas åt översynen av SNA, länkningarna mellan SNA-MPS samt frågorna kring samordning av ekonomisk statistik med SNA.

Åke Tengblad

Studiegrupp för livsmedels- och jordbruksstatistik

Den av FAO, ECE och Europeiska statistikerkonferensen gemensamt anordnade studiegruppen för livsmedels- och jordbruksstatistik höll sin trettonde session i Genève den 3–6 maj 1982. I mötet deltog representanter för ett tjugotal länder i Europa och Nordamerika. Sverige representerades av statistikchef Bernt Wilson, SCB. Avdelningschef Knut Medin, SCB, deltog i egenskap av konsult åt FAO.

Programmet för mötet innefattade en allmän genomgång av nuvarande förhållanden, utvecklingstendenser, problem etc. inom jordbruksstatistiken i berörda länder. Underlaget för diskussionerna var en av Knut Medin utarbetad rapport, "Review of Agricultural Statistics in Europe". Materialet hade insamlats genom enkäter till drygt tjugotalet öst- och västeuropeiska länder, EGs jordbruksstatistiska kontor samt USA och Kanada. Innehållet i rapporten disponerades vid mötet under följande rubriker:

- aktuella faktorer som påverkar behovet av jordbruksstatistik
- behov och tillgång av jordbruksstatistik
- utvecklingslinjer med hänsyn till bl.a. metodologiska och tekniska möjligheter
- studiegruppens fortsatta arbete.

Beträffande förutsättningarna för verksamheten inom det jordbruksstatistiska området redovisades från åtskilliga länder likartade tendenser. Resurserna har under de senaste åren begränsats. Uppdragsfinansiering av statistiken har i ökande utsträckning prövats i flera länder. Det finns tendenser till ökat motstånd mot försörjningen med primärdata från jordbrukare och andra uppgiftslämnare. Ett tilltagande intresse redovisades för användning av administrativa register som underlag för statistikframställning. Några länder har problem med bristande tillgång på statistiskt utbildad personal. Önskemål finns bl.a. därför om internationellt samarbete och stöd från internationella organ främst vad gäller metodutveckling.

Under rubriken allmänna aspekter på statistikproduktionen diskuterades bl.a. olika kvalitetsfrågor, kraven på snabbhet och

punktlighet i redovisningen samt behovet av samordning mellan jordbruksstatistiken och andra statistikgrenar.

Ett viktigt ämne vid mötet var frågan om världsjordbruksräkningarnas ställning inom den internationella jordbruksstatistiken. Diskussionen visade att räkningarna även i fortsättningen bedöms ha stor betydelse och att en väsentlig del av gruppens arbete kommer att ägnas åt förberedelserna för nästa världsjordbruksräkning omkring 1990. I åtskilliga länder existerar nu lantbruksregister, som vid sidan av eventuella administrativa ändamål används för framställning av jordbruksstatistik, bl. a. i samband med jordbruksräkningarna. I några länder planerar man å andra sidan att utnyttja räkningarna för att få fram stommen till ett lantbruksregister.

Studiegruppen kommer under de närmaste åren vidare att koncentrera sitt arbete på metodologiska frågor inom en rad områden som spelar en viktig roll i jordbrukets utveckling i de allra flesta länder. Hög prioritet skall sålunda ges åt arbetskraftsstatistiken för jordbruket. Frågor som sammanhänger med heltids/deltidsföretagen tilldrar sig stort intresse inte bara ur arbetskraftssynpunkt utan även med hänsyn till jordbrukets strukturella utveckling. Inom strukturstatistiken ägnas numera ett väsentligt arbete åt införande av typklassificeringssystem för jordbruksföretagen (s. k. typologi). En handbok eller monografi inom detta ämnesområde avses utarbetas under de närmaste åren. Studiegruppens program för de närmaste åren upptar vidare studier och möten beträffande prognoser inom vegetabilie- och animalieproduktionen, metodik för bokföringsundersökningar, samordning mellan jordbruks- och befolkningsstatistik.

Från svensk sida betonades särskilt värdet av fortsatt arbete inom arbetskraftsstatistiken, typologin och bokföringsundersökningarna.

Bernt Wilson

Möte om miljöstatistik

Ett möte kring frågor om "Framework for environment statistics" hölls i Genève den 4–6 oktober 1982 med deltagande från ett tiotal länder däribland Finland, Norge och Sverige. Deltagare från Sverige var avdelningschef Knut Medin och statistikchef Bengt Malmås, båda SCB.

Mötet hade föregåtts av ett informellt möte under 1981[1], då en inledande diskussion förts i flera av de frågor som nu stod på dagordningen. Det gällde bl. a. resultaten från en undersökning som sekretariatet utfört av olika principer för presentation av statistik på miljöområdet och som haft sin utgångspunkt i det kanadensiska s. k. STRESS-systemet. I studien hade belysts bl. a. hur innehållet i den svenska miljöstatistiska årsboken förhåller sig till strukturen i detta system. Mötet konstaterade att sekretariatets undersökning och statistikerkonferensens arbete över huvud taget med olika frågor om redovisningsstrukturer hade varit till god hjälp i försöken att beskriva det nya område som miljöstatistiken utgör. Det slogs emellertid också fast att de mer övergripande problemen nu fick anses vara tillräckligt väl behandlade och att de fortsatta insatserna i statistikerkonferensens regi i stället borde riktas in mot enskilda delstatistikområden och på behandlingen av viktiga metodfrågor.

Som andra punkt på dagordningen diskuterades en av sekretariatet utförd jämförande studie av de arbeten med s. k. naturresursräkenskaper som pågår i Norge och Frankrike. I anslutning till denna diskussion framkom att man i flera andra länder överväger att ta upp liknande arbeten.

Frågan om det innehållsmässiga sambandet mellan miljöstatistiken å ena sidan och nationalräkenskapssystemet och den socio-demografiska statistiken å andra sidan diskuterades också vid mötet. Utgångspunkten var här dels en sekretariatsrapport om kopplingen till det ekonomiska statistiksystemet (som diskuterats redan vid det informella mötet under 1981), dels en av Ungern utarbetad rapport om beröringspunkterna med den socio-demografiska statistiken. I diskussionen uppmärksammades ett antal

[1] Se närmare mötesreferat i Statistisk tidskrift 1981: 3.

områden där man i det fortsatta arbetet med miljöstatistiken borde vara medveten om att anknytningar finns till nationalräkenskapssystemet och den socio-demografiska statistiken. Några mera bestämda slutsatser om hur man praktiskt skulle gå till väga för att koppla samman statistiken på resp. områden var man dock inte beredd att dra vid mötet.

Vissa klassifikationsfrågor behandlades vid mötet mot bakgrund av att ett mera omfattande internationellt statistiksamarbete på miljöområdet förutsätter viss normering av klassificeringar och definitioner. Diskussionen ledde här till att mötet rekommenderade en närmare granskning av FN-systemet Classification of Functions of Government (COFOG) för att se var det kunde vara motiverat att söka åstadkomma underuppdelningar avpassade speciellt för miljöstatistikens behov.

I en av Kanada utarbetad rapport informerades om förberedelserna för en "State of Environment Report for Canada". Arbetet, som utförs i samarbete mellan Statistics Canada och berörda miljömyndigheter, utgår från den tidigare nämnda STRESS-ansatsen och bygger på en indelning av landet i femton s. k. ekologiska regioner.

Information lämnades också om det arbete som FNs statistiska kontor bedriver inom miljöstatistikområdet (med ekonomiskt stöd från FNs speciella miljöorgan UNEP). Det hittillsvarande arbetet har sammanfattats i en rapport "Framework for environment statistics", som innehåller en omfattande diskussion av olika frågor som aktualiseras då man börjar utveckla statistiken på miljöområdet. Från flera håll uttalade man vid mötet sin tillfredsställelse över de resultat som nåtts i det centrala FN-arbetet.

Det rådde enighet vid mötet om att arbetet inom delprojektet "Framework for environment statistics" fortsättningsvis borde bedrivas med starkt minskade insatser från sekretariatets sida. Med undantag för den nämnda genomgången av det s. k. COFOG-systemet blir arbetet närmast av bevakningskaraktär. Härigenom väntas det bli möjligt att intensifiera arbetet med metodfrågorna, som är ett av Sverige prioriterat område för det fortsatta internationella samarbetet. Man räknar vidare med att ett nystartat arbete med "miljöindikatorer" skall kunna få förhållandevis stora resurser.

Bengt Malmås

Informellt möte om kvalitetsredovisning m. m. (ECE)

Ett informellt möte om "Timeliness, Cost and Quality Attributes of Statistics" hölls i Genève 26–28 maj 1982. Mötet var en uppföljning av föregående års möte om metodproblem vid hushållsundersökningar. Sju länder samt ILO och EG var representerade. Från Sverige deltog Jörgen Dalén, SCB.

Mötet hade två huvudpunkter. Den första var "Guidelines for the presentation of the quality of statistics". Punkten var en uppföljning av en diskussion vid förra årets möte om detta ämne. Sverige hade författat det skriftliga bidraget till mötet betitlat "Some reflections on different approaches to the problem of presenting the quality of statistics", som jämförde Kanadas, USAs och Sveriges policies på området kvalitetsredovisning av statistik. Syftet med diskussionen var att enas om internationella principer, som kunde antas som "guidelines" för ECEs medlemsländer.

Mötet var framgångsrikt såtillvida som det enades om sammanfattande slutsatser som bl. a. innehöll en punkt där mötet ansåg att sammanfattande kvalitetsredovisningar skulle åtfölja varje förstagångsspridning av statistiska data. Nio punkter angavs som ett minimum för vad en sådan kvalitetsredovisning skulle innehålla.

Mötet rekommenderade CES att utveckla internationella "guidelines" för kvalitetsredovisning av statistik som innefattade mötets synpunkter i detta ämne. Ett förslag skulle föreläggas ett kommande möte.

Mötets andra punkt gällde Terminology and definitions of the various quality attributes of statistics. Ett papper från Spanien förelåg som underlag för diskussionen. Mötet informerades här om det arbete som utförs av en kommitté inom IASS (Internatio-

nal Association of Survey Statisticians). Mötet ansåg att detta arbete hade ett avsevärt intresse för de nationella statistiska myndigheterna och rekommenderade att konferensen skulle informeras regelbundet om de framsteg som kommittén gör och – om det är nödvändigt – hjälpa kommittén i dess arbete genom att sprida definitioner i utkast till nationella statistiska myndigheter för kommentarer.

Jörgen Dalén

Möte om utbildningsstatistik och utbildningsindikatorer (Unesco/ECE)

Inom ramen för Europeiska ekonomiska kommissionens (ECE) arbete hölls i Genève den 21–24 september 1982 i samarbete med Unesco det fjärde mötet om utbildningsstatistik och -indikatorer. På mötet var 19 länder representerade. Från Sverige deltog Leif Gouiedo (SCB). Dessutom var en representant från vardera Europeiska ekonomiska gemenskapen (EG) och OECD närvarande.

Till ordförande valdes R. Smulders, Nederländerna. På dagordningen stod huvudsakligen följande ämnen:

1) standardisering och internationell jämförbarhet av utbildningsstatistik
2) sociala indikatorer på utbildningsområdet
3) individualstatistik inom utbildningsområdet
4) prioriteringar i utvecklingen av utbildningsstatistik.

Standardisering av utbildningsstatistik

Detta ämne diskuterades med utgångspunkt i dels ett dokument som utarbetats inom Unesco-sekretariatet (CES/AC.23/27 ...), dels en rapport om vuxenutbildningen i Ungern 1970–1980. Allmänt konstaterades att den av Unesco utarbetade utbildningsklassifikationen – International Standard Classification of Education (ISCED) – huvudsakligen användes för

- statistikrapportering till internationella organisationer (Unesco, OECD, EG m. fl.)
- utveckling eller modifiering av nationella utbildningsstandarder
- utarbetande av handböcker och liknande över skilda utbildningssystem
- konstruktion av "nycklar" för översättning av nationella standarder till ISCED
- presentationen av statistik om utbildning i det "egna" landet.

Ytterligare ett användningsområde nämnde representanten från Nederländerna: där användes standarden för att registrera utbildningsnivån hos personal inom skilda administrativa sektorer[1]. Deltagaren från Finland omtalade att ISCED där användes för sammanfattning på engelska av inhemska utbildningsstatistiska resultat.

Diskussionen visade att det fanns behov av att få förtydliganden av flera delar av ISCED-systemet. För den skull önskade deltagarna att följande steg skulle tas inför den femte konferensen som man önskade ha inom tre år:

- Inom tolv månader skall medlemsländerna rapportera till Unesco vilka regler som tillämpas för att särskilja utbildningar på gymnasial och eftergymnasial nivå (ISCED-nivåer 5, 6 och 7).
- Inom den närmaste tvåårsperioden skall Unesco bilateralt diskutera förekommande avvikelser från ISCED.
- Utkast till nya internationella ISCED-handböcker av den typ som finns exempelvis för Storbritannien, Frankrike, de nordiska länderna, skall sändas till Unesco för synpunkter.
- Studier av användningen av ISCED på andra områden än elever är önskvärda, bl. a. för lärarstatistik.

Utbildningsindikatorer

Som bakgrundsmaterial till denna punkt på dagordningen hade sekretariatet ställt samman ett dokument (CES/AC. 23/28 ...), grundat på en enkät till medlemsländerna om förekomst av indikatorer avseende elever, lärare, utbildningskostnader, materiel och andra resurser.

[1] Vilket också kan anses gälla i Sverige genom att Svensk utbildningsnomenklatur (SUN) brukas för klassificering av skilda personalkategorier i offentlig (delvis också privat) tjänst. Det finns en översättningsnyckel SUN-ISCED.

Mötet informerades om det arbete som pågår inom statistikansvariga enheter i Förenta Nationerna, Unesco, OECD och EG i vad avser utvecklandet av sociala indikatorer. Bl. a. framgick det att Förenta Nationernas "indikatoransatser" under senare år knutits till arbetet med att utveckla en föreställningsram för socio-demografisk statistik (Framework for the integration of Social, Demographic and related economic Statistics, F.S.D.S.). Också de framsteg som gjorts för att mäta och "utvärdera" levnadsförhållanden beskrevs. Vidare fick deltagarna en redogörelse för de skilda aktiviteter som Unesco bedriver inom området för sociala indikatorer. När det exempelvis gäller indikatorer för utbildningsområdet, så refererade man till en 300-sidig bok som publicerats 1981 av International Institute for Educational Planning (IIEP) – ett organ, knutet till Unesco[1]. OECDs representant beskrev kort det projekt inom organisationen som syftade till att mäta välfärden; för närvarande bearbetas svaren på enkät till medlemsländerna med begäran om indikatorer för skilda företeelser inom utbildningen.

Vid den diskussion som följde, kom det fram att den ökade användningen av datorer har lett till en "sifferexplosion" som tvingar fram sammandrag av data för att informationen över huvud skall bli gripbar. Ett sätt att åstadkomma detta är att sammanfatta talen i form av indikatorer. Från flera håll framhöll man att tillgången på indikatorer var rikligare när det gällde data om "input" till utbildningssystemen än om "output" – här borde ett utvecklingsarbete komma till stånd!

Utbildningsstatistiska ID-system

Ett sekretariatsdokument (CES/AC.23/29 ...) och nationella rapporter från 11 länder (däribland Sverige) med beskrivningar av individualstatistiska system inom utbildningsområdet (omfattning, användning, design, presentationssätt etc.) togs till utgångspunkt för mötets diskussioner. På grundval av rapporterna konstaterade man att den vanligaste typen av ID-system avsåg elever och/eller lärare samt att ett fåtal länder hade sådana system för utbildningssökande, stocken eller flödet av examinerade. Man menade allmänt att de nationella rapporterna var av stort värde som informationskälla och att temat skulle tas upp igen på det framtida mötet om utbildningsstatistik (måhända i form av en sammanfattande – synoptisk – rapport).

Prioriteringsområden

Konferensen beslöt att rekommendera Unesco/ECE att ta initiativ till sammankallande av ett femte möte om utbildningsstatistik och -indikatorer inom de närmaste tre åren. Följande ämnen borde då behandlas:

a) Statistik om vuxenutbildning och utbildningsformer utanför de reguljära skolsystemen ("out-of-school education"). – Nationella rapporter bör utarbetas, varvid den ungerska studien om vuxenutbildning 1970–1980 kunde tjäna som förebild för utformningen av dessa rapporter.
b) Internationell jämförbarhet av utbildningsstatistik. – Här bör man ta som utgångspunkt den ansats i fyra steg som beskrivits i avsnittet ovan om standardisering av utbildningsstatistik.
c) Statistik om utbildningskostnader. – Flera medlemsländer arbetar med detta "statistikproblem". På nästa möte kan man därför få ett utbyte av erfarenheter.
d) Diskussion av metodologiska problem på det utbildningsstatistiska området. – Hit bör inräknas utbildningsindikatorer, individualstatistiska system[2], problem kring matrissystem, longitudinella studier, kohortundersökningar.

Leif Gouiedo

[1] Johnstone, James N.: Indicators of Educational Systems (Kogan Page Ltd., London/Unesco, Paris 1981).

[2] På mötet utarbetade en arbetsgrupp (deltagare från Storbritannien, Frankrike, Österrike, Sverige) ett utkast till en enkät för insamling av uppgifter om ID-system till nästa möte.

Nordiska statistiska utskott

Utskottet för nationalräkenskaper (NUNA)

NUNA-mötet hölls i Köpenhamn den 10–11 maj 1982. I mötet deltog Danmark, Finland, Norge, Sverige och – för första gången – även Island. Dessutom deltog Nordiska statistiska sekretariatet. Sverige representerades av statistikchef Åke Tengblad och avdelningsdirektör Nana Westerlund, båda från SCB. Mötet leddes av utskottets ordförande, Bengt Thage, Danmarks Statistik.

Nationella rapporter

En stående punkt vid utskottets möten är "nationella rapporter om nationalräkenskapsarbetet", till vilken samtliga deltagande länder hade bidragit med översiktliga dokument. Om utvecklingen av nationalräkenskaperna i de nordiska länderna kan särskilt omnämnas:

Utveckling av input/output

Dk. Årliga input-output-tabeller framställs och föreligger t.o.m. 1978. Investeringsmatriser för åren 1966–77 beräknas föreligga i slutet av 1982.

SF. Tabeller för 1979 publiceras under 1982. Utredning av input-strukturen år 1980 pågår.

S. Input-output-uppgifter för 1970–1980 i 1975 års priser håller på att framställas genom fram- resp. tillbakaskrivning av 1975 års tabeller med ledning av reviderade nationalräkenskapsdata. Samtidigt framställs tabeller för 1980 i löpande priser.

N. En strukturundersökning för byggnads- och anläggningsverksamhet har genomförts.

Utveckling av kvartalsräkenskaperna

N. Intensifiering, färdigt system hösten 1983.

S. Utbyggnad av redovisningarna även i löpande priser.

Övriga utvecklingsarbeten

Från *isländsk* sida redogjordes kort för de isländska nationalräkenskaperna och för några särdrag i den isländska ekonomin. Man arbetar bl.a. med övergången till det nya SNA-systemet[1]. Man har utvecklat metoder för produktionssideberäkningar i löpande priser och håller på att utarbeta metoder för fastprisberäkningar. Med hänsyn till Islands unika prisökningstakt är dessa metodstudier av mycket stort intresse.

Regionala produktionssideberäkningar har utförts i Finland och kommer att påbörjas i Norge. I Finland planeras dessutom regionala inkomst- och inkomstanvändningskonton.

Metodarbeten för fördelning på institutionella sektorer har påbörjats i Danmark och vidareutvecklats i Sverige. Utveckling av intäkts- och kapitalräkenskaper pågår i Norge och Finland. Danmark har utarbetat ett nytt beräkningssystem för preliminära nationalräkenskaper som medger redovisning av detaljerade uppgifter nio månader efter utgången av ett beräkningsår. Från svensk sida redogjordes för anledningen till korrigeringar av grundstatistiken – främst industrins s.k. lagervinster – som har medfört omfattande revideringar av nationalräkenskapsserierna.

Energibalanser anpassade till nationalräkenskaperna

Under denna punkt förelåg två dokument från Danmark, två från Sverige samt ett från Norge. Diskussionen rörde sig främst om hur de olika länderna lyckats insamla prisinformation samt hur "svinnet" på grund av energitransport och transformation skall beräknas.

Från svensk sida nämndes att konsumenterna börjat efterfråga totalbalanser och aggregerade energiflöden av typen "vad kostar en lastbil i energi räknat från det att järnmalmen bryts – osv. – till den färdiga lastbilen". Samtliga deltagare var eniga om att dessa frågor var de mest pådrivande för utvecklingsarbetet inom området.

[1] Helt genomförd i övriga nordiska länder.

Intäkts- och kapitalräkenskaper

Samtliga länder med undantag av Island hade lämnat dokumentation. Från isländsk sida meddelades att planer på utvidgade räkenskaper för den offentliga sektorn finns.

Det är intressant att konstatera att man i samtliga länder kommit igång med att bygga ut nationalräkenskapssystemen i rubricerad mening, men att endast de mest observanta konsumenterna – statsdepartementen, riksbankerna etc. – hittills visat större intresse för utbyggnaden.

Naturresursernas inverkan på kapitalräkenskaperna togs upp i sammanhanget. Men man enades om att det ingår så många osäkra och spekulativa element i naturresursberäkningarna, att de bör hållas utanför kapitalräkenskaperna.

Avgränsningen mellan s. k. institutionella sektorer diskuterades ingående. T. ex. att statliga affärsverk (SJ, Televerket, Vattenfallsverket etc.) hänförs till företagssektorn medan kommunal affärsdrivande verksamhet såsom avfallshantering, hamnverksamhet m. m. hänförs till den offentliga sektorn. I intäkts- och kapitalräkenskaperna har de olika länderna olika lösningar på problemet. I Norge t. ex. hänförs offentligt ägda banker till offentliga sektorn, i Sverige hänförs PK-banken till företagssektorn.

FNs förslag till ändring av sektorstrukturen behandlades och man ansåg att det har vissa analytiska fördelar framför nuvarande system, men att den praktiska tillämpningen skulle kräva betydande omläggningsresurser. Man enades om att en detaljerad internationell koordinering ej bör eftersträvas, eftersom förhållandena är så olika i olika länder.

Automatiska avstämningsmetoder

Under detta avsnitt förelåg dokument från Danmark och Sverige. Det konstaterades att – trots stora ambitioner – inget land kommit speciellt långt i sammanhanget.

Dilemmat är att nationalräkenskapsteorin förutsätter att differenser mellan produktions- och användningssideberäkning ej kan existera. Den statistiska verkligheten har visat sig vara helt annorlunda.

Efter lång diskussion blev man överens om att redovisningar som visar diskrepansen mellan produktionsside- och användningssideberäkning[1] är sundare än uppgifter utjämnade genom automatiska avstämningsmetoder.

Definitions- och tolkningsproblem i samband med SNA

Island bidrog med ett intressant dokument som behandlar många problem som kan anses utgöra nationalräkenskapernas primärproblem. Övriga länder lovade att sända dokumentation över hur problemen har lösts från fall till fall. Mötet var även enigt om att behandlingen av primärproblemen måste ha en framskjuten plats i ländernas utvecklingsarbeten.

Möjligheter för redovisning av lön under sjukdom

Mötet behandlade ett brev från Nordiska socialstatistikkommittén (NOSOSKO) till de nordiska chefsstatistikerna om särredovisning av löneutbetalningar vid sjukdom inom ramen för nationalinkomststatistiken.

Mötet konstaterade att dessa specialuppgifter inte insamlas i något av de nordiska länderna varför det skulle krävas ny primärstatistik från arbetsgivarna. Danmarks Statistik kommer emellertid att behandla frågan eftersom denna post ingår i EGs nationalräkenskapssystem.

Förslag till utskottets arbetsprogram 1982–1985

Man enades om att föreslå chefsstatistikermötet två utskottsmöten med nedanstående preliminära program. Från Islands sida lovade man att undersöka möjligheten av att förlägga ett av dessa möten till Reykjavik.

Möte 1983/84: Information från länderna om det nationella nationalräkenskapsarbetet
Revisionen av SNA
Intäkts- och kapitalräkenskaper
Byggnads- och anläggningssektorn
Handel och vinster (orienteringsämne)

[1] Tillämpas i Sverige.

Möte 1984/85: Information från länderna om det nationella nationalräkenskapsarbetet
Revisionen av SNA
Kapitalstocksberäkningar
Kvartalsräkenskaper
Sysselsättning och lön i nationalräkenskaperna (orienteringsämne).

Nana Westerlund

Europarådets befolkningskonferens

Vid Europarådets befolkningskonferens den 21-24 september 1982 deltog som delegater från Sverige statistikchef Gösta Guteland, SCB, professor Erik Bylund, Umeå universitet, avdelningsdirekör Arne Arvidsson, SCB, departementssekreterare Gunilla Bodin och avdelningsdirektör Sven Reinans expertgruppen för invandringsforskning (EIFO). Dessutom deltog som observatörer avdelningsdirektör Anne-Marie Bolander och byrådirektör Peter Springfeldt.

Konferensen syftade till att föra ut gemensamma karaktäristiska drag i den europeiska befolkningsutvecklingen och att ge information om den framtida verksamheten inom Europarådet på befolkningsområdet.

Konferensen var den tredje befolkningskonferensen i Europarådets regi. De tidigare hölls 1966 och 1971. Under de senaste tio åren har befolkningsutvecklingen kommit in på nya banor. Det har därför av många uppfattats som väsentligt att samla experter för att på nytt diskutera befolkningsfrågorna. Det betonades att det var viktigt att konferensen skulle präglas av diskussioner mellan experter. Endast ett begränsat utrymme ägnades åt slutsatser och rekommendationer, Europarådets kommitté för befolkningsfrågor (CDDE) kommer senare att diskutera erfarenheterna från konferensen.

Den demografiska situationen

Från svensk sida kan det vara värt att notera att diskussionerna om konsekvenserna av det minskade barnafödandet var mycket nyanserade. Även om det fanns skilda åsikter om orsakerna till de låga födelsetalen fanns det enighet i bedömningen att det viktiga är att nå en jämn befolkningsutveckling oavsett om den ger ökad, minskad eller oförändrad befolkning. Det är de kraftiga svängningarna i födelse- och dödstal och migration som skapar problem i samhällsplaneringen.

Kvinnornas dubbla roller ägnades stor uppmärksamhet. "Kvinnor önskar bli mödrar men vill inte lämna arbetsmarknaden" framhöll konferensens ordförande Norges tidigare statsminister Gro Harlem-Brundtland i sitt inledningsanförande.

Vikten av att var och en skall få välja fritt mellan olika roller betonades.

Befolkningsfrågorna i politiken

Få länder säger att de för en befolkningspolitik. Det framgår av den sammanställning av de rapporter från medlemsländerna som tagits fram inför konferensen. Många länder sätter emellertid in åtgärder för att förbättra barnfamiljernas levnadsförhållanden.

De skandinaviska länderna framförde i ett gemensamt uttalande att det råder stor osäkerhet om effekterna på befolkningsutvecklingen av olika politiska åtgärder. För att underlätta för barnfamiljerna bör man emellertid kunna utnyttja sådana medel som barnomsorg, förkortad arbetstid, finansiellt stöd och stimulans för fäder att ta ett större ansvar.

På den här punkten finns det emellertid åsiktsskillnader. Flera delegater framhöll t. ex. vikten av att ge stöd till hemarbetande kvinnor.

När invandringsfrågorna togs upp betonades vikten av att statistiken förbättras. Det kunde också noteras att de flesta länderna inom Europarådet i dag betraktar sig som invandrarländer.

Allmän slutsats

En allmän slutsats från konferensen är att vi med den utveckling som vi har måste ägna större uppmärksamhet åt befolkningsfrågorna. Effekterna på skilda områden inom samhället bör studeras. Till en del måste vi

också ändra vår inriktning av samhällsplaneringen. Hittills har den varit utformad för en ständig befolkningstillväxt. Nu måste den utgå från att folkmängden inte ökar längre.

Gösta Guteland

Seminarium om frågekonstruktion

Den 20–24 september 1982 hölls en "workshop" om frågekonstruktion i York, England. Den anordnades av Social and Community Planning Research (SCPR) Survey Methods Centre. Tonvikten var lagd på de problem som är förknippade med frågekonstruktion. För- och eftermiddagarna ägnades åt föreläsningar med diskussioner. Två kvällar avsattes till dels en diskussion om "Sexism in Survey Definitions" (ska begrepp som hushållsföreståndare, huvudman etc. överges?), dels en demonstration av hur man kan använda gruppdiskussionen som ett hjälpmedel för frågekonstruktion. Det sistnämnda var mycket intressant och vi hoppas kunna pröva metoden även här. Föreläsningarna behandlade följande ämnesområden:

- Öppna frågor kontra frågor med fasta svarsalternativ. Öppna och "slutna" versioner av samma fråga framkallar ofta olika svar. Vad beror det på och när händer det? Vilka överväganden bör frågekonstruktören ta innan han beslutar sig?
- Kontexteffekter i förkodade frågor. Föreläsningen innehöll bl. a. en översikt av de resultat som börjar komma från intervjuarvariabilitetsexperiment, split-ballot-studier och interaktionskodningar vad gäller frågeordningseffekter, svarsordningseffekter, användande av "ingen åsikt" och mittpunktskategorier etc.
- Attitydmätningar. Föreläsaren behandlade framförallt attitydmätningsproblem, men tog även upp värdet av socialpsykologiskt arbete i attitydteori och attitydmätningar. Olika teorier om attityders beskaffenhet har olika innebörd för frågekonstruktion och analys, speciellt för analys av samband och av attityder som leder till beteende. Några problem och möjliga praktiska lösningar till dessa presenterades. Fishbeins teori diskuterades närmare.
- Komplexa preferensfrågor. "The priority evaluator method" har utvecklats av SCPR i ett försök att identifiera det sätt på vilket människor byter ut en preferens mot en annan. Det finns även fler tekniker för att undersöka komplexa och motstridiga preferenser. Under föreläsningen diskuterades frågor och problem som hänger samman med preferensmätningar.
- Analys av svarseffekter. Föreläsningen behandlade skattning av svarsfelen, de korrelerade svarsfelen och biasen i surveyintervjuer. Vidare genomgång av studier av svarsattityder och motivation och en analys av några tekniker som dels ska underlätta den svarandes uppgift dels ska öka hans motivation.
- Förbättring av svarsvaliditeten. Föreläsningen innehöll bl. a. en beskrivning av forskning rörande utveckling och utprovning av tekniker som "lär" den svarande vad som fordras av honom för att genomföra uppgiften och även ökar hans förmåga att göra det. Innebörden av dessa tekniker för intervjuarutbildningen och supervisor-systemet diskuterades också.
- Faktafrågor. Vad är det vi kommer ihåg? En survey ställer ofta stora krav på den svarandes minne. Hur kan vi bygga in i formulärkonstruktionen några av de kunskaper som kommit fram genom experimentell psykologi vad gäller minne och glömska?
- Klassificeringsdata (bakgrundsdata). Några frågeområden (civilstånd, yrke, inkomst, m. fl.) förekommer i de flesta undersökningar. Trots den frekventa användningen varierar sättet de används på mycket. Föreläsningen handlade om en del av dessa variationer och de problem med definitioner och med datainsamling som lett fram till variationerna.
- Frågekonstruktionsproblem som speciellt rör telefonintervjuer. Föreläsningen handlade framförallt om förhållanden i USA och tog upp en översikt över den

forskning som pågår där.

Workshopen avslutades med en allmän diskussion på temat: Vilka är de mest framträdande problemen?

Gunilla Davidsson *Ingrid Friberg*

Expertmöte om miljöstatistik (FN)

I mötet, som hölls i New York den 20–24 september 1982, deltog ett tiotal personer, bl. a. undertecknad. Diskussionerna avsåg i första hand ett rapportutkast "Framework for Environmental Statistics", som i färdigt skick kommer att fungera som ett verktyg för att utveckla och organisera miljöstatistik – kanske främst i länder som befinner sig i början av sitt arbete på detta område. Kärnan i rapporten utgörs av en klassificering av miljöinformation dels efter händelser och verksamheter/miljöeffekter/följdeffekter och åtgärder, dels efter typ av miljö, med mark/vatten/luft/ekosystem/samhällen som viktigaste beståndsdelar. Användningen av detta schema hade med framgång prövats i tre länder (Dominikanska republiken, Fiji, Kenya). I den kommande versionen av rapporten avses att beskrivningar av hur man kan utnyttja schemat och tillhörande specificeringar skall få större utrymme. FNs statistiska kontor avser vidare att bl. a. arbeta med detaljerade riktlinjer för vissa viktiga områden inom miljöstatistiken.

Thomas Polfeldt

Konferens om dynamiska analysmöjligheter inom inkomststatistiken

Konferensen anordnades av "London School of Economics and Political Science" och ägde rum den 24–25 juni 1982 i London. Vid konferensen deltog representanter från Frankrike, Italien, Irland, Japan, Schweiz, Storbritannien, Sverige, USA och Västtyskland. Sverige representerades av Anders Björklund, Industrins utredningsinstitut, Björn Gustafsson, Göteborgs universitet, Carl-Axel Hammarlund, SCB och Anders Klevmarken, Göteborgs universitet.

Under de två dagarna presenterades och diskuterades PM där den gemensamma nämnaren var studier av olika ekonomiska förhållanden över tiden. Några av dessa PM innehöll modeller för denna typ av studier, det gällde såväl linjära som icke-linjära modeller. Bland de förhållande som studerades kan nämnas pensionsförhållande för tjänstemän under en 30-årsperiod (Frankrike), löneförhållande med stort urval och kort tidsrymd (Frankrike), livsinkomster (Västtyskland), jämförelse mellan ålder och lön i Japan, Sverige och USA (Sverige), ålder och förmögenhet (Frankrike), utbildning och lön (USA), problem med familjesammansättning vid analys av hushållsinkomster (USA), samband mellan inkomst och yrke för två generationer (Schweiz), lön och arbetsutbud (USA) samt nivå och utveckling av löner (Frankrike).

Carl-Axel Hammarlund

Världskonferens i programmeringsspråket APL

Association for Computing Machinery (ACM) är en sammanslutning för folk med intresse i computer science. Inom ACM finns ett antal s. k. "special interest groups" som var och en är inriktad på något delområde inom computer science. En av dessa grupper, SIGAPL, har programmeringsspråket APL som sitt intresseområde.

Varje år arrangerar SIGAPL en världskonferens där användare och utvecklare av APL träffas för att utbyta erfarenheter. 1982 var konferensen förlagd till Heidelberg i Västtyskland, 26–30 juli. Drygt 500 deltagare från ett 20-tal länder hade mött upp. Från SCB deltog Gösta Nilsson P/DBM.

Ca 80 författare till artiklar rörande APL hade inbjudits som föredragshållare. För att hinna med så mycket på så kort tid pågick presentationerna parallellt tre och tre. De ämnen som avhandlades kan indelas i följande grupper:

- språket APL, utvidgningar, standards
- programmeringsteknik
- statistiska matematiska tillämpningar
- databaser

- APL på mikrodatorer
- utbildning
- applikationer.

Dokumentation från samtliga föreläsningar, "APL 82 conference proceedings", finns hos enheten för databehandlingsmetoder vid SCB.

Under senare år har en ny och mera generell representation av datafält, "nested arrays", börjat implementeras i APL och utnyttjas i applikationer. Detta var ett ofta återkommande tema i föreläsningar och diskussioner. Standards har ännu inte fastställts men är under utarbetande och förväntas föreligga i slutet av 1983. Nested arrays öppnar helt nya och bättre möjligheter för datastrukturering och programmering i APL.

Ett annat intressant område för utvecklingen av APL var förslagen till förbättrad felhantering, "error trapping". Error trapping finns implementerat på många maskiner och kan användas för att undvika att slutanvändaren av system utvecklade i APL får obegripliga avbrott och felmeddelanden.

Andra ofta förekommande ämnen var APL på mikrodatorer och APL och relationsdatabaser.

Under konferensen pågick en utställning av maskin- och programvarufabrikanter och serviceföretag. Stor uppmärksamhet väckte förstås IBMs nya version av APL, "APL2", som innehåller nested arrays, error trapping, nya primitiva funktioner för stränghantering och en del andra förbättringar jämfört med tidigare APL-versioner på IBM-maskiner. APL2 är ännu ej kommersiellt tillgänglig och finns implementerat endast under operativsystemet VM/SP under CMS. IBM räknar dock med att när APL2 blir tillgängligt för marknaden, 1983-1984, ska den finnas även för andra operativsystem, exempelvis under OS/MVS under TSO. Tyvärr har man ännu inte gjort något åt filhanteringen. Förhoppningsvis kommer även denna att bli bättre.

APL 68000 kallas en APL som utvecklats för 16-bits mikrodatorer med Motorolas processor. Denna APL innefattar helt APLSV så när som på filhanteringen. Detta gör att system kan flyttas från stordator till mikrodator, vilket har testats och visat sig fungera. Mikrodatorerna arbetar vanligen med enklare och effektivare filhantering (komponentfiler). Primärminnesstorlekar på över 1 megabyte är vanliga vilket gör att workspacestorleken inte heller utgör någon begränsning.

Slutligen kan konstateras att intresset för och användandet av APL fortsätter att öka. Bara från Sverige deltog ett 30-tal personer från drygt 20 olika organisationer. Man kan förvänta sig att utvecklingen fortsätter då systemutvecklings- och programmeringstiden oftast är väsentligt kortare med APL än med andra programmeringsspråk.

Gösta Nilsson

Nordiskt olycksfallsforskningsseminarium 1982

Den 7–10 sept. 1982 hölls Nordiskt olycksfallsforskningsseminarium (NOFS-82). Platsen för seminariet var Danmark och det vackra Gilleleje på Själlands nordspets. Seminariet hade samlat drygt 40 deltagare. SCB representerades av byrådirektör Ewa Westman.

NOFS bildades för ungefär fem år sedan och var då helt inriktat på arbetsolyckor och möjligheterna att förebygga sådana. Under de senaste åren har seminariet utvidgats till att avhandla även olyckor inom andra områden. 1981 hade de s. k. hem- och fritidsolyckorna fått ett visst utrymme. Nu hade även trafikolyckorna tagits med – om än dåligt representerade vad gällde antalet trafikolycksfallsforskare.

Temat för NOFS-82 var "Olycksstatistikens möjligheter att användas i olycksförebyggande arbete". Som exempel på befintliga statistiksystem redovisades sjukhusregistreringen av hem- och fritidsolyckor i Odense, den officiella statistiken över vägtrafikolyckor i Norge och registreringen av arbetsskador i Sverige. Hur olycksstatistik kunnat användas inom de tre områdena i förebyggande arbete gavs också exempel på. Eftermiddagarnas gruppdiskussioner ägnades åt olika delfrågor inom seminariets tema.

Under programrubriken "korta inlägg" presenterade jag den täckningsgradsundersökning avseende vägtrafikolyckor som SCB för närvarande genomför. Deltagarna visade stort intresse för att få ta del av resultaten, som beräknas föreligga hösten 1983.

NOFS-82 kom att ytterligare stärka den inriktning som NOFS fått under senare år, nämligen att det är nödvändigt med ett nära samarbete mellan de tre olycksområdena. Inför nästa NOFS, som föreslogs äga rum våren 1984 i Finland, kommer man därför att se till att fler representanter för trafiksidan deltar.

Ewa Westman

Statistical Computing Project: Table Generation

Det sjätte mötet med "Table Generation Joint Group" ägde rum i Bukarest den 26–29 oktober 1982. Directia Centrala de Statistica arrangerade det hela i verkets byggnad (med fotograferingsförbud) i centrala Bukarest. Produktionen och datorcentralen (ICL) ligger utanför staden. I mötet deltog 18 personer från 9 länder. Sveriges representant var Sven Felme.

Gruppen har till första uppgift att granska, utvärdera och jämföra standardiserade tabelleringssystem som med någon framgång används av världens statistikbyråer. Hittills har engelska TAU, amerikanska TPL och svenska TAB68 beskrivits med hjälp av en omfattande kriterielista. Förhoppningen är att ytterligare kunna redogöra för åtminstone amerikanska CENTS. Utvärderingarna baseras på systemens dokumentation, användares erfarenheter och intervjuer om systemens inre uppbyggnad, som ju ofta är dåligt dokumenterad eller svåråtkomlig. Ett färskt exempel på det senare har vi inom TAB68, där det helt nyligen upptäcktes att originalsystemet ackumulerar till summacellerna på mikronivå, dvs. synkront med ackumuleringen till grundcellerna. Om man i stället bildar summorna först sedan grundcellerna är klara, på makronivå, ger ett exempel 40% tidsvinst.

Gruppens andra uppgift är att, om inget befintligt system bedöms utvecklingsbart mot gruppens globala krav, lägga specifikationer för ett eller flera heltäckande system. Ganska tidigt startade detta arbete, det är ju alltid roligare att hitta på eget, än att gräva i andras. Detta nya system kallas INTERTAB, och är nu hyggligt dokumenterat vad beträffar de yttre egenskaperna; något sämre, eller inte alls, vad beträffar de inre systemlösningarna. Systemet torde bli det sista i sin generation, ett epitafium över batchepoken i stordatorerna.

Gruppens tredje uppgift är mestadels självpåtagen; man beslöt nämligen i Bukarest att även försöka *tillverka* ett fungerande INTERTAB, i egen regi och med egen arbetskraft. Under år 1983 ställer sex länder nio manår till förfogande, och vidare har ECE stött projektet med 30000 dollar. Att implementera ett sådant här komplext system är ett vågstycke inom *en* agentur, så mycket djärvare då att dra igång "an international joint venture"!

Sverige är bara observatör i gruppen, och kan därför bara önska lycka till med projektets fortsättning.

Sven Felme

PL/1-GUIDE-möte

I detta möte, som hölls den 10–11 maj 1982 i Lyngby, Danmark, deltog från SCB Matts Bergwall och Bernt Svensson. Vid denna konferens deltog PL/1-användare från Danmark, Finland och Sverige.

Första dagen gjordes ett besök på IBM i Kongens Lyngby. Där demonstrerades GDDM, som är ett helt nytt system för interaktiv konstruktion av grafiska bilder. Exempel visades i färg på olika typer av stapeldiagram, kurvor och tårtdiagram. Kopplingen till PL/1 berördes.

Andra dagen ägnades åt själva sammanträdet. Bland de ämnen som behandlades kan nämnas: PL/1-utbildning, dels IBMs två nya kurser och dels Statskonsults uppföljning av release 4.0/6.0 samt dumpläsning.

Avslutningsvis bestämdes tid, plats och agenda för nästa möte.

Matts Bergwall *Bernt Svensson*

Seminarium om kriminalstatistik

Ett nordiskt seminarium om kriminalstatistik hölls den 14–17 september 1981 på Utö. Det var första gången ett seminarium inom detta ämnesområde arrangerades på nordiskt plan. Syftet med seminariet var att statistikproducenter, forskare och företrädare för myndigheter inom rättsväsendet skulle få tillfälle att träffas och diskutera olika aspekter på kriminalstatistiken.

I seminariet deltog 53 personer från Danmark, Finland, Norge och Sverige.

Programmet för seminariet hade utarbetats inom nordiska utskottet för kriminalstatistik (NUK) och godkänts av de nordiska chefsstatistikerna.

Programmet var indelat i följande fyra huvudområden:

1. Kriminalstatistiken under 1980-talet.
2. Nordisk kriminalstatistik 1950–80. Presentation av en ny publikation.
3. Kriminalvårdsstatistik.
4. Longitudinella studier och återfall i brott.

Som utgångspunkt för diskussionerna hade utarbetats ett antal rapporter. Dessa presenterades av resp. författare varpå en allmän diskussion följde. I det följande sammanfattas diskussionerna inom de olika huvudområdena.

1. Kriminalstatistiken under 1980-talet

Under denna punkt redovisade man från olika utgångspunkter vilka visioner man hade då det gäller utvecklingen under 1980-talet och vilka statistikbehov man förutsåg.

Från forskarhåll tecknades en samhällsutveckling som karakteriseras av bl. a. ett allt kärvare ekonomiskt läge, en allt oöverskådligare statsapparat, en ökande klyfta mellan stat och individ, en minskande informell social kontroll individerna emellan. Man förutsåg även en informationsexpansion på olika områden, fortsatt teknisk utveckling som bl. a. innebär ökade möjligheter att med exempelvis minidatorer analysera statistisk information.

Den nuvarande kriminalstatistiken ansågs behöva kompletteras med mer information om kontrollapparaten. Parallellt med utvecklingen av brottsligheten bör man följa utvecklingen av myndigheternas resursinsatser för kontroll. Frågan om vilka som blir offer för kriminalitet behöver belysas utförligare än hittills. Vidare saknas närmare information om de personer som begår brott. Här har utvecklingen i ett längre historiskt perspektiv varit att allt färre uppgifter om brottsförövarna statistikförs.

En särskild diskussion ägnades frågan om statistikens tillgänglighet. Behöver verkligen så mycket av informationen tas in i publikationer? För att tillgodose olika användares ofta vitt skilda behov kan det vara en fördel att i större omfattning än hittills lagra information i databaser som man kan vidarebearbeta. Vidare förutsågs kommande behov av att standardiserade socialindikatorer tas fram.

Företrädare för justitiedepartementen redogjorde för behoven av statistik för planering och uppföljning. Statistik används i ökad omfattning för styrning av resurser inom rättsväsendet. Viktiga krav på statistiken i dessa sammanhang är bl. a. att den är aktuell. Behov finns att kunna följa kriminalitetsutvecklingen igenom hela rättssystemet. Vilka effekter på antalet fängelseintagningar får exempelvis en lagändring som innebär straffskärpning? Dylika studier försvåras ofta av att det förekommer så många och skiftande redovisningsenheter inom kriminalstatistikens olika grenar (brottsanmälningar, brott, misstänkta personer, mål, ärenden, lagföringar, dömda personer etc.).

Pågående arbeten med att utveckla de administrativa rutinerna inom rättsväsendet leder successivt till att myndigheterna får bättre hjälpmedel för planering och analys. ADB-teknik utnyttjas på olika sätt. Mikrodatortekniken börjar användas allt mer.

En fråga som ägnades särskild uppmärksamhet var anpassningen av statistiken till olika användarkategorier. Hittills har i princip alla användare erbjudits samma utbud. I framtiden kommer statistiken i större utsträckning att behöva skräddarsys till olika specifika behov. All statistik behöver med nödvändighet inte publiceras. Det väsentliga är att kunskap om informationen sprids

och att statistiken finns tillgänglig då den behövs. Statistikproducenterna kommer i framtiden alltmera att få roller som informationsförmedlare än som rena sifferproducenter.

2. Nordisk kriminalstatistik 1950–78

Under denna punkt presenterades en provutgåva av en sammanställning över nordisk kriminalstatistik 1950–78. Denna hade tillkommit för att tillgodose föreliggande behov av jämförande data inom kriminalstatistikens område. Dylika data behövs för såväl kriminologiska som kriminalpolitiska jämförelser.

Rapporten hade sammanställts i Sverige i samarbete med de statistiska centralbyråerna i Danmark, Norge och Finland. Den omfattar ett urval av brottstyper och innehåller uppgifter om anmälda brott, uppklarade brott, uppklaringsprocent, antalet för brott lagförda personer samt påföljdernas fördelning.

Vid presentationen av rapporten redogjordes för ansatsmöjligheter, analystyper, val av brottstyper m. m. Vidare redogjordes för vilka tänkbara användningsområden som finns för publikationen.

Utöver de rena sifferbeskrivningarna innehåller rapporten en utförlig dokumentation av skillnaderna mellan ländernas lagstiftning. Vidare dokumenteras de rutiner som ligger till grund för resp. lands kriminalstatistik. En stor enighet rådde bland seminariedeltagarna om publikationens användning som handbok för den som vill jämföra kriminalitetsutvecklingen mellan de nordiska länderna. Några förslag till kompletteringar och förbättringar av innehållet framfördes.

3. Kriminalvårdsstatistik

Här redovisades från finsk sida tre projekt. Det första avsåg nordiskt samarbete inom fångvårdsförvaltningen med syftet att snabbare än hittills få fram jämförbara uppgifter om antalet fångar i de nordiska länderna. Hittills har arbetet lett till att man framställt ett tiotal tabeller. Dessa presenterades och kommenterades. Seminariedeltagarna fann tabellerna vara av mera allmänt intresse och det föreslogs att tabellerna borde ges en större spridning exempelvis genom publicering i någon facktidskrift.

Ett annat projekt avsåg den finska fångvårdsstatistiken och dess användning. Statistiken är innehållsmässigt sett relativt lika den svenska motsvarigheten. Såväl intagningar, avgångar och beläggningsuppgifter redovisas löpande. En brist i redovisningen är bl. a. att uppgifter om fångarnas brottsbakgrund i stort sett saknas.

Särskild uppmärksamhet ägnades frågan om varför Finland under hela 1900-talet haft markant högre antal fångar än övriga nordiska länder. Delvis förklaras detta av att man i Finland har hårdare straff för t. ex. rattfylleri. Vidare dömer domstolarna generellt sett ut längre fängelsestraff jämfört med övriga länder. En annan förklaring är att avvikande beteende i många fall sanktioneras med icke-straffrättsliga frihetsberövanden och att praxis härvidlag varierar mellan länderna.

Från svensk sida redovisades ett försök att finna empiriskt baserade mått på frihetsberövande påföljder. Här pekades bl. a. på svårigheterna att definiera vilka typer av påföljder som lämpligen bör ingå i en nordisk jämförelse. Statistiken över frihetsberövanden är bristfällig i samtliga länder och det bedöms inte sannolikt att man inom överskådlig tid kommer att ha en fullständig sådan i samtliga länder. Däremot efterlystes någon form av planlagd specialredovisning beträffande vissa former av frihetsberövanden. En sådan skulle möjliggöra uppföljning av förändringar i tillämpningen av regelsystemet.

4. Longitudinella studier och återfall i brott

Under denna punkt diskuterades olika problem som är förknippade med försök att åstadkomma en statistik över återfall i brott. I Sverige har man sedan slutet av 1960-talet, då ADB infördes hos SCB, producerat återfallsstatistik löpande. Till en början omfattade statistiken enbart den grövre brottsligheten men numera ingår även lindrigare brott.

För att kunna producera återfallsstatistik behövs data för ett antal år med bl. a. upp-

gifter om tidigare brottslighet. Detta är nödvändigt för att kunna definiera debutantbegreppet. Särskilda svårigheter sammanhänger med att kriminalitetsuppgifter saknas för ej straffmyndiga personer. Även mätproblemen erbjuder svårigheter. Hur långa uppföljningsperioder skall väljas? Vilka återfallskriterier skall väljas?

De största svårigheterna gäller valet av populationer. Statistiken syftar framför allt till att beskriva dimensionerna individ, brott och påföljd. I det nya återfallssystem som byggts upp inom SCB har informationen lagrats så att man kan välja ut delpopulationer med såväl brott, påföljder som individer som enhet. Härigenom kan systemet användas för följande olika syften:

- beräkna återfallsfrekvenser för delpopulationer
- belysa genomströmning av individer i rättssystemet
- belysa brottskarriärer och återfallsmönster
- beskriva myndigheters praxis
- söka belysa påföljdssystemets relativa effektivitet.

Som exempel på hur det nya återfallsregistret kan användas presenterades en ansats med s. k. födelsekohortstatistik.

Från dansk sida redogjordes för fängelsestatistikregister, dels på riksnivå och dels på lokal nivå. Jämfört med liknande register i Sverige innehåller de danska motsvarigheterna mer s. k. mjukdata.

En särskild programpunkt ägnades åt modellbaserade flödesstudier av rättssystemet. Dylika ansatser har gjorts vid statistiska institutionerna i Stockholm resp. Umeå. Den första ansatsen innebär att man med hjälp av en stokastisk modell kan beskriva en slumpmässigt vald individs observerade brottsbana. Modellen kan användas för såväl analyser av brottsbeteenden som i prognossyfte. Den andra ansatsen innebär utveckling av en kalkylmodell med vars hjälp man kan beräkna vilka konsekvenser kriminalpolitiska åtgärder får bl. a. på anstaltsbeläggningen.

Sammanfattningsvis kan konstateras att seminariet har varit givande, inte minst för oss som producerar kriminalstatistik. En utförlig dokumentation av seminariet har sammanställts av Nordisk statistisk sekretariat (Tekniske rapporter nr 28).

Torbjörn Israelsson

NORDFORSKs kontaktforum för FoU-statistik

NORDFORSKs (Nordiska samarbetsorganisationen för teknisk-naturvetenskaplig forskning) kontaktforum för FoU-statistik är en samarbetsgrupp mellan producenter av FoU-statistik i Norden. Dess arbete är inriktat på att finna gemensamma metodologiska lösningar, att skapa en förbättrad kommunikation mellan statistikproducenter och -konsumenter samt att ha gemensamt uppträdande i internationella sammanhang, såsom OECD och UNESCO.

Kontaktforumet har haft två möten under 1982. Vårmötet ägde rum den 14–16 juni och var förlagt till Island, Laugarvatn. Höstmötet hölls den 2–3 december i Stockholm.

Vid båda mötena var samtliga nordiska länder liksom NORDFORSK representerade. Representationen från Sverige under vårmötet bestod av Esbjörn Turesson, SCB samt undertecknad. Under höstmötet, som för svenskt vidkommande inte var förenat med tjänsteresa utomlands, var representationen från Sverige utökad och bestod, förutom av de två som deltog i vårmötet, även av Tage Berglund, SCB och Lennart Palm, STU (Styrelsen för teknisk utveckling).

Kontaktforumets ordförande under innevarande mandatperiod är Kari Ajer, NTNF, (Norges Teknisk Naturvitenskapelige Forskningsråd), Norge medan Erik Edvardsen, NTNF Norge är sekreterare.

Arbetet under året har bl. a. resulterat i publiceringen av en s. k. FoU-trend med rubriken: FoU-insatserna i Norden ökade internationellt sett. Trenden utkom i augusti.

Därutöver har gjorts en sammanställning av de nordiska ländernas redovisning av resp. lands FoU-insatser år 1979. Sammanställningen kommer inom kort att publiceras under titeln: Nordisk FoU-statistik for

1979 og Statsbudgetanalys 1981. Publiceringen ombesörjs av NORDFORSK.

Det viktigaste arbetet under året har bestått av en revidering och aktualisering av "Retningslinjer for analyse af statslige bevillinger til forskning och utveiklingsarbejde" – dvs. manualen för de årliga beräkningarna av de statliga FoU-insatserna så som de kommer till uttryck i budgetpropositionen.

En etapp i detta arbete uppnåddes under junimötet till vilket en inventering hade företagits av problem förknippade med möjligheterna att uppnå jämförbarhet mellan de olika ländernas statsbudgetanalyser.

Esbjörn Turesson har i detta sammanhang varit koordinator och huvudansvarig för arbetets genomförande.

En viktig del i denna revidering avser en översyn av NORDFORSKs klassifikationssystem i samhällsrelevanta ändamål. Det beslöts att denna del av arbetet skall genomföras nästa år.

Det beslöts också att manualen skall ges ut i ny utgåva med de ändringar och tillägg som forumet beslutat införa till följd av det genomförda arbetet.

Vidare togs beslut om innehållet i kommande FoU-trender, bl. a. en om FoU inom elektronikindustrin samt en om det statliga engagemanget i industrins FoU.

I övrigt diskuterades ett gemensamt nordiskt agerande i samband med ett i februari planerat OECD-möte angående uppläggningen av en publikation: Science Indicators II.

Birgitta Bergström-Balkeståhl

Internationell kartografisk konferens

The International Cartographic Association (ICA) höll sin elfte konferens i Warszawa den 29 juli–4 augusti 1982. I konferensen deltog 371 personer från 37 länder. En av de 14 deltagande svenskarna var Erik Liljegren, SCB.

Den politiska situationen i Polen fick direkta konsekvenser för konferensen genom den bojkott som amerikaner, kanadensare och flera fransmän genomförde genom att helt enkelt stanna hemma trots att de anmält sitt deltagande. Detta medförde många ändringar i dagordningen och flera av de intressantaste punkterna fick helt utgå. Det för en statistiker mycket intressanta ämnet om datorritade tematiska kartor drabbades givetvis hårt eftersom de ledande kanadensarna uteblev. En av de mer intressanta föredragningarna blev nu presentationen av THEMAK, som är en programvara för kartritning på elektrostatiska plotters som utvecklats vid Freie Universität i Berlin av Bollman och Freitag. Med programvaran kan man rita samtliga vanliga typer av kartor. Den kanske främsta fördelen med THEMAK är den korta inlärningstiden, endast 2–3 veckor. THEMAK är dessutom inte beroende av något speciellt fabrikat på hårdvaran. Under samma ämne presenterade professor David Rhind från London University en blivande tematisk atlas för EG-området. Atlasen kommer att framställas med programpaketet GIMMS, som även finns på SCB. Rhinds uppsats tar upp en rad problem förknippade med hierarkiska datastrukturer och nivån på den geografiska uppläsningen samt problem som indelningsändringar medför. Samma problem finns även i Sverige. Jämför t. ex. ändringarna i kommunindelningen under 1970-talet. Rhinds problembeskrivning och hans sätt att lösa dessa är klart intressanta för SCB.

Konferensens höjdpunkt var givetvis presentationen av det svenska bidraget "Computer-assisted Population and Census Mapping for Local and Regional Planning data, methods and techniques used in Sweden", författad av Rystedt-Szegö-Wastesson samtliga från centralnämnden för fastighetsdata (CFD). Uppsatsen behandlar såväl samarbetet mellan kartografer-statistiker och lantmätare i producentledet som de tekniska framställningsproblemen. Uppsatsens tyngdpunkt består dock av en detaljerad redogörelse för hur befolkningskartor använts i en konkret planeringssituation i detta fall hur kompletteringstrafiken i Hultsfreds kommun planlades. Det svenska bidraget fick ett mycket gott mottagande.

Trots den massiva militära närvaron på Warszawas gator blev även den icke-sche-

malagda konferenstiden med informella kontakter med övriga deltagare mycket givande. Det starkaste minnet från konferensen är emellertid de öppet trotsiga demonstrationerna mot militärledningen som den polska civilbefolkningen genomförde dag och natt på Segertorget mitt i Warszawa.

Erik Liljegren

Databehandling i statistiken (COMPSTAT 82)

Den 30 augusti–3 september 1982 träffades ca 600 personer i Toulouse för att byta åsikter om databehandling i statistiken (COMPSTAT 82). De flesta nationaliteter var representerade. Såväl näringsliv som offentliga institutioner hade skickat dit folk. Från SCB deltog Aino Widlund och Lars Nordbäck, S/SYS-S, för att informera om AXIS och för att skaffa närmare information om standardprogramvara för statistikbearbetning.

AXIS på COMPSTAT 82

På COMPSTAT 80 i Edinburgh hade SCB en monter, vilket var tungrott på två personer. 1982 satsade vi på schemalagda, väl annonserade aktiviteter, med möjligheter att följa intressanta föreläsningar m. m. emellan. Vi började emellertid med en helt oplanerad kupp. Stundens ingivelse. I förväg hade vi skickat ned 38,5 kg dokumentation av olika valörer. I handbagaget bar vi ytterligare material. Dels affischer, nya och fräscha, dels ca 150 blad med utdrag ur tidningen Sweden Today, där en artikel om och en annons för AXIS var med. Vi kommer av någon anledning ut på balkongen till salen där invigningen av COMPSTAT 82 högtidligen skall genomföras. Väskan med material börjar kännas tung, sedan den kompletterats med två tunga proceedingsböcker. Vi slår oss ned och tittar ut över den av delegater till bredden fyllda salen. Plötsligt står lösningen helt klar vad den tunga väskan anbelangar. Släng ut 150 reklamblad om AXIS som flygblad, så är det klart. Personerna som sitter där nere är ju dem som vi vill nå. AXIS in the air. Nåväl, en kortare konferens om förslaget menade att inriktningen var bra medan genomförandet var något burdust. Beslöts att vi i stället skulle ställa oss vid salens två utgångar och dela ut de 150 bladen så långt de räckte. Bladen gick åt som smör i solsken och lästes ivrigt under första rasten på COMPSTAT 82.

Den ovan nämnda affischen sattes upp på anvisad plats i affischsessionen. SCB hade den ärtigaste affischen, precis som förra gången på affischsessionen COMPSTAT 78 i Leiden, då med TAB68.

COMPSTAT-konferensen bestod förutom av affischsessionen och utställningsmontrar också av föreläsningar och demonstrationer. Föreläsningarna gick i upp till fyra parallella serier. De bestod dels av halvtimmespass för invited och contributed papers, dels av 10 minuters short communications. Alla pass följdes av 5 minuters frågestund. Till AXIS hade vi fått ett 10 minuters pass på onsdagen just före det att det sociala programmet skulle starta på eftermiddagen. Väl förberedd och med bilder dimensionerade efter den stora föreläsningssalen skissade Aino vad AXIS står för. Budskapet gick fram till de ca 50 personer som var närvarande och ställde frågor. Aino annonserade även demonstrationen av AXIS kl 11.00 på torsdagen. Demonstrationen anslogs även i den hall där utställningen fanns. Där fanns också kaffeservering, konferensservice och souvenirförsäljning. Anslagen bestod av AXIS-affischerna med demonstrationstiden angiven. I hallens högtalarsystem kungjordes likaså starten för demonstrationen. Före själva terminalpasset beskrev Aino AXIS för att ge en grund till praktiken. Demonstrationen gjordes via telenätet till Örebro-maskinen och gick med TV-kamera ut till 4 TV-apparater och en stor TV-skärm. Därmed kunde alla tydligt se vad som gjordes med AXIS. Många frågor ställdes. Även frågor om pris där man gärna ville ha detta ännu mer specificerat. Auditoriet bestod av folk från mycket stora företag såsom oljebolag och läkemedelsföretag samt statistikbyråer på nationell och internationell nivå, totalt 40 personer.

Dokumentationen som nämnts ovan exponerades vid föreläsning, demonstration och i ett konferensbibliotek. Det mesta var slut redan på torsdagen sånär som på AXIS-broschyrer som vi hade ganska många av. Dessa lämnades för exponering på den Biometri-konferens som alltid brukar följa direkt efter COMPSTAT-evenemangen.

AXIS på export ?

Varför denna exponering av AXIS utanför SCB och till och med utanför Sverige ? Ja, man kan ha många synpunkter på detta. Här är några:

- Vi har på SCB gjort en produkt som av t. ex. en östtysk centralbyråpresentant på bruten engelska beskrevs som "a wonderful tool". Det är klart vi måste ut och berätta om det verktyget.
- Det finns ingen anledning varför vi på internt utvecklad programvara skall ställa lägre krav än på köpt programvara. Det betyder att vår egen programvara skall hålla sådan klass ur alla avseenden att den skall gå att sälja.
- Om vi nu skall sälja programvara, så gäller det att på ett tidigt stadium öppna ögonen på eventuella kunder. Det har nämligen visat sig att tiden från det att en programvara annonseras till dess den anammas av en stor organisation är ganska lång.
- AXIS' fördelar jämfört med andra system för tidsserielagring är dess metadatahantering. Den är grunden för en god ordning i obegränsade datalager. Den är grunden till AXIS' lättanvändhet. Den är grunden till den goda texthanteringen i AXIS. Metadatahanteringen är också den kostnad som en organisation som installerat AXIS får betala för fördelarna med AXIS. Metadatahanteringen kräver utbildning och erfarenhet hos databasadministratörerna. Med AXIS-ansatsen ställs problemen med ordningen och dokumentation av och möjlighet till åtkomst av tidigare producerat material på sin spets. På många ställen ute i världen står man nu inför starten av projekt för att lösa just dessa problem.
- Eftersom AXIS tycks vara det första systemet av denna typ och det kostat mycket att utveckla, ligger det nära till hands att anta att andra ekonomiskt sinnade organisationer gärna vill undvika att själva göra samma insats för något som redan finns.
- Om AXIS installeras på flera ställen i världen, så skulle vi underlätta den internationella statistikhanteringen.
- Om flera AXIS-installationer går in i databasnätverk som EURONET, så skulle många få tillgång till nationella data över gränserna på ett enhetligt sätt. Det skulle minska marknaden men också underlätta för informationsjägarekonsulter.

Skall vi spänna bågen eller skall vi packa ned den. Det måste vi bestämma i höst.

TAB68 erövrar Asien och Stilla Havet

Sedan något år korresponderar jag med Mr Dekker på Statistical Institute for Asia and the Pacific (SIAP) i Tokyo. Han har efter utvärdering av olika program för tabellproduktion kommit fram till att TAB68 är det system som passar bäst för utvecklingsländerna i området. TAB68 är effektivt i maskinen, vilket spelar roll i länder där hela statsförvaltningens databehandling ligger på en maskin. TAB68-språket är på en nivå som kan klaras av personalen som har stora problem med konventionell programmering. Mr Dekker föreslog därför att någon från SCB på SIAPs bekostnad skulle resa till Tokyo och få TAB68 att fungera på deras FACOM-160 från Fujitsi, vilket är en IBM-kopia. Dessa FACOM-maskiner finns också i ett antal utvecklingsländer i området. Efter fullgjort värv i Tokyo skulle resan gå vidare till 4–5 andra länder. Totalt högst 6 veckor. Mr Dekker återkommer med närmare detaljer.

CEI och standardprogrammen vid SCB

Sista dagen i Toulouse besökte vi servicebyråföretaget CEI tillsammans med Dagny Mellgren från Barcelona. Dagny kände till att CEI löst sina problem med ojämna svarstider vid blandad beläggning på centrala stordatorer. Förslaget väcktes sedan vi beskrivit vår syn på hur vi ser på standardprogramutvecklingen vid SCB. Om

den kan kort sägas att preparering av styrinformation för och dokumentering av standardprogramsystem bör göras interaktivt på datakraft nära användaren. Den lokalt framställda korrekta styrinformationen skickas därefter över till en central maskin som ligger nära de stora datalagren. Detta skulle ge kontroll över svarstiderna och en renodlad bearbetningssituation i den centrala maskinen. Dagny var då helt på det klara med att det borde vara intressant för oss att besöka CEI, när vi nu ändå var där.

Vad man gjort hos CEI var att man tacklat problemet med kundernas ojämna terminalbelastning på deras tre Bouroughsmaskiner genom att erbjuda lokal datorkraft. Utrustningen som de själva sätter samman är en kombination eller ett s.k. cluster av 8 st. 8 bits processorer för olika funktioner under en 16 bits processor för kommunikation med stordatorn. Till detta cluster kan kopplas olika kraftfulla arbetsstationer med mer eller mindre minne. En eller fler funktioner i kombinationen skulle kunna vara tillhållet för den lokala programvara som beskrivs ovan. Även om vi kanske inte stannar för CEI, så visar ändå deras lösning att möjligheter finns att få kontroll över svarstider även om man ibland vill använda sig av kraften i en stordator.

Lars Nordbäck

SCBs yttrande över förslag till ämnesbeskrivning vid prövning av ledigblivande professur i statistik, särskilt offentlig statistikproduktion

Statistiska centralbyrån (SCB) har beretts tillfälle att inkomma med yttrande till Universitets- och högskoleämbetet (UHÄ) över i remissen redovisade förslag till ämnesbenämning och ämnesbeskrivning.

SCB anser att den uppfattning som verket redovisade 1980-12-16 i skrivelse (Dnr 1980: 580) till Samhällsvetenskapliga fakultetsnämnden vid Stockholms universitet fortfarande gäller. Där sades bl. a. följande:

"Den nuvarande professuren i statistik, särskilt offentlig statistikproduktion, vilar på grundvalen att den skall verka för ett utökat kunskapsutbyte mellan högskolan och det statistiska tillämpningsområdet. Det dubbelriktade kunskapsutbytet skall ge insikter om de för tillämpningsområdet mest angelägna frågorna samt peka på de mest akuta problemen som påkallar en snar lösning.

Kunskapsutbytet mellan högskolan och det statistiska tillämpningsområdet, särskilt den offentliga statistikproduktionen, kan dock ytterligare stärkas om banden med SCB görs formella. Tjänsten skulle innefatta ett nära samarbete med SCB, speciellt enheten för statistiska metoder, i syfte att initiera, förmedla och följa visst projekt- och utvecklingsarbete inom statistikproduktionen. Relationerna mellan högskolan och SCB kan också stärkas genom tillkomsten av gemensamt anordnade seminarier med ämnen hämtade från både det teoretiska och det praktiska statistiska arbetsfältet. Det torde sålunda vara lämpligt att knyta en del av professurens tjänstgöringsskyldighet till SCB, t. ex. inom ramen för vetenskapliga rådet i vilket ingår ledamöter från såväl högskolan som SCB. När professuren i statistik, särskilt offentlig statistikproduktion därför nybesätts bör i befattningsbeskrivning eller motsvarande stipuleras en tjänstgöringsskyldighet vid såväl högskolan som SCB, förslagsvis 50% tjänstgöringsskyldighet på vartdera hållet."

SCB vill tillstyrka det framlagda förslaget till ämnesbenämning och ämnesbeskrivning och anser det vara en klar fördel att man i tjänsten inbegriper arbetsfält som "statistikanvändning, tolkning och utvärdering".

Slutligen anser SCB det är nödvändigt att innehavaren av tjänsten skall utföra 50% av sin tjänstgöring vid SCB och inte 25% som föreslagits. Anledningen härtill är att det krävs ingående studier av SCBs olika statistiska problemkomplex på platsen och i nära kontakt med berörda befattningshavare. Det är inte möjligt att inom ramen "för drygt en dags tjänstgöring per vecka" på ett meningsfullt och effektivt sätt uppta och i forskningsbara termer omvandla de stora och komplicerade statistikfrågor som SCB brottas med.

SCBs yttrande över gymnasieutredningens betänkande "En reformerad gymnasieskola" (SOU 1981:96)

Endast sådana aspekter tas upp, som anknyter till SCB i dess roll som central förvaltningsmyndighet för den statliga statistikproduktionen.

SCBs gymnasieskolstatistik omfattar sökande/intagnastatistik, närvarostatistik och avgångsstatistik. Dessutom görs uppföljningsundersökningar av avgångna från grund- och gymnasieskolan. Närvarostatistiken grundas på uppgifter om elever och klasser som sammanställs och rapporteras av skolstyrelser/utbildningsnämnder. Övrig gymnasieskolstatistik och även uppföljningsundersökningarna grundas på individuppgifter. Dock gäller för avgångsstatistiken att denna för närvarande endast omfattar individuppgifter vad gäller avgångna från linjer. Den anpassning och förändring som behövs av de olika statistikdelarna beskrivs översiktligt i det följande.

SCB utarbetar också prognoser och analyser rörande utflödet på arbetsmarknaden och arbetsmarknadens behov av gymnasieskolutbildade och över behovet av lärare i gymnasieskolan. Hur denna verksamhet påverkas berörs också i det följande.

SCB anser och vill framhålla att gymnasieutredningen på ett förtjänstfullt sätt grundligt orienterat sig om och utnyttjat befintlig statistik samt även brukat sig av möjligheten att genom uppdrag till SCB och universitetsinstitutioner komplettera sitt statistiska underlag. Utredningen har också utförligt dokumenterat och i SOU- och DsU-serierna och på annat sätt redovisat sitt statistiska underlag och sina analyser.

Beträffande statistikbehoven kan allmänt konstateras att dessa, såvitt nu kan bedömas, torde öka om utredningens förslag genomförs. T.ex. torde de successivt inträffande valen av inriktning på studierna, som utredningens förslag innebär, behöva belysas statistiskt. Nuvarande statistik måste också i relativt stor utsträckning anpassas på olika sätt. De system som finns för statistikens framställning måste också anpassas och i viss utsträckning göras om.

Sökande/intagnastatistiken

Utredningens förslag kommer om det genomförs att kräva förändringar i statistiken över sökande och intagna då sök- och valprocessen och den yttre skolorganisationen förändras avsevärt. När det gäller urvalsreglerna som skall gälla vid intagningen till gymnasieskolan hänvisar utredningen till nu pågående försök innebärande bl.a. intagning på höstterminsbetyg. SCB vill peka på att det för närvarande pågår en utvärdering av det nya intagningssystemet som infördes fr.o.m. läsåret 1982/83. SCB deltar i denna utvärdering som skall redovisas senast 1 mars 1983.

En flexiblare skolorganisation som utredningen föreslår med en mjuk övergång mellan grund- och gymnasieskolan samt den decentralisering av skoladministrationen som för närvarande genomförs medför att informationstillgången troligtvis kommer att bli allt viktigare för att sök- och valprocessen till och i gymnasieskolan skall fungera effektivt.

För att klara prognosverksamhet, planering, dimensionering och information till SYO-konsulenter och allmänhet avseende den av utredningen föreslagna sök- och valprocessen till och i gymnasieskolan m.m. krävs troligtvis att varje valtillfälle följs upp. Enligt förslaget kommer eleverna att välja utbildningsområde, tillvalsämnen, studieprogram, koncentrerade yrkestekniska kurser (kyt-kurser) m.m. vid olika tillfällen. De större skolenheterna skall även kunna göra intagning mer än 1 gång per år. Statistiken över dem som söker och intages till gymnasieskolan omfattar för närvarande bara ett valtillfälle och endast studievägar om minst 1 termin som startar på hösten. Det nuvarande systemet för framställning av statistik om sökande och intagna går troligtvis ej att enbart anpassa till den nya sök- och valprocessen som utredningen föreslår utan ett nytt system måste konstrueras. Produktionskostnaderna kommer att öka om varje valtillfälle skall följas upp.

SCB vill peka på vikten av att det skapas *ett* enhetligt informationssystem som kan ge såväl de regionala som centrala organen den statistik dessa behöver om sökande och intagna till gymnasieskolan.

Utredningen förordar att man bör ha god information om de elever i grundskolan som ej söker till gymnasieskolan. Uppgifter om dessa kan och bör ingå i ett "sökande/intagnaregister".

Samarbetet mellan kommunala vuxenutbildningen (komvux) och gymnasieskolan bör ske på tre sätt enligt utredningen, nämligen genom samverkan, samplanering och samordning. Sökande/intagnastatistiken kan t. ex. beröras av en eventuell samordning vid intagning av vuxna och ungdomar till gemensamma grupper placerade antingen i komvux eller i gymnasieskolan för studievägar där slutmålet på studievägen är detsamma för de båda skolformerna.

Närvarostatistiken

Den uppgiftsinsamling som idag görs för närvarostatistiken är nära kopplad till behov av beslutsunderlag som finns hos skolöverstyrelsen (SÖ) och länsskolnämnderna. I vilken omfattning dessa behov kommer att förändras om utredningens förslag genomförs är för närvarande inte klarlagt. Det bör utredas gemensamt av SÖ och SCB sedan beslut fattats med anledning av utredningens förslag.

Avgångsstatistiken

Den nuvarande avgångsstatistiken omfattar uppgifter om antalet elever som avgår med eller utan slutförd lärokurs på studievägar om minst 8 veckor samt betygsstatistik för de 2-, 3- och 4-åriga linjerna. Här uppstår behov av anpassning till de nya begreppen korta och långa studieprogram i stället för nuvarande studievägar m. m. Betygsstatistiken måste också anpassas till vad som kan komma att gälla vid en övergång till s. k. kursrelaterade betyg.

Avgångsstatistik för en gymnasieskola enligt utredningens förslag bör omfatta dels uppgifter om antal avbrott inom varje studieprogram och antal byten mellan olika studieprogram, dels betygsstatistik för samtliga studieprogram, vilket torde innebära viss volymökning jämfört med nuvarande avgångsstatistik.

Vissa hjälpregister

Vissa register inom SÖ och SCB som för närvarande används som hjälpregister vid arbetet med olika delar av gymnasieskolstatistiken, i första hand studievägs- och skolorganisationsregistren, påverkas av utredningens förslag och behöver förändras. Även detta bör utredas gemensamt av SÖ och SCB.

Uppföljningsundersökningar av avgångna från grund- och gymnasieskolan

En stor del av det statistiska underlag gymnasieutredningen använt i sina analyser rörande gymnasieskolans funktioner härrör från SCBs uppföljningsundersökningar bland grund- och gymnasieskolavgångna. Dessa undersökningar är ett nödvändigt komplement till den övriga gymnasieskolstatistiken för att få insyn i och kunskap om hur gymnasieskolan fungerar för individ och samhälle.

Uppföljningsundersökningar av avgångna från grund- och gymnasieskolan med hjälp av postenkät genomförs för närvarande löpande enligt ett i samråd med statistikanvändarna fastställt schema där undersökningsgrupp och uppföljningsperiod väljs olika år från år.

I samband med ett genomförande av en reformerad gymnasieskola synes det angeläget att dessa uppföljningsundersökningar ges ökade resurser så att dels varje undersökning (grundskoleuppföljning eller gymnasieskoluppföljning med viss uppföljningsperiod) kan genomföras med tätare tidsintervall jämfört med nuläget, dels varje undersökning göras med så stort urval att önskvärd detaljeringsgrad i redovisningen erhålls.

Resultaten från uppföljningsundersökningarna av avgångna från gymnasieskolan kan redovisas så att man skulle kunna urskilja hur avgångna från olika studieprogram, inkl. kyt-kurser, klarade sig på arbetsmarknaden eller i fortsatta studier. Sådana uppgifter är nödvändiga för att man skall kunna se hur utredningens intentioner härvidlag uppfylls. Det framkommer också genom dessa undersökningar underlag som skulle göra det möjligt att se om tidigare

köns- och socioekonomiskt grundade studiemönster förändras i en ny gymnasieskolorganisation.

De mjukare övergångar från grundskola till gymnasieskola, som utredningen föreslår med den nya studieorganisationen, kan också vid behov följas upp med sådan information som uppföljningsundersökningar kan ge. Genom mer omfattande och tätare grundskoleuppföljningar än vad nuvarande resurser hos SCB tillåter kan man erhålla sådan information.

Speciell hänsyn kan i undersökningarna tas till de elevgrupper utredningen nämner, som redan i samband med grundskoleslutet kan förväntas få svårigheter med arbete och studier, och som enligt förslaget ska ges möjlighet till speciella studieprogram.

I det fall intagningen till gymnasieskolan kommer att ske 2 eller 3 gånger/år bör uppföljningsundersökningarna anpassas till detta, vilket innebär flera likartade undersökningar varje år.

I samband med en reform av gymnasieskolan följer självklart en förändring av frågeinnehållet i undersökningarna för att detta skall passa i en förändrad verklighet. SCB väntar sig starka krav från ansvariga utbildningsmyndigheter, och även andra intressenter, att i undersökningarna få med frågor av utvärderingskaraktär rörande gymnasieskolans innehåll och inre organisation m. m. på motsvarande sätt som för närvarande.

SCB räknar med att ansvariga i samband med försöksverksamhet och genomförande av reformen kommer att fastställa en plan för hur utvärdering såväl på mikro- som makronivå skall genomföras och ser därvid gärna att lämpliga delar av detta arbete som hittills också i framtiden ges som uppdrag åt SCB att genomföra vid sidan av, eller kopplat till, de löpande uppföljningsundersökningarna.

Prognosverksamheten

De prognoser som berörs av utredningens förslag gäller behovet av lärare och behovet av utbildade inom olika yrkesområden. Utredningsförslagen medför inte några väsentligt ändrade förutsättningar för prognosarbetet. Då förutsätts att det även framgent kan tas fram statistik i samma omfattning som för närvarande som beskriver valen av studieinriktning i gymnasieskolan. Sådana uppgifter behövs som ingångsvärden i prognosarbetet.

Prognosernas användningsområde kan dock komma att vidgas till följd av förslagen. De prognoser över arbetsmarknadens nyrekryteringsbehov av utbildade som SCB gör kan komma att utnyttjas på nya sätt till följd av utredningens förslag till ny studieorganisation och nya former för övergången mellan grundskola och gymnasieskola. Av utredningens huvudbetänkande framgår att man vill möta dimensioneringsproblemen med bättre information om arbetsmarknaden till eleverna både i grundskolan och gymnasieskolan. Den nya läroplanen för grundskolan innebär samtidigt att studie- och yrkesorienteringen bakas in i den reguljära undervisningen i större utsträckning. SCBs prognoser över behovet av utbildade bör kunna användas i undervisningen på grundskolans högstadium och gymnasieskolan för att bidra till den förbättrade arbetsmarknadsinformation man eftersträvar.

De studieorganisatoriska förändringar utredningen föreslår har stor betydelse för hur behovet av olika lärarkategorier utvecklas. Här kan SCBs lärarprognosmodell utnyttjas för att beräkna konsekvenserna av utredningens förslag för lärarutbildningsbehovet i framtiden.

I ett reformuppföljningsprogram kan de kalkyler som ligger till grund för prognosarbetet utnyttjas för att belysa utfallet av reformen. Här kan nämnas beräkningar av övergångsfrekvenser till fortsatta studier efter genomgången gymnasieskola på olika studievägar, val av studieinriktning och yrke bland pojkar resp. flickor samt valmönster bland elever med olika social bakgrund.

Översyn av gymnasieskolstatistiken

SCB anser att i proposition med anledning av gymnasieutredningens förslag bör tas upp medel för uppdrag åt SCB att i samråd med SÖ göra en översyn av gymnasieskol-

statistiken med utgångspunkt i de förändringar av statistiken som propositionen kan medföra och därvid samråda med statistikens övriga intressenter. Av översynen bör framgå vilken förändring av nuvarande statistik och vilken ny statistik som kan behövas samt vilka förändringar av systemen för statistikens framställning som krävs och vilka volymförändringar som är att vänta. Översynen bör också visa vilka kostnader som är förknippade med de olika förändringarna.

Sammanfattning
De i det föregående framförda synpunkterna kan sammanfattas på följande sätt:

- SCB anser att gymnasieutredningen på ett förtjänstfullt sätt utnyttjat statistisk information som underlag för sitt arbete och sina förslag.
- Om utredningens förslag genomförs torde behoven av gymnasieskolstatistik öka i förhållande till nuläget.
- Nuvarande gymnasieskolstatistik – sökande/intagnastatistik, närvarostatistik och avgångsstatistik – och uppföljningsundersökningarna av avgångna från grund- och gymnasieskolan samt de prognoser som görs inom detta utbildningsområde berörs mer eller mindre av utredningens förslag och måste, om förslagen genomförs, anpassas på olika sätt.
- De system med omfattande ADB-stöd som finns för statistikens framställning och vissa register som utnyttjas som hjälpmedel i detta sammanhang berörs också och måste om utredningens förslag genomförs anpassas och i viss utsträckning göras om.
- Uppdrag bör ges till SCB att i samråd med SÖ göra en översyn av gymnasieskolstatistiken med utgångspunkt i de förändringar av gymnasieskolan som beslutas på grund av utredningens förslag.

Studieresa till USA för miljöstatistik

Vid Stanford-universitetet i Kalifornien finns sedan nästan tio år ett särskilt projekt för statistiska problem i samband med luftkvalitetsmätningar samt tolkning och redovisning av dem. Under sommarterminen 1982 vistades jag där, delvis på ett forskningsanslag från Stanford, delvis i tjänsten, för att sätta mig in i problemen, själv bidra med synpunkter bl. a. vid onsdagsseminarierna, och dra ut konsekvenser av betydelse för den svenska luftkvalitetsstatistiken. Mitt arbete koncentrerades till metoder för att definiera parametrar som på ett relevant sätt ger en översiktlig beskrivning av luftkvalitet. Väsentliga utgångspunkter är de standarder som fastställts i USA och bakomliggande medicinska och biologiska undersökningar. Generaliseringsmöjligheter och mätmetoder – bl. a. nya idéer om laserbaserade mätningar efter samma principer som vid tomografi, dvs. röntgenundersökning av hjärnan – måste också beaktas. Resultaten är förslag till parametrar i den svenska statistiken. En rapport med detta innehåll är under utarbetande.

Efter sommarterminens slut gjorde jag under en vecka en mängd besök hos olika befattningshavare vid Environmental Protection Agency i Washington, D. C. och nära Raleigh, N. C., och vidare vid U. S. Geological Survey och The Conservation Foundation. Syftet var främst att skaffa en bild av insamling och användning av data i miljövårdsarbetet, i synnerhet på luftsidan. Bland det som nu händer på det miljöstatistiska området i USA märks;

- några större undersökningar för att mäta personliga doser (''exposures'') av vissa luftföroreningar, och vidare arbete på modeller och simuleringsstudier för doser, samt samband mellan exponeringstidens fördelning och hälsoeffekter (''impact'')
- en undersökning av fiskeförhållanden m. m. i floder, där även några vattenkvalitetsfrågor ingår
- diskussioner om trendberäkningar, såväl för luft- som vattenkvalitet. Bl. a. har USGS genomfört och publicerat en del beräkningar, och presidentens Council on Environmental Quality har gett ut en vacker men föga spridd volym med titeln ''Environmental Trends''.

Mera generellt är organisationen av miljövård och miljöinformation i stöpsleven – kraftiga anslagsminskningar har genomförts. Miljövård och styrning står i fokus, medan statistik – bl. a. för översikter, uppföljning och utvärdering – får litet utrymme. Under 1982 har dock två böcker med nationella miljööversikter utkommit, med mera diskussioner och analys än statistik, men man tror sig inte ha råd att fortsätta med sådana sammanställningar. Som svensk fick jag ofta höra kommentarer om försurning, som nu är mycket aktuellt även i USA.

Thomas Polfeldt

Konsultation vid Ungerns statistiska centralbyrå samt besök i Berlin och Bratislava

Under våren 1982 gjordes den andra av tre konsultationer vid Ungerns Central Statistical Office (C.S.O.). Konsultationen ägde rum 4 mars–2 april 1982, föregången av närvaro vid ett RAPID-möte i Berlin 1–3 mars 1982, besök vid Statistisches Zentralamt i Berlin 4 mars 1982, följt av ett besök vid Computing Research Center i Bratislava den 13–14 april 1982.

Huvudsyftet var att installera några program som utvecklats i Sverige. Ett annat syfte var att biträda med råd för användande av RAPID för den Ungerska Çensus Databasen i *Budapest*.

RAPID är en databashanterare som utvecklats vid Statistics Canada. RAPID användes i dagens läge vid flera statistiska centralbyråer och liknande institutioner. Ungern har begärt och fått konsultation vid installation av RAPID i tre faser. Första och tredje fasen av konsultationen gjordes av John Pieper från Statistics Canada. Den andra gjordes av undertecknad.

I den andra fasen installerades ytterligare program utöver dem som installerades i första fasen av konsultationen. Arbetet var fyllt av problem. Problemens orsaker var flera. En lista över uppkomna problem följer. Listan kan ses som en ledtråd till underlättandet av samarbete med andra datorinstallationer.

- De bibliotek som kom från SCB laddades från tape till skivminne typ 2314. Vissa problem uppkom beroende på det faktum att biblioteken kom från skivminne typ 3330 och hade större blockstorlek än som ryms på ett spår på 2314. Sensmoralen blir att inte ta till större blockstorlek än nöden kräver för bibliotek som är avsedda för distribution.
- En rutin som avgör om ett program exekverar i förgrund eller bakgrund fungerade inte på grund av beroende av operativsystem. Ungerska C.S.O. använder operativsystem SVS medan vi använder MVS (olika versioner av ett operativsystem).
- I SARA[1] upptäcktes några problem. SARA fungerade inte beroende på subprogram som var beroende av operativsystem. Eftersom dessa subprogram endast behövdes för att logga användningen av SARA beslöts att ta bort loggningen. Slutanvändaren märker ingen skillnad.
- Vid C.S.O. kan en TSO-användare normalt använda 128K till 256K minne. Som en konsekvens av detta måste mycket stora program som SARA och RAPUPDT länkas i overlay-struktur för att reducera minnesbehovet vid exekvering. TSO-versioner av SARA och RAPUPDT länkades i overlay-struktur och identifierades med ett suffix 'T' (SARA01T och RAPUP01T). *Mycket stora program bör distribueras i overlaystruktur för att underlätta installation.*
- För att underlätta användandet av RAPID installerades SCBCALL – en 'command processor'. Den gör att man inte behöver 'STEPLIB' i logonproceduren. Två fördelar uppnås med detta:
  1. TSO arbetar effektivare eftersom inte 'STEPLIB' behöver sökas för att finna ett TSO-kommando.
  2. Det är mer flexibelt för applikationsprogrammerare att "konkatenera" passande bibliotek när ett applikationsprogram körs under TSO.

[1] Standardprogram som används för att extrahera data från flera sammankopplade RAPID-relationer.

- Under konsultationen uppträdde skivminnesfel som försvårade arbetet. Ett skivminnesfel upptäcktes den 17 mars. Vid denna tidpunkt kunde vi inte erhålla en ny skivpacke så skivpacken återinitierades och återställdes till status före skivminnesfelet. Delar av disken dumpades till tape. Nästa dag, den 18 mars upptäcktes fel i RAPIDs laddmodulsbibliotek. Biblioteket återställdes till den version som gällde vid min ankomst. Den 19 mars upptäcktes fel i en objektmodul. Den 23 mars återkom ett skivminnesfel. Vid detta tillfälle var felet till och med synligt så nu var en ny skivpacke ett måste. Den nya skivpacken initierades och återställdes till det läge som gällde den 17 mars före skivminnesfelet. Ytterligare en gång upptäcktes fel i ett laddmodulsbibliotek. För att kunna garantera en felfri disk vidtogs följande åtgärder:
  1. En backup skapades för varje datasätt på skivpacken.
  2. Skivpacken återinitierades.
  3. Backup-datasätten kopierades tillbaka till skivpacken.

  För att undvika onödiga problem *är det tillrådligt att skapa backup av datasätt vid regelbundna tidpunkter*. Vid slutet av konsultationen installerades nya IBM 3330 skivminnen så att alla RAPID-bibliotek flyttas till 3330 skivpackar.
- Ett mycket svårt problem upptäcktes i SARA. Ett fel som varierade från gång till annan upptäcktes. Av detta skäl antog vi att felet hade sin grund i någon oinitierad variabel. Skillnaden i vårt och det ungerska operativsystemet är att vi använder minne som av operativsystemet initieras till binära nollor. Felet kunde efter mycken ansträngning spåras till användandet av 'DO UNTIL' istället för 'DO WHILE' (PL/1).

På grund av alla hårdvarufel kunde endast minimal tid ägnas åt diskussioner om det praktiska användandet av RAPID. Hårdvarusituationen förändrades emellertid med att nya IBM 3330 skivminnen installerades i slutet av mars.

Ytterligare RAPID software installerades vid C.S.O. Computing Centre. Programmen testades mot de RAPID-relationer som laddades under den första konsultationsfasen. Installation belyste några krav inför det fortsatta samarbetet inom RAPID-gruppen i Statistical Computing Project.

- För att effektivisera samarbete måste vi kommunicera mera. Det kommer att göra det enklare att använda RAPID på ett vettigt sett.
- En person från mottagande datorinstallation bör före installation besöka en installation som redan använder RAPID för att där analysera RAPID (program, kommandoprocedurer, JCL etc.) och avgöra vilka väsentliga skillnader mellan installationerna som finns och kan orsaka problem.
- Regler för generell programutveckling bör utvecklas. Den kanadensiska programmeringsstilen verkar vara en bra grund att starta från.

Förutom deltagande i möten i *Berlin* ägnades tid åt de problem som östtyskarna har med installation av RAPID på en EC 1040 maskin. Denna maskin är kompatibel med IBMs 360-serie. En annan organisation, Leitzentrum für Anwendungsforschung, har tillgång till en maskin EC 1055. Denna maskin är kompatibel med IBMs 370-serie. På denna maskin fungerar RAPID. Tillsammans diskuterade vi problem i samband med användning av RAPID.

Besöket i *Bratislava* ägnades åt diskussioner om en handbok för RAPID-användare som Dusan Prazenka utvecklar. Dessutom höll undertecknad ett föredrag: 'Requirements in a multiuser environment'.

Eyvind Björk

Studiebesök vid Statistics Canada och Bureau of the Census samt ASAs årsmöte 1982

Via New York och Montreal besöktes Statistics Canada i Ottawa den 10 augusti. Övriga närvarande från SCB var Gösta Forsman, Göran Råbäck och Carl-Erik Särndal. Samtal med kanadensiska metodstatistiker skedde bl. a. om problem vid deras area-

sampling i flera steg och med varierande urvals-sannolikheter. Via Niagarafallen åkte Gösta Forsman och jag vidare till Washington D. C. där vi den 12 och 13 augusti besökte US Bureau of the Census. Samtal och diskussioner skedde med metodstatistiker enligt följande ämneslista:

a. Estimation of change and rotating panel designs
b. Non-response
c. Error models
d. Small domain estimation

Intrycken från studiebesöken på andra sidan Atlanten är att kvalificerat metodutvecklingsarbete pågår och att komplexa urvals- och estimationsplaner används i de löpande statistikgrenarna.

Delvis beror detta på frånvaron av registerramar utom i företagsstatistiken. Trots denna skillnad jämfört med SCB finns det anledning att, vid genomgång av tidigare arbete inom något område, ta reda på vad som är gjort på Censusbyrån och Statistics Canada.

Den 15 augusti anlände vi till Cincinnati, vid Ohiofloden, där American Statistical Associations (ASA) 142a årsmöte avhölls. Medarrangörer var också Biometric Society och Institute of Mathematical Statistics. Sammanlagt 2 300 delegater deltog i konferensen den 16–19 augusti. Övriga närvarande från SCB var Göran Råbäck och Carl-Erik Särndal. Programmet var digert med 189 sessioner, varav flera pågick parallellt. Varje session bestod av ca åtta föredrag. I stor utsträckning bevistades sessioner arrangerade av Section of Survey Methods (delsektion i ASA). I en sådan session presenterade undertecknad SCB-rapporten "Optimal Estimation of Change in Sample Surveys" (PM 1982-05-11 av G. Forsman och T. Garås). Konferensen gav intryck av att amerikanarna är mycket flitiga med papper och penna och statistiska bidrag kom från alla tänkbara grupper av statistiker. Samtliga konferensbidrag, där en stor del är relevanta för SCB-problem, utkommer i ASA-proceedings 1982.

Tomas Garås

Statistiska centralbyråns styrelse

Sedan ett par förändringar företagits i sammansättningen av SCBs styrelse, har denna följande sammansättning fr. o. m. 1 januari 1983:

Generaldirektör Lennart Nilsson (ordf.), riksdagsman Jan Bergqvist, direktör Ola Carlsson, utredningssekreterare Sven Hegelund, riksdagsman Anders Högmark, förbundsdirektör Sven Järdler, sekreterare Tuve Lindeberg, departementssekreterare Agneta Norén, sekreterare Ingvar Ohlsson, professor Uno Zackrisson samt departementsrådet Svante Öberg. – Personalrepresentanter är byrådirektör Kerstin Björkman Werenfels och avdelningsdirektör Folke Pantzare.

Summary in English

Statistics demands in the future

by L. Nordström

A wise man has said that about the future we know nothing. Probably he was right, although many persons consider themselves skilled in forecasting and prognostication.

What kind of statistics do we need in a long-range perspective? The question implies the necessity of giving priority to some types of statistics. In a community with unlimited resources, the answer would otherwise be "as much statistics as possible, covering as many fields as possible". Such an unlimited production of statistics evidently presupposes a future preparedness to answer questions, as yet unforeseen, from planners, researchers, politicians and other statistics users.

However, the resources are not unlimited, and a very extensive statistics production will not be possible even in the 1980's. Consequently, an order of priorities has to be established, and this order obviously must be based on ideas about the future demand pattern for statistics. The central problem therefore is to try to find out which questions the future statistics users will want to answer.

To illustrate this problem, the author has made the following sub-division of the post-war era:

Period	Designation	Statistics demand
1945–1960	Period of disorder	Few demands for statistics
1960–1970	Period of planning	Great demands for basic material, time series, economic variables
1970–1980	Period of remorse	Social variables, crisis statistics
1980–1990	Period of change	???

The table serves to illustrate that each time period, with its specific characteristics, demands its own particular kind of statistics.

The period of planning in the 1960's required great amounts of time series data, since everybody assumed the growth trends to continue. Time series could then be used to justify an uncontestable social expansion.

During the 1970's most curves developed breaks, and many problems appeared which had not previously been considered, particularly problems of a social nature. Time series provided little help in solving these new problems, and new statistics demands appeared. There was a steep increase of Gallup-type surveys.

At present, in the 1980's, we find ourselves in a situation where we know that

- economic growth will be slow
- great structural changes will occur
- Sweden will be increasingly influenced by external factors in an increasingly complex world
- social problems will increase.

This picture of the future means that there will be entirely new statistics demands. Some examples are:

1. The 1980's will be characterized by rapidly developing events, often not anticipated, such as new industrial crises. The handling of these crises requires speedily available statistics of a type that cannot be specified in advance.

2. Unemployment will remain a great problem during the entire 1980's. Statistics that quickly and accurately indicate where new employment crises are maturing, which kinds of measures have positive and negative effects, and how the Swedish labour market in general functions, will be of great importance.
3. Sweden's increasing involvement in international business has sofar taken place without any sufficiently reliable monitoring of the process. In the future there will be more extensive demands for statistics on trade currents, ownership constellations across national borders, intercompany transfers in multinational companies etc.
4. In spite of a material standard of living that is among the highest in the world, large groups in Sweden will be afflicted by considerable social problems. Statistics indicating the population's state of health in a wide perspective will then be considered essential.

International Statistics

by Aage la Cour

The paper describes the institutional framework within which the work on international statistical co-operation is carried out. The role of the various bodies engaged in the essential task of improving comparability of statistics across the countries is explained.

In geographical terms the structure of the United Nations statistical system at world and regional levels is set out together with a description of the machinery linking the large number of organisations concerned.

At the centre there is the United Nations Statistical Office in New York supported by statistical organisations in each of the main geographical regions and by the statistical authorities in agencies and organisations dealing with special subject matters such as labour, health, tourism etc.

Other organisations enter into this overall framework;

- the International Statistical Institute and the International Association for Research into Income and Wealth to quote two examples at the world level
- and regional and sub-regional bodies such as the Nordic Statistical Secretariat, the Organisation for European Cooperation and Development, the European Community and the Council for Mutual Economic Assistance.

Examples of the importance of international statistical co-operation are given, with particular emphasis on the work on international classifications, and on national accounts. The relevance of this work to international negotiations, particularly, as regards trade and other flows between countries, is illustrated. The technical co-operation on statistical matters with developing countries is a growing aspect of international statistical co-operation and for this work and for the other tasks that are mentioned the role played by EUROSTAT (the Statistical Office of the European Communities), supported by the statistical services of the Member States, is explained.

Looking ahead the paper foresees the need for even greater co-operation over the coming years. Closer economic and political links between countries and groups of countries will call for improved information flows.

International telecommunication networks and growth of statistical databases will facilitate the flow of information, but international statistical norms will need to be respected and improved in order that the data can be readily used and understood in its international format.

Classes and electoral behaviour in Sweden

by P. Cricenti

Political voting behaviour in Sweden is analyzed from the point of view of the voters' objective class association, and the position of various socio-economic strata in a political conflict model is indicated.

The article starts with a discussion of voting in relation to socio-economic class. The author criticizes the traditional way of studying class structure and voting, i.e. the use of the Alford voting index with its dichotomization of the social classes into one working class and one middle class. The criticism is based on the fact that a homogeneous class, the working class—with basic interests in common—is compared with a heterogeneous class, the middle class—without basic interests in common. The Alford index is used in the article to check the differences in socialist voting percentages between workers/industrial workers on the one hand and farmers, senior officials/executives and medium-level salaried employees on the other (Tables 2, 3 and 5). Table 4 compares the percentage of non-socialist votes among senior officials/executives with the percentage of non-socialist votes among medium-level salaried employees. The results indicate (i) that the class voting index number between industrial workers and farmers has risen (Table 2), (ii) that the percentage of socialist-voting senior officials/executives has increased from 4% in 1956 to 23% in 1979 and that consequently the voting index number between the industrial workers and this group has dropped considerably (Table 3), and (iii) that the voting index number for medium-level salaried employees as compared with workers (Table 5) and senior officials/executives (Table 4) has remained at approximately the same level.

The changes in voting behaviour by occupational status have had a heavy influence on the Liberal Party, the Centre Party and the Conservative Party, among which the non-socialist voters oscillate. In respect to the Social Democratic Party, the most important changes have taken place in the group of senior officials/executives, where the percentage giving their votes to the Social Democrats has increased from 4% in 1956 to 20% in 1979, and in the farmer group, where the percentage has dropped from 14% in 1956 to 7% in 1979 (Table 10).

Table 11 and Fig. 4 indicate that the Swedish electorate is found in two political dimensions: The left-centre dimension and the right-centre dimension.

In the left-centre dimension the political struggle is for the large wage-earning population. The major occupational groups in this dimension are industrial workers, agricultural workers and medium-level salaried employees. The left-wing parties (the Social Democrats and the Communists) and the centre parties (the Centre and the Liberals) compete about the votes of these groups. In the right-centre dimension we find businessmen and entrepreneurs as well as a substantial number of employees, mostly senior officials/executives. In this dimension the political struggle is carried on by the Conservatives and the centre parties.

The Swedish National Forest Survey

by B. Hägglund and B. Ranneby

In Sweden National Forest Surveys have been carried out since 1923. The main task of the survey is to continously provide information for the planning and monitoring of the utilization of the forest resources on national and regional levels. Variables of great importance are, among others, area of forest land, stem volume per hectare forest land and volume of fellings during the latest season.

The first National Forest Survey (1923–29) was carried out as a strip survey. Since the observations from two adjacent points are highly correlated, the additional information obtained from the second point is sparse. At the second National Forest Survey (1938–52) the observations were concentrated to circular plots, located some hundred meters apart. These two surveys were performed successively by county.

In 1953, when the third National Forest Survey started, the so called tract system was introduced, i.e. sample plots were located systematically along the perimeter of squares (tracts) and a systematic sample of tracts covering the country was investigat-

ed every year. Thus it was possible to estimate, for the whole country, the volume of fellings during the latest season. The size of a tract was adjusted to constitute a day's work for a survey crew.

The fifth National Forest Survey was completed in 1982. This survey (as well as the fourth) is very similar to the third. As the optimal design of the tract is dependent on the variations in the population, Sweden is divided into different regions. For each of these regions the size of the tract is adjusted as well as possible. If the correlation between the plots is too great, then, for a given side-length of the square, the variance of the tract mean might increase with an increasing number of sample plots.

Developments within both forest management and research have actualized a new method for the sixth Swedish National Forest Survey which starts in 1983. This new method will comprise both temporary and permanent plots. With the design chosen, the growing stock is since the first remeasurement estimated with higher precision than if only temporary plots had been used. However, the great gain in precision is achieved in the estimation of changes.

Furthermore the field equipment will be modernized. For instance, the survey crews will use a micro-computer in the field work, which means that the number of persons in the survey crews can be reduced.

Litteratur

Nyare svensk och utländsk litteratur

Nyförvärv (urval vid SCBs bibliotek juli–september 1982)

Enskilda arbeten

Böhm, Franz, Graphen in der Datenverarbeitung: Anwendung des Markierungsmodells in Technik und Ökonomie / von Franz Böhm und Gotthard Weise. – 1. Aufl. – Leipzig: Fachbuchvlg, cop. 1981. – 238 s. – (Mathematik für Ingenieure) NBN 99-0329695-9 (2D/673)

Bühler, Walter Kaufmann, Gauss: a biographical study / W.K. Bühler. – New York; Berlin: Springer, cop. 1981. – viii, 208 s. : ill. S. 162–166: A survey of secondary literature. – S. 167–175: An index of Gauss's works ISBN 0-387-10662-6 (2D/680)

Cleary, James P. The professional forecaster: the forecasting process through data analysis / James P. Cleary, Hans Levenbach. – Belmont, Calif.: Lifetime learning publ., 1982. – 402 s. :diagr., tab. ISBN 0-534-97960-2 (2D/661)

Computational statistics: [Wolfgang Wetzel zur Vollendung seines 60. Lebens jahres] / ed.: Herbert Büning, Peter Naeve. – Berlin: de Gruyter, 1981. – 348 s. Med ty. och eng. text ISBN 3-11-008419-8 (2D/668)

Danmark. Danmarks statistik, Personstatistik på registergrundlag / [Danmarks statistik]. – København, 1982. – 180 s. ISBN 87-501-0535-3 (2D/660)

Elseth, Gerald D., Population biology / Gerald D. Elseth, Kandy D. Baumgardner. – New York: Van Nostrand, cop. 1981. – 623 s. ISBN 0-442-26235-3 (3C/698)

Gitlow, Howard S., Stat City: understanding statistics through realistic applications / Howard S. Gitlow. – Homewood, Ill.: Irwin, 1982. – 335 s. – (Irwin series in quantitative analysis for business, NBN 99-0125316-0) ISBN 0-256-02654-8 (2D/684)

Gourieroux, Christian, Théorie des sondages / Christian Gourieroux. – Paris: Economica, 1981. – 272 s. – (Collection Economie et statistiques avancées. Série: Ecole nationale de la statistique et de l'administration économique et Centre d'études des programmes économiques, ISSN 0154-196X) ISBN 2-7178-0369-6 (2D/675)

International meeting on Bayesian statistics, 1 (Valencia 1979) Bayesian statistics : proceedings of the first international meeting held in Valencia (Spain) May 28 to June 2, 1979 / ed. by J. M. Bernardo ... – Valencia : Univ. P., 1980. – 647 s. ISBN 84-370-0172-2 (2D/667)

Mendenhall, William, Mathematical statistics with applications / William Mendenhall, Richard L. Scheaffer, Dennis D. Wackerly. – 2. ed. – Boston, Mass. : Duxbury P., 1981. – 686 s. ISBN 0-87872-279-3 (2D/666)

Kingman, John Frank Charles, Mathematics of genetic diversity / J. F. C. Kingman. – Philadelphia, Pa. : Society for industrial and applied mathematics, 1980. – 70 s. – (Regional conference series in applied mathematics, ISSN 0097-4455;34) ISBN 0-89871-166-5 (2D/678)

Kpedekpo, G. M. K., Social and economic statistics for Africa : their sources, collection, uses and reliability / G. M. K. Kpedekpo, P.L. Arya. – London: Allen & Unwin, 1981. – x,259 p. Includes bibliographies and index ISBN 0-04-310011-2 (2D/662)

Kurtz, Thomas G., Approximation of population processes / Thomas G. Kurtz. – Philadelphia, Pa.: Society for industrial and applied mathematics, 1981. – 75 s. – (Regional conference series in applied mathematics, ISSN 0097-4455;36) ISBN 0-89871-169-X (2D/679)

Patchett, Isabel S., Statistical methods for managers and administrators / Isabel S. Patchett. – New York : Van Nostrand Reinhold, cop. 1982. – 365 s. ISBN 0-442-23124-5 (2D/672)

Piganiol, B., Statistique et économétrie: corrélation et regression, modèles économétriques, tests usuels, techniques de prévision court terme / par B. Piganiol. – 2. ed. – Paris, 1978. – 180 s. – (Mémentos Dalloz, NBN 99-0152144-0) ISBN 2-247-01138-1 (2D/674)

Pratique de l'analyse des données. ... 3: Linguistique & lexicologie / par H. Abdi ... – Cop. 1981. – 565 s. : diagr., tab. ISBN 2-04-010776-2 (2D/618)

Seber, G. A. F., Estimation of animal abundance and related parameters / G. A. F. Seber. – 2. ed. – London : Charles Griffin, 1982. – 654 s. : tab. ISBN 0-85264-262-8 (2D/664)

Statistikvägledning för energiplanering / [utg. av] Statistiska centralbyrån, Regionalstatistiska kansliet. – Stockholm : SCB, 1982. – 68 s. NBN 99-0334848-7 (2D/677)

Statistiska centralbyrån, Industrins kapacitetsutnyttjande : en ny ekonomisk statistik från SCB. – Stockholm : SCB, Enheten för industristatistik, [1982]. – 18, [10] bl. : diagr., tab. NBN 99-0332486-3 (2D/676)

Topics in applied multivariate analysis / ed. by Douglas M. Hawkins. – Cambridge : Cambridge University Press, 1982. – 362 s. ISBN 0-521-24368-8 (2D/683)

UBIS-projektet, Rapport från UBIS-projektet / Utredningen rörande uppgiftslämnarfrågor och bortfallsproblem i statistikproduktionen, [Statistiska centralbyrån]. – Stockholm : SCB, 1981. – 335 s. + sammanfattning (85 s.). NBN 99-0327989-2 (2D/669)

UBIS-projektet, Bibliografi över bortfallslitteratur mm/ Utredningen rörande uppgiftslämnarfrågor och bortfallsproblem i statistikproduktionen, [Statistiska centralbyrån]. – Stockholm : SCB, 1981. – 111 s. NBN 99-0327987-6 (2D/670)

Vachanija Nikolaj Nikolaevič, Probability distributions on linear spaces / N.N. Vackania. – New York : North Holland, cop. 1981. – 123 s. – (North Holland series in probability and applied mathematics, NBN 99-0184583-1) Orig:s titel: Verojatnostnye raspredelenija v linejnych prostranstvach ISBN 0-444-00577-3 (2D/665)

Bokanmälningar

Ghiselli, E., Campbell, J. and Zedeck, S., Measurement Theory for the Behavioral Sciences. W. H. Freeman and Co Ltd. Oxford 1981. 494 s. £ 11:70.

Boken ger en översikt av olika mätteorier och förfaranden inom beteendevetenskaperna. Fastän författarna försöker att ge exempel och metoder från alla beteendevetenskaper så skiner deras förankring inom psykologin igenom och en psykolog/pedagog känner sig nog mera hemma i boken än en sociolog.

Författarna har valt att grundligt gå igenom mätteorier och förfaranden på många områden och att trots detta försöka göra en lättläst bok som endast ska kräva ett minimum av förkunskaper. De anger själva en elementär statistikkurs plus en kurs i psykologi som en lämplig förkunskapsnivå. De förefaller ha lyckats någorlunda med sin svåra uppgift, men ansatsen har också medfört nackdelar, nämligen

- trots bokens tjocklek och innehållsliga bredd så saknar man en del mera avancerade mätteorier och ansatser som inte kunnat medtas på grund av ovannämnda val av uppläggning, t. ex. Cronbachs generaliserbarhetsmodell
- grundligheten, som i sig naturligtvis är ett plus, medför att läsaren måste arbeta sig igenom ganska många sidor för att få en viss mätteori eller modell förklarad och boken kan därför av en del upplevas som omständlig.

Med ovannämnda begränsningar så verkar boken ha bredd och en god täckning av olika mätmodeller inom främst testpsykologin. Förutom en grundlig, elementär genomgång av reliabilitets- och validitetsmodeller så innehåller boken ett bra material som täcker grundläggande statistiska begrepp och metoder på området liksom olika aspekter och principer vid mätförfaranden inom beteendevetenskaperna. Dessutom innehåller boken kapitel om:

- konstruktion av sammansatta variabler (kap. 7)
- multipel regressionsanalys och differentiell vägning (kap. 11)
- psykologisk skalning av personer och stimuli (kap. 12)
- grundläggande begrepp och förfaranden vid testkonstruktion (kap. 13).

Boken verkar stilistiskt sett lyckad och den är väl genomarbetad med bra lästips och litteraturanvisningar. Det finns också en ordlista där alla facktermer förklaras. Boken, eller delar av den, bör vara användbar som kursbok för studerande inom beteendevetenskaperna med en svag bakgrund i kvantitativa metoder.

Lars R. Bergman

Social rapport om ojämlikheten i Sverige, levnadsförhållanden 1977/1978. Rapport 22 i SCBs serie om levandsförhållanden, Liber Förlag, Stockholm 1981. 172 s. Ca kr. 98.40.

ULF är SCBs "undersökningar av levnadsförhållanden". Sedan 1974 intervjuas årligen ca 10000 individer i åldern 16–74 år om sina levnadsomständigheter. I rapport 22 presenterar man för första gången på sex år "en översikt över välfärdens fördelning" på alla de välfärdskomponenter man kartlägger. Målsättningen har varit att skapa "en uppslagsbok för jämlikhetsdebatten". Rapporten är författad av Joachim Vogel och han har hämtat underlaget från i första hand 1977 och 1978 års undersökningar.

Det föreligger stora skillnader i välfärd mellan olika befolkningsgrupper i 70-talets Sverige. Detta framgår klart av rapporten. Så har t. ex. högre tjänstemän avgjort bättre arbetsförhållanden, en högre materiell levnadsstandard och en högre livskvalitet[1] än befolkningsgenomsnittet enligt praktiskt taget alla de mått man uppställt. Samtidigt har

[1] Livskvalitet indikeras här av sju förhållanden: 1. Får inte lära sig nya saker i arbetet. 2. Instrumentell inställning till arbetet. 3. Ej aktiv i någon organisation. 4. Läser ej politiska artiklar. 5. Ensamboende utan familj. 6. Nedsatt arbetsförmåga. 7. Trötthet.

arbetare, och då särskilt kategorin biträdespersonal (de utan yrkesutbildning), sämre arbetsförhållanden, lägre materiell levnadsstandard och lägre livskvalitet än befolkningsgenomsnittet. Man kan också urskilja en rad befolkningsgrupper, t.ex. arbetslösa, extremt lågutbildade och undersysselsatta, hos vilka anhopningen av "onda" levnadsförhållanden är särskilt markant.

För läsaren kan det emellertid vara svårt, ibland omöjligt, att värdera storleken på och betydelsen av de skillnader hon/han finner i materialet eftersom jämförelser med andra undersökningar och över tid helt saknas. Att pensionärer har större hälsobrister än unga kan nog betraktas som ett ganska naturligt förhållande, men hur mycket sämre hälsa framstår som oundviklig? Har skillnaderna mellan gamla och unga förändrats under senare år? Är skillnaderna större eller mindre i andra länder? Frågeställningar av den här typen belyses inte i rapporten.

Utan dessa jämförelser tvingas man dock konstatera att rapporten inte tillfört jämlikhetsdebatten något nytt. Det är t.ex. ingen nyhet att biträdespersonal har sämre levnadsförhållanden än högre tjänstemän. På detta finns detaljerade mått sedan slutet av 60-talet och SCB gör ju egna undersökningar sedan 1974. Det är bara att beklaga att tidsjämförelser på grundval av ULF-materialet tillsvidare fått anstå. Nu kommer rapportens bidrag till jämlikhetsdebatten att ligga i det faktum att man förmått presentera avancerad socialstatistik på en rad områden på ett enkelt och lättfattligt sätt. I den meningen bör den kunna fungera som en uppslagsbok. Det är därför värt att peka på några av de grepp som använts.

I rapportens inledande kapitel preciseras det välfärdsbegrepp som bildat utgångspunkt för resultatsredovisningen och som omfattar elva *välfärdskomponenter*:

- Utbildning.
- Sysselsättning.
- Arbetsmiljö.
- Ekonomiska resurser.
- Boendeförhållanden.
- Transporter och kommunikationer.
- Fritid.
- Ensamhet och gemenskap.
- Politiska resurser.
- Trygghet och säkerhet.
- Hälsotillstånd.

Varje komponent ägnas ett kapitel där resultaten presenteras i tabellform utifrån fyra *perspektiv*; det demografiska, det socialpolitiska, det regionalpolitiska och arbetslivsperspektivet.

Tabellerna har genomgående samma uppbyggnad. Generellt uppmärksammas s.k. onda förhållanden, dvs. andelen som saknar något eller som har ett visst problem, och det är avvikelserna från befolkningsgenomsnittet i procent som anges. För de tre senare perspektiven (se ovan) redovisar man dessutom de för ålder och kön standardiserade avvikelserna från populationsmedelvärdena.

I respektive perspektiv indelas befolkningen i redovisningsgrupper enligt olika kriterier. I det *demografiska perspektivet* sker denna indelning efter kön, ålder resp. position i livscykeln. I det *socialpolitiska perspektivet* avgränsar man en rad "problemgrupper" med mycket dåliga resurser på någon välfärdskomponent. I det *regionalpolitiska perspektivet* indelas befolkningen efter bosättningsort länsvis och efter stödområde. I *arbetslivsperspektivet* slutligen, har indelningen skett efter socio-ekonomisk grupp, sysselsättningsvolym resp. utbildningsnivå.

Sammantaget är det en väldig mängd data som presenteras och överskådligheten är väl knappast omedelbar i komponentkapitlen. Men den konsekventa presentationen efter perspektiven och den genomgående användningen av procentandelar och standardvägning gör dock att materialet, trots kraven på fullständighet, är relativt lätt gripbart. Det ovan skisserade tillvägagångssättet är emellertid inte utan problem.

För det första ligger det bakom valen av perspektiv och indelningen av undersökningspopulationen i redovisningsgrupper antaganden om orsakssamband. Hur indelningen sker blir därför av central betydelse för resultattolkningen. Valen av redovisningsgrupper bör därför vara väl motiverade och avgränsningarna entydiga.

Det är i rapporten svårt att se vad som motiverar att gruppen invandrare i förväg utsetts till "problemgrupp" i det socialpolitiska perspektivet. Hur kan det på ett teoretiskt meningsfullt sätt vara möjligt att avgränsa en i sig så heterogent sammansatt grupp?

För det andra är de rent tekniska avgränsningarna inte alltid lika självklara som när det gäller t. ex. kön och ålder. Rapportförfattaren sätter själv frågetecken kring avgränsningen av de "extremt lågt utbildade" och i Statistisk tidskrift 1982:2 kritiseras en annan rapport i serien om Levnadsförhållandena för hur man avgränsat låginkomstfamiljer[1]. Frånvaron av en metodredovisning blir här och på andra ställen en allvarlig brist eftersom läsaren inte ges någon möjlighet att utifrån själva rapporten bedöma bevisvärdet i de presenterade resultaten.

För det tredje medför den konsekventa presentationen av alla resultaten att enskilda uppgifter om materiell levandsstandard, t. ex. innehav av stereoanläggning, kan trivialiseras eller misstolkas. Sett isolerat kan innehav av stereo tolkas som uttryck för konsumtionspreferens, eller livsstil, sett som en indikator bland många på ett hushålls ekonomiska resurser blir uppgiften till meningsfylld socialstatistik. Rapportförfattarens delvis vacklande hållning härvidlag skapar utrymme för den trivialiserande kritik Lars Söderström riktar mot rapporten i uppsatsen "Hur ojämlika är vi? – Kommentar till fördelningsstatistiken"[2].

I det näst sista kapitlet "Sammanfattande analys" söker författaren blottlägga de processer som producerar de observerade skillnaderna i välfärdens fördelning. Schematiskt uttrycks de grundläggande orsakerna ungefär så här: Svåra uppväxtförhållanden ger dålig utbildning och senare i livet jobb med arbetsmiljöproblem och stor risk för arbetslöshet. Dessa arbeten är också sämre avlönade och ger sämre möjligheter till självförverkligande i arbetet. Sammantaget blir alltså både de materiella levnadsförhållandena och livskvaliteten sämre. Förklaringsvärdet i denna "orsakskedja" torde emellertid vara mycket begränsat. Sambanden underkastas heller inte någon empirisk prövning. Något mått t. ex. på hur mycket av variationen i utbildning som förklaras av skillnader i uppväxtförhållanden anges inte.

Så i rapportens sista kapitel används en i sammanhanget ny presentationsform – "välfärdsprofiler" – som ger möjlighet för läsaren att jämföra flera redovisningsgrupper på en rad välfärdskomponenter samtidigt. Komponenterna har denna gång sammanfattats under rubrikerna: uppväxtförhållanden, arbetsförhållanden, levnadsstandard och livskvalitet. Även om antalet grupper och komponenter som samtidigt kan redovisas är begränsat ger "profilerna" god överblick och lämpar sig väl i sammanhanget.

Det huvudintryck läsningen av rapport 22 ger blir med nödvändighet delat. Å ena sidan innehåller rapporten en rad brister som skapar osäkerhet kring tillförlitligheten i en del av de resultat som presenteras. Å andra sidan är rapporten trots sina brister, ändå ett viktigt bidrag till arbetet mot en fungerande social rapportering tack vare den enkla, konsekventa och överskådliga presentationen. Kan man behålla samma klarhet och helhetssyn i kommande rapporter med tidsjämförelser blir resultaten mycket intressanta, ty:

"Frågan om vad som är en riktig och rättvis fördelning av välfärden minskar inte i betydelse i en tid av ekonomisk åtstramning. Snarare är det då särskilt viktigt att bevaka fördelningsfrågorna." (Ur rapportens förord.)

Joakim Palme

*

[1] Bokanmälningar: Låginkomstfamiljerna – vilka de är och hur de lever 1975/76, recenserad av M. Tåhlin, se också replikskifte Eklind – Tåhlin. *Statistisk tidskrift 1982:4.*

[2] Söderström, L., Hur ojämlika är vi? – Kommentar till fördelningsstatistiken i: *Hur ojämlika är vi.* RATIO, Kungälv, 1982. Att sedan Söderström bygger delar av sin kritik av direkta missförstånd av rapportens syften och innehåll är en delvis annan fråga.

Replik till Joakim Palme

1. Anmälaren anser att rapporten "inte tillfört jämlikhetsdebatten något nytt".

Rapporten redovisar situationen 1977/78 för samtliga välfärdskomponenter i ett sammanhang. Den *sammanför information* från ett tjugotal tidigare publicerade SCB-rapporter och utvidgar dessutom redovisningen till helt *nya problemgrupper*. En så fullständig presentation av ojämlikheten (många grupper och indikatorer) har inte publicerats tidigare i Sverige (och så vitt jag vet inte heller i andra länder). Avsikten med denna rapport har varit att *stimulera den allmänna debatten* kring välfärdsfördelningen. Den hade tidigare inte tillgång till en så "väldig mängd data" där man förmått "presentera avancerad socialstatistik på en rad områden på ett enkelt och lättfattligt sätt" (anmälarens formulering).

2. Anmälaren saknar jämförelser med andra undersökningar, jämförelser över tid och jämförelser med andra länder. Jämförelser med andra undersökningar över tid och mellan länder ingår i andra SCB-rapporter. Att även ta med dessa aspekter framstod emellertid helt orealistiskt. För det första är möjligheterna till mera heltäckande tidsjämförelser och internationella jämförelser för tillfället fortfarande rätt begränsade. För det andra skulle en sådan utvidgning av rapport nr 22 ha sprängt ramen för rapporten (enhetlighet i redovisningen, jämförbarhet, tolkning). Heltäckande jämförelser över tid finns ännu inte, men har nu påbörjats i ett program för åren 1979–1985. Den första rapporten har publicerats (rörande arbetsmiljö, rapport nr 32 i serien Levnadsförhållanden). Internationella jämförelser ingår också i vårt arbete på SCB, i samarbete mellan de nordiska ländernas välfärdsmätningar. En nordisk komparativ rapport, med utblick även utanför Norden publiceras om några månader.

3. Anmälaren kritiserar frånvaron av metodredovisning. Rapporten skiljer sig i det avseendet från andra rapporter i serien Levnadsförhållanden. Det är frågan om en översiktsrapport (som har ambitionen att vara en "uppslagsbok"). En mera fullständig metodinformation skulle också ha sprängt ramen för rapporten, om man betänker den "väldiga mängden av data" som finns.

4. Anmälaren hävdar att förklaringsvärdet för orsakskedjans dåliga uppväxtförhållanden – dålig utbildning – dålig arbetsmiljö samt arbetslöshet "torde vara mycket begränsat". För den som sysslat med analyser av dessa samband är nog detta en överraskande kommentar. Undersökningarna tyder definitivt på samband av den typen. Att demonstrera detta är kanske en av de viktigaste uppgifterna för den sociala rapporteringen.

Joachim Vogel

Social Trends 11, Central Statistical Office, London 1981. HMSO PO Box 569, London SE1 9NH. 257 s., £16.50.

Sedan 1970 utkommer Central Statistical Office i London årligen med Social Trends (ST): Syftet är och har varit att skapa ett bättre underlag för den offentliga debatten kring sociala och ekonomiska frågor. Som social rapportering har ST kommit att bli stilbildande även för andra länder. Trots att man med den senaste och elfte utgåvan brutit med den traditionella formen är ST fortfarande ett av de intressantaste försöken att fylla den "lucka i den demokratiska processen" som avsaknaden av ett adekvat faktaunderlag utgör när allmänhet och politiker skall bilda sig en uppfattning om "hur det är".[1]

Det var under andra hälften av 1960-talet som uppmärksamheten hade riktats mot den ekonomiska tillväxtens och de snabba sociala förändringarnas baksida. I Sverige arbetade man på "låginkomstutredningen", i Storbritannien förberedde man utgivningen av den första ST. Denna första utgåva kom att bestå av två relativt fristående delar; den ena med lite längre artiklar som behandlade olika aspekter på social förändring, den andra med ett omfattande tabell-

[1] För en diskussion om den sociala rapporteringens roll i den demokratiska processen se: Johansson, S., Mot en teori för social rapportering, Institutet för social forskning, Stockholm 1979.

och diagrammaterial med kortfattad kommenterande text – "a social indicators part". Statistiken omfattade de områden där politiken antogs ha ett direkt inflytande och man identifierade s. k. social concerns. Dessa fick utgöra den indelningsgrund varefter man organiserade presentationen av materialet. Kapitelrubrikerna blev följande: Public expenditure, Population and environment, Employment, Leisure, Personal income and expenditure, Social security, Welfare services, Health, Education, Housing, Justice and law.

Under 70-talet har sedan ST ändrats i flera avseenden men utan att den ursprungliga tudelningen rubbats. Sedan 1971 finns viss internationell statistik inkluderad. 1972 införde man "social commentary" – ett försök att sammanfatta och diskutera förändringar på en rad områden sammanhållet. 1974 började man med "calendar of events" som innebar att man följde upp viktiga lagändringar och politiska beslut på de områden ST avsåg att spegla. Dessutom har indelningen efter social concerns och stora delar av tabellmaterialet successivt förändrats.[1]

Också i 1981 års utgåva har delar av tabellmaterialet ändrats och förnyats. Men den verkligt viktiga förändringen ligger i att de analyserande artiklarna och "social commentary" helt tagits bort. Man kan visserligen argumentera för att det finns andra fora för den typen av material och att man med en ansträngd budget bör prioritera på det sättet som gjorts. Faktum kvarstår dock att man går miste om ett mycket värdefullt lässtoff som knappast kommer att replikeras någon annanstans.

Kvar är alltså tabellmaterialet, den kommenterande texten och "calendar of events". Kapitelindelningen efter social concerns är densamma som i 1980 års utgåva: Population Household and families, Social groups, Education, Employment, Income and wealth, Resources and expenditure, Health and public safety, Housing, Environment, Transports and communications, Leisure, Participation, Law enforcement. Sammanlagt omfattar dessa kapitel fler än 330 tabeller, diagram och figurer.

Det är onekligen en mångfasetterad bild av utvecklingen i Storbritannien som står att utläsa ur detta omfattande tabellmaterial och det låter sig svårligen sammanfattas. Jag skall här nöja mig med att göra ett par iakttagelser:

– Befolkningens storlek har i stort sett inte ändrats sedan början av 70-talet. Fertiliteten har sjunkit under samma period men ligger på en nivå över den i Västtyskland och ganska nära den i Frankrike. Den äldre delen av befolkningen har ökat sin andel.

– Inkomstfördelningen har enligt den statistik som presenteras inte genomgått några dramatiska förändringar även om en viss utjämning tycks ha skett under efterkrigstiden. Under samma period har reallönerna ökat väsentligt även om en stagnation har inträtt under 70-talet.

– Arbetslösheten har ökat kraftigt sedan mitten av 70-talet med endast en mindre nedgång 1978/79. Antalet långtidsarbetslösa fördubblades mellan 1975 och 1980. I början av 1970-talet fick de flesta arbetslösa sin försörjning genom den allmänna arbetslöshetsförsäkringen, i slutet av samma decennium är de fler som får lita till endast socialhjälp, än de som får arbetslöshetsersättning – med eller utan tillägg av socialhjälp.

Brittisk officiell statistik är traditionellt av god kvalitet. ST är inget undantag. Tabeller, diagram och figurer bjuder en ganska stor variation och synes väl anpassade till det material de avser att presentera. Här har nog svenska socialstatistiker fortfarande en del att lära. Men det faktum att ST inte baseras på någon separat undersökning speciellt avsedd för den här typen av publikationer medför en del nackdelar.

En brist är att man inte tillämpar något konsekvent välfärdsperspektiv och följaktligen inte formulerar något välfärdsbegrepp. Det blir därför omöjligt att relatera den statistik som presenteras till individer eller särskilda befolkningskategorier. En

[1] Se vidare Thompson, E.J., Social Trends: The development of an annual report for the United Kingdom. *International Social Science Journal, Vol. XXX, No. 3, 1978.*

annan svaghet är att de använda källorna inte tillåter någon jämförelse mellan t. ex. olika socio-ekonomiska grupper på mer än ett par av de social concerns man identifierat. Vidare har de olika undersökningarna utförts olika år och redovisningen av förändringar blir haltande om man vill jämföra utvecklingen på olika områden. Man lyckas heller inte, där det borde vara möjligt, presentera ordentliga tidsserier. Ibland tycks de år som redovisas vara helt slumpmässigt valda.

För att CSO skall kunna utveckla sin sociala rapportering krävs det nog att man utformar för ändamålet särskilda undersökningar. Krav på detta har rests i den socialpolitiska debatten, bl. a. från forskarhåll.[1] Härvid borde den svenska ansatsen för social rapportering med de s. k. levnadsnivåundersökningarna utgöra ett intressant alternativ. Med den typen av undersökningar som grund är det möjligt att komma förbi en del av de brister som föreligger med den modell man nu tillämpar. Det blir möjligt att jämföra olika befolkningsgruppers levnadsförhållanden på en rad områden sammanhållet.

Avslutningsvis skall bara påpekas att den nu förda prispolitiken har lett till att ett exemplar av ST 11 kostar £16.50. Denna summa är t. ex. ungefär hälften av vad en pensionär får i statlig pension *per vecka*. Att en bred allmänhet kommer att få tillgång till materialet förefaller därför uteslutet.

Joakim Palme

Hur jämställda är vi? Levnadsförhållanden 1975/77. Rapport 20 i SCBs serie om levnadsförhållanden. Liber Förlag, Stockholm 1980. 235 s. Kr. 90:-.

Denna rapport beskriver skillnader mellan män och kvinnor på områden utbildning, förvärvsarbete, arbetsplatsförhållanden, utbildningens effekter på arbete och inkomst, hushållsarbete, fritidsaktiviteter, sociala kontakter och hälsa. Vid redovisningen tar man fasta på olika skeden av familjecykel för kvinnor och män, socioekonomisk grupp och utbildning. Redovisningen beskriver situationen för individer utom i fråga om hushållsarbete som redovisas för gifta eller samboende par.

De föga överraskande resultaten kan sammanfattas: Kvinnor har lägre utbildning än män oberoende av ålder eller social bakgrund. Den markanta ökningen av kvinnors förvärvsarbete som ägt rum kan tillskrivas deltidsarbetande kvinnor. Kvinnors huvudsakliga verksamhetsområden är undervisning, sjuk- och socialvård medan tillverkningsindustrin sysselsätter mest män. Samboende män med barn i åldern 7–18 år har längsta arbetstider. Ju högre utbildning kvinnor har desto högre är deras förvärvsintensitet.

Heltidsarbetande män har högre inkomster än heltidsarbetande kvinnor även när båda har utbildning på samma nivå. Något påtagligt ekonomiskt utbyte får kvinnor av sin utbildning först om de har eftergymnasial utbildning. Hushållsarbetet är mycket ojämnt fördelat mellan män och kvinnor. Om mannen arbetar heltid (vilket gäller för 96% av anställda män) har omfattningen av kvinnans förvärvsarbetstid ingen större betydelse för hushållsarbetets fördelning.

Kvinnor läser veckotidningar och gör handarbeten medan män ägnar sig åt idrottsevenemang och olika utomhusaktiviteter på sin fritid. Män är mycket sällan ensamstående med barn. I allmänhet har kvinnor mera kontakt med sina grannar än män men vid småhusboende försvinner skillnaderna mellan män och kvinnor vad gäller umgänge med grannar. Fler män än kvinnor umgås med sina arbetskamrater såväl i som utanför arbetet. Huvudsakligen förklaras detta av att män är förvärvsarbetande oftare än kvinnor och att förvärvsarbetande kvinnor oftare än förvärvsarbetande män arbetar utan kontakt med arbetskamrater. När båda arbetar heltid är den fackliga organisationsgraden densamma för både män och kvinnor. Hälsomässigt skiljer sig de båda könen obetydligt från varandra. Regelbunden medicinsk behandling är

[1] Walker, A., Introduction: Public Expenditure, Social Policy and Social Planning: Public Expenditure and Social Policy, ed. by Alan Walker, London 1982.

emellertid vanligare bland kvinnor i alla åldersgrupper.

Rapportens bidrag till jämställdhetsdebatten

För den som är intresserad av likheter och olikheter mellan mäns och kvinnors levnadsförhållanden eller hur svenska kvinnors förhållanden ser ut från olika aspekter utgör rapporten en behändig uppslagsbok där de till största delen tidigare publicerade uppgifterna har samlats i en volym. Är man däremot intresserad av hur svenska män och kvinnor tillsammans har det, får man nöja sig med uppgifter om hushållsarbetets fördelning mellan män och kvinnor i gifta/samboende par.

Det är i första hand två generella begränsningar i rapporten som jag vill ta upp. Den första gäller i vilken *omfattning* kvinnors förhållanden belyses. Den andra gäller den *undersökningsenhet* som i rapporten tillämpas som grund för jämställdhetsdiskussionen. Exemplen hämtar jag ur rapporten.

Ambitionen i kap. 4 om utbildningens betydelse för arbete och inkomst är bl. a. att besvara frågan: "Är en hög utbildning ett lika effektivt konkurrensmedel för kvinnor som för män när det gäller att få en god lön eller ett stimulerande och intressant arbete?" Analysen begränsas emellertid till att endast gälla de heltidsarbetande kvinnorna med motiveringen att ".. högre befattningar erbjudes ... i mindre utsträckning i form av deltidsanställningar. ... Eftersom deltidsanställningar dessutom är mycket vanligare bland kvinnor än bland män är det särskilt betydelsefullt med en sådan begränsning..." Med denna inte speciellt väl underbyggda motivering utesluter man 40 procent av de förvärvsarbetande kvinnorna ur analysen. Resultatet blir att man inte vet om det faktiskt är *utbildningen* som ger effekt för arbete och inkomst eller om det i stället är det förhållandet att nästan hälften av kvinnorna åtminstone för en period inte ställer sig till arbetsmarknadens förfogande på heltid som är det avgörande.

Men även de deltidsarbetande kvinnorna har ekonomiskt utbyte av sin utbildning. Därför är det ingalunda oviktigt att lyfta fram deras förhållanden speciellt som de utgör en så stor del av den kvinnliga arbetskraften. De frågor man bl. a. ställer sig är: Motsvarar deltidsarbetande kvinnors arbetsuppgifter den utbildning de har eller går betydande resurser i form av outnyttjad utbildning och yrkeskvalifikationer samhället förlorade på grund av t. ex. brist på barntillsyn eller utbud av kvalificerat arbete på deltid? Finns det skillnader mellan de heltidsarbetande och de deltidsarbetande kvinnorna i detta hänseende? Hur återspeglas detta i lönen för de båda grupperna?

Att "högre befattningar" inte erbjudes i form av deltidsanställningar är högst sannolikt ett faktum som motsvarar verkligheten i dag men knappast ett förhållande som för all framtid bör accepteras som ett ovillkorligt faktum.

Generellt har man använt *individen* som undersökningsenhet. Enda undantaget är såsom tidigare nämnts, hushållsarbetets fördelning mellan män och kvinnor i gifta/samboende par. Nu är emellertid jämställdheten en mycket vidare fråga än vem som sköter markservicen. Mycket viktigare är t. ex. vilka möjligheter kvinnor resp. män har i ett livsperspektiv vad gäller utbildning, arbete, inkomst och deltagande i samhällslivet. Kärnfrågan är hur män och kvinnor *tillsammans* ordnar sina liv.

Den fråga i 1977 års ULF som belyser yrkeskarriärer i ett längre perspektiv och som ställdes för båda parterna i gifta eller samboende par har tyvärr inte bearbetats alls. Man har försummat den nästan unika möjligheten att belysa hur mannens resp. kvinnans yrkes-, utbildnings-, och karriärval och -möjligheter påverkas av familjesituationen och dess förändringar under en längre tidsperiod. Är det alltid så att det är kvinnor som får ge avkall (eller "väljer" att göra så) på sin yrkeskarriär när de får barn och under den tid barnen är små?

När man sedan behandlar fritidsavsnittet konstaterar man bara att kvinnor läser veckotidningar och gör handarbeten medan män går på idrottsmatch och ägnar sig åt andra utomhusaktiviteter. Den egentliga jämställdhetsaspekten har inte beaktats, dvs. hur mannen resp. kvinnan som har ett gemensamt hushåll och barn tillsammans

disponerar över egentlig fritid. Med egentlig fritid menar jag då frihet från ansvar över det som händer eller kan befaras hända innanför hemmets väggar. Det finns skäl att misstänka att kvinnor oftare än män är hänvisade till sådana aktiviteter som möjliggör för dem att samtidigt finnas till hands för barnen. Ensamstående kvinnor har annat aktivitetsmönster vilket är ett tydligt tecken på att det inte är något generellt "kvinnligt" beteende att tillbringa all fritid i hemmet utan snarare en fråga om bristande jämställdhet mellan män och kvinnor vad gäller att disponera över sin tid.

Den ovanstående diskussionen anger bara ett par exempel på hur kvinnors situation och hur jämställdhetsfrågan borde belysas. Den kan och bör sedan utvidgas till att gälla övriga livsområden såsom deltagande i fackliga och politiska aktiviteter osv. Den utformning som denna jämställdhetsrapport har fått är till övervägande del bestämd av männens situation och verklighet och ger ingen fullständig eller rättvisande bild av vare sig kvinnors situation eller jämställdheten mellan de båda könen.

Tarja-Liisa Leiniö

*

Replik till Tarja-Liisa Leiniö

Visst vore det intressant att som Tarja-Liisa Leiniö efterlyser få veta mer om "hur svenska män och kvinnor har det tillsammans" samt att få information om just kvinnornas situation. Eftersom rapporten baserar sig på ett tidigare insamlat material utan det primära syftet att belysa just jämställdheten har emellertid möjligheterna varit begränsade. Vårt syfte har ej heller varit att främst belysa kvinnornas situation utan huvudsakligen att belysa jämställdheten mellan könen på olika områden.

Att jämförelsen av arbetets kvalifikationsnivå resp. av arbetsinkomsten för män och kvinnor på samma utbildningsnivå begränsats till att omfatta heltidsanställda kritiseras av T-L L och visst kan man hålla med om att det hade varit värdefullt om dessa uppgifter redovisats även för de deltidsarbetande kvinnorna. Men kritiken hade säkert varit hård om inte en särredovisning av heltidsanställda män och kvinnor också ingått. En begränsning av analysen till de heltidsanställda innebär naturligtvis inte att man "för all framtid ... som ett ovillkorligt faktum accepterar att högre befattningar huvudsakligen erbjuds i form av heltidsanställningar", vilket T-L L antyder. Men om man i dag vill belysa frågan huruvida män och kvinnor i arbetslivet har lika god avkastning på sin utbildning bör man kanske göra en sådan begränsning så länge som män huvudsakligen arbetar heltid och det dessutom finns skäl att anta att arbetsgivare tar hänsyn till arbetstidens omfattning vid tillsättning av högre tjänster.

Att en högre utbildning inte ger kvinnor *samma* möjligheter som män att erhålla de mer kvalificerade arbetena och högre inkomsterna beror högst sannolikt, som T-L L säger, på just "det förhållandet att nästan hälften av kvinnorna åtminstone för en period inte ställer sig till arbetsmarknadens förfogande på heltid". Det är väl den tolkning som de flesta ansluter sig till. Oavsett hur det statistiska grundmaterialet i rapport nr 20 behandlats hade man emellertid inte kunnat visa det berättigade i T-L Ls tolkning. I rapporten förs dock en ingående diskussion av olika tolkningsmöjligheter. Det påpekas därvid bl. a. att skillnaderna mellan män och kvinnor i andelen med mer kvalificerade arbeten även kan bero på strukturella olikheter mellan de yrkesområden som män och kvinnor vanligen är verksamma inom eller med andra ord – andelen högre befattningar kan vara lägre inom exempelvis sjukvård än inom verkstadsindustrin. Detta har inte kartlagts eller beaktats.

Sammanfattningsvis har begränsningen till heltidsanställda gjorts för att det bedömts som speciellt intressant att se om det finns skillnader i arbetslivsframgång mellan män och kvinnor även då de har lika hög utbildning *och arbetar heltid*. Till skillnad från T-L L anser jag att den motiveringen är väl underbyggd.

T-L L anser också att man i rapport 20 borde ha utnyttjat uppgifterna i ULF om individernas yrkeshistorik för att belysa hur mannens och kvinnans yrkesliv och utbildningsmöjligheter påverkas av familjesitua-

tionen under en längre tidsperiod. Tyvärr är en sådan analys inte möjlig. För det första innehåller yrkeshistoriken endast uppgifter om yrken/sysselsättningar som personen arbetat inom under minst två år. Uppgifterna har nämligen insamlats för att ge information om exponering för olika arbetsmiljöer. För det andra har vi inte i ULF-materialet tillgång till någon familjehistorik över huvud taget.

Avslutningsvis konstaterar T-L L att fritidsavsnittet borde ha kompletterats med information om hur mannen och kvinnan med gemensamt hushåll disponerar sin fritid. Hur självständiga är deras aktiviteter i förhållande till familjen och barnen? En sådan analys vore värdefull och intressant. Tack vare den hushållsansats (intervju med såväl urvalspersonen som den samboende partnern) som tidigare tillämpats i ULF är det fullt möjligt att närmare studera hur fritiden fördelas inom hushållet. För att fullständigare kunna belysa jämställdheten är det således i detta och i flera andra sammanhang nödvändigt att följa T-L Ls råd och göra analysen med hjälp av information om hushållet, dvs. både mannens och kvinnans levnadsförhållanden måste kartläggas. Sådana analyser är dock komplicerade och resurskrävande och kunde därför inte beaktas vid utarbetandet av rapport 20 för vilken ambitionen var att relativt snabbt tillhandahålla en publikation om jämställdhet baserad huvudsakligen på redan befintligt material.

Lars-Gunnar Andersson

Shapiro, S.S. and Gross, A.J., Statistical Modeling Techniques. Marcel Dekker, Inc., (Vol. 38). New York 1981. 384 s. $ 29:75.

Bokens syfte är att ge studerande i naturvetenskapliga och samhällsvetenskapliga ämnen en introduktion av statistiska modelltekniker användbara för att analysera komplexa system. Boken är indelad enligt följande huvudkapitel:

1. Basic concepts of statistical models
2. Concepts of statistical theory
3. Models för measurement: continuous case
4. Models for measurement: discrete case
5. Empirical models
6. Testing model assumptions
7. Analysis of systems.

I kap. 1 definieras grundläggande modellbegrepp. Ett system definieras här som de objekt eller skeenden som man vill studera. En genomgång av grunderna i statistisk teori görs i kap. 2. I kap. 3 och 4 beskrivs modeller som kan användas som idealiseringar av saklogiska samband i ett system. Exempel på sådana modeller är exponential-, Weibull-, gamma- och Poisson-fördelningarna. Empiriska modeller kallas sannolikhetsfördelningar som är lämpliga att använda då man på bästa sätt vill approximera ett givet empiriskt material utan ambitionen att förstå de saklogiska systemsambanden.

Exempel på sådana fördelningar är den generaliserade lambda-familjen och Pearson-familjen. Detta behandlas i kap. 5. Kap. 6 behandlar test av modellantaganden, t.ex. χ^2- och Kolmogorov-Smirnov-tester. För att förstå komplexa system kan man exempelvis använda Monte Carlo-simuleringar. Resultat från sådana simuleringar leder sedan fram till val av modell. Detta beskrivs i kap. 7.

Resultaten i boken beskrivs med kortfattade matematiska motiveringar tillsammans med ett stort antal exempel och referenser. Vad boken i huvudsak består av är beskrivningar av ett stort antal statistiska modeller samt testmetoder för modellantaganden. Detta ger en något stereotyp bild av problemlösning med modellansatser. Man saknar diskussioner om vilka förutsättningar som ska vara uppfyllda för att kunna lösa problem med modellantaganden samt om modellers robusthet att beskriva studerade system då dessa förändras över tiden. Systemdefinitionen i kap. 1 innehåller ej tidsdimensionen explicit.

Den sammantagna bedömningen av ovanstående är att boken ej kan rekommenderas som en introduktion i ämnet. Boken kan dock med fördel användas som en handbok över vissa statistiska modeller samt testmetoder för modellantaganden.

Tomas Garås

Utkom från trycket 17 mars 1983

STATISTISK TIDSKRIFT

UTGIVEN AV STATISTISKA CENTRALBYRÅN

Ansvarig utgivare: Lennart Nilsson. Redaktör: Thore Grönqvist

Adr: Karlavägen 100. Postadress: 115 81 Stockholm

Telefoner: 08/14 05 60, ankn. 4260 (redaktören personligen)
och 4265 (Helen Lidén)

Prenumeration genom bokhandel eller hos Liber distribution, Prenumerationsorder, 162 89 Stockholm

Utkommer med 5 häften om året

Pris ca 105:– per år eller ca 20:– per häfte (inkl. moms)

Tredje följden Årg. 21 Nr 2 1983

Innehåll

Artiklar

Daniel Thorburn:
On Methods for Disclosure Control in Longitudinal Studies 93

Ronnie Andersson:
Kodning av arbetsställe vid folk- och bostadsräkningen 1980 (FoB 80) 102

Olle Sjöström:
Vad är statistisk metod? 109

Stig Tegle:
Deltidssysselsatta i SCBs arbetskraftsundersökningar 1965–1982. Synpunkter och förslag 121

Meddelanden

OECD-möten:
Möten om sociala indikatorer 134
Nordiska statistiska utskott:
Utskottet för näringsstatistik (NUN) 135
Sextonde nordiska statistikermötet 139
ILOs 13e arbetsmarknadsstatistikerkonferens 141
Möten om kulturstatistik 151
Möte om kulturstatistik och seminarium om kulturindustri 153
Internationella havsforskningsrådet 154
Kommitté för yrkeskod för statistiker (ISI) 154
Konferens om den aktuella befolkningsutvecklingen i Europa 155
Konferensen "NordDATA 82" 155
Arkivmöten 155
Programmeringsmöte i Östberlin 156

Statistical Computing Project: seminarium i RAPID-frågor 157
COM-tekniken vid Norges Sentralbyrå 157
SCB hos Census Research Group 157
Nya metoder för folk- och bostadsräkningar 158
Statistiska föreningens verksamhet år 1982 160
Personalnotis 163

Summary in English 164

Litteratur

Bokanmälningar:
Vinod, Hrishikesh D. & Ullah, Aman, Recent Advances in Regression Methods (Martin Ribe) 167
Hájek, J., Sampling from a finite population (Claes Cassel och Ingrid Lyberg) .. 170
Öhrvik, J., Robusteffektiv lägesestimation [Akad. avh.] (Marianne Frisén) 173

STATISTICAL REVIEW

PUBLISHED BY STATISTICS SWEDEN, STOCKHOLM, SWEDEN

Five issues a year

Third series. Vol. 21 No. 2 1983

Contents

Articles

Daniel Thorburn:
On Methods for Disclosure Control in Longitudinal Studies 93

Ronnie Andersson:
Coding of Establishments in the 1980 Population and Housing Census 102, 164

Olle Sjöström:
What is statistical method? 109, 164

Stig Tegle:
Part-time employment in the Swedish Labour Force Surveys 121, 165

Notes

OECD-meetings:
Meeting on social indicators 134
Inter-nordic statistical committees:
Committee on food statistics 135
Sixteenth meeting of the nordic statisticians 139
13th ILO-conference of the labour market statisticians 141
Meeting on cultural statistics 151
Meeting on cultural statistics and seminar on industrialized culture 153
The International Council for the Exploration of the Sea (ICES) 154
Committee on professional code for statisticians (ISI) 154
Conference on Current Population Problems in Europe 155
The conference "Nord-DATA-82" 155
International Council on Archives: Meeting 155
Meeting on programming in the German Democratic Republic 156
Statistical Computing Project: Seminar on RAPID-questions 157
Visit at the Central Bureau of Statistics in Norway on COM-technique 157
Visit at Census Research Group in London 157
Future Population and Housing Census: New methods 158
Swedish statistical society: Activity in 1982 160
Personnel note 163

Summary in English 164

Literature

Book reviews:
Vinod, Hrishikesh D. & Ullah, Aman, Recent Advances in Regression Methods (Martin Ribe) 167
Hájek, J., Sampling from a finite population (Claes Cassel, Ingrid Lyberg) 170
Öhrvik, J., Robusteffektiv lägesestimation [Akad. avh.] (Marianne Frisén) 173

On Methods for Disclosure Control in Longitudinal Studies

by Daniel Thorburn[1,2]

0. Summary

In longitudinal studies the same subjects are studied at different time-points. For privacy reasons, it may be unwise to retain individual data over the whole period of time studies. Different methods are discussed to protect the privacy of the subjects without losing all the merits of longitudinal studies. We first compare this situation with that leading to randomized response and use some ideas from that technique. Next we discuss how tables can be constructed so that longitudinal studies may be facilitated. The methods are illustrated by examples in which reasonable protection can be achieved by a loss in efficiency of between 50% and 80%.

1. Introduction

In many situations it is essential to be able to relate the behavior of the same individual at different times. This can be achieved by a longitudinal study: a study whose principal aim involves the actual measurement of individuals at several well defined occasions. (Goldstein, 1979, p. 2) The term "panel study" is sometimes used in the social science literature. In this paper we are mainly concerned with pure longitudinal studies where every individual should be measured on every single occasion. The methods can, however, be used also in other so-called mixed longitudinal studies. More material on longitudinal studies can be found in the books by Goldstein (1979) and Wall & Williams (1970).

In longitudinal studies the individual data are usually kept from one time-point to another. This procedure may be unsatisfactory, however, since the risk of invasion of privacy may be too large. There are two aspects to this risk. First, if the data are very personal or intimate, it may be too risky to keep the files during the intermediate time. Second, the combination of data from two different studies can also increase the risk. During the last decades new rules and legislation have been developed to protect individuals from the misuse of data files.

Different techniques have been developed so that longitudinal studies can be used without making the risk of invasion of privacy too large. Work has been done mainly in two directions. The first direction is to use technical devices to protect the files; examples are keys or passwords, coding, unidentification or file-splitting. In this paper we discuss the second approach in which the data is changed so that the original data cannot be recovered.

[1] Dep. of math. stat. Univ. of Lund.

[2] Work reported herein was supported by the Swedish Council for Research in the Humanities and Social Sciences, contract no. F 177/81, and by Stiftung Volkswagenwerk.

2. Techniques similar to randomized response

Randomized response was originally developed for personal interviews (Warner 1965). It is based on a very simple idea. The following description is taken from Anderson (1975):

Suppose one wants to estimate the proportion π of a population that has some sensitive characteristic A. The sampled persons are asked to draw and answer a question from a pack of cards with two types of questions:

Are you an A-individual?

and

Are you an A^*-individual?

where A^* denotes "not A". The proportions of cards are p and $1\text{-}p$, respectively, where p is a number $\neq 1/2$ determined by the statistician. Since the interviewer does not know which question is answered, he cannot infer the attribute of the respondent. The proportion π can still be estimated, because the probability of a yes-answer is a linear function of π.

This technique can be generalized to variables which are not dichotomous. If a person has the property A_i, he is asked to give a random answer Z which is drawn from the distribution f_i. There is no restriction on the type of variable that is studied: it may be quantitative or qualitative. The only restriction is that the interviewer must not know from which distribution the answer is drawn.

Lanke (1975) and Leyseffier (1975) have used the conditional probability $P(A/Z)$ to compare the protection given by different schemes in the simple yes-or-no case. Anderson (1975) introduced the expected conditional information from A_i given Z for the same purpose in the general case.

This paper deals only with the situation in which the same individuals are included at all time-points in the study. When there is non-response, special measures must be taken which resemble those used in the standard situation. Non-response usually leads to a bias and this will be the case for the present methods too.

Another situation in which the samples change between time-points is, for example, in studies using rotating panel designs. If the sample is divided into parts according to their entry and planned exit times, the methods presented here can be used for each part separately.

The idea behind randomized response is to replace the data with less sensitive variables which still contain some information on the subject. A similar technique can also be used in the present situation. Suppose that a register contains a variable A which is considered sensitive and must be removed. Assume further that it is important to be able to do longitudinal studies later on. To solve this problem, the sensitive variable is replaced with a new variable Z_i, whose distribution depends on the true value of the original variable A. This is done for all individuals in the file.

Example 2.1. Suppose that a file contains 20 individuals and that the register is

$(Y_{1,1}, \ldots, Y_{1,20}) =$

(11011 10111 11010 11111). (2.1)

Next replace each 1 with 0 with probability 1/4 and each 0 with 1 with the same probability. The new register may look like the following:

$(Z_{1,1}, \ldots, Z_{1,20}) =$

(11001 00100 11110 11111). (2.2)

If this new Z-register is stored until the time of the next survey, all individuals are protected:

$P(Y_{1,i} = 0 \mid Z_{1,i} = 0) \simeq 0.43,$

$P(Y_{1,i} = 0 \mid Z_{1,i} = 1) \simeq 0.08.$

Suppose that the following data is observed in the second round:

$(Y_{2,1}, \ldots, Y_{2,20}) =$

(11011 10101 11011 11101). (2.3)

The transition probabilities $p_{jk} = P(Y_{2,i} = k, Y_{1,i} = j)$, $(j = 0,1;\ k = 0,1)$ can now be estimated by setting the observed frequencies of transitions between Z_1 and Y_2 equal to their expected values:

$$\frac{3}{4}\pi_1 p_{11} + \frac{1}{4}\pi_0 p_{01} = 11/20,$$

$$\frac{3}{4}\pi_0 p_{00} + \frac{1}{4}\pi_1 p_{10} = 3/20,$$

where $\pi_j = P(Y_{1,i} = j)$ is estimated from the Y_1 or Z_1 variables. The estimates turn out to be

$$\hat{\pi}_0 = 1 - \hat{\pi}_1 = 4/5,$$

$$\hat{p}_{00} = 1 - \hat{p}_{01} = 7/8,$$

$$\hat{p}_{11} = 1 - \hat{p}_{10} = 29/32.$$

These estimates are, of course, not as good as if the non-randomized series (2.1) and (2.3) had been used. E.g., it can be shown that Var $(p_{00}^*) \approx 1.0$. When the transition probabilities are estimated from (2.1) and (2.3) the corresponding variance is 0.05. The loss of efficiency is thus almost 19/20 in the estimate of p_{00}. The loss is smaller but still considerable in the estimate of p_{11}.

The randomized response technique and its merits and disadvantages are discussed by Anderson (1975).

It is possible to randomize in other ways when the whole sequence is known. In some situations it may be desirable to construct the randomization in such a way that it gives the same estimates as the original sequence for each time-point. For example, one can select a fixed number of pairs randomly and change their values. The result may be (with 5 pairs: (4,14), (5,17), (6,18), (11,13), and (15,19)):

(11011 10111 01111 11101) (2.4)

The protection afforded by this randomization is

$P(Y_{1,i} = 0 \mid 0) = 11/19 = 0.58,$

$P(Y_{1,i} = 0 \mid 1) = 4/19 = 0.21.$

The so called non-reversible privacy transformations are other similar methods (Reiss, Post, Dalenius (1982)).

The protection is thus not as good with the scheme (2.4) as with (2.2). On the other hand, one's ability to estimate the transition probabilities is a little better. The measures and methods developed by Lanke (1975), Leyseffier (1975) and Anderson (1975) can be used to compare different schemes for privacy protection and to estimate the changes between time periods. As Anderson (1975) points out, an increase in the protection must be weighed against a decrease in the efficiency of estimation. This will not be discussed further here. We merely mention that the technique can also be used with more than two observation periods, but the calculations will be more complicated and the efficiency even smaller.

The protection described in this section may have a psychological disadvantage. It can be described in the following words (deliberately chosen to sound unfavorable). "For each person in a register there is stored a value which has a large probability of being incorrect (or may be transformed into such a value). It is unsatisfactory to have a register in which somewhat more than a quarter of the persons are classified as, say, tax-evaders". This is only a problem in the choice of words, but it may be felt to be a real problem for persons without scientific training.

3. Tables constructed to facilitate later studies

Before the data from the first round of a study is destroyed it may, of course, be analyzed statistically and tables may be constructed. These are usually made with the sole purpose of illustrating the situation at that time. They may, however, also be constructed with another purpose, viz. to facilitate longitudinal studies. To my knowledge nothing has been written on how such tables should be constructed.

Let us first give a simple example which for clarity deals with only one time-point.

Example 3.1. Suppose that there is a register with $N = nk$ persons. For each person the sex and annual income are recorded. The persons are randomly divided into n groups with k members in each. A table is now constructed with n cells. For each cell the number of females X_ν and the sum of the incomes Z_ν are given, $\nu = 1, \ldots, n$.

Given this table, the average income for males (W_m) and females (W_f) can be estimated in the following way. It is easily seen that

$$E(Z_\nu|X_\nu) = kW_m + X_\nu(W_f - W_m).$$

We assume that $\mathrm{Var}\,(Z_\nu|X_\nu) = k \cdot s^2$ (this assumption is not necessary but yields simple calculations). Standard linear regression can now be applied to give

$$W_f^* = \frac{\bar{Z}}{k} + (1-p)\frac{\Sigma_1^n (X_\nu - \bar{X})(Z_\nu - \bar{Z})}{\Sigma (X_\nu - \bar{X})^2}$$

$$W_m^* = \frac{\bar{Z}}{k} - p\frac{\Sigma_1^n (X_\nu - \bar{X})(Z_\nu - \bar{Z})}{\Sigma_1^n (X_\nu - \bar{X})^2}$$

where $\bar{Z} = \frac{\Sigma_1^n Z_\nu}{n}$, $\bar{X} = \frac{\Sigma_1^n X_\nu}{n}$,

and p is the relative proportion of women. Omitting the correction for a finite population, it follows that

$$\mathrm{Var}\,(W_f^* - W_m^*) = \frac{ks^2}{\Sigma (X_\nu - \bar{X})^2} \approx \frac{s^2}{np(1-p)},$$

$$\mathrm{Var}\,(W_f^*) \approx \frac{s^2}{nkp}(1 + (k-1)(1-p)),$$

$$\mathrm{Var}\,(W_m^*) \approx \frac{s^2}{nkp}(1 + (k-1)p). \qquad (3.1)$$

From formula (3.1) it follows that the decrease in the variance of $W_f^* - W_m^*$ is proportional to k. The precision is thus proportional to the number of groups and independent of the group size.

The same technique can be used in longitudinal studies if the second and the first samples can be grouped in the same way. Example 3.1 can be adapted to illustrate this if we posit that the data on sex is collected in the first round and the annual income data in the second. The next example also illustrates this idea.

Example 3.2. Suppose that a register contains data about 10 000 individuals. The first record tells whether or not the person has been exposed to a certain carcinogenic substance. In the next round data is collected on whether or not cancer has developed 30 years later. The persons are randomly divided into 1000 groups with 10 persons in each. In the first round the number of exposed individuals in each group is tabulated, and in the second the number of cancers developed in each group is tabulated. After that a longitudinal study is made by comparing the different groups.

Table 3.1. The number of persons exposed and non-exposed, with and without cancer

Exposed	Cancer		
	Yes	No	Total
Yes	523	457	980
No	944	8076	9020
Total	1467	8533	10000

If the data had not been grouped, it could have been presented as in table 3.1. When the data is grouped the resulting table will look like table 3.2 instead.

Table 3.2. The number of groups according to the number of persons exposed and the number of cancers

Number of exposed	Number of cancers					Total
	0	1	2	3	4–	
0	140	132	68	24	6	370
1	65	133	110	49	15	372
2	20	46	63	35	21	185
3	4	11	22	12	9	58
4–	1	3	2	5	4	15
Total	230	325	265	125	55	1 000

If we test independence between the two factors with a χ^2-test, we obtain $\chi^2 = 1\,300$ with one degree of freedom from table 3.1. A standard χ^2-test in table 3.2 would give 157 with 16 degrees of freedom. This test is too weak, however, since we know that the relation should be linear. A better test in this situation is based on Fisher's z-variable. It is $z = 1/2\log((1+r)/(1-r)) = 0.3850$. Its normalized square, $(n-3)z^2 = 148$, has an approximate χ^2-distribution with one degree of freedom if the variables are independent.

Note that the χ^2-value for table 3.1 is almost k times as large as that for table 3.2. This is no coincidence. The theoretical coefficient of correlation between the two 0–1 variables (cancer/no cancer and exposed/not exposed) is exactly the same as that between the number of exposed and the number of cancers in the same group, regardless of the group size k.

The previous examples have dealt with one and two time-points. In the next example a longitudinal study with T time-points will be discussed. It is a little more complicated, but when assessing the merits of different protection schemes in longitudinal studies it is necessary to consider sampling schemes of this type. The technical details and formulas of this example are given without a proof. They can be omitted at a first reading.

Example 3.3. Let the value for every person follow an individual straight line, but with a random error:

$$Y_{t,i} = a_i + b_i(t - \frac{T+1}{2}) + \varepsilon_{t,i},$$

$i = 1, \ldots, N$, $t = 1, \ldots, T$,

where $\varepsilon_{t,i}$ is normal with mean 0 and variance σ^2. Further, the pair (a_i, b_i) is considered to come from a normal distribution with mean (a, b) and variance $\begin{pmatrix} \sigma_{aa} & \sigma_{ab} \\ \sigma_{ab} & \sigma_{bb} \end{pmatrix}$. The object of the study is to estimate these parameters.

If the data for each individual were known for all periods, we could estimate

$$a_i^* = \frac{1}{T}\Sigma_1^T Y_{t,i},$$

$$E((a_i^* - a_i)^2) = \sigma^2/T,$$

$$b_i^* = \Sigma_1^T Y_{t,i}(t - \frac{T+1}{2}) / \Sigma_1^T (t - \frac{T+1}{2})^2,$$

$$E((b_i^* - b_i)^2) = \sigma^2 / \Sigma_1^T (t - \frac{T+1}{2})^2,$$

$$\sigma_{(i)}^{*2} = \frac{1}{T-2}\Sigma(Y_{t,i} - a_i^* - b_i^*(t - \frac{T+1}{2}))^2.$$

It also follows that these estimates are normal:

$$\begin{pmatrix} a_i^* \\ b_i^* \end{pmatrix} \varepsilon N\left\{\begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} \sigma_{aa} + \sigma^2/T & \sigma_{ab} \\ \sigma_{ab} & \sigma_{bb} + \sigma^2/\Sigma_1^T (t - \frac{T+1}{2})^2 \end{pmatrix}\right\}.$$

The problem is now reduced to a standard two-dimensional problem. Its solution can be written down immediately

$$a^* = \frac{1}{N}\Sigma_1^N a_i^*,$$

$$\text{Var}(a^*) = \sigma^2/(NT),$$

$$b^* = \frac{1}{N}\Sigma_1^N b_i^*,$$

$$\text{Var}(b^*) = \sigma^2/(N\Sigma_1^T(t-\frac{T+1}{2})^2),$$

$$\sigma^{*2} = \frac{1}{N}\Sigma_1^N \sigma_{(i)}^{*2},$$

$$N(T-2)\sigma^{*2}/\sigma^2 \varepsilon \chi^2(N(T-2)),$$

$$\sigma_{aa}^* = \frac{1}{N-1}\Sigma_1^N(a_i^*-a^*)^2 - \frac{\sigma^{*2}}{T},$$

$$\text{Var}(\sigma_{aa}^*) = \frac{2(\sigma_{aa}+\frac{\sigma^2}{T})^2}{N-1} + \frac{2\sigma^4}{T^2(T-2)N},$$

$$\sigma_{bb}^* = \frac{1}{N-1}\Sigma_1^N(b_i^*-b^*)^2 - \frac{\sigma^{*2}}{\Sigma_1^T(t-\frac{T+1}{2})^2},$$

$$\text{Var}(\sigma_{bb}^*) = \frac{2(\sigma_{bb}+\frac{\sigma^2}{\Sigma_1^T(t-\frac{T+1}{2})^2})^2}{N-1} + \frac{2\sigma^4}{(T-2)N(\Sigma_1^T(t-\frac{T+1}{2})^2)^2},$$

$$\sigma_{ab}^* = \frac{1}{N-1}\Sigma_1^N(a_i^*-a^*)(b_i^*-b^*),$$

$$\text{Var}(\sigma_{ab}^*) = \frac{1}{N-1}((\sigma_{aa}+\frac{\sigma^2}{T})(\sigma_{bb}+\frac{\sigma^2}{\Sigma_1^T(t-\frac{T+1}{2})^2})+\sigma_{ab}^2).$$

Next let the N persons be divided into n groups of equal size k. The group means $Z_{t,\nu}$, $\nu = 1, \ldots, n$ will then follow a straight line plus an error term:

$$Z_{t,\nu} = a_\nu^* + b_\nu^*(t-\frac{T+1}{2}) + n_{t,\nu},$$

where $n_{t,\nu}$ are independent normal with mean 0 and variance σ^2/k. Further, the pair (a_ν^*, b_ν^*) is from a normal distribution with mean (a,b) and variance $\frac{1}{k}\begin{pmatrix}\sigma_{aa} & \sigma_{ab}\\ \sigma_{ab} & \sigma_{bb}\end{pmatrix}$.

The same analysis can now be used as for the ungrouped case. If the estimates from the *grouped* data are marked with + instead of *, one gets after some calculations

$$a^+ = a^*, \qquad b^+ = b^*,$$

$$\text{Var}(\sigma^{+2}) = k \cdot \text{Var}(\sigma^{*2}),$$

$$\text{Var}(\sigma_{aa}^+) = (\frac{2(\sigma_{aa}+\frac{\sigma^2}{T})^2}{n-1} + \frac{2\sigma^4}{T^2(T-2)n}) \approx k\,\text{Var}(\sigma_{aa}^*),$$

$$\text{Var}(\sigma_{bb}^+) = (\frac{2(\sigma_{bb}+\sigma^2/\Sigma_1^T(t-\frac{T+1}{2})^2)}{n-1} + \frac{2\sigma^4}{(T-2)n(\Sigma_1^T(t-\frac{T+1}{2})^2)^2}) \approx k\,\text{Var}(\sigma_{bb}^*),$$

$$\text{Var}(\sigma_{ab}^*) = \frac{(\sigma_{aa}+\sigma^2/T)(\sigma_{bb}+\sigma^2/\Sigma_1^T(t-\frac{T+1}{2})^2)+\sigma_{ab}^2}{n-1} \approx k\,\text{Var}(\sigma_{ab}^*),$$

It is seen that the precision is a function of the number of groups and is independent of the group size in the estimate of σ, σ_{aa}, σ_{bb} and σ_{ab}. This means that the grouping leads to a decrease in efficiency with a factor $1/k$.

In all the previous examples it turned out that the precision or efficiency was proportional to the number of groups and independent of the group size k. If the respondent is protected by a group size k, this means that the sample must be k times as large in order to be able to draw the same conclusions. This can be seen to hold more generally. An example is the following proposition, the proof of which is trivial.

Proposition 3.1. Let $Y_{t,i,h}$ be the observation on variable h at time t for person number i. Further, let $\sqrt{k}\ Z_{t,v,h}$ be the group total of variable h at time t in group v. Then the covariance matrix of $(Y_{t,i,h})_{\substack{1\le t\le T\\ 1\le h\le H}}$ equals the covariance matrix of $(Z_{t,v,h})_{\substack{1\le t\le T\\ 1\le h\le H.}}$

In this section we have so far discussed only simple random grouping. If the grouping is not simple random but a more controlled random method is used, the loss in efficiency can be made smaller with the same protection. This is illustrated in the next example.

Example 3.4 (continued from example 3.2). Let us make the grouping so that each group contains 0 or a fixed number . . 5, say . . of exposed individuals. The resulting table will look like table 3.3.

Table 3.3. The number of groups according to the number of persons exposed and number of cancers

Number of exposed	Number of cancers								
	0	1	2	3	4	5	6	7-	Total
0	308	284	151	42	14	5			804
5	2	12	34	57	47	33	7	4	196
Total	310	296	185	99	61	38	7	4	1 000

The same calculations as after table 3.2 yield $r = 0.6641$ and $(n\text{-}3)z^2 = 638$. This may indicate that this grouping procedure is about four times as efficient as simple random grouping. (The exact figure is 4.5, as will be seen later.)

If the groups were even more different from each other the efficiency would be even higher. The most efficient grouping is 98 groups with only exposed persons and 902 without any exposed. Unfortunately, this procedure is not satisfactory since it gives no protection at all. The efficiency must be balanced against the protection given by the scheme, exactly as with the randomized response technique.

From formula (3.1) in example 3.1 it is seen that the variance is proportional to $k/\Sigma(X_v - \bar{X})^2$. A corresponding result also holds for the situation in examples 3.2 and 3.4. The grouping should thus be done so that $E(\Sigma(X_v - \bar{X})^2)$ is as large as possible subject to the protection restrictions. In examples 3.2 and 3.4 the X-variables correspond to the number of exposed persons. After some simple calculations it can be shown that the efficiencies of the methods of tables 3.1, 3.2 and 3.3 are in the following proportion: 10: 1: 4.5.

If more than one variable is collected at each time, different criteria are best for different variables. The analysis can then be a little more complicated, since one must in some cases correct the result for the influence of the grouping procedure. If the data are collected at more than two time-points similar difficulties may appear. If the grouping is simple random no such problems arise.

We shall finally mention some aspects of the protection of the different schemes. For the scheme of example 3.2 we have

P(person i has been exposed I the grouped data) ≤ 0.5
for all i and not more than 0.3 for all but 1.5% of the persons (if the randomization gives an extreme result the protection might be worse). For the scheme of example 3.4 it also holds that

P(person i has been exposed I the grouped data) ≤ 0.5.

The protection is thus the same in both schemes for the persons with the smallest protection, but on the average the protection is better in the scheme of example 3.2. However, the protection must be weighed against the efficiency of the scheme. A scheme like that in example 3.4 is usually a better choice.

The protection in the scheme in example 3.3 where the variables are continuous can be given as follows:

$$E((Y_{t,i}^{*} - Y_{t,i})^2) = \frac{k-1}{k}(\sigma_{aa} + \sigma_{bb}(t-T)^2 + \sigma^2)$$

where $Y_{t,i}^{*}$ is the best possible estimate of $Y_{t,i}$, given $Z_{t,\nu}, \nu = 1, \ldots, n$. The protection is thus quite good already for small group sizes.

One advantage of this type of protection scheme compared to the randomized response techniques of section 2 is that it can be generalized more easily to longitudinal studies with more than two rounds.

4. Conclusions

It has been shown that there are many techniques for doing longitudinal studies that yield reasonable protection of the privacy of individual persons. The efficiency can be made higher the less protected the subjects are. A reasonable protection can usually be obtained when the efficiency is between 0.2 and 0.5.

Much remains to be done to develop the best schemes for different types of studies and questions. In this paper we have only given some examples. For longitudinal studies with more than two sampling periods the grouping technique is simpler than that based on randomized response. Furthermore, the efficiency of the grouping techniques does not decrease with the number of periods. Finally, we note that although the simple random grouping is not optimal, the calculations based upon it are simpler when there are more than two sampling times.

5. Acknowledgement

I want to thank professor Tore Dalenius for his enthusiasm.

6. References

[1] Anderson, H. (1975): Efficiency versus protection in randomized response. Ph. D. thesis, Department of Math. Statistics, University of Lund, Sweden.

[2] Anderson, H. (1976): Estimation of a proportion through randomized response. Int. Stat. Rev. 44, pp. 213–217.

[3] Anderson, H. (1977): Efficiency versus protection in a general randomized response model. Scand. J. Statist. 4, pp. 11–19.

[4] Dalenius, T. and Klevmarken, A. (eds.) (1976): Proceedings of a symposium on personal integrity and the need for data in the social sciences. Swedish Council for Social Science Research.

[5] Goldstein, H. (1979): The Design and Analysis of Longitudinal Studies. Academic Press, London.

[6] Lanke, J. (1975): On the degree of protection in randomized interviews. Report no. 2 of the research project Confidentiality in Surveys, Department of Statistics, University of Stockholm, Sweden.

[7] Leyseffier, F. W. (1975): Respondent jeopardy in randomized response procedures. Technical report, Department of Statistics, Florida State University.

[8] Rao, C.R. (1965): Linear statistical inference and its applications. John Wiley & Sons, New York.

[9] Reiss, S. P., Post, M.J. and Dalenius, T. (1982): Non reversible privacy transforms, Report PRT 2/4, Dep. of statistics, Brown University.

[10] Thorburn, D. (1982): Estimating transition matrices from aggregate stock data, Techn. rep. 1982:7, Dep. of math. statistics, Univ. of Lund, Sweden.

[11] Wall, W. D. and Williams, H. C. (1970): Longitudinal Studies and the Social Sciences, Heineman, London.

[12] Warner, S.L. (1965): Randomized response: a technique for eliminating evasive answer bias. J. Amer. Statist. Assoc. 60, pp. 63–69.

Kodning av arbetsställe vid folk- och bostadsräkningen 1980 (FoB 80)

av byrådirektör Ronnie Andersson

1. Inledning

Kodning beskrivs ofta på följande sätt (se t. ex. Lyberg (1981)). En population består av N individer, vilka betecknas I_1, $I_2, \ldots, I_N$. Var och en av dessa individer tillhör en och endast en av, säg, C varandra uteslutande kategorier. Kodning kallas den operation som hänför individerna till resp. kategori.

Varje kategori identifieras med en kod. Hur koderna bör konstrueras beskrivs i exempelvis Felme (1980). Mängden av koder kallas nomenklatur eller kodschema. Oftast kodas i en undersökning flera variabler för varje individ. I 1975 års folk- och bostadsräkning (FoB 75) kodades t. ex. variablerna hushållsställning och sysselsättning för alla uppgiftsskyldiga (ca 6,4 milj. individer)

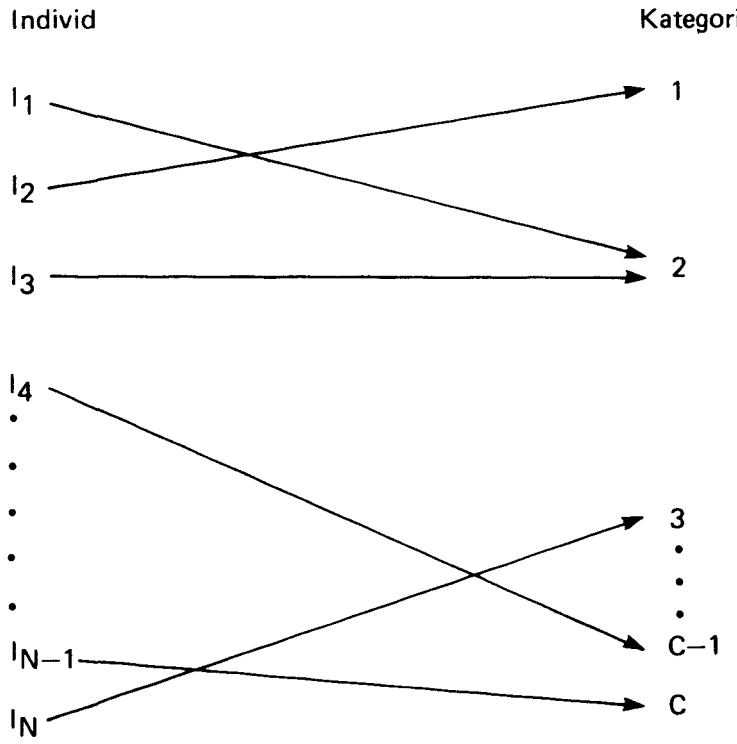


Fig 1.

samt variablerna yrke, yrkesställning, näringsgren, anställningens art och arbetsplatsens belägenhet för alla förvärvsarbetande (knappt 4 milj. individer). Kodningen i FoB 75 beskrivs i Ekwall och Andersson (1975) och i Andersson (1977 b).

I FoB 80 användes en i FoB-sammanhang helt ny kodningsteknik för variablerna näringsgren, sektortillhörighet (motsvarar ungefär anställningens art i FoB 75) och arbetsplatsens belägenhet.

Syftet med denna artikel är att beskriva denna kodningsteknik; kodning av arbetsställe eller – som den ibland kallas – löpnummerkodning. En något utförligare beskrivning av löpnummerkodning finns i Andersson (1982).

2. Principerna för löpnummerkodning

I stället för att koda varje individ $I_1, \ldots, I_N$ i populationen med koder för variablerna $v_1, \ldots, v_n$ utnyttjas en annan population med individerna $F_1, \ldots, F_P$ som redan är eller enkelt kan bli kodade avseende variablerna $v_1, \ldots, v_n$. Varje individ $F_1, \ldots, F_p$ identifieras med ett unikt identitetsnummer (löpnummer). Därefter knyts individerna I_i ($i = 1..N$) ihop med individerna F_i ($i = 1..P$) genom att de "ursprungliga" individerna $I_i, \ldots, I_N$ kodas med löpnummer till någon F_i ($i = 1..P$).

I och med att kopplingen via löpnummer är klar kan man maskinellt överföra koderna $(v_1, \ldots, v_n)$ till alla I_i som tillhör resp. F_i. För att det hela skall vara ekonomiskt

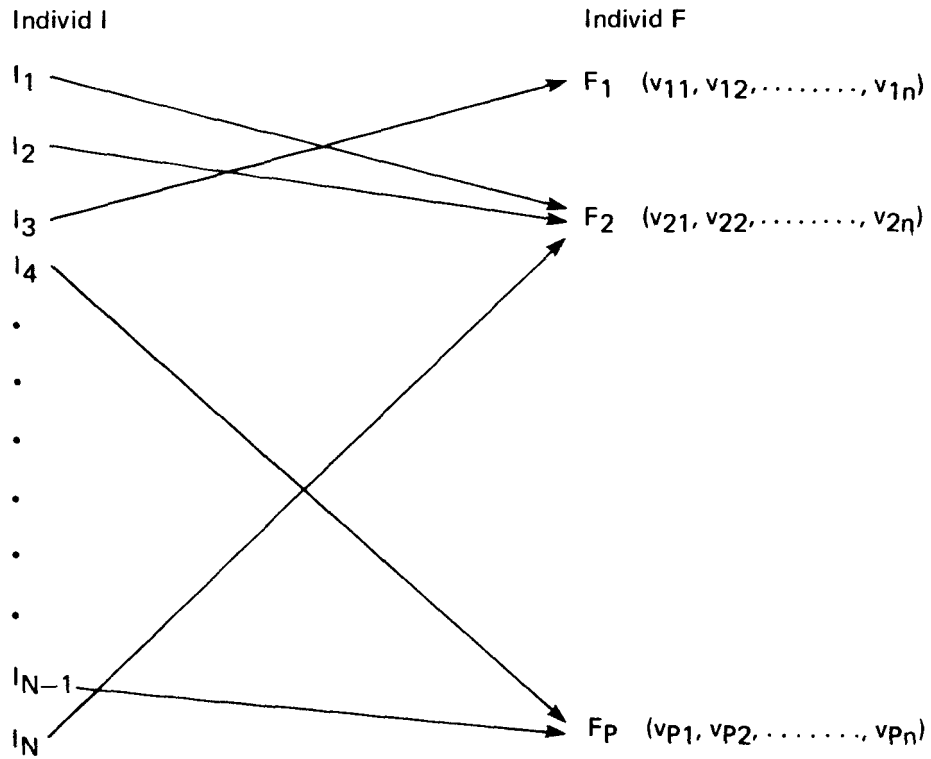


Fig. 2.

lönsamt bör P vara betydligt mindre än N. Dessutom bör F_i redan vara kodade eller enkelt kunna kodas.

3. Löpnummerkodningen i FoB 80

Löpnummerkodningen i FoB 80 bygger på det faktum att alla personer som arbetar vid samma arbetsställe skall ha samma koder för ett flertal variabler t. ex.

- näringsgren
- sektortillhörighet
- arbetsplatsens belägenhet.

Löpnummerkodningen i FoB 80 innebär att varje förvärvsarbetande person knyts till resp. arbetsställe med ett löpnummer. Därefter kan man maskinellt via löpnumret påföra personposterna den näringsgrenskod m. m. som gäller för resp. arbetsställe.

Löpnummerkodningen i FoB 80 utfördes av de kommunala granskningsorganen (GO). Varje kommun hade ca ett halvt år före räkningens genomförande fått en listning av arbetsställena inom resp. kommun utifrån det centrala företagsregistret (CFR), som finns vid SCB. Granskningsorganens uppgift var att rätta och komplettera arbetsställena i registret. Sedan skickades listorna till SCB för uppdatering av arbetsställeregistret.

Vid räkningstillfället fick kommunerna nya utlistningar av arbetsställeregistret (s. k. kodningsregister).

Dessa kodningsregister fick kommunerna i tre versioner, dels sorterade efter adress och dels efter företagets namn resp. benämning. Varje arbetsställe hade sitt specifika löpnummer. Registren beskrivs mer detaljerat i avsnitt 5.

GOs kodning av löpnummer gick till på följande sätt. GO jämförde beskrivningen av arbetsstället på FoB-blanketten (fråga 7–9) med motsvarande uppgifter i arbetsställeregistret. Då GO hittat rätt arbetsställe i registret skrevs löpnumret med "optiska" siffror på FoB-blankettens kodningsdel.

Om GO inte hittade "arbetsstället" i registret, skrev GO en tilläggsblankett för detta arbetsställe. Dessa arbetsställen fick tilläggslöpnummer och detta nummer kodades också på blankettens kodningsdel. Tilläggsblanketterna skickades till SCB och ursprungsregistret uppdaterades med dessa tillägg.

Kodningsdelarna, på vilka GO bl. a. kodat löpnummer, lästes optiskt vid SCB. Därefter genomgick det inlästa beståndet ett granskningsprogram. I granskningsprogrammet kontrollerades bl. a. löpnumren. Sista siffran i löpnumret är en checksiffra. Checksiffran användes bl. a. för att upptäcka felskrivningar av GO och substitut[1] vid inläsningen.

Rejecten[2] samt de fel som upptäcktes i granskningsprogrammet rättades med hjälp av bildskärmsterminaler. Totalt rättades ca 750 000 blanketter, vilket är ca 15 % av det totala antalet blanketter. Ca 2/3 av felen var reject. Efter terminalupprättningen samkördes personbeståndet med arbetsställeregistret via löpnumret och koderna för näringsgren, sektortillhörighet och arbetsplatsens belägenhet påfördes personposterna.

Övriga rutor i kodningsdelen (se fig. 3) användes för mera komplicerade fall t. ex. personer med växlande arbetsplats och utpendlare till andra kommuner.

Löpnummerkodning är ett effektivt sätt att koda men kräver tre saker:

- att rätt löpnummer kodas
- att arbetsställeregistret har god täckning och att koderna i registret är korrekta

[1] Substitut vid optisk läsning innebär att maskinen tolkar en siffra fel, t. ex. kan en slarvigt skriven etta läsas som en sjua.
[2] Reject innebär att den optiska läsaren inte kan läsa ett tecken.

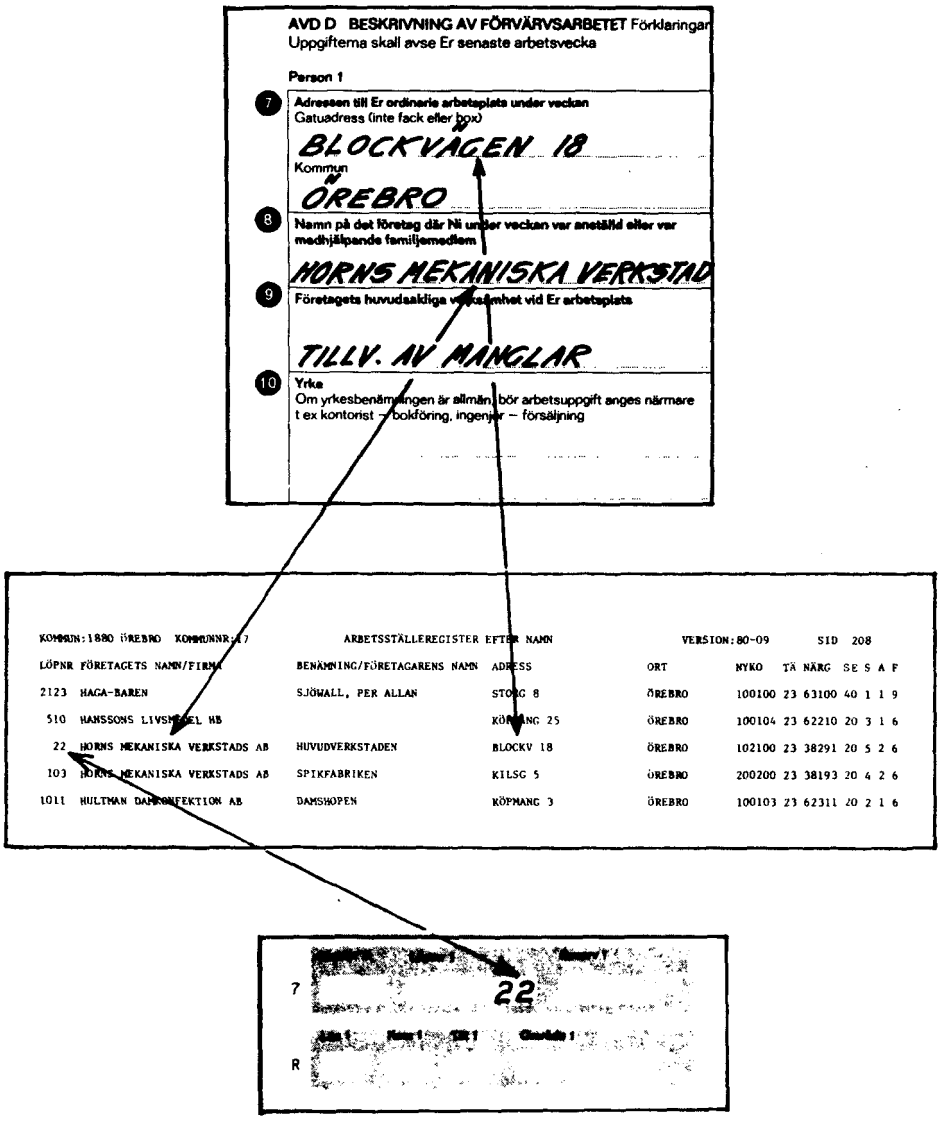


Fig. 3.

– att de "optiska" siffrorna skrivs enligt gällande regler (om man använder sig av optisk läsning).

Att utbilda GO i 279 kommuner att koda korrekt är mycket svårt, speciellt när det gäller mer komplicerade fall som personer med växlande arbetsplats, utpendlare m. m. Checksiffran i löpnumret är nödvändig för att upptäcka åtminstone en del av felen.

Att få arbetsställeregistret (ca 550 000

poster) någorlunda korrekt kräver också stora resurser. Inför FoB 80 gjordes stora insatser för att förbättra arbetsställeregistret. Bl. a. kompletterades registret med arbetsställena inom offentlig sektor, vilka tidigare inte funnits i registret. Därutöver förbättrades registret avseende bl. a. adresser, koder för arbetsplatsens belägenhet m. m.

Det tredje kravet angående optiska siffror fungerade bra. Rejectfrekvensen låg mycket nära den planerade (10 %).

4. Konstruktion av löpnummer

Löpnummerkodningen i FoB 80 skulle uppfylla flera, delvis motstridiga krav:

- GOs arbete skulle minimeras, givet visst informationsinnehåll och kvalitet i kodningen. Detta innebär bl. a. att antalet siffror som behövs för löpnummerkodningen bör minimeras.
  Antalet siffror påverkar givetvis förutom GOs arbetsbörda även rejectfrekvensen vid den optiska inläsningen.
- Det måste vara enkelt att löpnummerkoda. Kodningsevalveringen vid FoB 75 (se Andersson (1977 b)) visade att GO inte klarar ett mera "invecklat" system.
- Löpnummerkodningen borde utformas så att de flesta fel GO gör och fel som uppstår vid den optiska läsningen skall upptäckas och rättas. Helst bör åtminstone en del av felen gå att rätta maskinellt.

Den lösning vi så småningom fastnade för innebär att arbetsställena löpnumreras efter storlek (det största arbetsstället fick löpnummer 1, det näst största löpnummer 2 etc.) inom kommun. Dessutom numrerades kommunerna från 1 till 279 efter antalet inpendlare i FoB 75. Vi gjorde alltså en "nyckel".

"Kommunnummer"	*Län-Kommun*
1	0180
2	1480
.	
.	
.	
.	
279	2505

GO behövde *inte* koda kommunnummer om arbetsplatsen var belägen i hemkommun. "Blankt" i kommunnummer betydde alltså arbetsplatskommun = bostadskommun. Kommunnummer i stället för kommunkod sparar åtskilliga siffror vid kodningen. Dessutom beräknades en checksiffra (modul 10 metoden) för varje löpnummer (inkl. kommunnummer).

Checksiffran användes för att:

- upptäcka vissa kodningsfel, t. ex. att GO glömt koda kommunnummer för utpendlare
- upptäcka av GO felskrivna löpnummer samt substitut vid den optiska inläsningen
- maskinellt rätta reject på en siffra i löpnumret
- upptäcka fel vid annan dataregistrering av löpnumren (t. ex. vid stansning av tillägg).

Våra erfarenheter från den här konstruktionen av löpnummer är att den fungerade tillfredsställande.

5. Arbetsställeregistren

Sen tidigare FoBar hade vi ganska dåliga erfarenheter av centrala företagsregistret (CFR). Bl. a. var täckningen dålig och det var svårt att hitta det som trots allt fanns i registret (se Andersson (1977 a)).

Löpnummerkodning enligt FoB 80-modell ställer ännu större krav på arbetsställeregistret än konventionell kodning och därför inleddes ett samarbete mellan FoB

och CFR i januari 1979 i syfte att förbättra CFR och därmed arbetsställeregistret.

Under 1979 och fram till september 1980 lades stora resurser ned på att förbättra CFR. De viktigaste insatserna var:

- CFR kompletterades med den offentliga sektorns arbetsställen.
- Arbetsställena kompletterades med s. k. geokoder, dvs. koder för arbetsställets belägenhet.
- Täckningen i CFR har förbättrats avsevärt. Totalt har i samband med FoB 80 dataregistrerats 70000 tillägg, varav ca 40000 tilldelats organisationsnummer och lagts in i CFR.
- CFR har kompletterats med firmanamn och verksamhetsadress (i stället för hemadress) för personliga företag.
- Även om ett arbetsställeregister innehåller samtliga arbetsställen är detta till föga nytta om det är svårt att hitta i registret. Därför lades ett stort arbete ned på redigering, sortering och lay-out i registret.

I september 1980 fick varje GO s. k. kodningsregister, dels för egen kommun, dels för näraliggande kommuner dit man – enligt FoB 75 – hade många utpendlare. Kodningsregistren erhölls i tre sorteringsordningar:

- efter namn
- efter adress
- efter benämning.

Sorteringsfältet skrevs i annan stilsort än övrig information. Antalet exemplar av eventuella register till varje kommun varierade mellan 2 och 100, beroende på antalet granskare i resp. GO. Totalt printades på laserskrivaren ca 1 milj. sidor.

Enligt GO-instruktionen skulle GO vid löpnummerkodningen kontrollera att koderna för näringsgren, sektor, arbetsplatsområde samt arbetsplatstätort var korrekta för arbetsställen som utnyttjades vid kodningen. Eventuella fel skulle rättas i en s. k. registerstomme. Efter avslutat GO-arbete skickades registerstommen in till SCB. Totalt registrerades ca 150000 rättelser, de flesta rörde de s. k. geokoderna.

Tilläggen och rättelserna kördes ihop med ursprungsregistret och bildade arbetsställeregistret i FoB 80.

Kodningsregister, registerstomme och tilläggsblankett finns avbildade i Andersson (1982).

Vid samkörningen mellan arbetsställeregistret och personbestånd fick vi nästan inga omaka personposter, däremot ganska många omaka arbetsställen, vilket innebär att det i arbetsställeregistret förekommer ganska många arbetsställen där det enligt FoB inte finns någon förvärvsarbetande.

6. Kvaliteten på löpnummerkodningen

Någon evalvering av löpnummerkodningen har inte gjorts. Följande synpunkter är grundade på våra erfarenheter av materialet.

När det gäller den stora gruppen förvärvsarbetande, de som har en fast arbetsplats i hemkommunen har löpnummerkodningen fungerat bra. Problemen har uppstått för följande kategorier:

- utpendlare till annan kommun
- personer med växlande arbetsplats
- personer med fast arbetsplats som är skild från "huvudarbetsstället", t. ex. städpersonal, personer med hemarbete, skogs- och byggnadsarbetare.

Vidare har felkodningar ofta förekommit då "samma" företag har mer än ett arbetsställe på samma adress.

En variant av löpnummerkodning är att låta GO endast koda de fall där företagets namn och adress "exakt" överensstämmer mellan blankett och arbetsställeregister.

Övrig kodning tar SCB hand om. Ett sådant system skulle innebära att GO kodar ca 85 % av de förvärvsarbetande.

I sammanhanget bör påpekas att ett kodningssystem som är betydligt billigare än konventionell kodning måste ha rätt till en måttlig kvalitetsförsämring. Decentraliserad kodning medför utan tvivel risk för att kodningen blir av olika kvalitet i olika kommuner. Detta måste i görligaste mån rättas till vid den centrala bearbetningen.

7. Referenser

Andersson, R. (1977a): Utnyttjande av CFR vid folk- och bostadsräkningar. Stencil SCB.

Andersson, R. (1977b): Kodningskvaliten i FoB 75 – ekonomisk aktivitet. Stencil SCB.

Andersson, R. (1982): Kodning av arbetsställe vid folk- och bostadsräkningen 1980 (FoB 80). Metodinformation nr 82:2, Statistiska metoder, SCB,

Ekwall, T. och Andersson, R. (1975): Kodning via bildskärm i folk- och bostadsräkningen 1975. Statistisk tidskrift 1975:5, sid. 376–386.

Enheten för befolkningsstatistik (1980): Instruktion för granskningsledare.

Enheten för befolkningsstatistik (1980): Handbok för de lokala granskningsorganen.

Felme, S. (1980): Kodkonstruktion. Statistisk tidskrift 1980:6, sid. 472–480.

Lyberg, L. (1981): Control of the coding operation in statistical investigations – Some contributions. Urval nr 13, SCB.

Vad är statistisk metod?[1]

av fil. dr Olle Sjöström[2]

0. Inledning och sammanfattning

Metod har som sin ursprungliga betydelse "sätt att välja väg" eller "sätt att ta sig fram" (efter grekiskans meta, efter, och hodos, väg). I vårt vardagsspråk har det en liknande innebörd men kanske med tonvikt på genomtänkta och systematiskt utformade vägar för att uppnå ett mål, t. ex. att åstadkomma ett fysiskt arbetsresultat, att förmedla kunskaper eller att lösa ett problem.

Metod är således i ett visst sammanhang förknippad med *mål*. Funderingar över metod är förknippade med funderingar över målets art eller natur. Här berör vi också en grundbetydelse i ordet *teori* (av grek. theorein, begrundan) som t. ex. kan innebära djupare begrundan och utveckling mot en "lära" om både mål och metod.

1600-talet har av idéhistoriker och filosofer kallats "metodens århundrade". Det såg födelsen eller pånyttfödelsen av ett antal olika synsätt på och teorier om målet *att utveckla och uppnå kunskap*, liksom olika föreställningar om kunskapens natur. Kunskapsteori och vetenskapsteori fick en uppblomstring genom bl. a. filosoferna Francis Bacon (empirism) och René Descartes (rationalism), i sammanhang med genombrott för empirisk forskning inom såväl naturvetenskap som samhällsvetenskap. Det är i sammanhang med empirism och en spirande samhällsvetenskap som vi får placera samhällsstatistikens två lärofäder: filosofen, läkaren, juristen och historikern *Hermann Conring* och köpmannen och amatörforskaren *John Graunt*.

Conring startade (1660) en samhällsvetenskap som i mitten av 1700-talet fick namnet statistik. Graunts inflytelserika arbete (1662), grundat på sammanställningar och kritisk analys av dödslängderna i London, är en tillämpning av empirisk metodik i Bacons anda.

Om detta är ett historiskt och filosofiskt perspektiv på metod finns också ett "reduktionistiskt" synsätt på metod av senare datum. Metod kan komma att betyda redskapen, verktygen eller tekniken som utnyttjas, i stället för en helhetssyn på strategin för eller "gestaltningen" av dessa medel för att komma fram till ett mål. Ett vardagligt exempel kan illustrera risken för glidning av innebörd av metodbegreppet.

Om vi frågar en trädgårdsodlare vilken metod han tillämpar för att hålla rabatterna rena från ogräs kan han svara med en utläggning om arbetsuppläggning, bruk av verktyg, utnyttjande av personal etc. för att nå ett gott resultat. Men han kan också svara: "med hacka och spade och besprutning med XYZ", och utgår kanske ifrån att beskrivningen av verktyg och medel är en implicit beskrivning av metod.

Denna "instrumentalisering" av metodbegreppet kan återfinnas inom vetenskaper

[1] Ett diskussionsinlägg om behovet av vidgade perspektiv på samhällsstatistisk forskning, undervisning och tillämpning.

[2] Statistiska institutionen, Stockholms universitet.

som kallas metodvetenskaper och vi finner den i dag i statistiken. Det kunskapsteoretiska perspektivet är långt borta i våra läroböcker och vår undervisning och tycks inte ha någon plats i statistisk forskning. Därför saknas också detta perspektiv inom viktiga statistiska tillämpningsområden. En "metodstatistiker" är exempelvis i dagens språkbruk den som behärskar verktyg och tekniker – inte den som har förmåga att hantera undersökningssituationen i stort, från problemformulering till analys, tolkning och redovisning.

De olika perspektiv på statistisk metod som diskuteras i denna uppsats är:

a. metod i huvudsak som ett antal tekniker eller verktyg

Bl. a. vad som erbjuds i ett statistiskt programpaket, en samling tekniker i en lärobok (t. ex. Cochrans Sampling Techniques). I detta perspektiv ses tillämpningen som "praktisk statistik" eller "know how".

b. metod som sätt att "tänka statistiskt" och sätt/konsten att "bruka verktygen"

Detta perspektiv har en lång historisk tradition, som innebär en betydelseskillnad mellan metod och tekniker/verktyg.

c. ett kunskapsteoretiskt perspektiv på statistisk metod

Detta innebär – kort uttryckt – ett försök att mer grundläggande karakterisera perspektiv b. Det ger anknytningar till metoddebatt i våra grannvetenskaper och till dagens tvärvetenskapliga strömningar (t. ex. Sören Halldéns nu avslutade projekt "Kritisk skolning i högskoleutbildningen")[1] förutom till det ovannämnda historiska perspektivet.

Ytterligare en uppfattning kommer att nämnas:

d. metod som den matematik som tillämpas vid utvecklingen av verktyg

Här talas ibland om de dubbla identiteterna metod=matematik=teori. Detta är i huvudsak fråga om ett *matematiskt* perspektiv.

Denna diskussionsinledning pläderar för en ändring av perspektiv, från dagens tyngdpunkt i matematik och statistiska tekniker mot en mer medveten och kunskapsteoretiskt förankrad metodsyn. Jag tror att det är nödvändigt för att statistik, speciellt som samhällsvetenskap, skall kunna utveckla kontakter med andra vetenskaper och skapa fruktbara kontaktytor med olika slag av tillämpningar.

Det är också en nödvändighet att utveckla en teori om mål och metod för *utbildningen* i statistik – på olika nivåer och i samspel med andra samhällsvetenskaper – för att svara mot aktuella krav på ämnet.

Några genomgripande och snabba förändringar av perspektiv är inte att vänta. Jag försöker därför antyda några "evolutionära" utvecklingsvägar på något längre sikt och hoppas framför allt på att det blir en debatt om teori- och metodfrågor i statistik.

1. Ett inledande exempel om "mått"

Med vissa undantag vill jag karakterisera dagens statistiska läroböcker – svenska och engelskspråkiga – som spridare av perspektiv a) eller d). Statistisk metod betyder i huvudsak mängden av ett antal verktyg eller instrument. Detta meddelas inte tydligt utan ligger implicit i framställningen. Ser vi vidare till ett *kursutbud* av flera läromedel tycks verkan av detta utbud vara att metod betyder samlingen av de algebraiskt (matematiskt) något mer sofistikerade teknikerna. De algebraiskt enkla beskrivande måtten liksom olika redovisningstekniker (tabeller, diagram etc.) karakteriseras som

[1] UHÄ 1981: 34, Projekt 33.

"praktisk statistik" eller "beskrivande statistik".

Detta är – historiskt sett – en trivialisering av metodbegreppet i statistik. Den har skett i rask fart under de senaste ca 20–30 åren. Jag tror att det är en viktig orsak till att vårt ämne hamnat i det "bakvatten" som vi nu borde försöka ta oss ur.

Som exempel på undantag från dessa slag av läromedel vill jag nämna Wallis och Roberts lärobok [17]. Det är karakteristiskt att den svenska översättningen av de första delarna av denna bok ofta blev *kursivlitteratur* för att till slut försvinna. Ett annat undantag är Wallgren [16] som är ett intressant nyskapande försök på ekonomlinjen i Linköping. I samma anda är Göran Anderssons försök att skapa nytt liv i gymnasieundervisningen [1] och annan grundutbildning.

Som första exempel på skillnad mellan perspektiv a) och perspektiv b), där det senare innebär tänkandet och sättet att bruka verktygen, vill jag ta två alternativa utformningar av ett tentamenstal på en grundkurs i statistik, som bl.a. gäller "beskrivande mått".

Med *perspektiv a)* kan uppgiften se ut på följande sätt: "Nämn och definiera olika centralmått och spridningsmått". Svaret är en namnuppräkning och motsvarande definitioner.

Med *perspektiv b)* kan uppgiften formuleras på följande sätt: "Ge olika konkreta exempel på hur en *meningsfull* beskrivning kan ske genom lägesmått och spridningsmått. Minst två olika läges- resp. spridningsmått skall illustreras, vardera med ett eller flera exempel. Illustrera med både material och problemställning!" Den senare frågan förutsätter en undervisning som tar upp det meningsfulla sambandet mellan *problemställning* och *val* av beskrivande mått. Även detta kan givetvis besvaras med "katekessvar" (dvs. mer mekaniska kopior av inlärda exempel) men möjliggör också kreativa svar och därmed mer nyanserad poängsättning.

Detta senare exempel antyder karaktären av en bakomliggande metodundervisning enligt perspektiv b. Det visar dessutom på *en nödvändig svårighet*, nämligen att lösningar av uppgiften inte är så lätträttade som i det förra fallet. Denna obekvämhet med perspektiv b) torde ha bidragit till svårigheterna att hävda detta i statistisk grundundervisning, trots att här funnits en lång tradition.

Ett vanligt argument utifrån perspektiv a) är att det ändå är viktigast att lära in begreppen och teknikerna och att lära sig räkna. Det andra, tänkandet, kommer så småningom "på köpet". Det finns risker att "mjukt stoff" förstör undervisningen, hävdas ibland.

Jag har själv aldrig förstått dessa argument. Om inte tekniker *inlärs och upplevs* i ett sakproblemorienterat sammanhang blir de mycket svåranvändbara utantilläxor. Mina erfarenheter av olika utbildningskategorier visar att det är mycket svårt att i efterhand (i tillämpat statistiskt arbete) lära sig metod i meningen tänkande och applicering av statistiska verktyg, hur mycket statistisk teknik man än har lärt sig i en grundutbildning. En ensidig utbildning i olika slag av statistiska tekniker tycks också ofta ha samband med ointresse eller oförmåga att identifiera sakproblem i tillämpningen, vilket gör utbildningen föga användbar. En referens av intresse är här Biggs [2].

2. Tabeller

Utan tvivel hör presentation av hur statistiska tabeller kan utformas och läsas till det mest triviala – och ofta "träaktiga" – i

statistiska läroböcker. Tabellredovisning tycks också i dag ses som någonting "ute" och har i många sammanhang rent av dåligt rykte, t. ex. som digra bilagor i utredningar eller en bunt råtabeller från en dator.

När Wallis och Roberts [17] tog upp konsten att *läsa* en tabell blev detta avsnitt med rätta beundrat och ofta citerat. I anslutning till ett konkret exempel visar författarna hur väsentligheterna i en tabell kan sökas fram på ett metodiskt sätt. Men detta förutsätter också att tabellredovisningar sker på ett genomtänkt sätt.

En lärobok som tar tabellredovisningens metodik mer på allvar är Wallgren [16]. Med utgångspunkt i en frågeställning om sambandet kön–inkomst (däribland möjliga orsakssamband) visar författarna mycket instruktivt hur först en inledande beskrivning kan göras och sedan en successiv analys, där de olika stegen stöds av tabellredovisningar med kommentarer. Termer och olika tabellutformningar införs på detta sätt i ett problem- eller frågeställningsorienterat sammanhang, som skapar intresse och som innebär metod i den mening som perspektiv b) avser.

Här öppnar sig ett intressant historiskt perspektiv mot den syn på (teori om) redovisningens form och innehåll, som speciellt Schlözer utvecklar i sin lilla skrift 1804 [12]. Statistisk redovisning skulle styras av *ändamål* (Zweck, Telos) och utformas i form av ett *system*, där bl. a. tabellredovisning och verbala kommentarer ägnades speciell uppmärksamhet. "Tabellknektar" var på den tiden ett skällsord om de som alltför lättvindigt åstadkom tabellredovisningar.

3. Syner på den inledande planeringsfasen i en statistisk undersökning

En annan illustration på de två perspektiven, a resp. b, vill jag hämta från två välkända läroboksförfattare och välfrejdade tillämpade statistiker. Detta gäller förhållandet mellan sakproblemet (-en) och utformningen av en statistisk plan.

I Cochrans välkända lärobok Sampling Techniques (obs. titeln) är *allt* som sägs om detta följande mening:

"A clear statement of the problem is most helpful". [3] sid. 5.

Det bibringar föreställningen att statistikerna lugnt skall avvakta att "klienten" skall kämpa sig fram på egen hand till denna klara och för statistisk planering färdiga formulering av sitt problem.

W. E. Deming har i en uppsats [4] formulerat de metodiska problemen (enligt perspektiv b) på följande sätt (min översättning):

"... Statistikern måste arbeta med experter inom ämnesområdet och måste lära sig tillräckligt översiktligt om ämnesområdet för att arbeta med det och ge hjälp. Utbytet av idéer mellan statistiker och experter inom ämnesområdet kan gälla några veckor eller många månader, innan formuleringen av problemet i statistiska termer är någorlunda nära slutmålet. Detta grundläggande arbete är vanligen den svåraste delen av statistikerns medverkan."

Cochrans lärobok har en adekvat titel sett ur perspektiv b) och han uttalar sig *inte* om att ha skrivit om metod. Min mångåriga erfarenhet av hans bok är att den vållat mig en hel del problem i undervisningssammanhang, där den fordrar komplement i form av tillämpade metodövningar (laborativa uppgifter) och tydliga vägvisare i läsanvisningar m. m. I tillämpningsarbete har den varit utmärkt som uppslagsbok och inledande orientering om olika tekniker (med referenser).

En undervisning grundad på Cochrans lärobok (och andra liknande med uppläggning som "tekniksamlingar") utan komplement av metodisk karaktär kan leda till en olyck-

lig fixering vid tekniker och bibringar lätt en uppfattning om metod enligt perspektiv a. Detta händer alltför lätt om läraren saknar tillämpningserfarenhet.

Deming är en i detta avseende mer medveten författare. Hans mer kunskapsteoretiskt orienterade läroböcker och artiklar har en metodisk helhetssyn där sakproblemen och dess lösning ställs i fokus. Detta styr såväl hans pedagogiska framställning av planeringens problematik som av genomförandets och redovisningens. Se [4, 5].

4. Ett historiskt perspektiv

Pånyttfödelsen av intresset för grundläggande metodfrågor på 1600-talet, de olika framstegen inom naturvetenskaperna och intresset för samhällsfrågor, samhällsinflytande och statsmakt är en viktig bakgrund för att förstå statistikens födelse som mer medveten tillämpning och som *vetenskap om* denna tillämpning. Det är fråga om en gemensam grogrund med andra samhällsvetenskaper som statsvetenskap, juridik och nationalekonomi och – sedermera – geografi och sociologi.

Statskunskapstraditionen (systemtraditionen)

I uppslagsböckerna anges som en av statistikens två "stamfäder" den tyske läkaren, juristen och historikern Hermann Conring, som "på gamla da'r" skapade en helt ny vetenskap, NOTITIA RERUMPUBLICARUM, som kom att kallas "Staatenkunde" – eller på svenska – stats- eller samhällskunskapstraditionen.

Det var inte en metodvetenskap, utan en vetenskap om staten/samhället i *nuet* med syftet att följa utvecklingen av vad som på den tiden kallades "statsförnuftet", så som det realiserades på olika sätt i den tidens europeiska stater. Det skedde mot en bakgrund av en överblickande samhällsbeskrivning under rubrikerna:

causa materialis	samhällets resurser
causa finalis	statens/samhällets mål
causa formalis	förvaltning (medel)

Hur "statsförnuftet" *realiserades* kallades causa efficiens. (Vi känner igen nyare sådana slag av disponeringar i termer av systemanalys.)

"Obunden av teologiska förutsättningar skulle iakttagelse och kausal förklaring, som Francis Bacon hade ställt som krav på forskning, vara de grundläggande metoderna för den nya vetenskapen", säger en nutida biograf, Erich Wolf, i boken "Grosse Rechtsdenker" [20].

Politisk aritmetik (verktygstraditionen)

Den andra lärofadern är köpmannen John Graunt i London, som tidigt gjorde sig välbärgad och oberoende och ägnade sin tid åt empiriska studier, speciellt åt de varje vecka publicerade dödslängderna i London.

Efter några års arbete publicerade han sin berömda uppsats Natural and Political Observations upon the Bills of Mortality [7], som gav honom omedelbar berömmelse både i England och utomlands. Det var ett beundransvärt arbete både med avseende på källkritiskt noggrannhet och uppfinningsrikt modellbyggande. Och det ledde till uppmärksammade resultat om bl. a. en mängd demografiska förhållanden: Londons folkmängd, könsfördelning, åldersfördelning, m. m. Graunt blev omedelbart invald i Royal Society som nyligen stiftats med Francis Bacon (återigen) som "andlig fader".

Graunt startade inte en vetenskap, men hans arbete inspirerade till fortsatta uppfinningar av "fiffigheter" för att uppskatta nu-

meriska storheter relaterade till befolkning, handel, jordbruk och andra näringar. Det inspirerade också starten av *räkningar*, inte bara folkräkningar, och bidrog till en första uppblomstring av vad som kom att kallas "officiell statistik", dvs. av statsmakten framtagen och offentliggjord statistik – liksom också av hemlig eller "sekret" statistik.

Detta är också den intellektuella jordmånen för det stora svenska företaget, nämligen Tabellverket, som kom 1748, och som på den tiden var en unik totalstatistik över ett lands befolkning. Se den i dagarna utkomna biografin (av Erland Hofsten) om Wargentin och hans brev [18].

Dessa slag av "fiffigheter" kom att kallas "politisk aritmetik", inte för att den var politisk förarglig – snarare tvärtom – den var ett redskap för att statsmakten skulle få bättre grepp eller kontroll över sina resurser i krig och fred.

Däremot kom Conrings skapelse att utveckla sig i politisk riktning, i upplysningstidens anda, genom att de som följde Conring på ett antal olika tyskspråkiga lärostolar alltmer formulerade ett kunskapsspridande mål eller ett kunskapsfrigörande mål. Achenwall, som kom att kalla sitt ämne statistik, formulerar sig så här:

"Vi ägnar oss åt att från de omätliga massor av data, som man påträffar i samhället, noggrant välja ut sådana som visar både brister och fördelar hos ett samhälle, och vi skall noga utforska deras orsaker. Det betyder ... det som gör medborgaren fattig eller rik, nöjd eller missnöjd, vad som gör regeringen älskad eller hatad etc. ... kort sagt, vad som leder till grundlig insikt i en stats politiska författning; den som eftersträvar detta förtjänar att kallas statistiker."

Den siste "store" i denna tradition var August v. Schlözer, som förtydligade denna linje i en på den tiden inflytelserik liten bok med namnet Theori der Statistik [12], som var en teori över ämnets mål, och därav härledda metoder.

Schlözer talar f. ö. om behovet av system i forskning och pedagogik och vad gäller principerna för hur statistikern *väljer* sina data eller fakta. En uppradning av fakta eller tabeller är *inte* statistik, betonar han.

Schlözer kämpade också starkt för att den till stora delar hemliga statliga statistiken skulle bli offentlig och lär ha haft en avsevärd framgång. Han har kallats "den förste oppositionelle statistikern". Den första svenska statistiska årsboken av Carl af Forsell [6] är ett belysande uttryck för Schlözers statistiska program.

1800-talet gick in under en strid mellan företrädare för Conring-traditionen och vad de ibland elakt kallade "tabellknektarna". Därmed menade de de allt vanligare förekommande kvantitativa sammanställningar av data och den från politik och samhällsideologi avskilt arbetande statistikern, som nu också började få akademisk hemvist på olika lärostolar. Utvecklingen av nationalekonomin och geografin som universitetsämnen gjorde också att den gamla traditionen började förlora terräng.

Senare historiker har menat att dessa två traditioner inom statistiken egentligen inte var motsatta varandra utan *komplementära:* Conring-traditionen stod för djupet i medvetenhet – om man så vill ett systemmedvetande – den aritmetiska traditionen stod för uppfinningarna av nya verktyg, som gjorde det möjligt att allt bättre förverkliga en allsidigare välfärdsbeskrivning – som vi i dag kallar det.

Oberoende av dessa akademiska strider blev första hälften av 1800-talet en period som har kallats "entusiasmens era" (Westergaard), statistik blev t. o. m. en modeve-

tenskap för en tid. En stor mängd privata statistiska föreningar bildades, som bl.a. bekostade egna undersökningar. Statistiken gjorde intåg i medicin, biologi och andra tillämpningsområden – och den s.k. officiella statistiken fick ett uppsving, som skulle kulminera i andra hälften av 1800-talet. Här bidrog belgaren Adolph Quetelet som förde in statistiken även i den filosofiska och kulturella debatten. Se Westergaard [19].

Detta är något av bakgrunden till den vitala diskussionen kring mitten av 1800-talet om statistiken som vetenskap: är den en *ämnesvetenskap* med samhället som föremål eller är den en *metodvetenskap* som kan ha vilket ämne som helst till föremål? Frågan kunde besvaras dels principiellt, dels med hänsyn till de viktigaste praktiska arbetsfält och den akademiska tradition som förelåg.

Den senare delen av 1800-talet såg en utveckling och konsolidering av den officiella statistiken med de internationella statistiska kongresserna som viktigt instrument liksom tillämpningen av statistik inom både samhällsvetenskaper och naturvetenskaper. Först nu började sannolikhetsteori och -kalkyl få fotfäste i tillämpningen och började ses som en integrerad del av statistiken som vetenskap.

Vad menades då med statistisk metod? En intressant men till synes bortglömd kunskapsteoretisk syn finns redovisad i två böcker från denna tid, nämligen filosofen Christoph Sigwarts verk Logik (1878, 1893) och professorn i statistik, August Meitzens lärobok Geschichte, Theori und Technik der Statistik (1886, 1903).[1] Man kan uppfatta dem som försök att integrera de olika historiska utvecklingslinjerna, inklusive de nyare naturvetenskapliga tillämpningarna – en integrering av vad som skulle kunna betecknas som systemtraditionen (Staatenkunde) resp. verktygstraditionen (politisk aritmetik och dess efterföljare).

Sigwart svarar för vad jag ovan kallade den principiella synen på statistik som en generell metodvetenskap och Meitzen knyter an med hänsyn till den tidens viktigaste arbetsfält och akademiska tradition. Jag försöker här ge en antydan till deras kunskapsteoretiskt grundade metodsyn och hänvisar den intresserade läsaren till källorna.

Finns det verkligen något för oss intressant att hämta när vår tids "flaggskepp" i form av vad vi kallar formell inferens, beslutsteori, bayesiansk inferens etc. inte var utvecklade? Mitt svar är att vi så till den grad absorberats av det matematiskt intrikata och vissa mycket speciella sofistikerade tillämpningsaspekter att vi behöver en *återupptäckt* av de mer fundamentala grundvalar som vårt ämne bygger på och som bör få praktiska konsekvenser av hur vi angriper dagsaktuella samhällsstatistiska problem och hur vi utformar vår utbildning. Därmed kan vi också, tror jag, få möjligheter att skapa en kunskapsteoretisk brygga till andra (samhälls)vetenskaper och den debatt om metodfrågor som där blivit allt livaktigare de senaste åren. Ett litet försök till att skapa nyfikenhet för detta är följande korta referat.

5. Statistisk teori om mål och metod – Sigwart och Meitzen

5.1. Tidig inferensteori

Innan jag illustrerar det sagda med referat och citat vill jag också kortfattat nämna (fast det är värt en egen artikel) att Sigwart beskriver grundläggande teori bakom vad vi

[1] Båda verken finns i engelsk översättning. Se referenser [10 b resp. 13 b].

i dag kallar stokastiska modeller, med exempel från mätfelsteori och från demografi. Han diskuterar vidare kriterier för punktestimat (speciellt "maximum likelihood") och härleder även fördelningsfria konfidensintervall, fast givetvis med annan terminologi än dagens.

Han karakteriserar byggandet av stokastiska modeller som ett *hypotetiskt deduktivt förfarande*, där teorin ligger *bakom* t.ex. hypotesen om "lika sannolikhet" för olika utfall. En teori om massans fördelning i en tärning m.m. kan exempelvis leda oss till den subjektiva hypotesen att alla sidor i en tärning har samma sannolikhet.

Den deduktiva delen är matematisk och ger statistikern en tillgång till en viktig *teknik*. Vad gäller *statistisk* metod betyder detta att statistikern måste vara medveten om teknikens *förutsättningar* och således söka göra den hypotetiska delen *medveten*. Detta ställer, som det lilla exemplet ovan visar, i de flesta fall krav på *ämnesteori*. Statistisk metod innebär här att (av andra) mana fram och göra explicit sådan teori – eller eventuellt våga sig på att själv utveckla den.

5.2. Surveyteori eller en teori om den metodiska räkningen

Den egentliga *statistiska* teorin handlar enligt Schlözer och Sigwart om kunskapsbildning (som mål) och de olika vägar som befrämjar detta (metod) i ett helhetsperspektiv. Ett allmänt "effektivitetskriterium" formulerar Meitzen bl.a. på följande sätt:

"Det teoretiska perspektivet inriktar sig nödvändigtvis på det ändamålsenliga och utförbarheten i en undersökningsplan, på sättet att genomföra undersökningen, på säkerheten i och användningsmöjligheterna av resultaten och på de olika slutsatser som dras och den kritiska värderingen av dessa".

Var för sig är dessa moment metodfrågor, säger han, men teorin skall ge stöd för ett helhetsperspektiv, både för en enskild undersökning och – på ett vidare plan – för den statistiska tillämpningen *i ett samhällsperspektiv*. Teorin skall "dels utveckla grundtankarna i metoden till en lärobyggnad, dels måste teorin vara vägledande hur den statistiska tekniken skall utvecklas för att tillfredsställa de krav som en teoretiskt utvecklad kritik ställer".

Jag har kallat detta med det prosaiska namnet "effektivitetskriterium", eftersom denna term intar en så viktig plats i vad som i dag ofta kallas teori, men då med en mycket begränsad ekonomisk innebörd eller som t.ex. minsta varians-kriterium.

Meitzen behandlar under rubriken "Krav på utnyttjad teknik i undersökningar" bl.a. de företagsekonomiska kostnadsaspekterna, sätt att utforma anvisningar, organisationsfrågor inför och under genomförandet, olika skattnings- och redovisningssätt. Mycket av vad vi i dag, ofta utan större eftertanke, kallar metod betecknar således Meitzen som teknik.

Termen metod reserveras för det grundläggande *sättet att tänka* bakom olika tekniker. Under huvudrubriken "Grundprinciper för den metodiska räkningen" har han t.ex. följande rubrik och underrubriker:

Bestämning av begrepp

med bl.a. underrubrikerna:

- empiriska villkor för räkningen
- skillnaden mellan "att räkna" och en statistisk räkning
- villkor för mätbarhet
- räkningsenheten som begrepp.

Här diskuteras bl.a. med en grundlighet som jag inte tidigare träffat på vad vi kallar "population" och "individ" m.m.

Under huvudrubriken "Grundprinciper för den metodiska tolkningen" behandlar

Meitzen bl. a. hur *målinriktningen* (Zweck) *styr* valet av beräkningsförfaranden och presentation, vilka problemen är vid "bedömning av kausalitetsförhållanden", hur vad vi i dag kallar stokastisk modell kan införas, vilka krav dessa ställer (jfr ovan om en stokastisk modell som en hypotes) och vilka slag av slutsatser som kan dras (i "kritisk anda").

I ett följande huvudavsnitt drar han vissa slutsatser och reflexioner om hur metodiken får sina "samhälleliga" eller "kollektiva" uttryck t. ex. i olika typer av statistiska publikationer och material för framtida utnyttjande och vad detta ställer för frågor. (En underrubrik om detta är t. ex. "Innebörden av olika frågeställningar för en systematisk statistik".)

Vad menar då Meitzen med att tekniken skall underställas de krav som "en teoretiskt utvecklad kritik" ställer? Jag uppfattar framställningen så att det dels gäller den historiskt utvecklade källkritiken men också vad som kan läggas i uttrycket "kritisk förmåga". Jfr Halldén [8], kap. 1. Meitzen ställer statistiken som metodvetenskap bredvid de "etablerade" metodvetenskaper, som då benämndes filosofisk kritik, hermeneutik och logik, till vilka statistiken hade flera beröringspunkter.

Utan att hävda en doktrinär inställning företräder Sigwart och Meitzen en uppfattning, att statistiken som metodvetenskap djupast sett är en *tillämpad logik*. I förordet till sin lärobok säger Meitzen:

"Jag anser det viktigaste vara att klarlägga statistikens logiska karaktär och allvar (Ernst) och söka argumentera för detta. Jag tror att det nästan drömlika missbruk som statistiken råkar ut för bara kan bekämpas om de logiska kraven och de logiska gränserna för den statistiska kunskapen sätts i främsta rummet.

... Statistik och logik kan enligt min mening vinna lika mycket, om medvetandet om logikens stränga lagar sätts på levande prov i de omedelbara praktiska uppgifter som statistiken har."

Den fulla innebörden av detta uttalande är svår att helt klargöra i denna korta uppsats. Det är emellertid fråga om både induktiv och deduktiv logik.

Som avslutning vill jag nämna den grundläggande karakteristiken av statistik som en *teleologisk* ("mållogisk") vetenskap, som Sigwart framställer med stor stringens. Som metodvetenskap existerar den inte utan *telos* (Zweck, mål, ändamål). Såväl planering, genomförande, analys och tolkning styrs av telos och detta genomsyrar även Meitzens mer detaljerade framställning av statistisk teori, metod och teknik.

Som ett intressant och tankeväckande exempel på detta vill jag nämna Sigwarts framställning av metodiken i vad vi i dag kallar populations- och variabelavgränsning i planeringsfasen, där "strategien" vad gäller begränsningen av populationen, och hur *urvalet av variabler* sker, diskuteras på ett kunskapsteoretiskt djup, som jag inte har hittat någon annan motsvarighet till. ([13 a], sid. 394 f.) Detta är exempel där såväl deduktiva som induktiva logiska resonemang kommer in.

Detta teleologiska perspektiv gäller varje enstaka undersökning. Ett kompletterande teleologiskt perspektiv gäller det generella målet för statistiken som metodvetenskap, nämligen *målet att utveckla kunskap*. Detta gör t. ex. Meitzens framställning av detaljer i den statistiska *redovisningen*, som metod och teknik, mycket mer seriös än vad vi i dag är vana vid i statistisk litteratur. Även vad vi i dag kanske kallar "praktisk statistik" får på detta sätt en klart underbyggd teoretisk dimension, som inte tillåter trivialisering.

De inledande exemplen i denna uppsats illustrerar också detta teleologiska perspektiv.

6. Slutord

Detta försök till historiskt förankrade perspektiv på vad statistisk metod kan innebära, och vilka konsekvenser olika innebörder kan få, blir i detta diskussionsinlägg nödvändigtvis alltför grovt formulerade. Här krävs fortsatta studier och mer precist formulerade synpunkter. Det pekar också på behovet av en idéhistoriskt och kunskapsteoretiskt orienterad forskning.

Bakgrunden till detta diskussionsförsök är mina erfarenheter av olika slag av tillämpningar och en samtidigt bedriven lärarverksamhet i statistik liksom tidigare forskning *om* tillämpning [14]. Den historiska orienteringen har jag kommit in på under bl. a. studier och diskussion av olika nu utnyttjade läromedel i metodfrågor, som blev aktuella i samband med utformningen av kursinnehåll för systemvetenskapliga linjen (där samarbete mellan matematik, ADB, statistik och andra samhällsvetenskaper erfordrades vid utformningen av första året.

Viktiga kriterier för min tro på betydelsen av en ändring av perspektivet mot mer kunskapsteoretisk medvetenhet är *dels* den betydelse det har för olika slag av tillämpningar, enligt mina erfarenheter, *dels* att det ger förutsättningar för att ge ett mer intellektuellt givande innehåll i statistisk grundundervisning, t. ex. de s. k. deskriptiva delarna. Dessa har i dag i de flesta läromedel blivit starkt trivialiserade och stimulerar knappast till intresse för statistik.

Förutsättningarna för att utveckla grundundervisning och tillämpningar blir emellertid inte så goda om inte detta förankras i en statistisk *forskning* med mer kunskapsteoretisk medvetenhet än den nu ensidigt matematiskt och verktygsinriktade forskning som helt dominerar. Det innebär inget avståndstagande till något annat än ensidigheten. Kish [9] har här talat om låsningen vid vad han kallar de "dubbla identiteterna": teori=matematik=metod, dvs. det som rubricerats som perspektiv d) i inledningen. En annan liknande diskussion finns i Deming [4]. Hur ser sådan forskning ut? Det är en stor och svår fråga som jag antyder några tänkbara svar på.

Vi vet i dag mycket litet om de *faktiska* kunskaps- och informationsprocesser som statistik i form av redovisningar av olika slag (inkl. terminaluttag) och tillämpning av olika tekniker, i större eller mindre sammanhang, ger upphov till. Vi vet också mycket litet om vilka processer som leder fram till statistik och t. ex. statistiska databaser. I dag talas i samband med datorer om "man–machine-interaction" och avsevärd forskning intresserar sig för detta. Det var länge sedan statistisk vetenskap tog "man–statistics-interaction" på allvar.

Här behövs således en empiriskt orienterad samhällsvetenskaplig forskning, som statistik som metodvetenskap (hjälpvetenskap) måste söka främja i samarbete med andra samhällsvetenskaper.

I samband med en sådan forskning kan en kunskapsteoretiskt inriktad grundforskning ske t. ex. i samarbete med filosofi och idéhistoria. Det har framgått av denna artikel, att jag tror att ett studium av den statistiska doktrinhistorien kan ge viktiga idéer och lärdomar.

En sådan ny forskningsinriktning skulle också öppna för en anknytning mellan forskning och grundutbildning, som nu tycks ha gått förlorad.

7. Referenser

[1] Andersson, G. (1982): Lögn och sanning med statistik. Liber 1982.

[2] Biggs, J. B. (1971): Information and Human Learning. Scott, Foresman & Co 1971.

[3] Cochran, W. G. (1977): Sampling Techniques, 3rd ed. Wiley & Sons 1977.

[4] Deming, W. E. (1969): Boundaries of Statistical Inference. I samlingsverket New Developments in Survey Sampling, ed. N. L. Johnson m.fl. 1969.

[5] Deming, W. E. (1960): Sample Design in Business Research. Wiley & Sons 1960.

[6] Forsell, C. af (1978): Statistik öfver Sverige. Nyutgåva av 2:a uppl. 1833 med förord av Erland Hofsten. Bokförlaget Rediviva 1978. (Denna första statistiska årsbok i Sverige är intressant både som en kunskapsöversikt vid denna tid och som ett ambitiöst försök att realisera Staatenkunde-traditionens mål att ge *överblickande* samhällskunskap, genom såväl kvantitativa som kvalitativa redovisningar, med ett "ändamål", Zweck, som ledstjärna. Hur balansgången mellan bildningsmål och "vördnad för Konungen" skedde framgår på ett intressant sätt av bokens inledning.)

[7] Graunt, J. (1959): Natural and Political Observations upon the Bills of Mortality. London 1662. Svensk övers. i Sigma, del 3, sid. 996 f. Forum 1959.

[8] Halldén, S. (1981): Nyfikenhetens redskap. Studentlitteratur 1981. Hofsten, E. Se referens Forsell, C. af och Wargentin, P.

[9] Kish, L. (1978): Chance, Statistics, and Statisticians. JASA vol. 73, sid. 1–6, 1978.

[10] Meitzen, A. (1886, 1903): a. Geschichte, Theorie und Technik der Statistik. 2:a uppl. Stuttgart, Berlin 1903.
b. History, Theory and Technique of Statistics. Övers. till eng. av 1:a uppl. Amer. Academy of Political and Social Science 1891.

[11] Quinton, A. (1980): Francis Bacon. Oxford University Press 1980. I serien Past Masters.

[12] Schlözer, A. L. v. (1804): Theorie der Statistik. (Versuch den Begriff, die wesentlichen Teilen und die Methode der Statistik zu bestimmen.) Göttingen 1804. (Svår att få låna. Jag fann den i Oslo universitetsbibliotek och har en kopia av denna underhållande bok.)

[13] Sigwart, Ch.: a. Logik, band I och II. Tübingen 1878, 1893.
b. Logic I–II. Övers. till eng. Macmillan 1895.

[14] Sjöström, O. (1980): Svensk samhällsstatistik – etik, policy och planering. Akademilitteratur 1980.

[15] Tschrupow, A. (1922): Lehrbücher der Statistik. Bokanmälning av fyra läroböcker. Nordisk statistisk tidskrift, band 1, 1922. (Här diskuteras på ett intressant sätt statistikens och statistikerns relation till matematiken.)

[16] Wallgren, B. och A. (1982): Statistisk metodik. Kompendium, Linköpings universitet 1982. Utkommer under 1983 med titeln: Undersökningar och statistik – informationssökning, källkritik och analys. Natur och Kultur.

[17] Wallis, W. A.–Roberts, H. V. (1956): Statistics, A New Approach. The Free Press, New York 1956.

[18] Wargentin, P. (1983): Minnesskrift med sju originaluppsatser och biografi av Erland Hofsten. SCB, Urval (nr 15) 1983.

[19] Westergaard, H. (1969): Contribution to the History of Statistics. London 1932. Nytryck: Kelly, New York 1969. (Westergaards historia behandlar mycket summariskt frågor om teori- och metoduppfattningar. Bokens styrka ligger i beskrivningen av verktygens/teknikernas utveckling och av den officiella statistikens yttre utveckling.)

[20] Wolf, E. (1951): Grosse Rechtsdenker. (Uppsats om Hermann Conring.) Tübingen 1951.

Deltidssysselsatta i SCBs arbetskraftsundersökningar 1965–1982. Synpunkter och förslag

av fil.kand. Stig Tegle[1]

1. Introduktion

Deltidsarbetets utveckling har analyserats och diskuterats flitigt under senare år (1,2). Underlaget för dessa analyser har i de flesta fall varit SCBs arbetskraftsundersökningar (AKU). I AKU definieras deltidssysselsatt som den som under mätveckan normalt arbetar 34 timmar eller mindre.

I denna artikel skall AKUs definition och användning av deltidssysselsatt analyseras. Detta sker inledningsvis genom att 34-timmarsgränsen sätts i relation till heltidsarbetsveckans utveckling. Därefter redovisas, och analyseras, SCBs olika definitioner av deltidssysselsatt.

I det tredje avsnittet redovisas ett förslag till alternativ definition. Artikeln avslutas med några synpunkter och slutsatser om deltidsmåttets användning utifrån olika synvinklar.

2. 34-timmarsgränsen för deltidsdefinitionen

AKU har sedan 1965 använt sig av gränsen 34 timmar per vecka för att avskilja deltidssysselsatta från heltidssysselsatta. Det innebär att deltidssysselsatta definieras som de som arbetar denna tid *eller* mindre per vecka. Som namnet anger innebär deltidssysselsatt att man är sysselsatt en *del* (timmar) av något. I detta fall är det del av heltidsarbetsveckan eller den s.k. standardarbetsveckan enligt lag eller avtal. Under perioden 1965–1982 har denna sänkts från 45 till 40 timmar per vecka. Diagram 1 på sid. 123 visar utvecklingen av standardarbetsveckan[2], AKUs deltidsdefinition samt en alternativ definition av deltidsavgränsningen.

Som diagrammet indirekt visar ger AKUs definition upphov till snedvridning (bias) i uppskattningen av antalet deltidsarbetande över tiden. Fastlåsningen av deltidsdefinitionen till 34 timmar för hela perioden innebär en *över*skattning av antalet deltidssysselsatta under perioden 1973 och framåt, i relation till den tidigare perioden (1965–1972). Alternativt, vilket är samma sak, en *under*skattning av antalet deltidssysselsatta under perioden 1965–1972, i relation till den senare perioden.

En grov kalkyl av storleken på denna bias kan visas genom att beräkna *andelen* deltidsarbetande för olika definitioner för start- resp. slutåret 1965 och 1981. För gruppen "i arbete" visas denna bias i *tab. 1*, där deltidsandelen anges i procent av alla.

Jämförelsen i tab. 1 skall endast göras *inom* resp. år eftersom den alternativa del-

[1] Nationalekonomiska institutionen, Lunds universitet.

[2] För definition och utveckling av standardarbetsveckan, se (3). Standardarbetsveckan följer väl den *genomsnittliga* veckoarbetstiden (för de i arbete) enligt AKU för gruppen *anställda män*.

Tab. 1. Andelen deltidsarbetande i procent av alla i arbete 1965 och 1981 enligt olika definitioner på deltid

	Deltidsandel %			
	1965		1981	
	AKUs def. (1–34 t/v)	Alt.def (1–39 t/v)	AKUs def. (1–34 t/v)	Alt.def. (1–29 t/v)
Båda könen	17,9	24,2	31,4	23,7
Män	6,8	11,0	14,8	10,4
Kvinnor	37,9	48,0	52,2	40,3

Källa: AKU, Råtabeller, Årsmedeltal; tab. 20/19 (SNI 1–9) Notera att den alternativa definitionen är olika för de två åren (se motivation i texten nedan).

tidsdefinitionen är olika för de två åren. Den alternativa deltidsdefinitionen för 1965, 1–39 timmar per vecka, är gjord med tanke på jämförbarhet med relationen deltidsdefinition/heltidsarbetsvecka idag (1982). Dvs. deltidsdefinitionen som procent av standard- (heltids-) arbetsveckan är ungefär den samma 1965 som idag: 39/45 ≈ 34/40. En jämförelse mellan de två kolumnerna 1965 visar därmed graden av *under*skattning av andelen deltidsarbetande *1965* i jämförelse med *dagens* andel. *Under*skattningen varierar för de tre olika grupperna (samtliga, män resp. kvinnor) mellan 30–60%. Dvs. i AKU underskattar andelen deltidsarbetande under 60-talet i storleksordningen 30–60%.

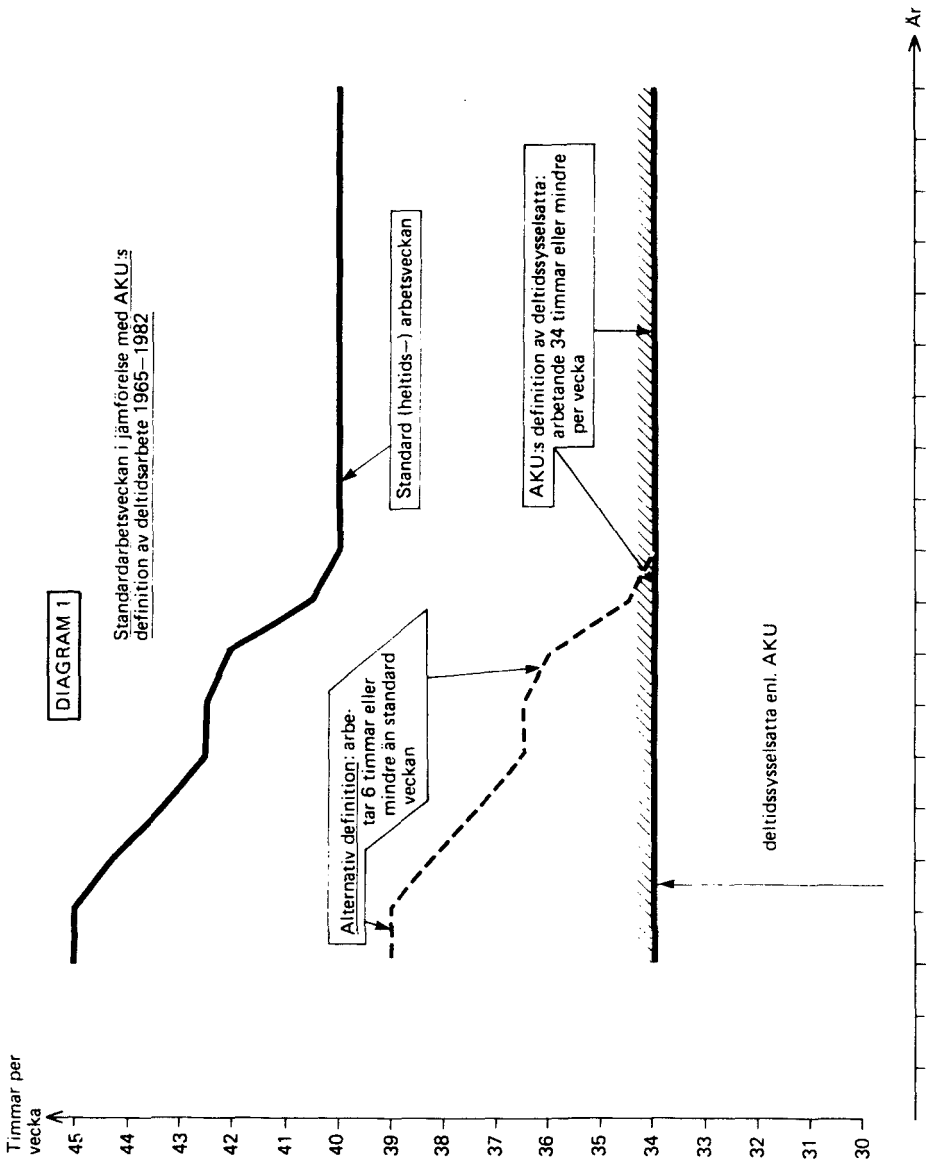
Spegelbilden till detta är jämförelsen mellan kolumnerna för 1981. Här är den alternativa deltidsdefinitionen, 1–29 timmar per vecka, satt så att jämförbarhet gäller mellan dagens deltidsdefinition applicerad på 1965 års standardarbetsvecka: dvs. 34/45 ≈ 29/40. Eftersom tab. 1 visar en högre andel deltidsarbetande enligt AKUs definition för 1981, innebär detta en *över*skattning av andelen idag i relation till andelen under 60-talet (1965). Överskattningen är i storleksordningen 30–40%.[1]

Någon redovisning eller diskussion av denna bias i AKUs skattningar, av såväl *andel* som *antal* deltidsarbetande, har så vitt jag vet inte skett vare sig av SCB eller av någon av de många användarna av AKUs deltidsuppgifter.

Vad är då orsaken till att SCB/AKU använt en fast definition av deltid trots av den ordinarie veckoarbetstiden ändrats? Den närmaste förklaringen är att 34-timmars gränsen helt enkelt kom med av "bara farten" när AMS 1959 importerade USAs motsvarighet till AKU, Labor Force Surveys.[2] I den sistnämnda finns 34-timmarsgränsen, men då är att märka att i USA var 40-timmarsveckan genomförd redan omkring 1950 (4). I Sverige genomfördes 40-

[1] Dessa över- och underskattningar grundar sig på beräkningar i tab. 1 på gruppen "i arbete". Vanligen talar man om deltids*sysselsatta*, vari ingår deltidsarbetande *plus* de frånvarande som vanligen arbetar deltid. Deltidsandelarna beräknade på gruppen sysselsatta ger lägre *nivåer* än de redovisade i tab. 1: för gruppen "båda könen" innebär det ingen skillnad i över- eller underskattningen medan det för "män" resp. "kvinnor" ger en mycket liten skillnad. Den stora över- resp. underskattningen av deltidsandelen kvarstår alltså, även om man gör beräkningarna på gruppen "sysselsatta".

[2] Under de första åren publicerades resultaten av AMS, för att sedan övertas av SCB.



timmarsveckan först 1973. Vad bör då göras åt denna över- eller underskattning? SCB/AKU bör nu behålla 34-timmarsgränsen. Vad SCB/AKU *bör* göra däremot är att redovisa (i någon specialpublikation) effekten av att använda olika definitioner för deltid. Detta kan enkelt ske genom att redovisa tidsserier för några olika definitioner, och för några olika undergrupper, enligt metoden i tab. 1.

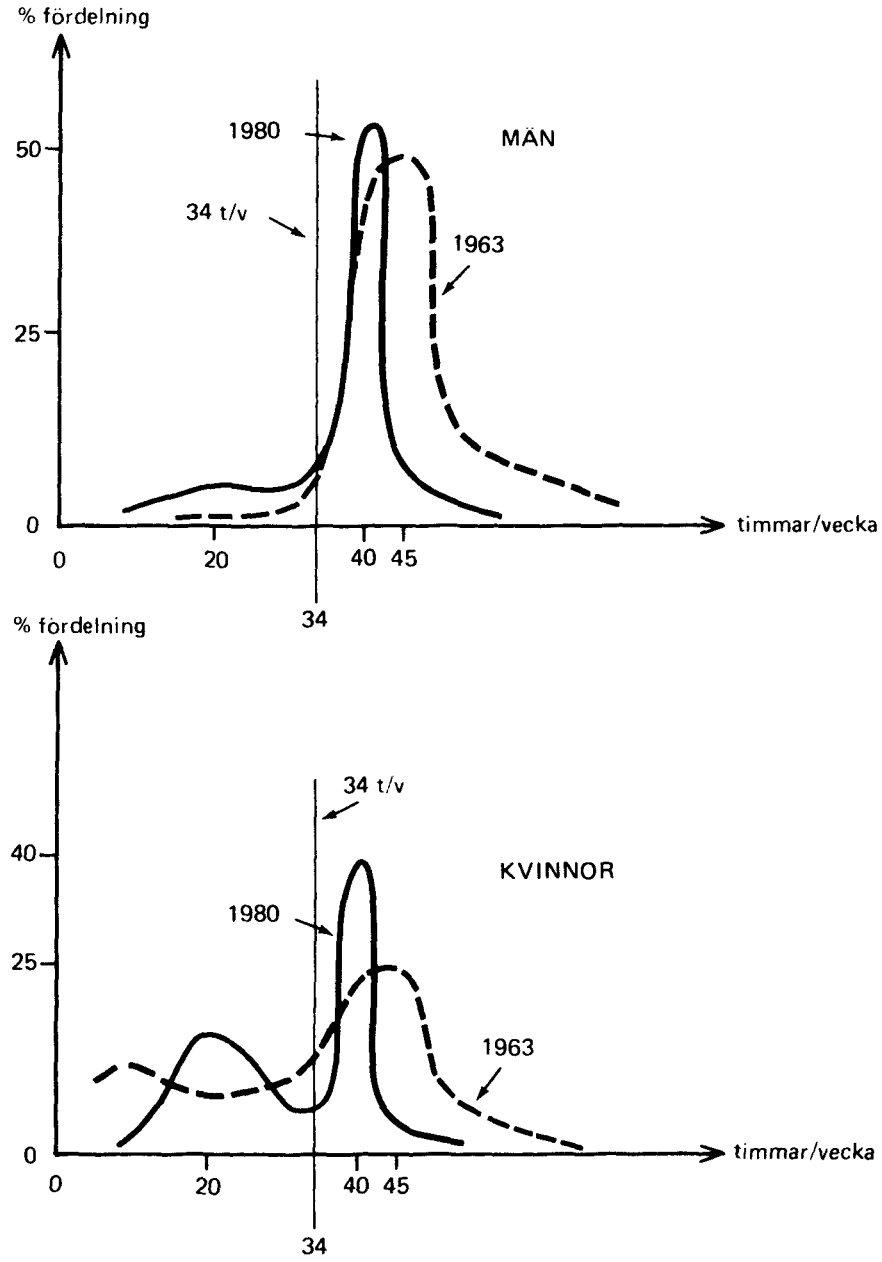


Fig. 1. Fördelningen av veckoarbetstiden för män resp. kvinnor 1963 och 1980 (gruppen "i arbete"). (Kurvorna är grovt ritade och bör inte användas för direkta mätningar i diagrammen. Analysen ovan gäller 1965 och framåt, medan figuren visar fördelningen 1963. Denna skillnad har dock ingen betydelse här.)
Källa: AKU: Råtabeller, årsmedeltal (tab. 20).

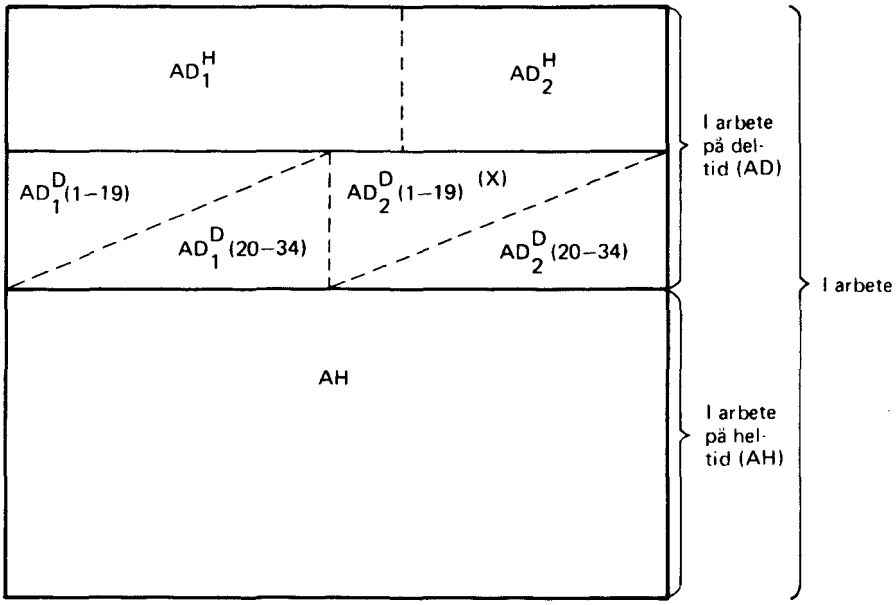


Fig. 2. Gruppen "i arbete" uppdelad med hänsyn till svar på frågor om arbetad tid

Förkortningarna i figuren betyder:

AH = i arbete på heltid.

AD^D = i arbete på deltid, *vanligen* arbetande deltid.

AD^H = i arbete på deltid, *vanligen* arbetande heltid.

AD_1^D (1–19) = AD_1^D där parentesen anger arbetstiden *i arbete*.

AD_1^D = i arbete deltid, vanligen arbetande deltid av *arbetsmarknadsskäl*.[1]

AD_2^D = i arbete deltid, vanligen arbetande deltid av *andra skäl*.[1]

Dvs. *siffran* anger *motivet* till "vanligen beteendet".

((x): denna grupp kan ytterligare, fr. o. m. 1970, delas upp i två undergrupper vad gäller längden på den vanligen arbetade tiden).

[1] Definitionen av "arbetsmarknadsskäl" resp. "andra skäl" skiljer sig något mellan de olika grupperna.
a) För gruppen som arbetar deltid och vanligen arbetar deltid, AD^D, gäller något av följande skäl som orsak till *"vanligen beteendet": Arbetsmarknadsskäl:* Dåligt med arbete, reparation av fabrik/maskin, materialbrist, driftsinskränkning. *Andra skäl:* (i) upptagen i eget hushåll, (ii) egen sjukdom, (iii) studier, (iv) full arbetsvecka mindre än 35 timmar, (v) annat skäl, samt (vi) vill inte ha heltid.
b) För gruppen som arbetar deltid men som vanligen arbetar *heltid*, AD^H, gäller något av följande skäl som orsak till att man arbetar deltid *under mätveckan:*
Arbetsmarknadsskäl: som ovan *plus* "börjat eller slutat ett arbete under mätveckan".
Andra skäl: skälen (i)-(v) ovan *plus* (vi) semester, (vii) dåligt väder, samt (viii) arbetskonflikt på egen arbetsplats.

Det här problemet berör också frågan om relationen mellan *arbetstidsförkortningarnas* effekter på den genomsnittliga veckoarbetstiden *och* på fördelningen av den arbe-

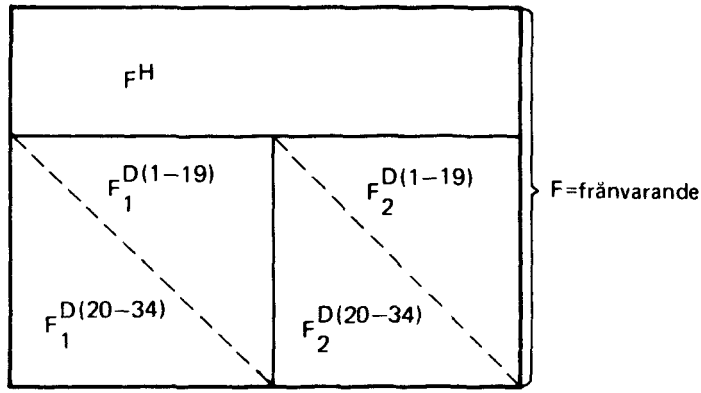


Fig. 3. Gruppen "frånvarande" uppdelad i olika underkategorier[1]

F^H = frånvarande, vanligen arbetande heltid

F^D = frånvarande, vanligen arbetande deltid.

F_1^D = frånvarande, vanligen arbetande deltid av arbetsmarknadsskäl.

$F_1^{D1-19)}$ = frånvarande, vanligen arbetande deltid 1–19 t/v av arbetsmarknadsskäl.

F_2^D = frånvarande, vanligen arbetande deltid av andra skäl.

"Arbetsmarknadsskäl" och "andra skäl" är för gruppen "frånvarande" definierade på samma sätt som för gruppen "vanligen arbetande deltid" (grupp a) i not 1 till fig. 2 ovan.

[1] Denna uppdelning av gruppen "frånvarande" är endast möjlig fr. o. m. 1970.

tade tiden. *Fig. 1* visar fördelningen av veckoarbetstiden för män resp. kvinnor 1963 och 1981 (för gruppen "i arbete").

I fig. 1 är också inlagt AKUs 34-timmarsgräns. Figuren visar att arbetstidsförkortningen, från 45 timmar/vecka 1963/1965 till 40 timmar/vecka 1980 (1973), inte endast sänkte den genomsnittliga veckoarbetstiden utan också påverkade fördelningens utseende. För män innebar förkortningen en kraftig koncentration till standardarbetsveckan, med en viss ökning av fördelningen för korta veckoarbetstider. För kvinnor var utvecklingen delvis samma, dock med en betydande koncentration *också* kring 20-timmar per vecka. Slutsatsen är att förändringar i standardarbetsveckan inte bara påverkar den genomsnittliga arbetstiden utan också dess fördelning, vilket inte är ointressant utifrån analysen av deltidsdefinitionen.

3. Definitioner använda och redovisade av SCB/AKU

I det följande skall de olika definitioner som SCB och/eller AKU använt, och redovisat, analyseras bortsett från problemen ovan med 34-timmarsgränsen.

Vi utgår följaktligen ifrån att 34-timmarsgränsen gäller utan att medföra några problem för någon definition.

När AKU redovisar antalet sysselsatta på deltid är detta ett *konstruerat* begrepp.

Dvs. antalet deltidssysselsatta, och alla sysselsatta i övrigt, fås inte fram som antalet intervjupersoner som svarar på *en* fråga. Med hjälp av svaren på *olika* frågor kombinerar AKU ihop dessa till en skattning av antalet sysselsatta, på deltid och heltid t. ex. Utgångspunkten är att gruppen "sysselsatta" utgör en summa av gruppen "i arbete" plus gruppen "frånvarande". Det innebär att den enda "direkta" uppgiften om veckoarbetstid AKU ger, är den som bygger på gruppen "i arbete"; dessa intervjupersoner svarar *direkt* på *en* fråga om hur många timmar de arbetat under mätveckan.

Men AKUs deltidsdefinition stannar inte här. Med hjälp av svaren på ytterligare frågor redovisar AKU sysselsatta på deltid och heltid i termer av "vanligen": dvs. "sysselsatta fördelade på *vanligen* arbetad tid" är den grundläggande redovisningen. Det innebär att AKUs deltidsdefinition är ytterligare, förutom tillägget av gruppen "frånvarande", konstruerad. Det är dock inget fel om bara AKU/SCB redovisar sina definitioner och beräkningsmetoder och anger motiv till ändringar i använda definitioner.

Så har dock *inte* varit fallet. Jag skall återkomma till detta, men först skall de underliggande begreppen redovisas.

3.1. Definitioner för arbetad tid i AKU

De uppgifter som samlas in från intervjupersonerna (ip) redovisas i fig. 2 och 3 för gruppen "i arbete" resp. för gruppen "frånvarande".

Med fig. 2 och 3 som bakgrund kan vi nu redovisa SCBs/AKUs använda definitioner för *antalet deltidssysselsatta:*

Definition 1: Redovisad i IPF 1973:7 (5) för perioden 1965–72

Den använda definitionen är följande: $AD^D + AD_1^H + F^D$.

I ord innebär denna definition att antalet deltidssysselsatta är lika med antalet i arbete på deltid men vanligen arbetande deltid (AD^D) *plus* antalet i arbete på deltid av arbetsmarknadsskäl, men vanligen arbetande heltid av arbetsmarknadsskäl (AD_1^H), samt *plus* antalet frånvarande som vanligen arbetar deltid (F^D). Dvs. man har tagit med alla som arbetade deltid under mätveckan *utom* dem som vanligen arbetade heltid men arbetar deltid på grund av *andra skäl* (AD_2^H). Motivet till detta är oklart, möjligen anser SCB att denna sistnämnda grupp mera *tillfälligt* arbetar deltid.

Definitionen i publikationen (IPF 1973:3) är svår att hitta och är implicerad i den löpande texten.

En ytterligare komplikation är att någon uppdelning av gruppen "frånvarande" inte är möjlig för denna tidsperiod varför SCB (på något okänt sätt)[1] skattat storleken på undergruppen "frånvarande vanligen arbetande deltid", F^D.

Definition 2: Redovisad i SM 1981:33 (6) för perioden 1970–75

Den använda definitionen är här följande: $AD^D - AD_1^D(20-34) + F^D - F_1^{D(20-34)}$.

I jämförelse med definition 1 har gruppen AD_1^H utgått, gruppen i arbete deltid 20–34 timmar, vanligen arbetande deltid av arbetsmarknadsskäl ($AD_1^D(20-34)$) likaså, samt har en ytterligare minskning av antalet skett genom att gruppen frånvarande vanligen arbetande deltid 20–34 timmar av arbetsmarknadsskäl ($F_1^{D(20-34)}$) också utgått.

Såväl motiv till denna definition som

[1] Det föreligger dock en arbetspromemoria (av M. Strandberg 1972-11-28, Prognosinstitutet) där metoden anges: uppgifter för 1970–72 har direkt antagits gälla också för 1965–69.

själva definitionen är helt omöjlig att finna i någon publikation av SCB/AKU.[1] Möjligen kan motiven till definition 2 vara att SCB velat ta bort grupper som "blivit tvingade" av arbetsmarknadsskäl att arbeta deltid.

Definition 3: Redovisad i SM 1978:32 (7) och SM 1981:33 (6) för perioden 1965–1975 resp. 1975–1980

Detta är den nu gällande definitionen och den ser ut enligt följande: $AD^D + F^D$.

Denna definition innebär en tillbakagång till definition 1, dock med den skillnaden att gruppen "i arbete deltid på grund av arbetsmarknadsskäl, vanligen arbetande heltid" (AD_I^H) inte ingår. Motiven även till denna definition är okända, men troligen grundade på antaganden om att grupper typ AD_I^H bara "tillfälligt" är deltidsarbetande.

Definitionen är svår att hitta[2] och det är mycket svårt att förstå att här finns en tidsserie, 1965 och framåt, med samma definition av deltidssysselsatt.

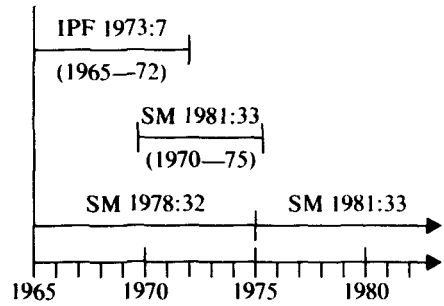
3.2. Sammanfattning

De av SCB/AKU använda definitionerna sammanfattas nedan tillsammans med uppgift om tidsperiod och publikation:

Def.1: $AD^D + AD_I^H + F^D$

Def.2: $AD^D - AD_I^D(20-34) + F^D - F_I^{D(20-34)}$

Def.3: $AD^D + F^D$



Relationerna mellan de olika definitionernas mått på antalet och andelen deltidssysselsatta (för båda könen) ges i tab. 2 på sid. 129.

Relationen mellan de olika definitionerna, vad gäller både antalet och andelen deltidssysselsatta, är

def 1>def 3>def 2.

Dvs. definition 1 ger alltid ett högre värde än definition 3, som i sin tur ger ett högre värde än definition 2.

Detta innebär att definitionen överskattar antalet (och andelen) deltidssysselsatta i relation till dagens definition (definition 3). Detta alldeles bortsett från att 60-talets AKU-genomsnitt är 4-månadersgenomsnitt (mot dagens 12-månadersgenomsnitt) och därför ytterligare överskattar antalet deltidssysselsatta. Jämförelsen mellan definition 2 och 3 vad gäller perioden efter 1970 skymmes dock inte av några sådana skillnader. För denna period gäller att definition

[1] Den "definition" som lämnas på sid. 9, SM 1981:33 ger ingen som helst vägledning, snarare verkar den förvirrande.

[2] För att få någon hjälp får man gå till SM 1978:32 (fotnot 3 och 4 i tab. 1) eftersom ingen hjälp står att få i den senare publikationen (SM 1981:33). *Eller* får man gå till figuren på försättsbladen i råtabellerna och härleda definitionen; figuren finns dock bara fr. o. m. 1976.

Tab. 2. Antal och andel deltidssysselsatta, båda könen (Se även diagram 2)

	Antal (100-tal)			% av totala antalet sysselsatta			
	Def 1	Def 2	Def 3	Def 1	Def 2	Def 3	Def 4[3]
1965	5 432		5 290	14,7		14,3	15,9
1966	5 932		5 760	15,9		15,4	17,3
1967	5 878		5 640	15,9		15,3	17,2
1968	6 370		6 160	17,0		16,5	18,5
1969	6 666		6 470	17,6		17,1	19,1
1970	6 969	6 560	6 807	18,1	17,0	17,7	19,4
1971	7 179	6 670	6 945	18,6	17,3	18,0	19,8
1972	7 409	6 860	7 194	19,2	17,8	18,6	20,4
1973	7 545[1]	7 000	7 364	19,4	18,0	19,0	21,0
1974	7 632[1]	7 100	7 454	19,3	17,9	18,8	20,9
1975	8 242[1]	7 740	8 088	20,3	19,1	19,9	21,9
1976	8 638[1]		8 496[2]	21,1		20,8	22,9
1977	9 284[1]		9 130[2]	22,7		22,3	24,5
1978	9 832[1]		9 675[2]	23,9		23,5	25,9
1979	10 261[1]		10 127[2]	24,6		24,2	26,7
1980	10 629[1]		10 470[2]	25,1		24,7	27,2
1981	10 869[1]		10 665[2]	25,7		25,2	27,9

[1] Finns ej redovisade.
[2] Finns redovisade (i SM 1981: 33, tab. 1) men svårt att identifiera som denna definition.
[3] Alternativ definition: se under punkt 4 nedan.

3 överskattar antalet och andelen deltidssysselsatta.[1]

Tab. 2 *kan* tas som intäkt för att säga att skillnaden mellan de olika definitionerna är små. Och det kan man hålla med om, även om en nedbrytning till undergrupper (män och kvinnor i olika åldrar t. ex.) troligen inte ger lika små skillnader.

Det är dock olyckligt att någon klar redovisning av använda definitioner och motiven till dessa inte gjorts från SCBs sida.[2] Den främsta nackdelen med detta är att konsumenterna av AKUs statistik kan få intrycket att definitionen av deltidssysselsatta är en absolut, statistisk fråga.

Analysen ovan av relationen standardarbetsvecka-deltidsdefinition och diskussionen om graden av "tillfällighet" i SCBs definitioner, visar dock att definitionen av deltidssysselsatt är en *social definition* med dess brister och fördelar. Den sociala dimensionen accentueras ytterligare av att begreppet "vanligen" spelar en central roll i definitionerna (detta skall diskuteras mera ingående nedan).

[1] För åren 1970–75 kan denna överskattning exakt beräknas från råtabellerna (inom parentes anges denna överskattning i relation till alla sysselsatta): 1970: 26 000 (0,7 %); 1971: 30 800 (0,8 %); 1972: 36 000 (0,9 %); 1973: 37 100 (1,0 %); 1974: 36 300 (0,9 %) samt för 1975: 35 000 (0,9 %). Siffrorna fr. o. m. 1976 kan ej beräknas ur råtabellerna. För 1976 anges dock 36 000 (SM 1981: 33, tab. 1, not 1).
SCB anger dock *olika* uppgifter för denna överskattning. I en PM (SCB, I/UI, L. Gellerstedt, "Deltidsdefinitioner m. m. i AKU", 1977-05-10) anges att *samtliga* sysselsatta som vanligen arbetade deltid på grund av arbetsmarknadsskäl fördes till gruppen heltidssysselsatta. I SM (1981: 33, tab. 1, not 1) anges att bara de som arbetade *20–34 timmar/vecka* fördes till heltidsgruppen. Denna sistnämnda definition stämmer med definition 3 ovan.

[2] Nyligen publicerade SCB en ytterligare definition av deltid: antalet "i arbete 1–34 tim. per vecka" i procent av "sysselsatta". (Se tab. 5: 7, sid. 92 i "Perspektiv på välfärden 1982". Levnadsförhållanden nr 33, SCB 1982. Observera att definitionen inte redovisas.)

Denna redovisning (fig. 2 och 3) av underlaget till definitionerna visar också att var och en kan sätta samman *sin egen definition* (inom ramen för AKUs frågor), beroende på vad syftet är. Denna egenskap hos statistikproduktionen är ytterst värdefull för dem som utnyttjar statistiken; SCB borde därför uppmuntra en uppläggning och presentation av statistiken som underlättar möjligheterna till sådana "egna" definitioner.

4. Begreppet "vanligen" samt förslag till ny definition

Förutom frågan till ip om hur många timmar ip *arbetade* under mätveckan, ställs också frågan hur många timmar ip *vanligen arbetar*. Denna sista fråga ställs till alla i arbete som arbetar deltid[1] samt till alla frånvarande. Svaret på frågan om vanligen arbetad tid kodas direkt i fyra arbetstids*klasser* (1–15, 16–19, 20–34 samt 35–t/v). Enligt intervjuanvisningarna får ip dessa fyra klasser upplästa för sig och väljer sedan vilken klass ip anser att begreppet "vanligen" motsvarar.

Enligt min mening är uppgiften om den *arbetade tiden under mätveckan* betydligt mera tillförlitlig (inom ramen för AKUs allmänna kvalitet) än uppgifterna om "vanligen" arbetad tid. Orsaken till detta är två. Dels är "i arbete"-måttet mera exakt (på 0,5 timmar när) i relation till "vanligen arbetad tid", vilket är en uppskattning av en typ av genomsnitt; dessutom skall "vanligen arbetad tid" inplaceras i tämligen breda klasser.

För det andra kan ip ha mycket diffusa och, mellan grupper och över tiden, varierande uppfattningar om vad som skall menas med "vanligen". Detta även om intervjuaren ger viss hjälp vid intervjutillfället. Det kan t. ex. vara så att *män*, i relation till kvinnor, överdriver heltidsarbete som "vanlig" tid eftersom män i sin sociala roll mera är "försörjaren". Likaså kan ip med "vanligen" i vissa fall mena den för tillfället rådande standard(heltids)arbetsveckan. Dvs. problemen ur validitets- och reliabilitetssynpunkt är betydande för begreppet "vanligen" *i relation till* måttet på arbetad tid under mätveckan.

Så vitt jag vet har SCB inte gjort någon undersökning av de här problemen vad gäller arbetskraftsundersökningarna.

4.1. Förslag till alternativ definition

För att minimera effekten av begreppet "vanligen" skall här lämnas ett förslag till ny definition av deltidssysselsatt. Den nya definitionen innebär också att begreppet "tillfällig" explicit analyseras.

Den här föreslagna definitionen får ses som ett *första förslag* i väntan på att SCB, eller någon annan, tar upp och analyserar de problem som berörts i den här artikeln.

Den nya definitionen, i enlighet med ordningen ovan kallad definition 4, ser ut som följer:

Definition 4: $AD+F^D-AD_2^H$ (sjukdom/semester).

I ord innebär definition 4 att deltidssysselsatta är de som arbetar deltid under mätveckan (AD) *plus* de frånvarande som vanligen (!) arbetar deltid (F^D) *minus* gruppen som arbetar deltid på grund av sjukdom och semester men som vanligen arbetar heltid (AD_2^H (sjukdom/semester)).

Definition 4 innebär, i relation till de andra definitionerna, en ytterligare ökning av antalet deltidssysselsatta eftersom fler grupper inkluderats i definitionen; definition 4 ger därför större antal och andel deltidssysselsatta än definition 1 ovan.

[1] Dvs. till gruppen "i arbete på *hel*tid" ställs *inte* någon fråga om vanligen arbetad tid.

Om vi jämför med fig. 2, ser vi att SCBs definitioner i de flesta fall (utom för definition 1) undantagit gruppen "i arbete deltid men vanligen arbetande heltid" (AD^H) från deltidsdefinitionen. Definition 4, å andra sidan tar med nästan hela denna grupp med undantag av en undergrupp till AD_2^H (dvs. AD_2^H (sjukdom/semester)).

Tab. 3. Procentuell fördelning av orsaker till deltidsarbete för gruppen "i arbete deltid men vanligen arbetande heltid" (AD^H): 1981, båda könen, samtliga SNI

Angiven orsak	% av alla svaren
Arbetsmarknadsskäl	8,2
Arbetskonflikt, dåligt väder	1,4
Egen sjukdom	32,0
Semester	22,0
Upptagen i eget hushåll	5,9
Studier	1,5
Hel arbetsvecka<35 t/v	19,6
Övriga skäl (bl. a. tjl.)	9,3
	100

Källa: AKU, Råtabeller, Årsmedeltal 1981, tab. 23.

Skillnaden i detta avseende kan sägas ligga i tolkningen av hur många "tillfälliga" deltidsarbetande det ingår i gruppen AD^H. Definition 4 implicerar att de flesta i denna grupp *inte* är att betrakta som tillfälligt deltidssysselsatta, medan SCBs definitioner implicerar motsatsen.

Hur ser då gruppen AD^H ut? I tab. 3 redovisas fördelningen 1981 av olika orsaker. ip uppgett till att han/hon arbetar deltid fastän ip "vanligen" arbetar heltid.

Tab. 3 visar fördelningen endast för ett år, (1981). Detta har dock ingen större betydelse eftersom den relativa fördelningen på orsaker har varit förvånansvärt konstant under 60- och 70-talen. Den enda skillnaden är att skälet "hel arbetsvecka<35 timmar" har ökat och "annat skäl" minskat, förutom en konjunkturvariation för andelen "arbetsmarknadsskäl". Någon större skillnad mellan könen föreligger inte heller.

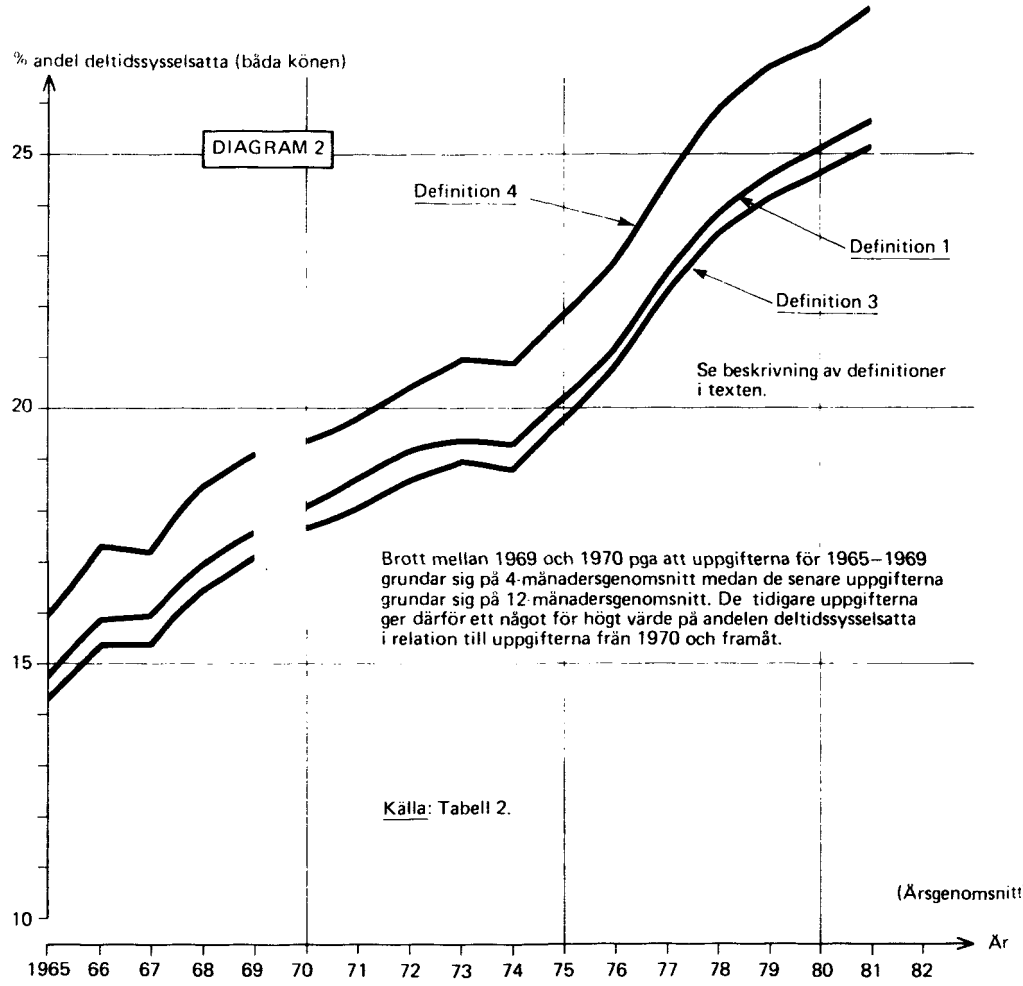
Frågan är nu: om vi inkluderar hela gruppen ovan (AD_2^H) i deltidsdefinitionen, vilken (eller vilka) orsak(-er) i tab. 3 skapar mest "brus" i definitionen? Och med "brus" menar vi "tillfällighet", dvs. ju mera tillfälligt ip är deltidsarbetande desto mera "brus".

Som sådant brus bör "semester" och "egen sjukdom" anses vara, vilket framgår av definition 4 där denna grupp är frändragen. Om ip är sjuk eller har semester, t. ex. en dag under mätveckan, innebär detta inte att ip är att anse som deltidsarbetande, givet att uppgiften "vanligen" är korrekt. Skälen "upptagen i eget hushåll", "studier", "hel arbetsvecka<35 timmar" samt "övriga skäl" får anses klart tillhöra gruppen av mera permanent deltidsarbete.

Återstår så "arbetsmarknadsskäl" och "arbetskonflikt och dåligt väder". Vad gäller "arbetsmarknadsskäl" består den till 10–20% av att ip är ledig mellan arbetsbyten, medan resten utgöres av orsaker som hänger samman med bristande efterfråga. Denna grupp är därför en deltidsgrupp även om orsaken nästan helt är "påtvingad" ip. Orsaken "arbetskonflikt och dåligt väder" slutligen, får anses tillhöra gruppen av mera tillfälliga orsaker, men är dock av mycket liten betydelse.

Summa summarum innebär inkluderandet av gruppen AD^H, minus gruppen som anger sjukdom/semester som orsak, i gruppen deltidssysselsatta, att vi får med grupper som i stort är mera permanent deltidssysselsatta.[1]

[1] När det gäller genomslag för ev. "brus" beror detta på hur stor gruppen AD_2^H (sjukdom/semester) är i relation till basgruppen (AD). För båda könen är denna grupp ca 10% av alla i arbete (AD), varför ev. brus slår igenom med *maximalt* en faktor 1/10. (För män är denna andel större, ca 25%, och för kvinnor betydligt mindre eller ca 5%).



Definition 4, som *andel* deltidssysselsatta av alla sysselsatta, är inlagd i diagram 2 tillsammans med två av SCBs tre definitioner.

Diagrammet visar en relativt stor följsamhet mellan kurvorna fram till mitten av 70-talet. Därefter ökar definition 4 relativt mer än de övriga två. Detta kan tolkas som att deltidsarbetets ökning har skett i *andra former* än vad som täcks in av de "traditionella" definitionerna (def. 1 och 3).

Som nämnts ovan är definition 4 ett *förslag* i väntan på att de berörda problemen tas upp till analys. Definitionen är, vilket framgår, inte helt fri från beroendet av begreppet "vanligen". Påverkan från den osäkerhet som detta begrepp skapar, har dock minskat betydligt i relation till SCBs definitioner. Vi har också, inom ramen för utrymmet för en artikel, explicit redovisat vad för *social* innebörd vår definition har.

5. Slutsatser

Den viktigaste slutsatsen av denna genomgång är att deltidsbegreppet inte är entydigt, och detta alldeles bortsett från den allmänna statistiska osäkerhet som finns i undersökningar av en urvalstyp som AKU. Definitionen av vem som är deltidssysselsatt är en i grunden *social* definition och inte någon statistisk sådan.

De av SCB/AKU använda definitionerna har både brister och systematiska fel. Det senare gäller främst relationen till heltidsarbetsveckan och därmed innebörden av begreppet "vanligen". Vad gäller bristerna gäller dessa främst att SCB/AKU inte explicit redovisat dessa. Det är i och för sig inte någon avgörande kritik mot en statistisk serie att den har brister. Vilka har inte det, med hänsyn till bl.a. de ekonomiska begränsningar som gäller? Kritiken bör dock vara relevant om bristerna inte redovisas på ett sådant sätt att användaren kan tolka statistiken utifrån olika synpunkter.

Allmänt kan man säga att arbetstidsmåtten i AKU är koncentrerade på utbudssidan: dvs. det är mängden timmar som *individen* önskar eller verkligen arbetar som är utgångspunkten. Ett exempel på detta är att antalet arbetade timmar under mätveckan mätes oberoende av antalet anställningar/arbetsgivare. Från efterfrågesynpunkt är det intressant att se utvecklingen av arbetade timmar *per arbetsgivare*, med uppdelning på ordinarie arbetstid och övertid t.ex. Här finns med andra ord *ytterligare* områden att analysera när det gäller arbetstidens ställning i arbetskraftsundersökningarna.

6. Referenser

1. Pettersson, M. (1981): Deltidsarbetet i Sverige, Arbetslivscentrum 1981.
2. Deltidsarbete 1974 (SOU 1976:7).
3. Tegle, S. (1983): Den ordinarie veckoarbetstiden i Sverige 1860–1980: En översikt, Meddelande 1983:86 Nationalekonomiska institutionen, Lunds universitet.
4. Dankert, C.E. et. al. (1965): Hours of Work, Industrial Relations, Research Association Publications No 32, New York 1965, (kap. 2 och 4).
5. Arbetskraftsresursernas utveckling 1965–1972 och 1972–1980. Analys och prognos. Information i prognosfrågor nr 1973:7, SCB 1973.
6. Arbetskraftsundersökningarna 1970–1980. Statistiska meddelanden, Am 1981:33, SCB 1982.
7. Arbetskraftsundersökningarna 1963–1975. Statistiska meddelanden, Am 1978:32, SCB 1978.

Meddelanden

OECD-möten

Möten om sociala indikatorer

Under hösten 1982 har OECD arrangerat två möten om sociala indikatorer. Det första mötet ägde rum 21–22 oktober och var ett s.k. expertmöte där användningen av sociala indikatorer i medlemsländerna och internationellt diskuterades. I detta möte deltog från Sverige som svensk representant Staffan Wahlström, SCB och som särskilt av OECD inbjuden expert Joachim Vogel, SCB.

Det andra mötet ägde rum 9–10 december 1982. Det var ett ordinarie s.k. working party möte. Sådana möten har arrangerats inom ramen för OECDs Social Indicator Development Program sedan början av 70-talet. Svensk delegat vid working party mötet var Staffan Wahlström, SCB.

Arbetet med sociala indikatorer inom OECD har hittills resulterat bl.a. i att OECDs högsta organ "Council" fastställt en "OECD List of Social Indicators". Den omfattar 37 indikatorer inom flera sociala områden. Höstens möten har behandlat hur denna lista av sociala indikatorer ska bättre kunna nyttiggöras såväl nationellt som internationellt. Samt om och i så fall hur OECD-arbetet skall drivas vidare.

Det ska understrykas att 70-talet inneburit ett starkt uppsving för välfärdsmätning och välfärdsforskning i många OECD-länder – inte minst i de nordiska länderna. Det framkom vid höstens OECD-möten att denna verksamhet nu fått en fast förankring i ett stort antal OECD-länder.

Förutom ett erfarenhetsutbyte vad beträffar nationell användning av sociala indikatorer redovisades vid oktobermötet två konkreta exempel på internationella jämförelser av välfärdsfördelningen. Dels redovisade Joachim Vogel arbetet med en nordisk social rapport. Ett arbete som genomförs av de statistiska centralbyråerna i Danmark, Finland, Norge och Sverige, med Sverige som samordnare och på uppdrag av Nordiska rådet. Dels redovisade Mr Alex Michalos, Social Science Federation of Canada, sitt arbete med att jämföra levnadsförhållandena i Kanada och USA – "The North American Social Report". Möjligheter och problem med den här typen av internationella jämförelser diskuterades.

Olika åsikter om framtiden för den s.k. sociala indikatorrörelsen kom fram i diskussionen vid de här OECD-mötena. En linje hade en pessimistisk grundton. Behovet av välfärdsbeskrivningar och välfärdsanalys kommer att minska i en värld där politikernas intresse för fördelningsfrågorna är i avtagande på grund av den ekonomiska stagnationen.

Bland skeptikerna fanns också de som menade att sociala indikatorrörelsen hade misslyckats med att få fram användbara och tillförlitliga indikatorer.

En annan linje – som undertecknad företrädde – gick ut på att behovet av fördelningsstudier snarare ökar när omprioriteringar och nedskärningar av de sociala trygghetssystemen blir nödvändiga. Beslutsfattare och experter måste veta hur förändringarna ska kunna genomföras så att de drabbar olika grupper så rättvist som möjligt och så att svaga grupper skyddas. Vad beträffar sociala indikatorrörelsens "misslyckande" så ligger det något i denna kritik när det gäller vissa områden t.ex. att mäta arbetsmiljö men absolut inte när det gäller tunga områden som sysselsättning, utbildning, inkomst och boende.

Företrädare för de två framtidsbedömningarna kunde enas i att informationsbehovet i hög grad är beroende av de politiska ambitionerna i resp. land.

Vid working party mötet i december kunde enighet nås om att OECD-arbetet med sociala indikatorer borde fortsätta de närmaste åren. Resurserna bör satsas på att implementera och "konsolidera" de redan utvecklade indikatorerna snarare än på ut-

veckling av nya indikatorer inom sociala områden som är dåligt täckta i den nuvarande OECD-listan av sociala indikatorer. OECD bör stimulera medlemsländerna att använda sig av OECD-listan nationellt. Men även internationella analyser bör stimuleras bl. a. genom att OECD med hjälp av statistik från medlemsländerna utarbetar sociala rapporter som beskriver den sociala utvecklingen i medlemsländerna.

Staffan Wahlström

Nordiska statistiska utskott

Utskottet för näringsstatistik (NUN)

Under verksamhetsperioden 1979–1982 har utskottet haft två möten: det ena i Köpenhamn dagarna 16–18 september 1980 och det andra i Oslo 19–21 januari 1982. I anslutning till dessa möten sammanträde en särskilt tillsatt expertgrupp för det praktiska fullföljandet av NUNs arbete med en gemensam nordisk nyckel mellan bransch- och varunomenklaturerna. Som svenska representanter i utskottets möten deltog avdelningsdirektörerna Oddmund Forsmo och Allan Ljungberg och i expertgruppen avdelningsdirektör Carl-Axel Carlström, samtliga från SCB.

Chefsstatistikermötet 1979 konstaterade vid granskningen av utskottets verksamhetsprogram att utskottet har ett mycket brett verksamhetsfält och att omfattande arbeten pågick sedan en längre tid avseende utveckling och vidare utbyggnad av gemensamma instrument på området för näringsgrensklassificering. Det var önskvärt att prioritera och eventuellt begränsa antalet dagordningspunkter samt fördela dessa mellan mötena så att en större koncentration kring fullförandet av de mest angelägna uppgifterna kunde förverkligas. Under Köpenhamnsmötet ägnade utskottet också stor uppmärksamhet åt dessa frågor. Man fann det ändamålsenligt för planeringen och beskrivningen av utskottets arbeten att indela verksamhetsfältet för utskottet i tre huvudområden: Utveckling av gemensamma nordiska statistikprodukter, samarbete om internationella statistikprojekt och informationsutbyte om nationella utvecklingsarbeten.

I de följande tre avsnitten redogöres för de frågor som behandlades på mötena under respektive huvudområde. I ett avslutande fjärde avsnitt beskrivs utskottets förslag inför chefsstatistikermötet 1982 till verksamhetsprogram för den kommande nya treårsperioden fram till och med 1985.

1. Utveckling av gemensamma nordiska statistikprodukter

Tre dagordningspunkter inom detta område har behandlats vid båda mötena: a) utformning och publicering av en NOR-NI Del 2, b) nyckel mellan de nordiska bransch- och varunomenklaturerna och c) tabeller i Nordisk statistisk årsbok. Den senast nämnda punkten återkommer som en stående punkt på utskottets möten. Mer omfattande utredningsarbeten som en förberedelse till att kunna presentera nya slag av gemensamma nordiska tabeller upptages dock under särskilda punkter; t. ex. vad gäller tabeller över inrikeshandel och utlandsägda företag i Norden.

a) *NOR-NI Del 2*. En första del (NOR-NI Del 1) av den gemensamma nordiska näringsgrensindelningen publicerades 1976 i serien Tekniske rapporter utgiven av Nordisk statistisk sekretariat. Den omfattade näringarna 2 och 3, dvs. Gruvor och mineralbrott samt Tillverkningsindustri. Utskottets fortsatta arbete med en andra del täckande resterande näringar 1, 4–9 hade vid diskussionen på senaste Oslomötet förts så långt att man kunde ta ställning till innehållet och slutliga utformningen av publikationen. Sverige åtog sig att fullgöra den avslutande överarbetningen av textdelarna och Nordiska statistiska sekretariatet den återstående genomgången av förteckningen över NOR-NIs grupper och undergrupper med nycklar till ländernas (Danmark, Finland, Norge och Sverige) nationella standarder.

Enligt planerna på Oslomötet räknades med att en tryckning av Del 2 kunde ske i slutet av år 1982 och utges i samma serie som Del 1. Möjligheten att ge ut en ny utgå-

va där de två delarna slogs samman till en enda med de möjligheter detta skulle innebära vad gäller uppdatering av del 1 och samlad redigering diskuterades även. En sådan lösning avslogs dock bl. a. på grund av osäkerheten om hur länge den gemensamma näringsgrensindelningen i nuvarande utförande kan bestå innan en revision aktualiseras som en följd av pågående internationella revisions- och harmoniseringsarbeten på området.

b) *Nyckel mellan de nordiska bransch- och varunomenklaturerna.* Vid det första mötet i Köpenhamn enades man i samråd med den för detta arbete utsedda expertgruppen om att skilja mellan dels en nyckel för gruppering av varor efter ursprung i produktionen, dvs. en gruppering i branschkaraktäristiska varukategorier (''varuaspekten''), och dels en nyckel för klassificering av produktionsenheter efter näringsgren (''branschaspekten''). Arbetet behövde i första hand koncentreras på att färdigställa den förstnämnda nyckeln. En sådan nyckel utgör ett instrument för en gemensam aggregering av detaljerad varustatistik i såväl produktions- som utrikeshandelsstatistiken till varukategorier knutna till NOR-NI.

För att komma fram till den senare nyckeln för branschklassificering erfordras tillägg (''bivillkor'') till nyckeln enligt varuaspekten avseende fall där hänsyn tages till att samma varukod i varunomenklaturen återkommer i definitionen för flera näringsgrenar. Då dylika hänsyn finns inarbetade i ländernas standarder i olika omfattning och efter skilda ofta ej lätt utskiljbara principer är en vidareutveckling av nyckeln till att fullt ut kunna täcka branschaspekten en komplicerad fråga, som utskottet fann nödvändigt att hänskjuta till ett framtida verksamhetsprogram.

Vid Oslomötet i början av 1982 hade expertgruppens arbete med inriktning på en gemensam indelning i NOR-NI-varukategorier, som ett resultat av en nyckel enligt varuaspekten, förts så långt att NUN kunde diskutera formerna för en slutlig sammanställning av en publikation. Sverige åtog sig att senare under 1982 lämna förslag till mötesdeltagarna avseende textdelen i publikationen. Vidare skulle kvarstående uppföljningsarbeten efter mötet avseende själva nyckeln jämte förteckningen med koder och benämningar på NOR-NI-varugrupper ske skriftligt och per telefon mellan experterna. Sverige åtog sig att fortsätta med att uppfylla sekretariatsfunktionen i expertgruppens arbete.

Vid det senaste mötet enades man också om att utöver att publicera ifrågavarande varugrupperingsstandard även skulle tillhandahålla denna i form av ett ADB-register lämpligt för den tekniska produktionsberedskapen.

Vid Oslomötet diskuterades även en skiss för användning av ovan nämnda ADB-register för sammanställning av en nordisk statistik baserad på ländernas grundmaterial i utrikeshandelsstatistiken. Flera alternativ för uppläggning av tabellinnehåll vid en sådan användning kunde tänkas. Frågan behövde utvecklas vidare och tas upp på ett senare möte.

Klarläggande ansågs också nödvändigt beträffande vilka effekter den gemensamma nyckeln kan få inte bara i användning för gemensam nordisk statistik utan även för ländernas nationella statistik.

c) *Tabeller i Nordisk statistisk årsbok.* Utskottet gick igenom ett antal förslag till ändringar i befintliga tabeller eller till införande av nya. Bland de förslag som antogs var en ny tabell över utlandsägda företag i Norden. Tabellerna avseende produktionsdata för tillverkningsindustrin skulle utökas i flera avseenden. Vidare skulle kartan i årsboken över olje- och gasfält i Nordsjön kompletteras med en tabell över oljeproduktionen och beräknade tillgångar av olja och gas. Ett tabellförslag med uppgifter om arbetsställen efter storlek inom detaljhandeln föll på grund av att olikheter i definitionen av arbetsställebegreppet gjorde att innehållet inte var jämförbart mellan länderna. Frågan föreslogs bli föremål för ytterligare utredning. Detta gällde även ett förslag till tabeller med vissa relationstal som kan belysa industrins kostnadsstruktur, räntabilitet och kapitalförhållanden.

2. Samarbete om internationella statistikprojekt

Viktigare frågor i samband med det internationella samarbetet inom OECD resp. FN avseende uppbyggnad av statistiken över multinationella företag har behandlats vid flera av utskottets möten redan före Köpenhamnsmötet 1980. Vid detta möte diskuterades den då nyligen framlagda rapporten som utarbetats av en utredningsgrupp tillsatt av Nordiska ministerrådet om "Multinasjonale selskapers rolle i de nordiske lands økonomi". Utskottet ansåg att en bättre samordning mellan de nordiska statistikverken genom NUN å ena sidan och Nordiska ministerrådet å den andra i uppläggningen av det fortsatta utredningsarbetet var nödvändigt. Vidare påpekades bl. a. att en klarare grund för statistiken på området skulle uppnås genom att i stället för det svårdefinierbara begreppet "multinationella företag" arbeta med hela populationen "utlandsäga företag".

När frågan om vidareutvecklingen av den gemensamma statistiken på området behandlades ånyo på det senare Oslomötet deltog en representant från Nordiska ministerrådets sekretariat. Det framgick att rådet fattat beslut om att det fortsatta arbetet med att belysa de multinationella företagens roll i de nordiska ländernas ekonomi skulle ombesörjas av Ämbetsmannakommittén för industri- och energipolitik. NUN framhöll igen betydelsen av att det kommande arbetet i kommittén sker i nära kontakt med utskottet så att klarhet erhålls vad beträffar tillgångar på relevant statistik och eventuella brister i denna. Från svensk sida underströks även vikten av att kommittén närmare undersöker på vilket sätt den verksamhet som OECD bedriver avseende utlandsägda företags roll bör och kan påverka arbetet på nordiskt plan.

3. Informationsutbyte om nationella utvecklingsarbeten

Beträffande de olika områden för vilka man inom utskottet kommit överens om att utbyta fördjupad information behandlades bl. a.

a) *Enhetsfrågor som anknyter till näringsstatistiken.* Vid båda mötena redogjorde länderna för skilda slag av arbeten med uppbyggnad, vidareutveckling och underhåll av de centrala företagsregistren. Enhetsfrågorna i sammanhanget är till viss del besvärliga och gäller såväl företags- som arbetsställeenheter. Bland dylika frågor kan särskilt nämnas principiella och praktiska problem med att komma fram till fungerande lösningar avseende enheternas existens över tiden.

Som underlag för bl. a. strukturförändringsstatistik har det intresse att – förutom att kunna identifiera reell nyetablering och nedläggning av enheter skild från fall med enbart byte av ägare eller juridisk form – även kunna särskilda regional omlokalisering (flyttning) och byte av branschtillhörighet på grund av ändrad verksamhetsinriktning. Sammanslagningar och delningar av enheter som medför ändringar i beståndet vållar också betydande problem för bestämning av vilka enheter som fortlever, födes eller dör.

På det senare mötet redogjordes från norsk sida för en utredning om enheterna i företagsregistret och vissa problem i sammanhanget för produktionsstatistiken. Svårigheter föreligger att entydigt avgränsa funktionella enheter i industristatistiken på grund av den ökade integrationen som sker på ägaresidan i norskt näringsliv och den allt vanligare förekomsten av nyetablerade enheter som inte har något (påtagligt) samband med nystartad eller redan etablerad produktionsverksamhet. Verksamheten inom sådana enheter är ofta mer jämförlig med den verksamhet som bedrivs inom interna hjälpavdelningar, dvs. aktiviteter som är inriktade på att tillgodose ett företags eller ett arbetsställes eget behov. Frågan belyste vissa problem i jämförbarheten i den nordiska statistiken och aktualiserade en förnyad översyn av enhetskomplexet. Det föreslogs att en sådan skulle ske varvid speciellt behandlingen av s. k. hjälpverksamhet borde bli föremål för genomgång vid kommande möte i utskottet.

b) *Statistik som belyser strukturförändringar.* Informationsutbytet på utskottets

möten avseende sådan strukturförändringsstatistik som gäller t. ex. nyetablering och nedläggning av produktionsenheter var så intimt knutet till enhetsfrågor och ländernas centrala företagsregister att detta beskrivs i föregående avsnitt. En annan fråga av stor betydelse för studier av inriktningen m. m. i strukturomvandlingen är den snabba utvecklingen av elektroniken. Från svensk sida lämnades en redogörelse över SCBs förstudie av statistikbehoven om elektronikens utnyttjande och effekter. I de övriga nordiska länderna hade man bl. a. av resursskäl ej kunnat realisera idéer om ett liknande utvecklingsarbete. Utskottet var intresserat av att följa upp det svenska arbetet och rekommenderade att frågan togs upp på ett kommande möte.

c) *Byggnads- och bostadsstatistik.* Under denna punkt informerades om det arbete som pågår i länderna och problem som är förknippade med att utveckla och ajourföra byggnads- och lägenhetsregistret. Från svensk sida behandlades även det pågående arbetet med en ny byggnadsproduktionsnomenklatur som är inriktat mot en uppläggning som ger en högre grad av flexibilitet i uppläggningen.

d) *Konjunkturstatistik.* Länderna informerade om olika publikationer innehållande konjunkturindikationer som belyser den kortsiktiga utvecklingen. Från svensk sida presenterades bl. a. de tre nya SCB-publikationerna "Indikatorer – ekonomisk månadsöversikt för Sverige", "Arbetskraftskostnader per producerad enhet – ett mått på industrins internationella konkurrenskraft" och "Industrins kapacitetsutnyttjande". I sammanhanget diskuterades även om det fanns behov av och möjligheter till att sammanställa och sprida väsentlig näringsstatistisk information om det aktuella ekonomiska läget och utvecklingen i de nordiska länderna genom en nordisk "konjunkturrapport". Utskottet fann dock att OECDs rapport "Indicators of Industrial Activity" täcker detta behov.

Under denna punkt behandlades även frågor beträffande nordiskt samarbete om databaser. Ett nordiskt kontaktmöte om databassamarbete främst utifrån statistikservice avseende tidsseriedata hade tidigare ägt rum i Stockholm varvid särskilda kontaktpersoner utsetts. Dessa kunde i första hand förväntas ta itu med frågor avseende teknisk omföring av data etc. NUN borde vara ett lämpligt forum för att i fortsättningen behandla frågor avseende innehåll i basen, jämförbarhet av data m. m.

e) *Utnyttjande av administrativa källor.* Från finsk sida presenterades ett dokument om statistikcentralens erfarenheter av administrativt källmaterial i statistikproduktionen och om pågående projekt angående rationalisering av datainsamlingen. Erfarenheterna från länderna visade att det för administrativt resp. statistiskt bruk ofta ställs olika krav på innehållet och systemens utformning. Problem uppkommer avseende definition och identifikation av enheter, definition och klassificering av egenskaper samt identifiering, överföring och användning av data. Det konstaterades att utskottet borde närmare studera vilka effekter nyttjandet av administrativt material får för jämförbarheten inom olika delar av näringsstatistiken.

f) *Energistatistiken.* Information lämnades även om pågående arbeten med energistatistiken både på nationellt och internationellt plan. Bland de frågor som diskuterades var huruvida det fanns skäl att ändra den gemensamma nordiska statistiken med avseende på innehållet i olika aggregat av energibärare, specifikationen av energibalansens flödestransaktioner eller metoderna för att beräkna den primära energin från vattenkraften. Syftet skulle vara att uppnå bättre anpassning till annan internationell statistisk, främst den som utarbetades av ECE. Erfarenheter utbyttes även angående de problem som föreligger att avgränsa och operationellt hantera fjärrvärmeverken i statistiken. Utskottet fann att ett fortsatt utbyte av information om utvecklingen inom energistatistiken borde ske vid framtida möten.

4. NUNs verksamhetsprogram 1982/83

I samband med diskussionen vid senaste mötet i januari 1982 av utskottets verksamhetsprogram för den kommande treårspe-

rioden behandlades även vissa frågor, som sekretariatet behövde få besvarade med anledning av att det inför chefsstatistikermötet i augusti 1982 skulle utarbeta förslag till det nordiska statistiska samarbetets framtida utformning. Brist på tid gjorde det ej möjligt att slutföra diskussionen vid utskottets möte, utan det bestämdes att skriftliga synpunkter skulle insändas till sekretariatet från de olika länderna.

Det ansågs att NUN har ett mycket brett område att arbeta med som ställer stora krav på insatser. Detta gäller särskilt som vissa frågor ökar i betydelse och fordrar insatser, som endast delvis uppvägs av att andra klaras av eller förlorar i betydelse. För att det skall bli ett meningsfullt arbete som förs framåt är det nödvändigt att koncentrera utskottsarbetet till färre frågor, samtidigt som man måste hålla sig informerad om utvecklingen inom området och därigenom lyfta fram nya angelägna arbetsuppgifter. För att göra detta möjligt krävs att vissa arbetsuppgifter bryts ut och behandlas i tillfälliga arbets(ad-hoc)grupper för framläggande av bearbetade underlag till utskottets behandling. Dessa förutsätts delvis kunna arbeta utifrån skriftliga kontakter. Med anledning av utskottets breda arbetsområde, nödvändigheten av att få en tillfredsställande kontinuitet i ärendenas behandling och kunna ta upp aktuella frågor, borde mötesfrekvensen vara ett per år.

Vad beträffar utskottets verksamhetsprogram 1982/85 ansågs det att man borde arbete med

Vidareföring av gemensamma arbeten avseende:
- Nyckeln mellan de nordiska bransch- och varunomenklaturerna (följa upp och avsluta arbetet, särskilt när det gäller användningen av nyckeln)
- Näringsstatistiska tabeller och andra frågor beträffande NSÅ
- Enhetsproblematiken i näringsstatistik (bl. a. hjälpverksamhet inom koncern/företag/arbetsställe)
- Uppföljning av utvecklingen av nordiska databaser (med avsikt att behandla frågor om innehåll och jämförbar statistik; i samarbete med en annan, för databasernas utveckling i huvudsak i tekniskt avseende, grundad nordisk samarbetsgrupp).

Fördjupat informationsutbyte beträffande:
- Byggnads- och bostadsstatistik
- Energistatistik
- Konjunkturstatistik
- Statistik som belyser näringslivsstrukturen
- Administrativt material i näringsstatistiken.

Genom förstudier undersöka förutsättningarna för att bedriva (t. ex. inom ramen för ad hoc-grupper) gemensamma utredningsarbeten inom följande områden:
- De utlandsägda företagens roll i Norden
- Utvecklandet av relationstal för industrin avseende räntabilitet, kapitalförhållanden etc.
- Pågående internationella revisions-, utbyggnads- och harmoniseringsarbeten avseende aktivitets- och varuklassificeringar (deras betydelse för den gemensamma nordiska statistiken och det agerande detta föranleder).

Oddmund Forsmo *Allan Ljungberg*

Sextonde nordiska statistikermötet

De nordiska statistiska föreningarna samlas vart tredje år till ett gemensamt möte. Nu var det den norska föreningen, som inbjudit till möte i Oslo den 12–14 augusti 1982. Mötets tema var: "80-årenes krav til statistikken og statistikerne". I mötet deltog ca 200 personer. De flesta deltagarna kom från värdlandet. Från Sverige deltog 50 personer.

Mötesprogrammet var mycket ambitiöst upplagt med ett stort antal föredrag, som väl täckte hela det statistiska verksamhetsfältet. Statssekreteraren i det norska finansdepartementet, som öppnade mötet, såg viktiga uppgifter för statistikerna i dagens samhälle. Här finns många nya uppgifter, tekniken ger nya möjligheter och den statistiska fack-kompetensen blir alltmer nödvändig. Statistikerna måste peka ut föränd-

ringarna i samhället och ge politikerna ett beslutsunderlag på rätt nivå. Han såg det statistiska samarbetet som något av ett mönster för nordiskt samarbete.

Flera av inläggen tog upp den snabba tekniska utvecklingen (databaser, terminaler m. m.) och de positiva utvecklingsmöjligheterna man här såg. En annan återkommande fråga var höga och växande krav på statistiken och statistikernas ansvar för en lättillgänglig statistisk allmän information samtidigt som mer specialiserade och för olika syften tillrättalagda presentationer måste fram. Statistikerna måste också verka för en bättre samordning. En obesvarad fråga var dock om de nödvändiga resurserna kommer att finnas och hur fort utvecklingen kommer att gå? Här följer några korta noteringar i form av refererade uttalanden från några av föredragen.

i. "Vet vi for mye om lite og for lite om mye – noen kjetterske tanker om statistikk" av Harald Hanssen-Bauer, Norge.

Vi har redan för mycket information. Besluten blir ändå inte bättre. Vad gör egentligen statistikerna? Det finns mycket kvantitativa data men dålig problemorientering. Mycket statistik verkar finnas för sin egen skull. Den statistiska relevansen är dålig. Det är svårt att hitta rätt – det behövs därför bättre information om statistiken. Både statistikerna och användarna har ett stort ansvar. Centralbyråernas service är dock i allmänhet god.

ii. "Vårt behov av statistik från olika databaser" av Gösta Guteland, Sverige.

Tre utvecklingslinjer kan skisseras: Databaser för beslutsunderlag, allmän information och för minskat uppgiftslämnande. Det blir lätt att följa statistikanvändningen och därmed att kanske ta betalt för statistikuttagen. Bearbetad statistik av allmänt intresse måste bjudas ut brett – en centralbyråroll. Vi får en ökad databasanvändning i framtiden. Ny teknik kan användas för spridning av statistik. Sekretessfrågor och höga kostnader kan bli problem. Statistikerna måste också på ett lättbegripligt sätt tala om vad de håller på med.

iii. "Årsbok eller dataskärm – om förmedling av statistik – vilken plats skall traditionella publikationer ha i framtiden" av Aarno Soivio, Finland.

Mängden publikationer förblir nog konstant, men de blir bättre konsumentanpassade av en ny typ, t. ex. temapublikationer. Allmänheten visar ett växande intresse för statistik. Databaserna blir ett snabbt växande medel för produktion och presentation. Centralbyråerna måste utöka sin service till alla användargrupper.

iv. "En brukers synspunkt på formidlingen av statistikk" av Olav Magnussen, Norge.

Distributionen av statistik är ett underutvecklat forskningsområde. Produktionens organisation leder inte till marknadshänsyn. Centralbyråerna, administrationen och forskarna bildar ett slutet system. Övriga intressegrupper måste också få möjligheter att påverka liksom att få hjälp av centralbyråerna. Allmänhetens behov tillgodoses inte idag. De traditionella publikationerna är t. ex. främst avsedda för administrativt bruk. En mer lättillgänglig publikation behövs för den vanliga användningen, t. ex. för skolan. Grafisk presentation måste utnyttjas bättre. Mer specialiserade publikationer behövs dock dessutom. Den tekniska utvecklingen ger nya presentationsmöjligheter.

v. "Statistikkproduksjonen i et dynamisk samfunn – 80-årenes utfordinger til statistikkerne og deres arbeidsplass. Organisering og oppgaver" av Tom Rishøy, Danmark.

Den dynamiska samhällsutvecklingen kräver anpassning och specialisering från centralbyråernas sida. Statistiken måste leva upp till sin nya roll. En ökad registeranvändning kommer att bli ännu mer nödvändig för en god statistisk beredskap. Centralbyråerna bygger databaser och framställer en mer förädlad statistik. Användarna utnyttjar databaserna via terminaler. Användarna behöver hjälp att tolka data via dokumentation i form av variabelkataloger, kvalitetsdeklarationer m. m.

vi. "Som man frågar får man svar?? Survey-undersökning i samhällsvetenskaplig forskning" av Anders Klevmarken, Sverige.

Behovet av mikrodata om hushållen är stort. Det finns inga alternativ till surveyundersökningar. Registerdata, som är insamlade för andra ändamål, kan aldrig ersätta surveys gjorda för forskning, utan endast komplettera och stödja dem. Aggregerade data kan inte heller ersätta surveydata. Den stora stötestenen är kostnaden. Forskarna måste också ta ett ansvar för t. ex. integritetsproblem. Väl förberedda surveys, och i synnerhet longitudinella surveys, kan ge mycket stora bidrag till våra kunskaper om samhället.

vii. "Utfordinger i arbeidet med å beskrive den økonomiske virkelighet" av Bent Thage, Danmark.

Vi är ofta alltför precisa i behandlingen av ekonomiska frågor i förhållande till den ekonomiska modellbeskrivningen av verkligheten. De internationella organisationernas krav på ändringar av definitioner och klassificeringar kan inte alltid sägas vara nödvändiga. Internationellt samordnad statistik är i mycket en illusion. Vi skall inte räkna med en onödigt hög precision. Samhällsekonomi är inte särskilt exakt. Statistikanvändare och statistikproducenter måste ta sitt ansvar.

viii. "Levnadsförhållanden och statistiken" av Jouko Kaartinen, Finland.

Statistikens gränser har under senare år vidgats allt mer och omfattar nu också t. ex. olika välfärdsmått. Dagens statistik är dock på intet sätt tillfredsställande på detta område. En del kan mätas, men begrepp och definitioner är svåra att precisera. De s. k. socialindikatorerna kan vara ett alternativ. Vi behöver dessutom olika subjektiva mått på levnadsförhållanden.

ix. "Statistik over landegrensene" av Aage la Cour, Danmark[1].

Det finns gamla traditioner och väl fungerande former för det internationella statistiska samarbetet speciellt vad avser standarder, klassificeringar m. m. Kraven på samordnad internationell statistik ökar också alltmer till följd av ett växande ekonomiskt beroende länder emellan. Det statistiska basmaterialet kan förväntas bli allt svårare att samla in. Nationella administrativa register kan bli svåra att samordnat utnyttja på internationell bas. Vi bör dessutom i framtiden sträva efter att finna en bättre fungerande samarbetsform mellan teoretiker och praktiker på det statistiska området.

Till sist gjorde Odd Aukrust en utmärkt sammanfattning av de olika föredragen. Han klassificerade föredragshållarna i sann statistisk anda. Utöver vad som sagts inledningsvis så konstaterade han bl. a. att väldigt få hade talat om kvalitetens och aktualitetens betydelse. Han konstaterade slutligen att utvecklingen nog inte går så fort.

Ett viktigt inslag i de nordiska statistikermötena är den sociala samvaron i samband med luncher, bankett m. m. Våra norska värdar hade verkligen ordnat förnämliga arrangemang. Första kvällen hade Oslo kommun mottagning för oss i Rådhuset. En intressant visning av Rådhuset och dess konstverk avslutade besöket. Andra kvällen är som traditionen bjuder vikt för banketten. Vi hade en mycket trevlig kväll med god mat och dryck på Norsk Sjöfartsmuseums restaurang Najaden. Avskedslunchen på Grand Hotell var en värdig avslutning på tre intensiva konferensdagar i Oslo, som är en trivsam sommarstad.

Efter att ha inbjudit mötesdeltagarna till nästa nordiska möte i Stockholm 1985 åkte vi hem med funderingar på hur vi på bästa sätt skulle lyckas föra traditionen vidare.

Anders Rickardsson

ILOs 13e arbetsmarknadsstatistikerkonferens

ILOs (International Labour Organisation) 13e arbetsmarknadsstatistikerkonferens ägde rum i Genève 18–29 oktober 1982. 156 delegater (varav 26 kvinnor) från 63 länder deltog tillsammans med representanter för arbetsgivar- och arbetstagargrupperna i ILOs styrelse. Vidare deltog, utöver tjänstemän vid ILO, representanter för 13 andra internationella organ. Från Sverige deltog

[1] Detta föredrag har publicerats i denna tidskrift 1983: 1, sid. 7 ff.

avdelningsdirektören Marianne Axling, arbetsmarknadsstyrelsen, departementssekreteraren Björn Jonzon, arbetsmarknadsdepartementet, avdelningsdirektören Elisabeth Lagerlöf, arbetarskyddsstyrelsen samt statistikcheferna Claes Lagerkvist och Margit Strandberg, båda SCB. Till ordförande för konferensen valdes mrs J. L. Norwood, USA och till vice ordförande mrs C. Jusidman, Mexiko, mr E. Ouaba, Benin samt mr K.C. Seal, Indien. Till rapportör utsågs mr D. A. Worton, Kanada. Sekretariatet leddes av mr R. Turvey, chef för ILOs statistiska byrå.

Dagordningen för konferensen upptog sex punkter, nämligen:

1. General Report.
2. Labour force, employment, unemployment and underemployment.
3. Occupational injuries.
4. Revision of the ILO October inquiry on occupational wages.
5. International coding of labour statistics.
6. Paid holidays.

För var och en av punkterna på dagordningen förelåg en rapport, utarbetad av ILOs statistiska byrå.

1. General Report

"General Report" innehöll en överblick över arbetet vid ILOs statistiska byrå under senare år – den 12e arbetsmarknadsstatistikerkonferensen ägde rum 1973. Byråns statistiska publikationer, Year Book of Labour Statistics och den kvartalsvisa Bulletin of Labour Statistics, har i ökad utsträckning datoriserats och därmed snabbats upp. Vidare behandlades utvecklingsarbetet avseende olika innehålls- och metodfrågor, bl. a. har nyligen publicerats en manual för lönestatistik och arbete har påbörjats beträffande den informella sektorn, arbetskraftskostnader, arbetskraftsomsättningen samt arbetskonflikter. Frågan om en revision av ISCO (International Standard Classification of Occupations) togs också upp liksom den tekniska hjälpen till u-länder på arbetsmarknadsstatistikområdet. Slutligen behandlades i rapporten det fortsatta arbetet för statistiska byrån.

Konferensen diskuterade bl. a. prisstatistiken, i synnerhet konsumentprisindex (KPI). Det föreslogs att ILO skulle se över den internationella standarden för konstruktion av KPI. Statistiska byrån påpekade i sammanhanget att byrån tillsammans med FNs statistiska kontor förberedde en manual för konstruktion av KPI med syfte att formulera en bra praxis på detta metodområde med särskilt hänsynstagande till utvecklingsländernas problem.

Under denna punkt behandlades alltså även frågan om en *revision av ISCO*. En konsult – Birgitta Ehrenström, Sverige – hade för ILOs räkning efter konsultationer med ett 15-tal länder och flera internationella organisationer utarbetat en rapport. Ett sammandrag av rapporten inkl. konsultens rekommendationer förelades konferensen. Den fullständiga rapporten avses publiceras i marsnumret av Bulletin of Labour Statistics.

I konsultens rapport och i diskussionen konstaterades samfällt att ISCO, som senast reviderades 1968, var i starkt behov av en förnyad och grundläggande revidering. Med utgångspunkt i vad som i övrigt framgick av konsultrapporten och diskussionen utarbetade en särskild arbetsgrupp en sammanfattande bedömning och föreslog ILO att vidta vissa åtgärder. Gruppens förslag antogs av konferensen. Det konstaterades att ISCO på grund av arbetslivets förändring inte längre är en bra bas för att beskriva yrkesstrukturen vare sig i i- eller u-länder. ISCO tillgodoser inte de tre ändamål den förväntas tillgodose, nämligen:

- att möjliggöra meningsfulla jämförelser mellan olika länder av arbetskraftens yrkes- och socio-ekonomiska struktur
- att vara en ram för förbättring av existerande nationella klassificeringar
- att utgöra riktlinjer och hjälpmedel för länder som för första gången utarbetar en yrkesklassificering.

Bland faktorer som borde beaktas vid en revidering var:

- att revideringen borde avslutas senast 1987 så att en ny klassificering kunde användas i 1990 års internationella folkräkningsomgång
- att en ny klassificering borde kunna upp-

dateras oftare än hittills. Samtidigt borde kontinuitetsbehovet tillgodoses
- att en yrkesklassificering idag tillgodoser ett brett spektrum av ändamål
- att nationella klassificeringar torde behöva ta hänsyn till ett flertal aspekter såsom erforderliga kunskaper, upplärning och erfarenhet, använda material, ansvarsnivå, näringsstruktur och socioekonomiska förhållanden m. m.
- att en internationell klassificering knappast kunde omfatta alla dessa aspekter.

Det rekommenderades mot denna bakgrund att ett internationellt klassificeringssystem skulle utvecklas, som med hänsynstagande till ett brett spektrum av faktorer begränsades till en indelning, som var användbar för internationella jämförelser. Därtill borde riktlinjer utarbetas för utvecklingen av mer detaljerade nationella klassificeringar så att dessa så långt möjligt kunde omvandlas till den internationella ramen.

Beträffande det fortsatta arbetet med ISCO rekommenderade konferensen att en expertgrupp tillsattes som, med utnyttjande av bl. a. de erfarenheter som dokumenteras i konsultrapporten, fick i uppgift att:
- planera ett internationellt yrkesklassificeringssystem och precisera elementen i detta
- ange vilka element i systemet som är relevanta för internationella jämförelser
- specificera form och innehåll i de riktlinjer som behöver utarbetas för de nationella elementen i systemet
- uppskatta tid- och resursåtgång för att fullfölja det arbete som specificeras enligt ovan.

Från statistiska byråns sida framhölls att byrån för närvarande hade varken finansiella – en grov kalkyl pekade på en kostnad på ca 1,5 milj. dollar – eller personella resurser för ett sådant revisionsarbete och att det därför var nödvändigt att man på nationellt plan agerade för tillskott från olika länder för ändamålet eller för omprioriteringar inom ILO; denna fråga fick vidarebehandlas av ILOs styrelse.

Under denna punkt behandlades som nämnts också *arbetsprogrammet för statistiska byrån*. I konferensdokumentet hade åtta programpunkter tagits upp för diskussion. Sju av dessa ansåg konferensen borde övervägas att tas upp på programmet; den åttonde – att utarbeta historiska serier av arbetsmarknadsstatistik – fick inget stöd. De sju punkterna, som i olika grad fick positivt stöd, var:
- Revidera och uppdatera tidigare konferensers resolutioner.
- Gå igenom nya områden med avseende på tillgänglig statistik och använd metodik på det sätt som semesterstatistiken (se avsnitt 6 nedan, "Paid holidays") har studerats. En rad tänkbara områden nämndes.
- Ordna seminarier för träning av statistiker från utvecklingsländer med hjälp av statistiker från mer avancerade länder.
- Justera nationell statistik för att öka tillgången på internationellt jämförbar statistik.
- Publicera fler säsongkorrigerade serier i Bulletin of Labour Statistics och medverka till att flera nationella serier blir säsongkorrigerade.
- Göra ytterligare information om källor och metoder för nationell statistik tillgänglig.
- Studera jämförbarheten i olika länders konsumentprisindex.

2. Labour force, employment, unemployment and underemployment

ILOs statistiska byrå hade sammanställt en rapport där man grundligt och på ett förtjänstfullt sätt diskuterade och redogjorde för bakgrunden till sina förslag till definitioner av olika arbetskraftsbegrepp. I rapporten framlades också ett utkast till ny resolutionstext, som skulle ersätta den resolution rörande definition av arbetskraft, sysselsättning och arbetslöshet som gällt sedan 1954 och delar av den resolution rörande definition av undersysselsättning m. m. som gällt sedan 1966.

De viktigaste nyheterna i den föreslagna resolutionen var en distinktion mellan å ena sidan den vid ett visst tillfälle ekonomiskt aktiva befolkningen – arbetskraften – och å andra sidan den vanligen ekonomiskt aktiva befolkningen. En annan distinktion

som föreslogs var att man skulle skilja mellan den betalda arbetskraften, dvs. med vår terminologi de anställda, och den övriga delen av arbetskraften, dvs. företagare, medhjälpande familjemedlemmar och självständiga yrkesutövare. Man föreslog vidare att den relativa arbetslösheten skulle beräknas separat för dem som söker anställning. Man föreslog också två definitioner för arbetslöshet. En standarddefinition, eller begränsad definition, som bygger på de kriterier som gällt hittills, dvs. för att räknas som arbetslös skall en person vara utan arbete, vara tillgänglig för arbete och söka arbete. Vidare föreslog man en utvidgad definition av arbetslöshet, där man släpper kriteriet att en person måste ha sökt arbete.

Man framförde också förslag om hur undersysselsättning skulle definieras och mätas.

Förslagen diskuterades först under två plenarsessioner. Man fortsatte sedan granskningen av resolutionstexten i en särskild kommitté under ledning av mr B. Grais, Frankrike. Någon eller några representanter från hälften av de närvarande länderna deltog i denna kommittés arbete. Sverige representerades där av Margit Strandberg, Marianne Axling och Björn Jonzon. Den föreslagna resolutionstexten reviderades avsevärt både till form och innehåll. Ett nytt förslag förelades plenarförsamlingen som ägnade ytterligare två dagars sessioner åt denna resolution.

Endast huvuddragen i den slutliga resolutionstexten refereras i det följande:

Definitionen av den ekonomiskt aktiva befolkningen knyter an till FNs definition av produktion av varor och tjänster inom ramen för nationalräkenskapssystemet. Där poängteras att all produktion vare sig det är för försäljning, för byteshandel eller för egen konsumtion skall medräknas. De personer som medverkar i denna produktion skall alltså räknas till den ekonomiskt aktiva befolkningen.

Förslaget att skilja mellan å ena sidan den *vanligen aktiva befolkningen* och å andra sidan *arbetskraften* eller den vid ett visst tillfälle aktiva befolkningen godtogs. Den vanligen aktiva befolkningen definieras som de personer vars huvudsakliga aktivitet under ett visst antal veckor eller dagar under en lång specificerad period, som de närmast föregående tolv månaderna, varit sysselsatta eller arbetslösa. I resolutionen anger man inte närmare under hur stor del av tolvmånadersperioden som sysselsättningen eller arbetslösheten varat. Arbetskraften mäts däremot i relation till en kort referensperiod som en vecka eller en dag.

De *sysselsatta* omfattar alla personer som antingen arbetar som anställda och därvid får lön i pengar eller i natura eller arbetar för egen eller familjens vinst, dvs. är företagare, självständiga yrkesutövare eller medhjälpande familjemedlemmar utan lön. De engelska termerna är "paid employment" och "self-employment". För båda grupperna gäller att de kan vara tillfälligt frånvarande från sitt arbete. För de anställda, som är frånvarande från arbetet, specificeras att de fortfarande skall ha en formell anknytning till sitt arbete. Denna anknytning skall uppfylla ett eller fler av följande kriterier: (i) lön fortsätter att utgå, (ii) en försäkran föreligger att den frånvarande får återvända till arbetet, (iii) den förflutna frånvarotidens längd, under vilken de anställda kan erhålla bidrag utan att behöva acceptera andra arbeten. – Även företagarna m. fl. kan vara tillfälligt frånvarande från arbetet av liknande skäl som de anställda och ändock räknas till de sysselsatta.

Medhjälpande familjemedlemmar utan lön skall hänföras till gruppen företagare m. fl. oavsett hur många timmar de arbetat under mätveckan. De länder som önskar använda sig av ett minimikriterium för arbetstidens längd bör kunna redovisa det antal som arbetar ett färre antal timmar. – I den tidigare resolutionen definierades medhjälpande familjemedlemmarna utan lön som sysselsatta endast om de arbetat minst 15 timmar under mätveckan, en definition som också har tillämpats i de svenska AKU (arbetskraftsundersökningarna).

De sysselsatta inom försvaret skall omfatta både de värnpliktiga och de fast anställda. – I de svenska AKU har de värnpliktiga som inte har någon anställning hänförts till gruppen utanför arbetskraften, me-

dan de som har en anställning klassats som sysselsatta i sitt ordinarie arbete, men tillfälligt frånvarande.

För operationella syften fastslås i resolutionen att den som arbetat minst en timme under referensperioden skall räknas som sysselsatt.

För att räknas som *arbetslös* skall fortfarande som nämnts ovan de tre kriterierna att vara utan arbete, att vara tillgänglig för arbete och att söka arbete gälla. Detta är definitionen för arbetslöshet. I situationer där konventionella metoder att söka arbete har begränsad tillämpning, där arbetsmarknaden till stor del inte är organiserad eller av begränsad omfattning kan sökkriteriet frångås. I sådana fall bör man utveckla test så att man kan pröva önskan och viljan att arbeta för dem som då skall räknas som arbetslösa.

Beroende på nationella omständigheter kan man också betrakta dem som är tillfälligt permitterade utan att de sökt arbete som arbetslösa. I sådana fall bör det vara möjligt att redovisa denna grupp separat.

Man påpekar att studerande, hemarbetande eller andra personer som har annan huvudsaklig sysselsättning än förvärvsarbete och som uppfyller kriterierna ovan för arbetslöshet under referensperioden, också skall räknas som arbetslösa. I de svenska AKU räknas inte heltidsstuderande som arbetslösa även om de söker ett deltidsarbete vid sidan av studierna.

Man definierar också begreppet *undersysselsättning*. En person är undersysselsatt när personens sysselsättning är otillräcklig jämfört med vissa specificerade normer. Man skiljer på synlig resp. osynlig undersysselsättning. Den osynliga undersysselsättningen är i första hand ett analytiskt begrepp. Symptomen härpå kan vara låga inkomster, underutnyttjade kunskaper eller färdigheter, låg produktivitet.

Operationellt sett bör man begränsa den statistiska mätningen till den synliga undersysselsättningen. Man skall då skilja på följande två element: antalet undersysselsatta resp. kvantiteten undersysselsättning.

Antalet undersysselsatta består av alla sysselsatta, alltså även tillfälligt frånvarande, som ofrivilligt arbetar mindre än som är normalt, och som söker eller är tillgängliga för ytterligare arbete under referensperioden. Detta ytterligare arbete kan mätas i t. ex. antal timmar under en vecka för att få fram kvantiteten undersysselsättning.

Man rekommenderar också att man bildar en kvot mellan å ena sidan summan av den tid som de arbetslösa önskar arbete och undersysselsättningen mätt i arbetstid samt å den andra sidan den totalt tillgängliga arbetstiden som också omfattar de sysselsattas vanliga arbetstid.

Man rekommenderar vidare ett annat analytiskt begrepp som inte använts i de svenska AKU, nämligen en sysselsättningskvot. Med denna kvot mäter man andelen sysselsatta i en viss befolkningsgrupp i relation till hela befolkningen i denna grupp.

I det ursprungliga förslaget till resolution hade man relativt ingående och detaljerade rekommendationer om olika slags klassificeringar som borde göras. Denna typ av klassificeringar tas inte upp i resolutionen utan kommer i stället att tas upp i den manual för statistik över den ekonomiskt aktiva befolkningen som ILO skall utarbeta. Där skall man också ta upp metoder för datainsamling och analys. ILO skall också utarbeta klassificeringar avseende arbetad tid och arbetslöshetens varaktighet.

Man nämner också att det finns en mängd speciella frågor och ämnesområden som kan fordra särskilda och mer ingående undersökningar än de regelbundna AKU. Som exempel härpå nämns övergången från studier till förvärvsverksamhet för barn och ungdom. Man bör tidvis också ha information om familjernas och hushållens arbetsmarknadsstatus i kombination med medlemmarnas ålder, antal barn m. m.

För att få förbättrad och mer detaljerad information om arbetskraften så att man bl. a. kan belysa multipla aktiviteter bör man göra tidsanvändningsstudier. Vidare bör statistik som belyser handikappades situation vad gäller sysselsättning och arbetslöshet samlas in.

Resolutionen avslutas med några påpekanden som berör kvalitetsproblem och

publiceringspolicy.

Förslaget till resolution kommer att föreläggas ILOs styrelse för beslut. Antas resolutionen kommer den som nämnts att ersätta 1954 års resolution och delar av 1966 års resolution.

De viktigaste avvikelserna mellan de definitioner som används i de svenska AKU och de nu föreslagna internationella rekommendationerna gäller sammanfattningsvis:

- Medhjälpande familjemedlemmar utan lön, där Sverige följt den tidigare rekommendationen, men som nu skall räknas som sysselsatta oavsett arbetstid. Man har inom ramen för AKU samlat in uppgifter om dem som arbetat mindre än 15 timmar i veckan även tidigare. Under 1981 uppgick de till 10 000 personer.
- Värnpliktiga som ej har någon anställning hänförs i de svenska AKU inte till arbetskraften. De uppgick 1981 till 19 000 personer. Rekommendationen innebär att denna grupp skall klassificeras som sysselsatt och tillsammans med övriga värnpliktiga, som har en civil anställning, föras till försvaret.
- Studerande som söker arbete har i de svenska AKU inte räknats som arbetslösa, inte ens om de sökt deltidsarbete vid sidan av studierna. Effekterna av en ändring i enlighet med rekommendationerna är svår att uppskatta. En bedömning är att en sådan grupp inte skulle överstiga 12 000. Detta baseras på att under 1981 fanns det i genomsnitt 12 000 studerande som enligt AKU velat och kunnat ta ett arbete. Det förefaller troligast att det är bland dessa som man kan finna studerande som också sökt arbete.

I samband med det arbete som för närvarande pågår inom SCB med bl.a. en översyn av definitioner och begrepp i AKU kommer man att diskutera och ta ställning till om man skall ändra de definitioner som avviker från de internationella normerna.

SCB lade vid årsskiftet 1982/83 fram en utredning om hur folkräkningarna i framtiden skall utformas när de skall baseras på existerande administrativa register, eventuellt med vissa kompletteringar – den s.k. FOBALT-utredningen. Det skulle föra långt att i detta sammanhang gå igenom de olika förslag som presenterats och ange på vilka punkter man följt resp. inte följt de nu föreslagna internationella rekommendationerna. Det kan emellertid nämnas att man på flera punkter måste föreslå lösningar så att man kommer att få avvikelser från dessa rekommendationer.

3. Occupational injuries

Resolutioner med rekommendationer om statistik över arbetsskador har arbetats fram vid flera tillfällen vid ILOs arbetsmarknadsstatistikerkonferenser, bl.a. vid den första (1923) och den tionde (1962). De rekommendationer som antogs 1962 var mycket detaljerade och tog bl.a. upp förslag till olika indelningar.

Det har emellertid visat sig att av de olika grenarna av arbetsmarknadsstatistiken är den som gäller arbetsskadestatistik fortfarande den mest heterogena. Detta beror bl.a. på att arbetsskadestatistik ofta är en biprodukt från ett administrativt system, t.ex. försäkringssystem, underlag för yrkesinspektionens arbete etc. Detta i sin tur gör att man haft svårighet att presentera statistik enligt de detaljerade rekommendationer ILO gav ut 1962. Ett flertal medlemsländer har därför bett ILO att förbättra rekommendationerna.

Som en förberedelse för den trettonde arbetsmarknadsstatistikerkonferensen sände statistiska byrån vid ILO ut ett frågeformulär till medlemsländerna angående de olika ländernas arbetsskadestatistik och vilka data man använde. Svaren från 47 länder sammanställdes och presenterades som ett underlag vid ett expertmöte i januari 1980. Från Sverige deltog Elisabeth Lagerlöf som en av tre experter och hon var också ordförande vid mötet.[1] Mötet resulterade i en rapport som utgjort underlag för förberedelserna av konferensen. Rapporten tog upp ett antal problem som t.ex. heterogeniteten beroende på skillnader i nationella begrepp, definitioner, insamlingskällor och -metoder, slag av rapporterade arbetsolyckor och hur väl olika grupper täcktes av

[1] Se Statistisk tidskrift 1980:4.

statistiken. Vidare berördes underlagen för olika mått som frekvens, svårhet, incidens osv. och även underlag för och behov av relativa mått som t. ex. att många länder använder relativ incidens snarare än relativa frekvensmått beroende på svårighet att få fram uppgift om antal arbetade timmar.

ILOs statistiska byrå har därefter överarbetat rapporten och lagt fram ett förslag till ny resolution med rekommendationer som är mer flexibla och därmed skulle eliminera en del av problemen med 1962 års rekommendationer. I förslaget påpekas också behovet av att bygga upp system som var direkt användbara för arbetarskyddets förebyggande ändamål inom de enskilda länderna och som också skulle möjliggöra internationella jämförelser. I rapporten om "Occupational Injuries" framgår det att syftet är att belysa registrering av arbetsolyckor för att producera aggregerad statistik vars användbarhet för förebyggande ändamål kan förbättras om de kan jämföras såväl inom som mellan länder.

Konferensens allmänna diskussion kring rapporten inleddes med att ett flertal delegater från olika länder komplimenterade ILO för rapporten och instämde i den generella ansats som ILO hade tagit i rapporten.

Innan man diskuterade rapporten i sin helhet inbjöds länderna att beskriva sina egna system för arbetsolycksfall. Så gjorde t. ex. USA, som presenterade sitt system. Det bygger på urval av ett antal företag som fyller i färdigställda tabeller. Vid sidan av detta används alla anmälningar till Workers Compensation. Sverige presenterade sitt system, som är ett kombinerat administrativt system och förebyggande system. Schweiz och Spanien presenterade sina system, som båda är försäkringsbaserade, och England slutligen sitt, som är baserat på rapporter både till Labour Inspectorate och till försäkringssystemet. (Gäller fram till 1982, därefter enbart rapporter till Labour Inspectorate).

Vid den generella diskussion, som sedan följde, framgick dock att en rad länder skulle få svårigheter att följa rekommendationerna. Det gällde t. ex. att bryta ned olycksfallen med dödlig utgång i sådana under 30 dagar resp. upp till 1 år, där många länder ville ha ytterligare tid för att följa olycksfallen. Även hur frånvaro från arbetet skulle beräknas var ett problem för vissa länder som saknar data av detta slag. Också uppgift om arbetad tid är svårt för många länder. Mycket få länder kan ge uppgift om yrke och då speciellt relativa frekvensmått som t. ex. antal olyckor per 1 000 personer i yrket.

Några delegater kritiserade att den föreslagna rekommendationen inte i tillräcklig grad tog hänsyn till ländernas problem och speciellt frågan om olyckor som inträffade i t. ex. jordbruksarbete eller handel där bara en person eller familjemedlemmar arbetade. Några delegater gick så långt som att ifrågasätta nyttan av statistik inom detta område. På ordförandens fråga om hur statistik kan användas förebyggande, gav delegaten från USA exempel på hur statistik kan användas för att prioritera yrkesinspektionens arbete, medan det från Sveriges sida gavs exempel på hur relativt enkla statistiska sammanställningar av olyckor från skogs- och jordbruk kan användas i förebyggande syfte.

Ett annat problem, som togs upp, var hur man skulle få information om arbetssjukdomar, som yppas många år efter exponering för ett ämne, t. ex. asbest, radon etc. System, som bygger på arbetsgivares rapportering, får problem eftersom många gånger exponeringen inte skett hos samma arbetsgivare. Sambandet med arbetet blir därför oklart. Samma problem gäller urvalsundersökningar, som enbart täcker sjukdomar som är lätt igenkända eller enkelt diagnosticerbara, när urvalsundersökningen sker.

En detaljerad genomgång gjordes av de föreslagna punkterna i resolutionen. En rad förändringar gjordes i texten men innebar inga stora skillnader gentemot det ursprungliga förslaget.

Några huvuddrag i rekommendationerna redovisas nedan.

Den föreslagna klassifikationen för arbetsolyckor framgår av fig. 1 på sid. 148.

Resolutionen syftar till att möjliggöra internationella jämförelser. Den indelning

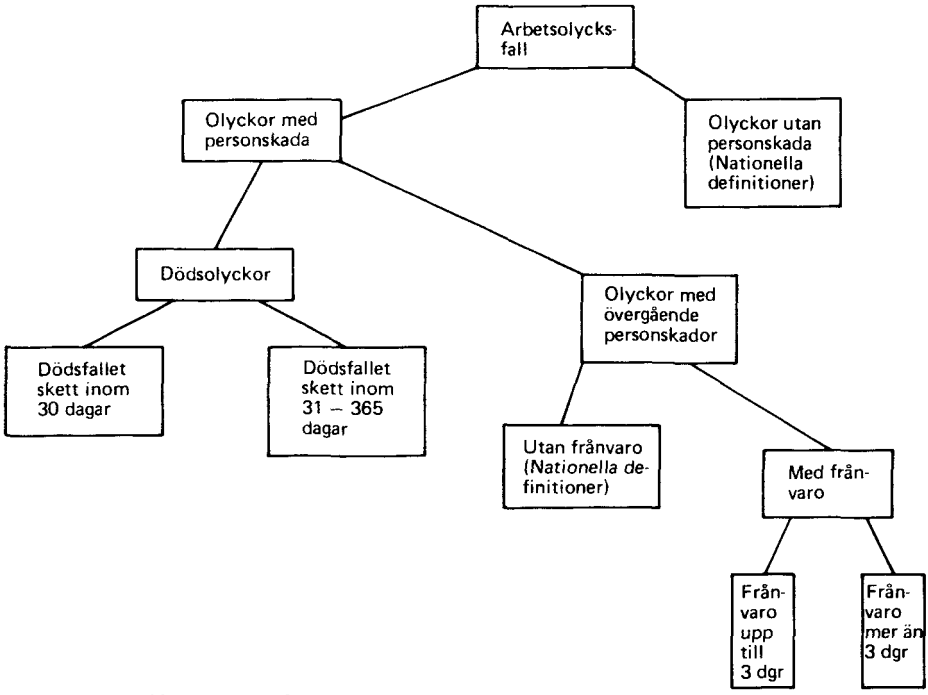


Fig. 1. Klassifikation av arbetsolyckor

som föreslås speglar detta förhållande. Uppdelningen av arbetsolyckor med dödlig utgång i två grupper (mer eller mindre än 30 dagar) dikteras av att en rad länder endast har tillgång till uppgifter av det ena slaget. Det gäller t. ex. länder, som har system som baseras på direktrapportering från arbetsgivaren, där denne har svårt att följa utgången av en olycka mer än 30 dagar. Gränsen ett år är satt för att möjliggöra en snabb statistik.

Uppdelningen av olyckor med övergående personskador i olyckor med mer eller mindre än tre dagars frånvaro beror på att en rad länder endast kan ge uppgifter om olyckor med mer än tre dagars frånvaro.

För Sveriges del, som har möjlighet att ge data av samtliga slag, torde det för nationella behov vara onödigt att använda den föreslagna klassifikationen, eftersom den enbart skulle göra statistiken mer svårförståelig och mindre användbar ur förebyggande synpunkt. Det är dessutom ganska enkelt att göra vissa specialkörningar när statistik skall levereras till ILO eller, om så önskas, i den engelskspråkiga sammanfattningen i årsstatistiken.

Ett visst problem kan uppstå vad gäller dödsolyckor, eftersom vi i Sverige har möjlighet att följa de skadade under längre tid än ett år. Som huvudregel gäller dock att de flesta skadade avlider inom en vecka från skadetillfället räknat (95 %). Under de senaste tre åren har enligt statistiken endast en olycka lett till döden efter ett år. Problemet är således ur statistiksynpunkt marginellt.

För arbetssjukdomar går inte de föreslagna reglerna att tillämpa. Så är ju inte heller meningen eftersom rekommendationen gäller arbetsolycksfall med personskador, vilket inkluderar dödliga olycksfall, olycksfall

med personskador och sjukdomar som uppstår på grund av arbetsolyckor.

4. Revision of the ILO October inquiry on occupational wages

ILO gör varje år genom en postenkät till medlemsländerna en undersökning av lönerna i ett antal yrken med oktober som referensmånad. 1980 lämnade 85 länder data. (Även detaljhandelspriser för ett antal konsumtionsvaror inhämtas, men denna del av oktoberenkäten berörs ej av revisionen och diskuterades endast marginellt av konferensen.) För närvarande omfattas 48 yrken, uttryckta i ISCO-termer, av enkäten och resultaten publiceras i Bulletin of Labour Statistics. En betydande utökning från år 1983 av antalet yrken till sammanlagt 159 i 41 olika branscher var den huvudsakliga innebörden i det förslag som förelades konferensen (Report IV).

En utökad yrkeslista efterfrågas av flera skäl och av olika grupper. Bl. a. erfordras en bredare täckning av sektorer och näringsgrenar, en bättre täckning av stora yrkesgrupper och ett bättre underlag för bedömning av lönesituationen för kvinnor och män. Uttalanden som stöder detta har gjorts av ILOs olika branschkommittéer och liknande organ, i ILOs World Employment Programme och i förberedelserna till FNs kvinnodecennium.

Diskussionen tog upp frågor om ändamålet med oktoberenkäten, som ansågs något oklart, mätmetoder och definitioner, tillgång till och kvaliteten i nationella data särskilt med hänsyn till internationell jämförbarhet. Många länder stödde huvuddragen i revisionsförslaget medan andra var tveksamma till värdet av enkäten. Från svensk sida togs bl. a. upp kvalitets- och jämförbarhetsaspekterna och anfördes att de borde prioriteras i det fortsatta arbetet. Samtidigt konstaterades att Sverige i stort sett skulle kunna leverera data också enligt en utökad yrkeslista.

En särskild arbetsgrupp beredde vissa frågor ytterligare och kom fram till vissa slutsatser, vilka konferensen sedan i sin tur kunde ansluta sig till. Bl. a. ansågs att arbetet med att se över och bygga ut oktoberenkäten skulle fortsätta, men att det också var nödvändigt att – på grund av risken för att data missförstås – samla in och publicera uppgifter om innebörden av olika länders lönedata.

ILO borde vidare utarbeta en katalog över i olika länder tillgänglig löne- och förtjänststatistik och konsumentprisstatistik med detaljerad information om statistikens uppbyggnad och innehåll, använda metoder, definitioner, begrepp och annan information av betydelse för förståelsen och användningen av data. ILO borde också arbeta med att öka den internationella jämförbarheten i publicerade data.

I arbetsprogrammet för statistiska byrån borde, på grund av stora variationer mellan länderna vad gäller tillgänglighet och innehåll i data på detta komplicerade område, det fortsatta arbetet här prioriteras med sikte på att ta upp området för behandling vid nästa arbetsmarknadsstatistikerkonferens. Begränsade resurser var dock ett hinder för alltför omfattande insatser.

5. International coding of labour statistics

I rubricerade rapport (Report V) presenterades ett förslag till internationellt kodsystem för arbetsmarknadsstatistiken. Tanken var att statistiska tabeller inom området skulle karakteriseras i olika avseenden med en numerisk kod. Härigenom skulle man lättare hitta de tabeller och den information man sökte, särskilt om systemet datoriserades. Det internationella systemet skulle vara ganska grovt och kunna tjäna också som ram för mer detaljerade, nationella system. De olika komponenterna i det föreslagna kodschemat var Member Country, Subject, Categories, Quantifiers och Timing, och varje komponent föreslogs kodas med 1 till 3 siffror.

Diskussionen gav ett starkt stöd för tanken att få ett system att lättare hitta i statistiken. Det system för arbetsmarknadsstatistiken som skisserades ansågs av många vara ett steg i rätt riktning även om åtskilliga tekniska detaljer återstod att lösa. Från svensk sida stöddes också idén, men den föreslagna utformningen ansågs av svenska experter ha en alltför rigid och hierarkisk

struktur. En mindre strukturerad modell rekommenderades, för vars utarbetande olika länders erfarenheter på datalogiska området och från andra försök med metainformation i statistiken borde kunna utnyttjas.

Från flera länder, däribland Sverige, men inte minst från representanterna för andra internationella organisationer underströks vikten av samarbete och samordning med liknande arbeten inom andra statistikområden. Stora delar av det föreslagna schemat hade nämligen en sådan generell karaktär att de lätt kunde anpassas och utnyttjas på ett bredare område. Den komponent vars innehåll var mest specifikt för arbetsmarknadsstatistiken var "Subject" (variabelområden) och en framförd idé var att ILOs arbete skulle koncentreras dit. Samtidigt var det viktigt att stödja det tagna initiativet, som också inrymt kontakter med andra internationella organ, och uppmuntra ett fortsatt tänkande på området.

ILOs statistiska byrå skulle, med det intresse och stöd konferensen visat, arbeta vidare med ett kodschema. De synpunkter man fått skulle utnyttjas och man räknade med fortsatt experthjälp i arbetet, t. ex. genom inhemska seminarier kring nya förslag och genom översändande av intressanta papper. Något särskilt möte i t. ex. en arbetsgrupp ansåg man däremot inte att det fanns resurser till. Man hoppades kunna komma fram med en mer genomarbetad rapport och lämpliga rekommendationer till nästa arbetsmarknadsstatistikerkonferens.

6. Paid holidays

Med "Paid holidays" – betalda lediga dagar – avsågs såväl betalda allmänna helgdagar o. d. som betald semester. Report VI med denna rubrik var en sammanfattning av statistiska byråns kartläggning av tillgången på nationell statistik på detta område och de metoder och begrepp som användes. Ett antal länder, däribland Sverige, hade bidragit.

Flera skäl till att ha statistik över betalda lediga dagar anfördes i rapporten. Det första var behovet att belysa ett viktigt, positivt inslag i arbetsförhållandena, som i ökad utsträckning bidragit till höjd levnadsstandard. Det andra var att data behövdes för denna avdragspost vid beräkningen av total årsarbetstid, viktig bl. a. för produktivitetsanalyser. Vidare gav statistiken ett bidrag till kalkylerna av arbetskraftskostnader. Statistiken hade således både en social och en ekonomisk aspekt.

Kartläggningen visade enligt rapporten en stor spridning mellan länder – och olika ländertyper – i såväl tillgången på som utformningen av statistiken. Det konstaterades också att statistik över helgdagar o. d. betalda eller obetalda, inte var huvudproblemet utan semesterstatistiken och rapporten koncentrerades till detta område. De källor som i varierande utsträckning utnyttjades i statistiken i olika länder var data från kollektivavtal, lagstiftning etc., arbetsgivarenkäter samt hushålls- och individundersökningar.

Konferensen uttryckte sin uppskattning över initiativet att studera de internationella aspekterna på statistiken över lediga dagar. För- och nackdelar med olika metoder att insamla data diskuterades. Olika begrepp och definitioner diskuterades också med utgångspunkt i ländernas erfarenheter. Det betonades samtidigt vikten av att ha data även över annan betald ledighet.

Sammanfattningsvis konstaterades utifrån den gjorda studien och den förda diskussionen att det för närvarande var svårt att sammanställa och publicera internationellt godtagbara data över betald semester och helgdagar. Statistiska byrån skulle, mot bakgrund av det intresse för området som finns både på ILO och på nationellt plan, i första hand arbeta med att samla och sammanställa information om använda begrepp, definitioner och metoder i olika länder i någon form av metadatabas. En första åtgärd skulle bli att se över och komplettera konferensrapporten (Report VI) och publicera den som en metodartikel i Bulletin of Labour Statistics.

7. Avslutning

Från konferensdeltagarna uttrycktes ett önskemål om att det inte skulle dröja alltför länge till nästa arbetsmarknadsstatistikerkonferens. De nio år som gått sedan förra

konferensen ansågs vara alltför lång tid och man föreslog att nästa skulle äga rum senast 1987. Generalsekreteraren skulle vidarebefordra konferensens uppfattning till ILOs beslutande organ och han hoppades att bl. a. knapphet på resurser inte skulle hindra att 14e arbetsmarknadsstatistikerkonferensen avhålls 1987.

Marianne Axling *Claes Lagerkvist*
Elisabeth Lagerlöf *Margit Strandberg*

Möten om kulturstatistik

Inom ramen för Unesco's expertmöten för kulturstatistik och -indikatorer har under hösten 1982 hållits möten i (1) Washington D.C. och (2) Paris. Från Sverige deltog Leif Gouiedo, SCB i bägge sammankomsterna på inbjudan av Unesco-kommissionen i USA resp. Unesco (Office of Statistics).

(1) Mötet i *Washington* hölls den 28 september–1 oktober 1982 med den arbetsgrupp som har i uppdrag att vidareutveckla ett av Unesco (i samarbete med Europeiska ekonomiska kommissionen, ECE) utarbetat förslag till ett kulturstatistiskt ramverk.[1] Föremålet för gruppens tillämpningsstudier är kulturområdena film, radio och TV. (Av tablån på sid. 152 framgår vilka grupper av deltagarländer som arbetar med andra kulturområden inom ramverket.)

Förutom Sverige ingår i arbetsgruppen Kanada, USA och Finland. Detta var gruppens tredje möte[2], där det främsta syftet var att diskutera erfarenheter och konsekvenser av de synpunkter på det förslag till indikatorer inom film[3], radio/TV som gruppen tidigare arbetat fram. Förslaget hade "remitterats" till skilda intressenter inom var och en av de fyra deltagarländerna. För Sveriges del hade i cirkulärskrivelse begärts yttranden från ett 30-tal myndigheter och organisationer (av vilka 24 svarat), som förmodats kunna tillföra information av betydelse för en förbättring av förslagen. I Finland hade man först kontaktat intressenterna per telefon, varefter man sänt förslaget för kommentarer. (Omkring 20 organisationer hade kontaktats, varav ett 15-tal lämnat synpunkter.) Kanada och USA har använt i stort sett samma ansats med initialkontakter per telefon och därefter utsändning av förslaget. Här skall inte redogöras för "remissinstansernas" kommentarer – det framgår av gruppens mötesrapport (ingår som nummer 35 i Unescos serie Current Surveys and Research in Statistics for Culture, CSR-C). För gruppens fortsatta arbete ledde diskussionerna bl. a. till att

a. den amerikanska representanten åtog sig att utarbeta förslag till ett klassifikationssystem för filmgenrer, huvudsakligen byggande på scheman som tillämpats inom (i) International Federation of Film Archives; (ii) USAs Film Archives Committee;
b. den kanadensiska representanten skall "ta fram" förslag till definitioner för de skilda begrepp och mått som gruppen föreslagit för användning inom sina kulturområden.
c. de finländska och svenska representanterna skall samfällt skriva en rapport med förslag till lämpliga klassificeringar av program- och ämnesinnehåll i radio/TV (innebärande eventuellt en vidareuppspaltning av vissa programkategorier – bl. a. kulturprogram, musik – i Unescos schema).

(2) *Parismötet* den 7–10 december 1982 var förlagt till slottet Bois du Rocher (ca 3 mil utanför huvudstaden). Här samlades medlemmar i den styrgrupp, bildad av Unesco för att utveckla allmänna och operativa riktlinjer samt koordinera verksamheten i de grupper som arbetar med F.C.S.-systemets skilda kulturområden. Detta var gruppens andra sammankomst, nu utvidgad med representanter från tio länder och en observatör från ECE. För närvarande är följande studiegrupper verksamma:

[1] "Framework for Cultural Statistics" (F.C.S.), som består av tio kulturområden, se Statistisk tidskrift 1981: 6, sid. 453, not 1.

[2] Statistisk tidskrift 1981: 4, sid. 297 f. och 1981: 6, sid. 453 ff. innehåller bakgrundsteckning och korta meddelanden från de två föregående mötena.

[3] Se exempel i Statistisk tidskrift 1981: 6, sid. 454.

Grupp	Kulturområden	Länder (motsv.) representerade
A	Kulturminnesvård; tryckta medier	Belgien; Frankrike; Rumänien; Holland; Norge
B	Musik; dramatisk konst	Italien; Österrike; Danmark; Ungern
C	Film; radio o. TV	Kanada; Finland; Sverige; USA
D	Socio-kulturella aktiviteter	Schweiz; Ungern; Nord. kultursekretariatet

(Ytterligare två grupper, E avseende Bildkonst och F Sport och spel; Natur och miljö beräknas kunna starta under 1983.)

Inför mötet hade Unesco-sekretariatet skrivit en bakgrundsrapport, innehållande en översikt över problem och förslag som från arbetsgrupperna överlämnats till styrgruppen för behandling. För att strukturera problemen till punkter på en dagordning "konstituerades" en kommitté på fyra personer med en representant från ett av länderna inom varje studiegrupp. Italien, Österrike, Sverige och Ungern valdes att ingå i kommittén; ledamöterna satte upp följande punkter på mötesordningen. Resultatet av diskussionerna redovisas.

Information: För att öka kunskapen om arbetet med F.C.S. skall Unesco göra en informationsbroschyr (målgrupp: myndigheter, organisationer i länder som är representerade i arbetsgrupperna). Vidare skall löpande information med lägesbeskrivningar kanaliseras på annat sätt. Här nämndes som ett lämpligt medium den av ECE nyligen startade The Statistical Journal.

Fasetter: Styrgruppen fäste uppmärksamheten på en av grundprinciperna för F.C.S.-systemet, nämligen att detta skall vara enkelt, använda ett fåtal dataelement och ge information som underlag för nationell kulturpolitik, samtidigt som det skall möjliggöra internationella jämförelser av nationella kulturkarakteristika. Allt detta förutsätter att utbildningsmässiga, sociala och ekonomiska aspekter – fasetter – bör beaktas. I tillägg, menade styrgruppen, skall man speciellt uppmärksamma artisters och andra kulturarbetares roll, förhållanden och karakteristika.

Systemkopplingar: Styrgruppen önskar att F.C.S. skall utvecklas mer i detalj och med större tyngd på det kulturella än vad som vore sannolikt ifall man enbart skulle se till resultat av det arbete som andra internationella grupper håller på med på det kulturella området. Å andra sidan är det önskvärt att man skapar kopplingar mellan systemen där så är möjligt. Lämpliga Unescodefinitioner och klassifikationer bör beaktas. Likaså bör tas tillvara arbetsresultat inom andra internationella organ – som exempelvis FN-projektet "Framework for the Integration of Social and Demographic Statistics" (F.S.D.S.).

Funktioner: Styrgruppen bedömer att den funktionella dimensionen i F.C.S.-matrisen (skapande, produktion, spridning, distribution etc.) tills vidare bör hållas flexibel till dess samtliga sex studiegrupper har "spikat" den funktionella indelningen genom förslagen till statistik och indikatorer för "sina" kulturområden. Först då är det lämpligt att styrgruppen ser övergripande på ramverkets funktionsstruktur.

Dataslag: Till existerande och kommande studiegrupper rekommenderar styrgruppen att använda den ansats som C-gruppen tillämpat för att bestämma vilka slags data som bör ingå i F.C.S. avseende film, radio/TV, dvs. att först komma fram till de ideala (från "teoretisk" synvinkel sedda mest önskvärda variablerna), därefter besluta måttenheter för dessa. Det slutliga valet av mått och begrepp kräver också att definitioner och klassifikationer bestäms.

Förväntan och tillfredsställelse: Styrgruppen rådde varje studiegrupp att fundera kring begreppen förväntan (expectation) och tillfredsställelse (satisfaction) för att därigenom komma fram till en bedömning av om det finns lämpliga mått för dessa "variabler" och, ifall så är fallet, utveckla erforderliga rekommendationer.

Amatörer: Dessa spelar en mycket viktig roll i kulturpolitiska sammanhang och det är därför av stor betydelse att i F.C.S.-systemet få en statistisk beskrivning av amatörverksamheten. I styrgruppen kände man dock inte till existensen av någon enhetlig

definition av begreppet amatör. Arbetsgrupperna bör därför noga studera amatöraktiviteterna inom varje funktion (skapande etc.) och ge meningsfulla och avgränsande beskrivningar för studerade kulturområden.

Nästa möte med styrgruppen avses hållas i januari 1984. En Unesco/ECE-konferens om kulturstatistik och -indikatorer[1] planeras till 1984/85. (Tentativa ämnen på dagordningen: klassifikationsproblem, lägesrapportering från studiegrupperna, etablering av en koppling mellan F.C.S. och F.S.D.S.)

Leif Gouiedo

Möte om kulturstatistik och seminarium om kulturindustri

Nordiska ministerrådets kultursekretariat, med säte i Köpenhamn, anordnade den 6–8 oktober 1982 ett seminarium om kulturindustri och nordisk kulturpolitik. Upprinnelsen till seminariet var Nordiska Rådets rekommendation till kultursekretariatet att undersöka kulturkonsumtionen i de nordiska länderna med hänsyn till den internationella masskulturindustrins påverkan. Seminariet hölls dels i kultursekretariatets lokaler, dels på Örenäs slott vid Landskrona. Före denna konferens höll kultursekretariatet ett möte i Köpenhamn den 1 september 1982 med representanter från de statistiska centralbyråerna i Norden – varvid diskuterades möjligheterna att få fram statistiska uppgifter till seminariet samt riktlinjer för arbetet med att sammanställa uppgifterna. Från SCB deltog Leif Gouiedo i bägge sammankomsterna.

Till seminariets deltagare kunde sekretariatet distribuera de av gruppen "nordiska kulturstatistiker" sammanställda uppgifterna kring tre teman med bäring på kulturindustri, nämligen *musik* (försäljning av fonogram, hushållens innehav av skiv- och kassettbandspelare, befolkningens aktiva musikutövande [körsång, spelning i orkester, eget musicerande]), *TV* (licenser, sändningstid, programinnehåll, "tittartid") och *dagspress* (antal, nettoupplaga). På förhand hade sekretariatet också sänt ut två "beställningsverk": (a) en skrift av Harry Schein, Kulturindustrins negativa verkningar, (b) en bok av Peter Duelund, Tilskuer eller deltager? Bägge publikationerna tar upp frågan om kulturstatistik. Schein diskuterar i första hand behovet av sådan statistik som kan ge informationer om kulturindustrins, företrädesvis den internationella industrins utbud. Han föreslår i korthet bl. a. följande åtgärder:

- Kulturstatistisk expertis redovisar nordisk kulturstatistik, anger på vilket sätt den skiljer sig mellan de nordiska länderna, bedömer om det finns praktiska fördelar med en komplettering av statistiken i de olika länderna resp. en samordning mellan länderna och beräknar kostnaderna för en sådan komplettering resp. samordning.
- Statistisk expertis undersöker med representanter för film, musik och litteratur möjligheterna att finna på samma gång enkla och meningsfulla värderingsprinciper som kan tillföra kulturstatistiken kvalitativa informationer – samt behovet av och kostnaderna för en sådan komplettering.
- Tillgängliga data rörande kultur- och fritidskonsumtionen (alltså inte utbudet) inventeras. Särskilt angelägna behov till komplettering av sådana data anges.
- Möjligheterna att punktbelysa samspelet mellan kulturindustrins agerande, kulturinstitutionernas verksamhet och kulturmönstret diskuteras, liksom kulturpolitikens möjligheter att påverka detta samspel.
- Sådana kulturpolitiska åtgärder som haft som uttryckligt syfte att påverka kulturmönstret inventeras. Metoder att utvärdera dessa åtgärder, bl. a. genom jämförande studier mellan de nordiska länderna diskuteras.

Duelund å sin sida vill ha kulturstatistik som

– kan tillföra kvalitativa informationer om produktion, distribution och konsumtion

[1] Den tredje i ordningen – den närmast föregående hölls i april 1979, se Statistisk tidskrift 1979: 4, sid. 303 f.

av film, musik, litteratur, konst, fonogram och videogram med avseende på aspekten "demokratisering av kulturen"
- kan ge kvalitativ och kvantitativ information om aktiviteter ur aspekten "den kulturella demokratin", samt
- belyser relationerna mellan dessa båda aspekter och deras samhälleliga betydelser.

Dessa krav på statistiska uppgifter är i vissa fall måhända tämligen lätta att tillgodose om man har tillräckligt med resurser av olika slag för statistikproduktion att tillgå. I andra fall är frågan mer komplicerad – det gäller exempelvis statistik om fonogrammarknaden, som domineras av multinationella företag, vilka av branschmässiga skäl ofta är obenägna att lämna ut uppgifter om produktion etc.

Av vad som ovan återgetts framgår att det finns flera områden som måste få en statistisk belysning så att det finns uppgifter att tillgå som underlag för beslut i nordisk kulturpolitik vis á vis kulturindustrin. Kulturstatistiker inom Norden får nu försöka driva ett sådant utvecklingsarbete vidare, i första hand inom ramen för det nordiska utskottet för utbildnings- och kulturstatistik (NUUK), i samarbete med bl. a. kultursekretariatet, utskotten för levnadsförhållanden och tidsanvändning (NUUL) samt socio-demografisk statistik (NUSD).

Leif Gouiedo

Internationella havsforskningsrådet (ICES)

Internationella havsforskningsrådets 70e årsmöte hölls i Köpenhamn 11–20 oktober 1982. Rådets statistiska kommitté sammanträdde den 13 oktober under ordförandeskap av H. Lassen från Danmarks Fiskeri- og Havsundersøgelser. Representanter för 18 medlemsländer deltog i mötet, där även observatörer från FAO, OECD och Eurostat var närvarande. Från Sverige deltog Bengt Lindfors, SCB.

De frågor som behandlades vid statistikkommitténs möte hade förberetts av en särskild arbetsgrupp vid ett sammanträde 9–11 juni 1982, i vilket Lindfors och Rune Lind, fiskeristyrelsen, deltog.

Kommittén godkände arbetsgruppens förslag att endast mycket tungt vägande skäl skulle föranleda ändringar i gränsdragningen mellan de olika statistiska redovisningsområdena (fångstområdena). Uppdelning på mindre områden inom tidigare uppdragna gränser skulle i första hand övervägas.

Vidare beslöts att den årliga rutinmässiga rapporteringen från varje medlemsland rörande viktigare ändringar i metodik för insamling och redovisning av fångstdata skulle slopas. I förekommande fall skall viktigare förändringar dock alltid rapporteras.

Statistikkommittén uppmanar medlemsländerna att redan i den preliminära rapporteringen av årsfångsterna lämna så fullständiga uppgifter som möjligt, eftersom rapporterna utgör viktiga underlag för beståndsuppskattningen. Å andra sidan bör alltför detaljerad redovisning undvikas; ICES databank är redan i alltför hög grad belastad med records som innehåller fångster på mindre än ett halvt ton. Sådan information ansågs oanvändbar för nämnda syfte.

Till ny ordförande för kommande treårsperiod valdes Kathryn Paine från Oceanografic Laboratory, Woods Hole, Mass., USA.

Sverige står som värdland för nästa årsmöte, som äger rum i Göteborg 10–19 oktober 1983.

Bengt Lindfors

Kommitté för yrkeskod för statistiker (ISI)

Den arbetsgrupp som tillsattes inom Internationella Statistiska Institutets kommitté för yrkeskod för statistiker fortsätter sitt arbete i enlighet med fastställd plan. Ett första komplett förslag kommer således att presenteras för ISI i september 1983 i Madrid då institutet samlas till sin ordinarie session. Som ett led i sitt arbete sammanträdde arbetsgruppen den 20–23 oktober

1982 i London för genomgång av det av medlemmarna framtagna underlaget. Det kunde därvid bl.a. konstateras att arbetet fortskridit planenligt men att frågeställningarna ofta är svårbehandlade och intrikata. Värdefulla synpunkter har inkommit från en del medlemmar på ett till dessa utsänt utkast.

I samband med London-besöket företogs också ett besök hos ett multinationellt datakonsultföretag i och för diskussion av ett visst samarbete inom ramen för SCBs uppdragsverksamhet.

Edmund Rapaport

Konferens om den aktuella befolkningsutvecklingen i Europa

The British Society for Population Studies anordnade den 16–18 september 1981 ett möte om befolkningsutvecklingen i Europa. Mötet hölls på universitetet i Exeter. Deltagarna kom från ett drygt tiotal länder i såväl Väst- som Östeuropa samt Nordamerika. Från Sverige deltog byrådirektör Peter Springfeldt, SCB.

Mötet var upplagt på sådant sätt att första sessionen ägnades åt en genomgång av födelsetalens utveckling i Väst- och Östeuropa. Den andra och tredje sessionen behandlade huvudsakligen orsaker bakom de ändrade födelsetalen. Medan, slutligen, den fjärde sessionen tog upp de befolkningspolitiska reaktionerna på befolkningsutvecklingen.

Som en utgångspunkt för forskningen om de sjunkande födelsetalen ligger det faktum att trenderna i födelsetalen och det trendbrott med sjunkande födelsetal som skedde 1964–65 ser förbluffande lika ut i hela Europa. Det visades t.ex. att samma trender och trendbrott som finns på nationell nivå i Europa även finns i olika italienska regioner. Detta är oberoende av om regionen ligger i det industrialiserade norra Italien eller i det mer rurala södra Italien. Ett annat trendbrott i Europa var att från 1975 började medelåldern för förstagångsmödrar öka.

Forskare från Östeuropa menar att det i de länderna har skett en förändring i den traditionella kulturen. Numera föregås en graviditet i allt större utsträckning av ett gemensamt beslut av makarna. Vad gäller s.k. differentiell fruktsamhet så har man funnit ett mönster som gäller för hela Östeuropa, nämligen att kvinnor med hög utbildning har få barn. Vid sessionen som behandlade befolkningspolitik rapporterade forskare från Östeuropa, där man ju en längre tid fört en pronatalistisk politik, att enligt deras bedömning har olika pronatalistiska åtgärder bara påverkat födelsetalen under en begränsad period av ett par år. Rapportören av befolkningspolitiken i Västeuropa inledde sitt anförande, och sin rapport, med att citera en boktitel av E. M. Remarque: "På västfronten intet nytt".

Peter Springfeldt

Konferensen "NordDATA 82"

NordDATA 82 anordnades i Göteborg den 21–23 juni 1982. Huvudarrangör var Svenska Dataföreningen och Svenska Samfundet för Informationsbehandling.

Syftet med evenemanget är att datoranvändare och systemkonstruktörer från näringsliv och förvaltning möts tillsammans med forskare och lärare från universiteten. Erfarenheter och idéer presenteras och diskuteras.

Konferensprogrammet bjöd på ca 150 föreläsningar inom 15 olika ämnesområden. Ungefär en femtedel av programmet behandlade icke datatekniska ämnen. Detta visar bl.a. att databranschen fått upp ögonen för behovet att diskutera och ta hänsyn till omvärlden utanför den slutna värld som dataavdelningar länge varit i många företag och organisationer.

År 1983 kommer NordDATA att arrangeras i Oslo.

Lars Berglöf

Arkivmöten

Inom International Council on Archives (ICA) som är en sammanslutning av arkiv och arkivarier i samarbete med UNESCO finns en särskild ADB-kommitté. Denna

kommitté sysslar med frågor som rör omhändertagandet av maskinläsbara data, användningen av datorer inom arkiven, utbildningen av arkivarier i ADB m. m. Den 9–12 november 1982 ägde den årliga sammankomsten rum i Toledo, varvid 15 komittémedlemmar från 12 olika länder träffades för att diskutera under året bedrivna arbetet, förslag till nya arbetsuppgifter m. m.

Ett stort arbetsområde är det som rör bevarande av maskinläsbara data, dvs., uppgifter på magnetband, skivminnen osv. Inom detta område behandlades frågor som rör sådana datas legalitet, kriterier för bevarandet resp. gallring, dokumentation, vård, kostnader i samband med användandet m. m. Kommittén samlar även in uppgifter om användningen av ADB i arkiven. Tanken är att publicera användningsområde, användare m. m. så att arkiv som funderar på att införa ADB-rutiner skall kunna ta del av andras erfarenheter. Resultaten från kommitténs arbete publiceras bl. a. i den tidskrift, ADPA, som ges ut några gånger per år och i en särskild skriftserie.

Den 14–18 september 1981 besökte jag IASSISTs årskonferens i Grenoble. IASSIST, (International Association for Social Science Information, Service and Technology) är en sammanslutning av personer och institutioner som arbetar med framför allt maskinläsbara primärdata inom det samhällsvetenskapliga området. Konferensen behandlade frågor som rör datoriseringens inverkan på den samhällsvetenskapliga forskningen. Olika teman behandlades i ett stort antal föreläsningar och diskussioner. Det som var av särskilt intresse för mig som arkivarie vid SCBs arkiv var de föreläsningar som behandlade dokumentationen till maskinläsbara data. Sammandrag av föreläsningarna finns samlade i en volym som finns att tillgå i SCBs arkiv.

Birgit Fredberg

Programmeringsmöte i Östberlin

Vid detta möte som hölls i Östberlin 20 september–1 oktober 1982 var utgångspunkten för diskussionerna en PM från Östtyskland, Ungern och Sverige. Det konstaterades att Sveriges förslag till samarbetsprojekt även täckte de problem som diskuterades i Östtyskland och Ungern. Från svensk sida deltog i mötet byrådirektör Eyvind Björk, SCB.

Det Svenska förslaget (Base Operators in Statistical Computing) diskuterades och några modifieringar gjordes till förslaget.

Basoperatorn SORT valdes som en första uppgift för gruppen. Huvudskälen till att välja Basoperatorn SORT är:

- Det är bäst att börja med en enklare uppgift.
- Utvecklingsgruppen måste samtidigt utveckla en acceptabel programmeringsstandard för fortsatt arbete.
- De utländska deltagarna måste vänja sig vid att använda de verktyg för programmering som vår dator installation tillhandahåller.
- Gruppen har som mål att presentera en fungerande komponent efter den första testfasen som äger rum hos SCB.

Basoperatorn SORT utformades och indelades i moduler. Modulerna programmerades i pseudokod och varje modul diskuterades.

Måndagen den 10 januari 1983 valdes som startdatum för den första testfasen. Tre veckor bedöms som nödvändigt för testfasen.

SCB ska invitera deltagarna från Ungern och Östtyskland.

Arbete som måste vara klart till 10 januari 1983:

Följande moment görs av Östtyskland:

- Koda och kompilera modulen BSSDRIV – Driver för Basoperatorn SORT.
- Koda och kompilera modulen BSSANAL – Analysera kontrollsatser.
- Koda och kompilera modulen BSSMSG – Skicka ett meddelande.
- Koda och kompilera modulen GSMSG – Skicka ett meddelande.
- Designa Basoperatorerna SELECT och AGGREGATE för vidare diskussioner.

Följande moment görs av Ungern:

- Koda och kompilera modulen BSSMD – Behandla metadata.
- Koda och kompilera modulen BSSCTAB

– Skapa metadata för output.
- Designa Basoperatorerna MINUS och UNION för vidare diskussioner.

Sverige, slutligen skall bidra med följande:
- Koda och kompilera de subrutiner som används allmänt i Basoperatorerna.
- Ändra design av JOIN och PROJECT så att de överensstämmer med den standard som accepterades vid planeringsmötet i Berlin.

Fördelning av arbetet med nästa deluppgift skall göras under testfasen i januari i Stockholm.

Eyvind Björk

Statistical Computing Project: seminarium i RAPID-frågor

Statistical Computing Project är ett projekt som startats av ECE med stöd av UNDP. I projektet ingår som en av fyra undergrupper RAPID Joint Group, som ägnar sig åt RAPID, en databashanterare utvecklad vid Statistics Canada. Gruppen har Sverige som s.k. "leading country". Den 4 och 5 oktober 1982 höll RAPID Joint Group ett möte i Ottawa i anslutning till ett öppet seminarium som arrangerades av Statistics Canada den 6–8 oktober 1982 i Ottawa. Sveriges representant var Märit Leringe, SCB.

RAPID Joint Group's möte leddes av Sverige och hade deltagare från Kanada, Italien, Storbritannien och Ungern. Närvarande var också SCPs projektkoordinator och tekniska rådgivare från Genève resp. Bratislava. Diskussionerna berörde bl.a. tekniska problem kring distributionen av RAPID och den handbok som skall bli ett resultat av arbetet inom gruppen.

På seminariet under de nästföljande tre dagarna tillkom sex personer från Bureau of Labor Statistics (BLS) i Washington samt ytterligare personer från Statistics Canada och andra statliga institutioner i Kanada. Även Brasilien skulle ha varit representerat, men gav återbud i sista stund. Vid seminariet presenterade Kanada flera av sina RAPID-applikationer och gav information om den nya version av RAPID som kommer under våren 1983. BLS beskrev en förbättring av kopplingen mellan RAPID och TPL, en kombination som de uppenbarligen använder flitigt. De visade också hur RAPID använts för en periodisk undersökning, där mängden variabler varierade mellan gångerna. Sverige och Ungern presenterade applikationer och en beskrivning av RAPID Joint Group's verksamhet.

Alla arrangemang kring mötet och seminariet, inte minst de sociala, var mycket trevliga. Detta bidrog till att diskussionerna kunde föras under lättsamma och konstruktiva former.

Märit Leringe

COM-tekniken vid Norges Sentralbyrå

Den 11 november 1982 gjorde undertecknad ett studiebesök vid Norges Statistiska Sentralbyrå i Oslo. Under besöket diskuterades COM-tekniken. Fördelar resp. nackdelar med utrustning "online" resp. "inhouse" vägdes därvid mot varandra. Med på besöket var representanter från Kodak, vilka levererat den utrustning som används av sentralbyrået. Jag fick även se en demonstration av utrustningen.

Lennart Österberg

SCB hos Census Research Group

Hur skall vi göra för att med allt större framgång marknadsföra våra statistikprodukter? Det är en aktuell fråga bland statistikproducenter på olika håll i världen. Hur just SCB säljer statistiken från 1980 års folk- och bostadsräkning fick jag själv möjlighet att redogöra för som inbjuden föredragshållare vid den engelska Census Research Group – en sammanslutning av statistiker/statistikanvändare från bl.a. universitet, centrala och lokala myndigheter och näringsliv.

Föredraget hölls i samband med organisationens årsmöte den 10 december 1982 i the County Hall i London. Av de omedelbara reaktionerna och senare tackbrev framkom att SCBs problem inom området

också har blivit aktuella för statliga statistikproducenter som t. ex. Office of Population Censuses and Surveys. Man var intresserad av att även framöver hålla kontakten för att informera varandra om lämpliga säljåtgärder och hyggliga säljresultat.

Knapphetens stjärna lyser som bekant även över Englands officiella statistikproduktion. Man har där, precis som hos oss, goda skäl att öka uppdragsverksamheten.

För att utnyttja tiden så bra som möjligt passade jag på att göra en del studiebesök under Londonveckan, 6–10 december 1982. En heldag ägnades åt Business Statistics Centre i Newport, som ligger en och en halv timmes färd med snabbtåg från London. Här ägnades samtalen åt marknadsfrågor, publikationer och bibliotekstjänst.

Jag besökte även Central Statistical Office och min "counter-part" där. Diskussionerna kretsade främst kring olika former av samarbete med massmedia. En sådan form är Viewdata (datavision). Om detta visste man också en hel del vid Central Office of Information. Där fick jag tillfälle att ta del av den senaste utvecklingen inom området: Viewdata har inte visat sig bli det som man en gång hade räknat med.

Central Office of Information – nästan ett "informationsministerium" – är, vid sidan om de statistiska myndigheterna, särskilt matnyttigt för en samhällsinformatör att hälsa på. En halv dag ägnades åt en lärorik genomgång av en rad varierande, pågående och planerade samhälleliga informationskampanjer.

Mina engelska vänner hade också ombesörjt att jag under en eftermiddagstimme kunde följa en nog så hetsig debatt mellan premiärministern, Mrs. Margaret Thatcher och oppositionsledaren Mr. Michael Foot, från Strangers Gallery of the House of Commons. Men det är ju en annan historia ...

Lennart Törnqvist

Nya metoder för folk- och bostadsräkningar

Statistiska centralbyrån överlämnade i januari 1983 till regeringen slutrapporten från ett utredningsarbete om alternativa metoder för framtida folk- och bostadsräkningar (FOBALT). Enligt regeringsdirektiven har FOBALT undersökt olika metoder för att framtida folk- och bostadsräkningar skall kunna genomföras utan insamling av uppgifter från allmänheten via blankett.

I Danmark beslöt statsmakterna redan i mitten av 1970-talet att bygga upp ett system med registerbaserad statistik, i vilken även registerbaserade folk- och bostadsräkningar ingår. Den första FoB av denna typ sammanställdes 1981. I Norge och Finland pågår utredningsarbeten motsvarande FOBALTs med sikte att skapa registerbaserade ersättningar för FoB.

FOBALTs slutrapport, som består av en huvudrapport och en detaljrapport remissbehandlas för närvarande (våren 1983) av ett 80-tal instanser. Remisstiden löper ut den 1 juni 1983. Statistisk tidskrift avser att senare återkomma med en utförligare presentation av de nya FoB-metoderna, om och när dessa blivit beslutade av statsmakterna. I det följande lämnas en sammanfattning av den metod som utgör utredningens huvudalternativ.

FOBALT slår fast att en utveckling i riktning mot en registermodell bör ske i steg på ett sådant sätt, att kvaliteten kan hållas under kontroll och att förutsättningar finns för ett så smidigt genomförande som möjligt. Följande åtgärder föreslås.

Sysselsättnings- och pendlingsstatistik

Fr. o. m. inkomståret 1985 startar en årlig regional sysselsättningsstatistik byggd på en kontrolluppgift som även innehåller uppgift om arbetsställenummer och överenskommen veckoarbetstid. Detta innebär att den kontrolluppgift som lämnas av arbetsgivarna i januari 1986 föreslås vara utbyggd med dessa uppgifter.

Hushålls- och lägenhetsstatistik

En fullgod lösning på detta område är ett system med lägenhetsnummer i folkbokföringen och ett lägenhetsregister. Därför föreslår FOBALT att beslut snarast fattas om en särskild utredning om folkbokföring på lägenhet och möjligheterna att skapa ett

byggnads/lägenhetsregister. Detta utredningsuppdrag bör vara avslutat före utgången av 1986. I avvaktan på vad resultatet av denna utredning kan leda fram till, föreslås följande åtgärder:

– Hösten 1985 genomförs en insamling från allmänheten av uppgifter om hushållssammansättning och lägenhetens beskaffenhet (begränsad FoB 85); samordning sker därvid med en eventuell allmän fastighetstaxering (AFT 86).

– Med utgångspunkt i den begränsade FoB 85 etableras år 1985 ett statistiskt hushållsregister som även innehåller data om lägenhetsstorlek.

– För att möjliggöra löpande uppdatering av detta register fr. o. m. FoB-tidpunkten föreslås att flyttningsanmälan ändras fr. o. m. 1 juli 1984 så, att därav framgår till vem eller med vem man flyttar samt antal rum i inflyttningslägenheten.

– Med utgångspunkt i den begränsade FoB 85 etableras även ett statistiskt bostadsbyggnadsregister.

– För att möjliggöra uppdatering från detta register fr. o. m. FoB-tidpunkten genomförs förbättringar i kommunernas löpande rapportering till SCBs bostadsbyggnadsstatistik fr. o. m. 1985.

– Utredningen konstaterar dock, mot bakgrund av de olika undersökningar som genomförts, att en viss osäkerhet för närvarande råder om hur kvaliteten i hushålls- och bostadsbyggnadsregistren kommer att utvecklas under en fortlöpande uppdatering. Därför skall ingå i planerna att man under 1987 och 1988 genomför en kontrollundersökning på urvalsbasis.

– Resultaten från denna kontrollundersökning får bli vägledande för vad som skall göras 1990. Vid mer begränsade kvalitetsbrister kan kompletteringar göras på olika sätt, genom utnyttjande av fastighetstaxeringsmaterial, enkäter o. d. Resultatet av kontrollundersökningen och av det utvecklingsarbete om folkbokföring på lägenhet m. m., som föreslås, bör skapa förutsättningar för att ta ställning till en mer långsiktig utveckling.

– Om resultaten från 1987/88 års kontroll pekar på nödvändigheten därav, bör en ny kontrollundersökning genomföras 1992/93 och på motsvarande sätt ligga till grund för ställningstagande till vad som skall göras 1995.

– Denna lösning innebär att man under tiden efter 1985 inte kommer att kunna beskriva sambandet mellan hushåll och lägenhet med ett totalmaterial annat än vad gäller trångboddhet och boendetäthet (utan koppling till andra data om lägenheten och byggnaden än lägenhetsstorlek). Därför föreslås att särskilda undersökningar bör genomföras för denna beskrivning, eventuellt i anslutning till de ovan nämnda kontrollundersökningarna och/eller de större bostads- och hyresundersökningar som SCB med viss intermittens genomför. Härvid bör de statistiska skattningsmetoder som utredningen prövat och för vilka föreslås att utvecklingsarbetet intensifieras, vara ett viktigt hjälpmedel.

Yrkesstatistik

Frågeformuleringen och anvisningarna för yrkesuppgiften på till försäkringskassorna inkommande inkomstanmälningar bör förbättras fr. o. m. 1984.

Under 1985 och därefter vart femte år dataregistreras yrkesuppgifterna i samtliga under året till försäkringskassorna inkomna inkomstanmälningar.

Det förutsätts att överenskommelse träffas mellan SCB och arbetsmarknadens parter om tillgången till underlagen för lönestatistiken som kompletterande underlag för yrkesstatistik.

Utbildningsstatistik

Under 1985 byggs ett utbildningsregister upp baserat på uppgifter från FoB 70, socialstyrelsens medicinalpersonalregister, SCBs lärarregister, högskoleregister och gymnasieskoleregister.

Uppdateringen av detta register sker med hjälp av SCBs löpande individbaserade statistik om fullbordade gymnasiala och eftergymnasiala studier. För att erhålla en fullständigare täckning i uppdateringen förutsätts dels att SCB får tillgång till AMS uppgifter om arbetsmarknadsutbildning, dels att individrapporteringen från skolorna till

SCB utvidgas till att omfatta även gymnasieskolans specialkurser från år 1984.

Register-FoB
Vart femte år med början 1985 kopplas uppgifter från de olika delområdena inkl. inkomst samman till en register-FoB.

Hans Erik Altvall

Statistiska föreningens verksamhet år 1982

Föreningens årsmöte ägde rum den 18 februari. Styrelsen har under året sammanträtt 10 gånger, därav avser ett möte den gamla styrelsen.

Styrelsen har under året planerat den konferens, som genomförs samma dag som årsmötet 1983. Konferensens tema är: Den statistiska allmänutbildningen eller statistik är inte bara en fråga för experter.

Under året har inkl. årsmötet sammanlagt 10 föreningsmöten hållits.

De olika föreningsmötena tog upp följande ämnen:

1. I samband med årsmötet den 18 februari presenterades och diskuterades den s. k. AU-gruppens förslag om högskoleutbildning för: "Ekonomilinjen, förvaltningslinjen och linjen för personal- och arbetslivsfrågor." Avdelningsdirektör Jean-Pierre Zune, UHÄ och sekreterare i AU-gruppen, presenterade inledningsvis huvudinnehållet i utbildningsförslaget och då speciellt statistikämnets plats och omfattning i de olika utbildningslinjerna. Här kan bl. a. nämnas att statistiken har fått en speciallösning i form av ett fördjupningsalternativ inom ekonom- och förvaltningslinjen (samhällsekonomiskt utredningsarbete). Professor Uno Zackrisson, statistiska institutionen vid Umeå universitet, gav några kommentarer till AU-gruppens förslag. Han sade bl. a. att utredningsarbetet i stort sett har varit bra. Värdet av statistik integrerat med andra ämnen finns med. Statistik borde dock finnas med i flera av fördjupningsalternativen. Han pekade också på de goda erfarenheterna från Umeå, där statistiken är väl integrerad i ekonomlinjen.

2. Den 1 april. – Konsumentprisindex: mätare på det allmänna prisläget. Överdirektör Lennart Fastbom, SCB, redogjorde inledningsvis för metoder och problem i samband med beräkningar av konsumentprisindex (KPI). Han nämnde bl. a. att KPI är SCBs mest efterfrågade statistikprodukt (ca 75 förfrågningar per dag). KPI har en stor användning i bl. a. avtal, ekonomisk analys och som en allmän mätare på prisläget. Ett antal delindexar beräknas samt det s. k. "avtalsindex". Arbetet med KPI leds av indexnämnden, som bl. a. behandlar kvalitetsfrågor, varuurval och utformningen av olika delindexar. En viktig utgångspunkt i arbetet är att klara objektivitetskravet. KPI får inte kunna manipuleras eller ifrågasättas. Några synpunkter på KPI och vilka konsekvenser användningen av denna index får, gavs av fil. dr Olle Sjöström, statistiska institutionen vid Stockholms universitet, departementsrådet Sune Davidsson, ekonomidepartementet och fil. kand. Olof Fahlén, Dagens Industri. Olle Sjöström sade bl. a. att KPI är mest efterfrågat men minst ifrågasatt, mätfelet är okänt och KPI kan manipuleras. Det senare då genom att priskontrollera representantvaror. Sune Davidsson pekade på betydelsen av en mätning som inte kan ifrågasättas. Pensionsregleringen är t. ex. en mycket viktig tillämpning. Här finns en risk att indexreglera för mycket i samhället. Index får inte påverkas av priskontroll utan endast av urvalet av varor. Olof Fahlén konstaterade bl. a. att SCBs pressmeddelanden numera är mycket bättre. Det är dock ett missbruk att jämföra allt med KPI.

3. Vad gör Tidningsstatistik med dotterföretag? Direktör Bertil Klinte, Tidningsstatistik AB, presenterade den 14 april företagets statistiska verksamhet. Tidningsstatistik insamlar, bearbetar och distribuerar data om massmedia. Ett välkänt exempel på detta är fastställandet av dags- och fackpressens upplagor och spridning. Reklamstatistik AB är ett dotterföretag, som bl. a. tar fram en regionalt fördelad marknadsstatistik och gör reklammätningar. Sedan en tid ingår dessutom Institutet för

Marknadsundersökningar (IMU) i företaget.

4. Aktuariens roll i ett försäkringsbolag. Föreningen var den 12 maj inbjuden till Trygg-Hansa, där vi fick information om bolagets statistiska verksamhet av aktuarierna Olof Werner, Hans Andersson och Bert Langhed. Försäkringsinspektionen, som är tillsynsmyndighet, samlar in statistiska uppgifter från bolagen. Denna statistik kan sedan utnyttjas av försäkringsbolagen. Statistik i olika former spelar en stor roll i bolagens verksamhet. Premier sätts efter genomförda riskanalyser. Tillgänglig driftsstatistik är också ett viktigt beslutsunderlag. Försäkringsbolagen skall vidare arbeta efter den s.k. skälighetsprincipen, som bl.a. innebär att statistik skall utnyttjas.

5. En presentation av Lantbrukets utredningsinstitut (LUI) gjorde institutets ställföreträdande chef Olle Winfridsson den 10 juni. LUI bildades 1979 och är efterföljare till Jordbrukets utredningsinstitut. Verksamheten är uppdelad på sektioner för lantbruk, industri och marknadsfrågor. Några statliga medel utgår inte, utan LUI arbetar på kommersiella villkor. Verksamheten omfattar konsultarbeten, utredningar och utarbetande av prognoser inom lantbruksområdet.

6. Direktör Bo Bergström och statistikdirektör Rolf Frankenberg presenterade verksamheten vid Stockholms utrednings- och statistikkontor (USK) vid ett möte den 16 juni. De gav dessutom några aktuella exempel på genomförda statistiska undersökningar inom kommunen.

7. Kvinnor och barn. Till detta möte som hölls den 6 oktober, hade också en inbjudan sänts till medlemmarna i Svensk Demografisk Förening. Avdelningsdirektör Eva Bernardt, SCB, presenterade inledningsvis några av de hittills framkomna undersökningsresultaten från SCBs pågående undersökning. SCB har regeringens uppdrag att försöka klarlägga orsakerna till den vikande befolkningsutvecklingen i Sverige. Det kan bl.a. konstateras att barnafödandet nu sker senare och att därmed kvinnorna också föder färre barn. Kvinnorna har däremot en positiv attityd till att ha barn. Flertalet anser dock att det räcker med två barn. Steget från två till tre barn är för de flesta stort. I utredningen redovisas några områden som är av central betydelse för kvinnors livssituation och som kan ge förklaringar till födelsenedgången. Dessa områden är: familjebildningens ändrade karaktär, kvinnors förvärvsarbete, effektiv familjeplanering och den sociala bakgrunden. Utredningsarbetet fortsätter med bl.a. analys av de yttre omständigheter som kvinnor lever i. Mäns inställning till barn och barnafödande måste också kartläggas.

Professor Gunnar Biörck har under flera år aktivt arbetat med befolkningsfrågor i riksdagen. Han menade bl.a. att det mycket är en fråga om familjepolitik och familjens behållna inkomst. Vi kanske borde ha skatt efter försörjningsbörda. Effekten av olika politiska beslut är oklar och borde ses över. Chefredaktör Carl-Johan Åberg tog upp några samhällsekonomiska effekter av den nu vikande trenden. Vissa effekter kan vara bra på kort sikt, men övergår efter hand i sämre faser. Antalet pensionärer blir t.ex. allt större och därmed pensionsutgifterna allt högre. Dagens ekonomiska stagnation gör det också allt svårare. Han ansåg inte att de negativa samhällsekonomiska effekterna är skäl till befolkningspolitiska åtgärder. Problemen ligger på ett mer allmänpolitiskt plan. Vi måste dock upplysa om effekterna och komma fram till åtgärder.

8. Forskning och utredningar i varuhandeln: från säljstatistik till stöldmätningar. Folke Larsson, chef för Handelns Utredningsinstitut (HUI) och Handelns Forskningsinstitut, presenterade de båda institutens verksamhet vid en sammankomst den 22 oktober. HUI arbetar främst med konsultfrågor, bl.a. enkäter och datainsamlingar. Forskningsinstitutet svarar för den tillämpade forskningen och de mer långsiktiga metodfrågorna. Denna verksamhet är anslagsfinansierad. Instituten har ett bra samarbete med SCB och utnyttjar ofta SCB-material i analysen. Viktig verksamhet är prognoser på kort och lång sikt liksom struktur- och ekonomifrågor. Stölderna är ett stort problem för handeln. Det finns flera led i detaljhandeln där svinn eller stöld i

olika form kan ske.

9. Den 29 november. – Statistiken i samhällsdebatten: vem skall presentera och analysera det aktuella statistiska materialet och vad är SCBs roll? Generaldirektör Lennart Nilsson redogjorde inledningsvis för sin syn på SCBs roll och den omorientering som är på väg inom SCB. Det sker en utveckling från det s. k. tabellverket till något som mer påminner om ett informationsverk. SCB kan få en mer aktiv roll i samhället. Här kan bl. a. nämnas arbetet mot en ökad användarorientering, de s. k. budorden att på olika sätt bättre utnyttja den kunskap som finns inom SCB. Det senare kan ske genom att på vissa viktiga områden genomföra mer analys, t. ex. genom lite mer text i de vanliga publikationerna. SCB vill försöka ge bättre helhetsbilder av olika samhällsfrågor.

Utredningschef Jan Bröms, SACO/SR, sade bl. a. att statistikerna måste kunna sakområdet för att kunna genomföra en analys. SCB måste vara försiktig med att ge sig in i debatten. Det vore farligt att mäta responsen med antalet tidningsrubriker. Analysgrupper kanske kan arbeta på eget ansvar. Det finns en risk med selektiv försäljning av statistisk information. Man borde istället sträva efter en bred offentlig publicering.

Fil. dr Robert Eriksson, SOFI, förklarade att SCB inte får misstänkas. Information från SCB måste vara objektiv. En erfarenhet är att SCB är expert på statistik men mindre på innehåll. Även om analys till en viss del är en SCB-uppgift, så måste SCB lämna ut material för analys av externa institutioner och liknande. Materialet måste dock alltid göras tillgängligt för forskningen. Journalister och t.ex. lärare måste alltid få tillrättalagda tabellpresentationer med statistiska översikter.

Fil. dr Ulf Jacobsson, SAF, menade att SCBs analysroll måste ses i förhållande till den monopolställning som verket har. Denna ställning är dock inte självklar. Myndigheter vill ha en koppling mellan produktion och användning. SCBs centraliseringsfördel blir allt mindre. Det som är kvar är främst objektiviteten, som är SCBs livsvillkor. SCBs förtroendekapital är lätt att förstöra. Enskilda personer får inte fritt agera i SCBs namn.

Regelbundet genomförd analysverksamhet inom SCB behöver inte vara så problematisk som mer ad hoc genomförda analyser. Statistikproduktion ger ingen särskild analyskompetens och ej heller någon distributionsfördel. Statistiken måste alltid vara åtkomlig för alla och presenterad på ett begripligt sätt.

10. Statistikutredningens arbete diskuterades den 9 december. Landshövding Lennart Sandgren, utredningens ordförande, lämnade en lägesrapport från utredningens arbete. En omfattande kartläggning av behoven av statlig statistik håller nu på att avslutas. Flera av rapporterna är redan färdiga. Denna del av utredningen skall avslutas i mars 1983. Utredningen diskuterar också formerna för den fortlöpande översynen av den statliga statistiken. Hittills har endast beslutsordningen i departementet studerats. Statistiken har nog här idag en alltför blygsam plats. Alternativa modeller diskuteras. Utredningen har ännu inte tagit ställning till enskilda statistikgrenar. Behovskartläggningarna visar på ett stort behov av statistik. Vi är också av tradition flitiga statistikanvändare. Departementssekreterare Mac Murray, utbildningsdepartementet, ansåg att vi är dåliga på att förutse informationsbehov. Den politiska processen är dessutom dynamisk varför vi sällan har rätt information vid beslutstillfället. Det tar lång tid att lägga upp en statistik. Genom den ofta korta beslutsprocessen i regeringen hinner man inte ta fram aktuell statistik. Sektionschef Arne Myrbeck, Kommunförbundet, pekade på teknikens stora betydelse för en annorlunda roll för statistiken. Anpassningen till användarnas behov kan bli bättre. RSDB-tekniken är här ett bra exempel. Produktionsrollen förändras dock också. Det sker t. ex. en viss förflyttning mot den regionala och lokala nivån. Ett fortsatt samspel mellan SCB och kommunerna ger säkert den största nyttan. Det finns stora behov av samordning och jämförelser.

Statistikdag för statistikstuderande

Statistiska föreningen och Svenska Statistiksamfundet inbjöd den 24 mars samtliga statistikstuderande (mer än 500 studenter) till en informationsdag på temat.

"Vad gör en statistiker? Vilken arbetsmarknad finns? Vilka kunskaper behövs eller varför du skall läsa statistik!"

Statistiker från förvaltning och näringsliv var inbjudna för att berätta utifrån sina erfarenheter och sin arbetsplats om hur det är att arbeta som statistiker. Tyvärr kom inte mer än sex studenter, varför en del av programmet inställdes.

*

Antalet medlemmar i föreningen uppgick den 31 december 1982 till 580, varav 12 hedersledamöter, 80 frimedlemmar (över 65 år) och 29 ständiga medlemmar. Under året har 58 nya medlemmar invalts, medan 10 har avgått.

Som ordförande har under året fungerat riksdagsman Kurt Hugosson, och som sekreterare byrådirektör Anders Rickardsson. Vid föreningens årsmöte den 7 februari 1983 valdes följande nya styrelse:

Ordförande:	Erik Ruist (nyval)
Vice ordf.:	Lennart Nilsson
Sekreterare:	Anders Rickardsson
Skattmästare:	Cecilia Henriksson (nyval)
Övriga:	Olle Sjöström
	Kerstin Malmberg
	Gösta Larsson

Övrig verksamhet

Föreningen har utarbetat ett remissyttrande på AU-gruppens slutbetänkande "Ekonomlinjen, förvaltningslinjen och linjen för personal- och arbetslivsfrågor", (UHÄ-rapport 1981:23). I yttrandet pekades bl. a. på att det behövs breda grundläggande kunskaper i statistik. Detta är ett absolut nödvändigt komplement i alla samhällsekonomiska utbildningslinjer. Det är t. ex. kombinationen statistik och ekonomi/sociologi/statsvetenskap som är speciellt värdefull i det praktiska yrkeslivet. Yttrandets slutsats löd enligt följande: "Statistiska Föreningen vill med hänvisning till ovanstående få en helhetssyn på statistikens roll i den samhällsekonomiska utbildningen. Vi vill också få ökade och tydligt angivna statistikinslag i basblocket (upp till 20 poäng) i alla tre linjerna. Fördjupningsalternativet samhällsekonomiskt utredningsarbete bör bli markerat som ett huvudalternativ för den kvalificerade statistikutbildningen och därmed en bas för forskarutbildningen i statistik".

Dansk Statistisk Förening har under hösten firat sitt 60-årsjubileum. Vi var inbjudna att deltaga i firandet. Svenska Statistiska Föreningen representerades av Sten Borg som uppvaktade med kulram och änglaspel.

En noggrann genomgång av medlems- och adressregistren har genomförts under året. Nya registerkort har upprättats. Någon ny medlemsförteckning har dock inte tryckts.

Föreningen har även deltagit i det 16e nordiska statistikermötet, som hölls i Oslo den 12–14 augusti 1982. Ett särskilt referat från detta möte publiceras på annan plats i denna tidskrift.

Personalnotis

Tidigare chefen för avdelningen för individstatistik (I-avd.) Nils Elmhammer har förordnats till chef för avdelningen för central administration (C-avd.) fr. o. m. 22 april 1983. Han efterträder avdelningschef Ulf Wetterberg.

Summary in English

Coding of Establishments in the 1980 Population and Housing Census

by R. Andersson

In the 1980 Population and Housing Census, a coding technique entirely new in Swedish censuses was used for the variables industry, institutional classification, and location of the place of work (establishment).

The aim of the article is to describe this coding technique, i.e. coding of establishments or, as it is sometimes termed, serial number coding. The technique is based on the fact that all persons who work at the same establishment are to receive identical codes for several variables, e.g.

- Industry
- Institutional classification
- Location of the place of work.

The serial number coding in the 1980 Census means that each economically active person is linked to his/her place of work (establishment) by means of a serial number. Via this number the industry code, etc. pertaining to the establishment is then automatically added to the record for each individual. This serial number coding in the 1980 Census was carried out by the local census offices in each municipality.

The article describes:

- Principles of serial number coding (section 2)
- Practical serial number coding in more detail (section 3)
- Construction of a serial number (section 4)
- Establishment registers (section 5)
- Quality of the serial number coding (section 6).

As regards the quality of the serial number coding it should be mentioned that no evaluation study has been carried out. The views expressed below are based on our practical experience of the material.

For the great majority of the economically active, i.e. those who have a fixed place of work in the municipality where they are domiciled, the serial number coding has worked well. The problems that have arisen have been related to the following categories:

- Commuters to other municipalities
- Persons without a fixed place of work
- Persons with a fixed place of work which is not identical with the "principal establishment", e.g. cleaning personnel, persons working in their homes, lumberjacks and construction workers.

It should be pointed out that in a coding system considerably less expensive than conventional systems, it is reasonable to accept a slightly lower quality. Decentralized coding of this type inevitably entails a risk of quality variations between different municipalities. As far as possible, an adjustment for this variation has to be included in the central processing.

A more detailed description of serial number coding can be found in: Andersson, R., Coding of Establishments in the 1980 Population and Housing Census. Information on Methods No. 82:2, Statistics Sweden, 1982 (in Swedish).

What is statistical method?

by O. Sjöström

"Method" (from Gr. meta along, and hodos way) has as its original meaning "seeking ways", and presupposes a *goal*. In everyday language we mean by "method" consciously and systematically chosen ways to accomplish an end, for instance: getting

something done by means of some well-planned manual work, or solving an intellectual problem through systematic thinking.

This article is a polemic against the present use of the word "method" in statistics to mean "instruments" or "techniques", a use which seems to be implicit or explicit in most of todays education, applied statistics and in research.

A historical view of statistics as a social science shows that we seem to have lost contact with an important tradition formulating a theory of statistics as a *science of method* in an epistemologically more conscious way than we do today. Of special interest is the work of Ch. Sigwart and A. Meitzen at the end of the 19th century. Ref. [10] and [13]. This particular view seems to have been unnoticed by, for instance, Westergaard [19].

It is argued that "method" in a more fundamental sense is of relevance when dealing with many of todays' serious problems in social statistics. For instance: developing improved statistical planning, analysis and implementation–educating (and attracting) students–developing perspectives (theory) for what in Sweden is called "production" and "consumption" of social statistics.

The discussion centres around the following different views of statistical "method":

a. "method" as a collection of "techniques".

This is a relatively new "instrumental" meaning.

b. "method" as "ways of thinking" and "ways of make techniques useful".

This meaning has a very old tradition.

c. a more conscious epistemological view which analyses more fundamental aspects on perspective b.

A fourth view is also discussed:

d. an "applied-mathematics"-view that can mean the "double identity" Method ≡mathematics ≡ theory. Kish [9].

This article argues for moving from the restricted views a. and d. towards b. and c. This will be a difficult reorientation but it is important for education, applied social statistics and contacts within the social sciences. The opinions expressed coincide partly with those of Edgar Dunn in his book "Social Information Processing and Statistical Systems", (Wiley 1974) and W. E. Deming, ref. [4] and [5]. What is new is basing a critical view of present-day social statistics on a tradition going back to Francis Bacon, Hermann Conring as well as John Graunt, developed through the 18th and 19th centuries, culminating in an interesting theory (Sigwart and Meitzen).

Some proposals on statistical education and research are outlined for discussion.

Part-time employment in the Swedish Labour Force Surveys

by S. Tegle

This article discusses the definition of part-time employment 1965–1982 in the Swedish Labour Force Surveys (AKU), published by Statistics Sweden (SCB).

In section one the 34-hour per week limit of part-time employment is compared to the development of the full-time standard work week. This limit has been constant during the period in spite of the decrease in the standard full-time work week from 45 hours in 1965 to 40 hours per week 1973 on. Thus, the AKU estimates of the number, and proportion, of people in part-time employment is producing a negative bias for the period 1965 to 1973 in relation to the later period. Or, which is the mirror picture of the negative bias, a positive bias for the later period (1973–1982) in comparison with the earlier period before 1973. This bias is roughly calculated to be in the range of 30 to 60 %. The conclusion is that the SCB should produce more exact estimates of this bias using alternative limits as complements to the 34-hour limit.

In section three, it is shown that the SCB/AKU have defined part-time employment in, at least, three different ways. The definitions have been changed by interpreting the group "at work part-time and usually working part-time" differently in terms of what

should be ment by *temporary* part-time employment. Further, it is stressed that the SCB should make effort to clearly motivate, and state, the definitions used in order to facilitate the use of the produced statistics.

The use of the concept "usually" in the AKU questionnaire, together with a suggestion of an alternative definition, is analyzed in section four. The validity and reliability of the concept "usually" is questioned and it is concluded that it is better to use the concept "at work" as the base concept in constructing definitions of part-time employment. Along these lines, an alternative definition is suggested which gives a higher estimate of the number of part-time employed than the definitions used in the AKU. It is argued, that this higher estimate gives a better picture of the part-time employment development, especially from the middle of the 70's.

Finally, in section five, it is argued that the definition of who is employed part-time (or full-time) is a *social* definition and not a statistical one. The Statistics Sweden should make this evident by facilitate the users' possibilities of constructing, within the limits of the AKU questionnaire, their own definition of part-time employment.

Litteratur

Bokanmälningar

Vinod, Hrishikesh D. & Ullah, Aman, Recent Advances in Regression Methods ("Statistics" textbooks, Vol. 41), Marcel Dekker, New York 1981, 384 s. SFr 118:–.

Regressionsanalys är en vanlig metod att beskriva och utnyttja statistiska samband mellan olika variabler. Den klassiska modellen för linjär regressionsanalys kan man skriva som

$$y = b_0 + b_1 x_1 + \ldots + b_p x_p + u.$$

Här står y och $x_1, \ldots, x_p$ för olika observerade variabler; de senare, $x_1, \ldots, x_p$, kallas regressorer eller förklarande variabler. Vidare är u en variabel som inte kan observeras direkt och som betraktas som en slumpmässig felterm. Slutligen utgör b_0, b_1, ..., b_p de s. k. regressionskoefficienterna. Dessa är modellens parametrar, alltså okända storheter som man vill skatta utifrån de värden som de observerade variablerna antagit vid ett antal skilda tillfällen.

Praktiskt sett kan en regressionsanalys ha olika slags syften. Ibland kan regressionskoefficienterna som sådana vara innehållsmässigt intressanta, och då kan deras skattade värden ses som analysens resultat. En annan typ av tillämpning är s. k. prediktion. Detta innebär att man vill förutsäga vilket värde y kan väntas få i ett läge, då man vet eller antar att regressorerna $x_1, \ldots, x_p$ har vissa värden. Regressionsanalys med stort antal variabler har kommit att användas för bl. a. ekonometriska tillämpningar, dvs. ekonomisk–statistiska analyser.

Genom datorer och standardprogram har regressionsanalysen blivit mera tillgänglig som ett praktiskt medel vid analys av statistik. Önskade regressionsskattningar kan lätt räknas fram, även för stora material och modeller, och det hela kan genast göras om med andra material eller ändrad modell. Exempelvis kan man experimentera med att ta bort eller lägga till vissa regressorer. De stora fördelarna med denna ökade tillgänglighet skall inte förnekas, men samtidigt kan det bli en viss fara att man glömmer vad man egentligen håller på med. Att provköra modell efter modell på jakt efter den som ger mest hyfsat resultat är en tvivelaktig metod! Skall resultatet ha en mening bör det ligga någon tanke bakom modellen. Det kan också vara lätt att glömma att regressionsanalys med flera regressorer faktiskt fungerar på ett något komplicerat sätt. Det är i praktiken mycket svårt att mera åskådligt överblicka vad som egentligen "händer" i modellen. Bakom resultaten kan dölja sig fenomen som – om de avslöjas – kan förvåna även den som tror sig förstå teorin.

Det är dock inte bara de datatekniska hjälpmedlen som blivit bättre. På senare år har också regressionsanalysen som metod utvecklats. Den gamla klassiska ansatsen har byggts på och modifierats på många olika sätt, så att ett flertal nya tekniker skapats. På så sätt har man försökt komma till rätta med olika problem och begränsningar.

Den här boken ger en aktuell överblick över nyare metoder inom regressionsanalysen. Författarna har ekonometrisk bakgrund, men i huvudsak kan boken sägas vara allmänt inriktad på regressionsanalysen som statistisk metodik.

Ganska mycket av det som behandlas i boken har direkt eller indirekt att göra med en företeelse som kallas multikollinearitet. Regressionsanalysen som idé bygger ju på att det antas råda ett inte alltför svagt samband mellan å ena sidan variabeln y och å andra sidan regressorerna $x_1, \ldots, x_p$ tillsammans. Däremot är det mindre bra om det bland regressorerna sinsemellan finns några verkligt starka samband. Om så är

fallet sägs multikollinearitet föreligga. Då kan skattningen av regressionskoefficienterna bli mycket osäker, även om antalet observationer i underlaget verkar betryggande i sig. Osäkerheten kan utan vidare vara så stor att de skattade värdena får "fel" förtecken, t. ex. negativt i stället för positivt värde. Vidare kan tillägg av enstaka ytterligare observationer leda till stora hopp i värdena. När man använder regressionsanalysen för prediktion kan det dock hända att multikollinearitet inte får lika stora vådor som för regressionskoefficienterna. Det beror på att, under vissa villkor, felen i de sistnämnda kan ta ut varandra till mycket stor del när man räknar fram ett predikterat värde på y.

Här kan nämnas ett par frågor som boken inte tar upp annat än antydningsvis. För det första kan man fråga sig hur man på ett praktiskt sätt skall kunna belysa följderna av multikollinearitet vid prediktion. För det andra bör noteras att uppkomst av så stora fel, att till och med förtecknen på de skattade regressionskoefficienterna kan vara fel, även kan ha andra orsaker. Denna, på sätt och vis än mera förrädiska, situation kan uppstå om modellen är fel tänkt i förhållande till verklighet eller observationsmaterial.

Något som boken trycker på är vad man kan göra när man har eller misstänker multikollinearitet. En huvudmetod här är "ridge regression". Denna metod innebär att man räknar fram korrelationskoefficienterna som om samtliga regressorers varianser ökats med ett visst belopp k; kovarianserna antas däremot ha sina ursprungliga värden. Man kan säga att man antar en allmän försvagning av de verkliga sambanden mellan regressorerna, genom att varje regressor antas ha en större "egen" variation än den i själva verket har. De skattningar, som man får efter denna manipulering, kan ses som alternativa skattningar av den ursprungliga modellens regressionskoefficienter. Dessa "ridge-skattningar" blir inte väntevärdesriktiga, utan de får ett systematiskt fel ("bias"). Däremot får de en lägre skattningsvarians än de vanliga skattningarna. Man kan därför få en mer eller mindre påtaglig minskning i det totala felet (uttryckt som den s. k. medelkvadratavvikelsen, innefattande både systematiskt och slumpmässigt fel).

Om man vill använda ridge-skattningar, hur bör man välja talet k? I boken ges metoder för att behandla den frågan. I princip är det graden av multikollinearitet som avgör hur högt det lönar sig att gå med k, med tanke på de totala felen. Ett annat slags tillämpning är "ridge trace". Detta innebär helt enkelt att man ritar upp kurvor över regressionskoefficienterna som funktioner av k. Om kurvan svänger kraftigt, och eventuellt skär k-axeln, så är detta ett varningstecken. (Kurvorna har alltid k-axeln som asymptot.) Författarna rekommenderar att man rutinmässigt använder "ridge trace" som en dataanalytisk metod. – I boken behandlas också ett antal varianter på ridgemetoder.

Kan man då veta om ridge-skattningarna är de "rätta" i någon mening? I boken visas att man utifrån vissa modellansatser faktiskt kommer fram till skattningsformler som liknar ridge-skattningarnas. Det är då främst fråga om Bayesianska ansatser, som f. ö. är ett återkommande tema i boken. Enligt Bayesianskt synsätt skall modellparametrarna, alltså regressionskoefficienternas sanna värden, ses som variabler vilka slumpmässigt antagit sina värden. Dessa slumpvariabler antas ha en viss sannolikhetsfördelning. En svårighet med att använda Bayesianska skattningar praktiskt är att veta hur man skall välja denna fördelning, som kallas a priori-fördelning. Men oavsett det kan man nog konstatera att de Bayesianska metoderna har ett betydande teoretiskt värde. Andra metoder kan komma i en ny intressant belysning när man jämför dem med de Bayesianska skattningsformlerna.

Boken tar vidare upp Steinskattningar. Även här är det fråga om skattningar som tillåts ha ett visst systematiskt fel men som har mindre totalt fel än de vanliga. Här är det dock inte speciellt multikollinearitet man vill bemästra, utan även annan osäkerhet, som beror på sådant som att underlaget är litet. Själva idén är ganska enkel: den ursprungliga skattningen vägs ihop med ett

annat tal. Detta andra tal kan exempelvis väljas som medelvärdet av motsvarande skattningsvärden i ett antal andra, liknande fall. Metoden har i sig inget särskilt med regression att göra utan är en mera allmän skattningsmetod. Att den här kommer in som ett medel för regressionsskattning kan dock möjligen tänkas ha sina skäl. I en del tillämpningar kanske man vill laborera med ganska många regressorer även när man har dataunderlag som inte är alltför stora. Skattningarna av regressionskoefficienterna får då ofrånkomligen stora slumpfel, men Steinmetoden kan möjligen ses som en väg att reducera problemet.

Den klassiska teorin för regressionsanalys bygger strängt taget på normalfördelningen. Den slumpmässiga feltermen u i modellen antas nämligen vara normalfördelad. Om denna förutsättning inte är uppfylld så kan man med de vanliga skattningsformlerna få mycket stora fel. Detta gäller särskilt vid förekomst av utliggande värden, dvs. när enstaka observationer motsvaras av u-värden med särskilt stora absolutbelopp. Det nu sagda gäller även de flesta av de metoder som boken tar upp. I det sista kapitlet berörs dock helt kort robusta metoder. Dessa metoder har förmåga att fungera hyfsat utan att alltför precisa förutsättningar behöver antas gälla. Bl. a. störs de mindre av utliggande värden än vad andra metoder kan göra. Det pris man får betala är att om den "vanliga" metodens förutsättningar råkar vara perfekt uppfyllda, och alltså normalfördelning gäller, så får de robusta skattningarna något sämre precision än de vanliga. Men hur ofta har man i praktiken garantier för "perfekt" normalfördelning? De robusta metoderna synes ha mycket som talar för sig, men teorin har ännu inte hunnit utvecklas så mycket.

Detta var ett axplock bland de ämnen som boken tar upp. I det hela är det ett mycket stort antal metoder och varianter på metoder som beskrivs. Litteraturförteckningen omfattar ca 500 arbeten, varav dock inte alla åberopas i texten. Bland andra ämnen än de redan nämnda kan märkas:

- Modeller med bivillkor på regressionskoefficienterna
- Korrelerade feltermer
- Modeller med mer än ett regressionssamband.

För att läsa boken bör man vara något så när förtrogen med den vanliga regressionsanalysen, liksom med matriskalkyl och estimationsteorins grundbegrepp. I de båda första kapitlen ges dock viss repetition och komplettering av förkunskaperna. Ett av de sista kapitlen vänder sig särskilt till läsare med viss vana av ekonometrisk litteratur. Totalt ingår tretton kapitel. Kapitlen inleds oftast med en relativt tillgänglig orientering innan mera speciella eller tekniska detaljer tas upp. Till varje kapitel finns ett mindre antal övningsuppgifter. Med tanke på det omfattande och disparata stoffet får man säga att boken har god läslighet för den som har rätt förkunskaper. Självfallet kan dock inte allt bli lika detaljerat framställt i ett översiktsarbete av detta slag. Författarna lyckas hålla ihop ett material som bör ha ett inom sitt område ganska brett intresse, både teoretiskt och praktiskt.

Men några universallösningar får man inte begära. Många olika ansatser har kommit fram under senare år. Dock måste man konstatera att det knappast finns någon teori om hur man skall välja mellan de olika metoderna. Detta måste till stor del bli en omdömesfråga, eller som det sägs i boken (sid. 88):

> ... the job of a researcher is becoming more and more skilled, requiring some judgement and experience.

Kanske är författarna något för okritiska mot en del moderna idéer. Det finns knappast grund för att utan vidare tala om Steinskattningarnas "teoretiska överlägsenhet" (sid. 136). Man är onödigt raljant mot det gängse gamla kravet på väntevärdesriktighet (dvs. på att det systematiska felet skall vara noll; sid. 170). Detta krav bör visserligen inte av ren slentrian tas som en absolut dogm. Likväl är det ett högst naturligt krav på en skattning, och beroende på situationen kan det ha ett mer eller mindre starkt sakligt fog.

Trots mångfalden av olika metoder kan man i boken vagt tycka sig ana konturer av helheter. Kanske kommer det på lång sikt

att utkristalliseras en ny enhetlig grundsyn på regressionsfrågorna. Men det torde knappast kunna ske innan någon form av robust metodik blivit mogen att tas in som en självklar och integrerad del av teorin.

Martin Ribe

Hájek, J., Sampling from a finite population. (Vol. 37) Marcel Dekker Inc. New York 1981, 247 s. $ 27.50.

Professor Jaroslav Hájek (1926–74) var internationellt uppskattad för sina bidrag inom sannolikhetsteorin, parametrisk estimation, stokastiska processer, sannolikhetsurval m. m. Han var mycket mångsidig och förmådde också knyta ihop teorin inom olika områden.

I början av 1970-talet påbörjade Hájek den här anmälda monografin som slutligen sammanställdes av hans forna kollegor och studenter i Tjeckoslovakien.

Boken är olik traditionella läroböcker i (survey) sampling (t. ex. Cochran, 1977), vilket kanske framgår av titeln. Genom att ange att det handlar om urval från *ändliga populationer* antyds att författaren uppmärksammar de speciella problem (och möjligheter) som föreligger vid inferens från surveys jämfört med så kallad traditionell statistisk inferens. Boken är dock inte hållen renodlat i den anda som präglats av andra forskare i genren, som t. ex. Godambe, Joshi, Thompson, Ericson, Royall, Basu, Särndal m. fl.

Till skillnad från den andra (icke-traditionella) boken inom området (Cassel, Särndal & Wretman, 1977) har Hájek inte försökt att ge en översikt över kända forskningsresultat och utvecklingen inom området. (Referenserna är bara 35 till antalet och av naturliga skäl är alla utom fyra publicerade före 1970.) Hájek har i stället valt att följa och konsekvent genomföra den linje han intresserat sig för och arbetat med. Boken är således uppbyggd kring inklusionssannolikheter, Poisson sampling och rejective-successive sampling, Horwitz-Thompson estimatorn (som kallas "simple linear estimate") samt en regressionsmodell för att ta tillvara hjälpinformation.

Hájek har lyckats ge en avrundad och sluten framställning i den meningen att allt hänger bra ihop. Det finns inte många lösa trådar inom den ram som valts. Framställningen är stringent och logisk i bästa Hájekstil och man besväras inte av irriterande tryckfel.

Boken innehåller många elegant bevisade och intressanta resultat samt egna ideer och begrepp. Exempel på det sistnämnda är representativ strategi, entropi i samplingdesigner, tight approximation av MSE, samt approximation av estimationsvikter.

Tyvärr har överblicken gått förlorad genom att exempel, teorem och bevis staplats på varandra utan typografisk åtskillnad. Det är svårt att se vad som är väsentligt och nytt. De ibland ganska långa kapitlen saknar också rubricerade underavsnitt. Detta tillsammans med att Hájek använder "sina" begrepp enligt ovan gör att man knappast kan använda boken som uppslagsbok utan att först ha läst igenom den ordentligt.

Boken består av tre delar; Grundläggande begrepp, Urvalsförfaranden och Estimation.

Den första delen behandlar grundläggande begrepp i de 4 kapitlen: *Population, Bayesiansk ansats, Robust ansats* och *Mer om inferens*.

Definitionen av *population* är en aning originell. Begreppet *identifierbara enheter* borde ha poängterats mer, eftersom det är grundläggande för kommande inferens. I problem-och-not-avsnittet skyms tyvärr definitioner och användbara resultat av räknemässigt krävande övningsuppgifter och algebraiska omskrivningar av olika populationskaraktäristikor.

Den *Bayesianska ansatsen* (kap. 2) är inte Bayesiansk i strikt mening, dvs. att man utgår från en a priorifördelning och sedan via ett observerat stickprov och Bayes' sats kommer fram till en a posteriorifördelning. Den diskuterade ansatsen är i stället vad som ibland brukar kallas semi-Bayesiansk vilken ibland tillåts omfatta all inferens med hjälp av subjektiva sannolik-

heter. Det förefaller också (i detta kapitel men inte senare) som om författaren anser att superpopulationer nödvändigtvis måste behandlas under sannolikhetsspekulationer (med subjektiva sannolikheter), vilket är onödigt restriktivt.

I kap. 3 om *Robust ansats:* sannolikhetsurval definieras bl. a. inklusionsindikatorn, inklusionssannolikhet och Poisson sampling (som i princip innebär att varje enhet slumpas till urvalet med en viss sannolikhet och oberoende av hur slumpexperimentet utfaller för övriga enheter). Entropin används som ett mått på ett urvalsförfarandes "utspridning" och med detta mått visas att betingad Poission sampling maximerar utspridningen inom en klass av förfaranden.

Första delen av boken avslutas med kap. 4: *Mera om inferens* som kortfattat behandlar grundläggande begrepp som parameter- och utfallsrum, likelihood funktion, Bayesiansk inferens, linjära estimat (där vikterna tillåts vara stickprovsberoende), Hájeks definition av representativ strategi (som innebär att en strategi är representativ med avseende på x_i, $1 \leq i \leq N$, om estimatorn, baserad på hela populationen $(x_1,..x_N)$, estimerar totalen utan fel med sannolikheten 1).

Bokens andra del behandlar *urvalsförfaranden. Obundet slumpmässigt urval* (OSU), som har maximal entropi bland alla sannolikhetsurval med fix urvalsstorlek, klaras av på några sidor (kap. 5). En relativt utförlig beskrivning av *Poisson sampling* (kap. 6) motiveras av att andra användbara urvalsförfaranden, t. ex. rejective sampling, kan ses som betingad Poisson sampling. Poisson sampling ger ju ett urval där enheterna får önskade inklusionssannolikheter, men urvalsstorleken blir variabel. Det senare kan undvikas genom rejective sampling eller successive sampling.

Rejective sampling (kap. 7) innebär att man drar olika stickprov (med Poisson sampling eller genom urval med återläggning och specificerade dragningssannolikheter) tills man får ett stickprov med ett fixt antal, n, (distinkta) enheter (erhållna i följd). Approximativa (asymptotiska) formler för sannolikheten att få ett godkänt stickprov vid ett försök samt för sambandet mellan första ordningens inklusionssannolikheter och dragningssannolikheterna resp. andra ordningens inklusionssannolikheter ges med bevis. Med *Sampford-Durbins* metod (kap. 8) kan man exakt bestämma de dragningssannolikheter som ger de önskade inklusionssannolikheterna av första ordningen.

Successive sampling (kap. 9) innebär en sekvens av oberoende dragningar med konstanta specificerade dragningssannolikheter vilket pågår tills man fått ett fixt antal, n, distinkta enheter (duplikat, men inte hela påbörjade stickprov förkastas). Antalet dragningar blir färre men teorin blir svårare än vid rejective sampling. Hájek refererar till Roséns arbeten, (1972) men visar metodens asymptotiska egenskaper med en annan ansats.

Då det gäller *systematiskt urval* (kap. 10) påpekas svårigheten att bestämma andra ordningens inklusionssannolikheter och Hartley och Raos förslag (1962) kritiseras.

Stratifierat urval och *flerstegsurval* klaras vardera av på två fyrsidiga kapitel (11 och 12) men nämns sedan också som exempel på betingad *Poisson sampling* vilket behandlas utförligare i ett eget kap. 14. Problemet vid betingad Poisson sampling är att återevaluera inklusionssannolikheterna, vilka här prediceras med hjälp av linjär och kvadratisk regression.

I kap. 13 om *Olika slags metoder* (assorted methods) behandlas kvotvarianter (t. ex. den som ger unbiased kvotskattningar), permutation sampling och Rao-Hartley-Cochrans metod med urval av en enhet i slumpmässigt konstruerade grupper (Cochran et al., 1962), dock utan att tala om varför man använder dessa metoder.

Andra delen av boken avslutas med kap. 15 om hur man kan *korrigera ett stickprov,* varmed avses hur man skall transformera ett stickprov av storleken n' och inklusionssannolikheterna Π_i', $1 \leq i \leq N$, till ett stickprov av storleken n och inklusionssannolikheterna Π_i, $1 \leq i \leq N$. Det är svårt att inse innebörden av detta, eftersom det tidigare visats hur man (asymptotiskt) bestämmer dragningssannolikheterna för att få önskade inklusionssannolikheter.

Av andra delens totalt ca 100 sidor ägnas huvuddelen åt sådana generella urvalsförfaranden med olika urvalssannolikheter där Hájek själv bidragit till forskningen. Endast knappt 15 sidor ägnas åt "vanliga" urvalsförfaranden som OSU, stratifierat urval och flerstegsurval (flerfasurval nämns inte alls). Dessa kan ju också ses som specialfall av betingad Poisson sampling och dessutom behandlas de utförligare i bokens tredje del.

Bokens tredje del behandlar *estimationsmetoder* och först nu diskuteras hur man ska finna goda strategier (avseende urvals- och estimationsförfarande).

Inledningsvis behandlas i kap. 16 begreppen representativitet (igen), robusthet och tightness. En strategi är *representativ* med avseende på regressorn x_k, $1 \leq k \leq m$, om man erhåller populationstotalen för regressorn, när man sätter in dess stickprovsvärden i den enkla linjära estimatorn (se nedan). *Tightness* är en egenskap hos approximationer av MSE och en strategi är *robust* om MSE (-approximationen) har ett visst utseende som gör att den kan estimeras. Hájek diskuterar strategier nästan uteslutande under en regressionsmodell där undersökningsvariabeln antas bero linjärt av regressorerna (hjälpvariablerna). Nämnda begrepp kan då inplaceras naturligt i diskussionen.

I kap. 17 införs den *enkla linjära skattningen*, utan att det nämns att den är ekvivalent med den kända Horwitz-Thompson estimatorn (Horwitz och Thompson, 1952). Dess exakta variansformel, som ju innehåller inklusionssannolikheter av andra ordningen, anges. Dessutom föreslås en approximation för rejective, Sampford-Durbins och successive sampling.

Kap. 18 ägnas åt problemet att finna *optimala strategier* med hjälp av en superpopulationsmodell som förutsätter linjär regression och okorrelerade residualer. En kuriositet är att Hájek använder $EM(Y\text{-}Y)^2$ och inte det kanske mer vanliga $ME(Y\text{-}Y)^2$ (E och M betecknar förväntan i design- resp. modelldimensionen) som kriterium. I allmänhet spelar det dock ingen roll hur man gör, eftersom resultatet blir detsamma för icke-informativa designs.

Under den antagna regressionsmodellen och den enkla kostnadsmodellen, $C = \sum_i^N c_i \Pi_i$, visas att optimal strategi bl.a. innebär den enkla linjära estimatorn och att strategin är representativ med avseende på regressorerna. Representativiteten bör alltså säkerställas genom urvalsförfarandet, givet att alla urvalsförfaranden med fixa inklusionssannolikheter kostar lika mycket.

I praktiken kan det emellertid vara komplicerat eller dyrt att säkerställa representativiteten genom urvalsförfarandet. Det kan då finnas anledning att i stället säkerställa den (dvs. utnyttja hjälpinformationen) helt eller delvis vid estimationen. En sådan möjlighet är *regressionsestimation* (med specialfallen differens- och kvotestimation) som ger representativitet för alla urvalsförfaranden. Denna metod behandlas i kap. 19, där skattningsvikterna analyseras för olika urvalsförfaranden och approximationer till dessa vikter ges. Hájek visar också att precisionsförlusten blir liten om man utnyttjar hjälpinformationen vid estimationen i stället för vid urvalet, förutsatt att antalet regressorer är litet i förhållande till stickprovsstorleken.

I kap. 20 och 21 behandlas estimation vid *stratifierat* urval och *flerstegsurval*. Här ges variansformler och formler för optimal urvalsallokering.

Boken avslutas med *A general approach* (kap. 22), som i huvudsak behandlar skattning av MSE, samt *tillämpningar vid OSU* (kap. 23).

Sammanfattningsvis är detta en bok för den som är intresserad av stickprovteorins grunder och inferensproblem. För den matematiskt orienterade statistikern torde Hájeks stringens och logiska uppbyggnad samt anknytningen till statistisk inferensteori vara mycket tilltalande. Man skall dock vara medveten om att begreppsapparaten till stora delar avviker från den som används av survey statistiker. Poisson sampling beskrivs t. ex. inte i någon av de vanligaste läroböckerna i sampling.

För den praktiskt arbetande surveystatistikern räcker inte boken som enda källa.

För detta ändamål saknas flera viktiga områden som t. ex. ramproblem, redovisningsgrupper, bortfall och mätfel och de vanligaste urvals- och skattningsmetoderna är nog alltför komplicerat beskrivna. Den ambitiöse samplingpraktikern som vill vidga sina vyer och utveckla sina metoder kan dock ha stor nytta av boken. Vid t. ex. analys av bortfall och mätfel räcker det ju inte med OSU-grunden och fixpopulationsansatsen, utan man måste tillgripa modellbaserad inferens och kanske göra antaganden om stokastiskt svarsbeteende (som kanske kan ses som Poisson sampling som genererar svarsstratum?).

Boken är krävande om man vill gå igenom alla bevis, men det är ju inte nödvändigt. Den dåliga typografin kan man ju själv också avhjälpa genom att göra lämpliga understrykningar. Det är mödan värt att läsa boken, den innehåller många intressanta resultat, även om den inte är heltäckande.

Referenser

[1] Cassel, C. M., Särndal, C. E., and Wretman, J. H. (1977): Foundations of inference in survey sampling. New York, John Wiley & Sons.

[2] Cochran, W. G. (1977): Sampling Techniques, third ed. New York, John Wiley & Sons.

[3] Cochran, W. G., Hartley, H. O., and Rao, J. N. K. (1962): On a simple procedure of unequal probability sampling without replacement. J. Roy. Stat. Soc. Ser. B24, sid. 482–491.

[4] Hartley, H. O., and Rao, J. N. K. (1962): Sampling with unequal probabilities and without replacement. Ann. Math. Stat., 33, sid. 350–374.

[5] Horwitz, D. G., and Thompson, D. J. (1952): A generalization of sampling without replacement from a finite universe. J. Amer. Stat. Ass., 47, sid. 663–685.

[6] Rosén, B. (1972): Asymptotic theory for successive sampling with varying probabilities without replacement, I, Ann. Math. Stat., 43, sid. 373–397: II, sid. 748–776.

Claes Cassel *Ingrid Lyberg*

Öhrvik, J., Robusteffektiv lägesestimation. [Akad. avh.] Research Report 1982: 17, Statistiska institutionen, Stockholms Universitet. 167 s. + XII.

Den statistiska termen "robust" infördes på 1950-talet, men långt tidigare hade många statistiker arbetat med en medvetenhet om behovet av robusta metoder. Sedan 1960-talet har det varit stor forskningsaktivitet kring robusta metoder. Den s. k. Princetonrapporten (Andrews et al. 1972) där 68 olika robusta lägesestimatorer undersöks har i hög grad stimulerat debatten kring behovet av robusta metoder och har genom den väl dokumenterade metodiken varit ett mönster för simuleringsstudier inom detta område.

"Robustheten" kan bl. a. gälla fördelningsantaganden, outliers och beroende. Både i Princetonrapporten och i Öhrviks arbete behandlas robusthet med avseende på fördelning och speciellt undersöks symmetriska fördelningar som är mer "långsvansade" än normalfördelningen. Lägesestimationen avser symmetripunkten för symmetriska fördelningar.

I litteraturen nämns ofta tre huvudgrupper av robusta lägesestimatorer: Dels metoder som bygger på rangtest, dels metoder av maximum-likelihood-typ och dels estimatorer som är linjärkombinationer av de storleksordnade observationerna. De metoder som diskuteras i avhandlingen har anknytning till den sista huvudgruppen av estimatorer.

En linjärkombination, θ^L, av de ordnade observationerna kan skrivas

$$\theta^L = \sum_{i=1}^{h} \omega_i X_{in}$$

där w_i är en vikt och X_{in} är observation nummer "i" bland de n stycken storleksordnade observationerna. Det enklaste exemplet inom denna klass är stickprovsmedelvärdet som erhålles om $\omega_i = {}^1/_n$ för alla i (se fig. 1).

Trimmade medelvärden erhålles om en andel p_0 av de största och samma andel av de minsta observationerna får vikten noll

och resterande observationer får den vikt, c, som krävs för att $\sum_{i=1}^{n} \omega_i = 1$ (se fig. 2).

Trimningsproportionen $p_0 = 0$ (vilken ger stickprovsmedelvärdet) ger minst varians för normalfördelningen, medan ett stort värde på p_0 är fördelaktigt för mer långsvansade fördelningar. Värdet på p_0 väljs enligt vissa metoder adaptivt dvs. med hjälp av stickprovet. Fig. 3 visar det fall då vikten har värdet noll för andelen p_0 av de minsta och största observationerna, värdet c_1 därnäst för andelarna p_1 och värdet c_2 för resterande mittobservationer. Det finns då 3 stycken oberoende parametrar (p_0, p_1 och c_1) och $k=2$ stycken viktvärden förutom värdet noll. Fig. 4 visar en finare indelning med $k=3$ stycken viktvärden förutom värdet noll.

Avhandlingen behandlar en estimator för fallet $k=2$ där p_0 och p_1 fixerats och c_1 väljs med hjälp av stickprovet. Estimatorn kan också ses som en adaptiv linjärkombination av två trimmade medelvärden. Metoder av denna typ har även beskrivits i Jaeckel (1971), Andrews et al. (1972), Johns (1974) och Sacks (1975).

–En studentiserad version som kan användas för att ge konfidensintervall även för relativt små stickprov föreslås i avhandlingen. Detta ökar väsentligt den praktiska nyttan av estimatorn. Den föreslagna estimatorn jämförs med stickprovsmedelvärdet, med trimmade medelvärden samt med en variant av den föreslagna estimatorn. Egenskaperna för punkt- och intervallestimatorn undersöks och visas, dels analytiskt

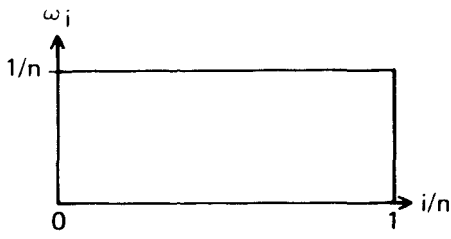


Fig. 1. Aritmetiskt medelvärde

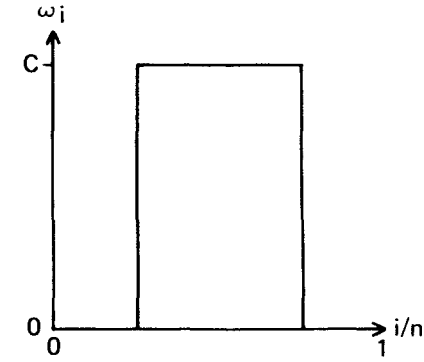


Fig. 2. Trimmat medelvärde

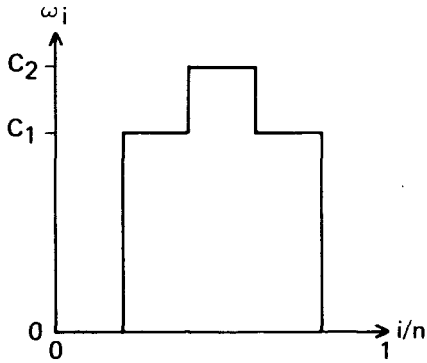


Fig. 3. Linjärkombination av två trimmade medelvärden

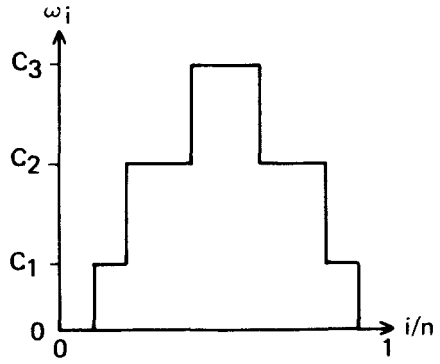


Fig. 4. Linjärkombination av tre trimmade medelvärden

och dels med en väl kontrollerad simuleringsstudie, vara tillfredsställande.

Referenser

Andrews, D. F., et al. (1972): Robust estimates of location: Survey and Advances, Princeton, N.J.: Princeton University Press, 1972.

Jaeckel, L. A. (1971): Some flexible estimates of location. Ann. Math. Statist. 42, sid. 1540–1552.

Johns (1974): Nonparametric estimation of location. J. Ann. Statist. Ass. 69, sid. 453–459.

Sacks, J. (1975): An asymptotically efficient sequence of estimators of a location parameter. Ann. Statist. 3, sid. 285–298.

Marianne Frisén

Utkom från trycket den 29 juni 1983

STATISTISK TIDSKRIFT

UTGIVEN AV STATISTISKA CENTRALBYRÅN

Ansvarig utgivare: Sten Johansson. Redaktör: Thore Grönqvist

Adr: Karlavägen 100. Postadress: 115 81 Stockholm

Telefoner: 08/14 05 60, ankn. 4260 (redaktören personligen)

och 4265 (Helen Lidén)

Prenumeration genom bokhandel eller hos Liber distribution, Prenumerationsorder, 162 89 Stockholm

Utkommer med 5 häften om året

Pris ca 105: – per år eller ca 20: – per häfte (inkl. moms)

Tredje följden Årg. 21 Nr 3 1983

Innehåll

Artiklar

P. W. Wargentin 1717–1783. "Den svenska statistikens fader" 181

Lene Skotte:
Befolkningsstatistikkens udvikling i Danmark indtil 1849 182

Göran Strengell:
Nordiska och finländska tankar med anledning av 200-årsminnet av Pehr Wargentins död 184

Gerd Lettenström:
Befolkningsstatistikens utveckling i Norge 188

Mogens Nygaard Christoffersen:
Simpel sammenligning af indkomstoplysninger 191

Ann-Charlotte Ståhlberg:
Avgiftsprocenten i ATP 197

Bo Rennermalm:
The Advantage of good Methods. – Considerations on the Application of Markov Models 201

Tarja-Liisa Leiniö:
Invandrare som brottsoffer 209

Meddelanden

SCB får ny chef 225
Europeiska statistiska möten:
Miljöstatistiska indikatorer (ECE) 226
OECD möten:
Expertmöte om FoU-statistik 226
Utbildnings- och sysselsättningsåtgärder för ungdomar i 16–19 årsåldern 227
Det nordiska chefsstatistikermötet 1982 228
IIASAs seminarium om informationssystem för integrerad regional utveckling 237
Statistical Computing Project 238
SCBs yttrande över rapport 2 från UHÄs datareferensgrupp "Förstärkning av den administrativt inriktade utbildningen inom dataområdet" 239
SCBs yttrande över länsdemokratikommitténs betänkande "Vidgad länsdemokrati" (SOU 1982:24) 241
SCBs yttrande över betänkandet "Information om arbetsmiljörisker" (SOU 1982:30) 242
Studiebesök i Danmark och Norge med anledning av SCBs fördelningsmodellprojekt 246
Det finsk-ungersk-svenska samarbetet kring miljöstatistiska metodfrågor 248

Summary in English 249

Litteratur

Bokanmälningar:
Pehr Wargentin – den svenska statistikens fader (Ingvar Holmberg) 253
Lärare för livet? (Björn Steneberg) 254
Bergman, L. R. och Wärneryd, B., Om datainsamling i surveyundersökningar (Per Kempe) 255
Brändström, A. & Sundin, J. (ed), Tradition and Transition (Rune Tryggveson) 258
Hofsten, E., Indien. Om samhället och befolkningsproblemen (Ingvar Holmberg) .. 261
Hofsten, E., Demografins grunder (Ingvar Holmberg) 263
Rolski, T., Stationary Random Processes Associated with Point Processes (Martin Ribe) 263

STATISTICAL REVIEW

PUBLISHED BY STATISTICS SWEDEN, STOCKHOLM, SWEDEN

Five issues a year

Third series. Vol. 21 No. 3 1983

Contents

Articles

P. W. Wargentin 1717–1783. "The Father of Swedish Statistics" 181, 249

Lene Skotte:
Befolkningsstatistikkens udvikling i Danmark indtil 1849 (Without English summary) 182

Göran Strengell:
Nordiska och finländska tankar med anledning av 200-årsminnet av Pehr Wargentins död (Without English summary) 184

Gerd Lettenström:
Befolkningsstatistikens utveckling i Norge (Without English summary) 188

Mogens Nygaard Christoffersen:
Information on income obtained from an interview survey compared with information given to taxation authorities 191, 249

Ann-Charlotte Ståhlberg:
ATP-Contributions 197, 251

Bo Rennermalm:
The Advantage of good Methods. – Considerations on the Application of Markov Models 201

Tarja-Liisa Leiniö:
Immigrants as Victims of Crime 209, 252

Notes

New head of Statistics Sweden appointed 225

European statistical meetings:
Methodological problems in environmental statistics (ECE) 226

OECD-meetings:
Expert meeting on FoU-statistics 226
Educational and employment measures for youth 16–19 years old 227
The 1982 meeting of the heads of the national statistical offices of the Nordic countries 228
IIASAs seminar on information systems for integrated regional development ... 237
Statistical Computing Project 238
Statistics Sweden:
Comments on the report no. 2 by the National Swedish Board of Universities and Colleges "Intensification of the administrative aimed education within the data processing sector" 239
Comments on the report "Extended County Democracy" by the Committee on County Democracy (SOU 1982:24) 241
Comments on the report "Information on risks in the working environment" (SOU 1982:30) 242
Study visit in Denmark and Norway concerning SCBs project of a distribution model 246
Co-operation agreement between Finland, Hungary and Sweden in the field of methods for environment statistics 248

Summary in English 249

Literature

Book reviews:
Pehr Wargentin – den svenska statistikens fader (Ingvar Holmberg) 253
Lärare för livet? (Björn Steneberg) 254
Bergman, L. R. och Wärneryd, B., Om datainsamling i surveyundersökningar (Per Kempe) 255
Brändström, A. & Sundin, J. (ed), Tradition and Transition (Rune Tryggveson) . 258
Hofsten, E., Indien. Om samhället och befolkningsproblemen (Ingvar Holmberg) 261
Hofsten, E., Demografins grunder (Ingvar Holmberg) 263
Rolski, T., Stationary Random Processes Associated with Point Processes (Martin Ribe) 263



P. W. Wargentin 1717–1783

"Den svenska statistikens fader"

Det är i år 200 år sedan Pehr Wilhelm Wargentin – den svenska statistikens fader – avled. För att högtidlighålla minnet härav har SCB publicerat en minnesskrift (Urval nr 15). Skriften innehåller dels sju originaluppsatser av Wargentin ur Kungl. Svenska Vetenskapsakademiens handlingar för åren 1754, 1755 och 1766, dels en nyskriven uppsats av Erland Hofsten om Pehr Wargentin och grundandet av den svenska befolkningsstatistiken, om Tabellverkets förhistoria samt om Wargentin som person och som demograf[1].

Därtill arrangerade SCB ett seminarium den 19 januari 1983 om Wargentin, då ovannämnda skrift presenterades av Erland Hofsten. Dessutom hölls föredrag av professor Erik Bylund (Umeå) om "Befolkningsstatistiken i historisk-geografiskt perspektiv",

[1] Boken är recenserad på sid. 253 i detta häfte.

och av professor Sune Åkerman (Umeå) om "Befolkningsstatistikens roll i den historisk-demografiska forskningen". Parallellt med detta seminarium hade det nordiska statistiska utskottet för befolkningsstatistik ett möte i Stockholm. De fyra representanterna för övriga nordiska länder i detta utskott lämnade i samband med Wargentinseminariet korta anföranden om befolkningsstatistiken i resp. land. Tre av dessa anföranden publiceras här nedan.

Befolkningsstatistikkens udvikling i Danmark indtil 1849

af Lene Skotte[1]

Ved denne sammenkomst til minde om Pehr Wargentin vil det danske bidrag være ganske korte rids over den danske befolkningsstatistiks historie og udvikling. Da vi i Danmark på mange måder – som sædvane er – halter langt efter Sverige, må dette rids nødvendigvis gå betydeligt længere frem i tiden end Wargentin-epoken i Sverige gør, nemlig til Grundloven 1849.

Inledningsvis vil jeg citere min kollega, kontorchef i Statens Statistiske Centralbureau Axel Holck, som i år 1900 i sin beskrivelse af Dansk Statistiks Historie 1800–1850 siger følgende:

"Gennem en Skildring af Statistikkens Fremadskriden vil der også falde lidt Lys hen over Menneskets skiftende Videlyst, dets samfundsinteresserede Trang til at klargøre sig og forstaa den Mekanisme, hvoraf det selv er et Led, til bogholderimæssig at kontrollere Statsmaskineriet. Samfundsborgeren, der skal lovgive eller interesseret leve med i Lovgivningsarbejdet, maa give Statistikken en Hylde i sin Haandbogsamling. Statistikkens Liv staar i ret naturlig Forbindelse med Væksten i Samfundets politiske Selvbevidsthed, og det er derfor ingen Tilfældighed, naar større Deltagelse i det offentlige Liv og statistiske Fremskridt gerne gaa Haand i Haand. Danmark fik sit Tabelkontor – et velment, men ganske vist ikke vellykket Forsøg – under den økonomiske Frihedsæra i Slutningen af det attende Aarhundrede, sin Tabelkommission, da Provinsialstænderne gennemførtes, og sit statistiske Bureau samtidig med Grundloven".

En meget betydningsfuld del af de nævnte statistikinstitutioners opgave var af befolkningsstatistisk art, så derfor er en væsentlig del af befolkningsstatistikkens udvikling sammenfaldende med den officielle generelle statistikorganisations udvikling.

Til at præge og føre denne udvikling frem kan vi i Danmark – som i andre lande – pege på en lang række personer. Men vi har ikke en statistikkens fader – endsige moder – som Wargentin, men netop mange, og af forskellige professioner: læger, historikere, geografer, jurister, statsøkonomer, polytek-

[1] Kontorchef vid 1 kontor (befolkningsstatistik), Danmarks Statistik, Köpenhamn.

nikere. Befolkningsstatistikkens udviklingslinie fremstår som et produkt af mange små og store indsatser af disse mennesker, som har opfyldt statistikbehov hos statsmagten eller i det private næringsliv – eller som har dyrket den statistiske virksomhed con amore.

Endelig skal vi ikke glemme – når vi nu taler om statistik – tilfældighedernes betydning for en statistiks fremkomst: Og også de organisatoriske rammer – eller mangelen på samme – har haft betydning.

Så statistikken – og så sandelig også befolkningsstatistikken – kan nok siges at have udviklet sig gennem mange faktorers brogede samvirke.

Grundlaget for den danske statistik over fødsler, dødsfald og vielser er den gejstlige registrering – kirkebøgerne – fra forordningen af 1646. Denne forordnings bestemmelser er gennem de to følgende århundreder blevet strammet, sådan at registreringen gradvist er blevet mere ensartet gennemført og optegnelserne mere pålideligt ført.

På grundlag af kirkebøgerne har præsterne ført årlige lister over tilgang og afgang i befolkningen. Listerne var i begyndelsen uensartede og blev ført med forskellige forsinkelser. Fra 1735 blev listerne ført i fastere form og det blev formentlig muligt at sammentælle til landsdækkende tal, idet biskopperne i Danmark og Norge skulle indsende lister over fødte og døde fra sognepræsten til Kancelliet. Fra 1775 indførtes standardskema for fødte, døde og viede og der blev foretaget opdeling på køn. Fra 1835 registreredes døde på 5 års aldersgrupper samt døde under et år.

I hele denne registrering manglede jo det, som man i Sverige forlængst havde: nemlig mandtal- altså bestandsopgørelser. Vi havde nok af og til mandtalsopgørelser for visse områder, men de må siges at være særdeles upålidelige.

De manglende bestandsopgørelser blev først mulige med de egentlige folketællinger.

Den første folketælling blev afholdt i 1769 og omfattede Danmark, Norge, Island, Færøerne samt de sydlige besiddelser. Den næste tælling i 1787 omfattede kun Danmark og Færøerne, mens tællingen i 1801 tillige omfattede Norge og Island. Endelig omfattede tællingerne i 1834 og 1840 Danmark og Island. Efter 1840-tællingen blev det ved kgl. resolution bestemt, at der fremover skulle afholdes folketællinger hvert 5. år.

De to første tællinger var i deres gennemførelse primitive, såvel i indsamlings- som i bearbejdelsesfasen, og resultaterne ret tvivlsomme – ikke mindst fordi der i befolkningen var en udbredt frygt for, at tællingen skulle danne grundlag for skatteudskrivning. Karakteristisk er endvidere, at tællingsresultaterne blev holdt hemmelige ud fra hele tidens tankegang om, at et lands folketal o.l. oplysninger var statshemmeligheder.

Alligevel er indførelsen af folketællingstraditionen et særdeles betydningsfuldt led i den danske befolkningsstatistiks udvikling, fordi en folketælling giver mulighed for at belyse befolkningsstrukturen på et givet tidspunkt ud fra mange forskellige synsvinkler, og fordi rækken af tællinger muliggør belysning af ændringer i denne struktur. Specielt skal her peges på, at alle danske folketællinger giver oplysninger om befolkningens erhvervsforhold og om civilstandsforhold.

Med hensyn til organiseringen af statistikbearbejdelsen skal nævnes, at mens den gejstlige traditionsbundne registrering af befolkningens bevægelser gradvist blev centralt organiseret og styret, så var de op-

gørelser af mere eller mindre statistisk karakter, som statsmagten behøvede, blevet gennemført ved ad hoc oprettelse af det nødvendige tællingsapparat. Når opgørelsen var afsluttet nedlagdes tællingsapparatet. Det gælder også de første to folketællingers gennemførelse.

Men med regeringens ændrede indstilling til statistik i slutningen af 1700-tallet åbnes låget på klem for statshemmeligheder som folketal o. l. og hermed opstår behovet for et officielt permanent organ til belysning af statistiske størrelser som folketal og -sammensætning m. v.

Efter svensk forbillede oprettes derfor 1797(−1819) det dansk-norske tabelkontor. Kontoret blev imidlertid en skuffelse: 1801-folketællingen, der var dets væsentligste opgave, viste sig at være alt for stor, og kontorets besætning var ufuldstændig.

Da Norge og Danmark blev adskilt i 1814 faldt yderligere et væsentligt grundlag for kontoret bort.

Herefter oprettedes med stændervæsenet og folkerepræsentationens fordring på statistiske oplysninger til brug for skøn over de sager, de skulle udtale sig om tabelkommissionen i 1834. Kommissionen bestod indtil 1848, hvorefter det statistiske bureau oprettedes i året for Grundlovens tilblivelse, 1849.

I tabelkommissionens regi bearbejdedes folketællingerne 1834, 1840 og 1845, og i forbindelse med disse folketællinger publiceredes fødsels- og dødstallene m.v. for de mellemliggende år. Først da, dvs. ca. 100 år efter at Sverige råder over fuldstændige oplysninger om folketallet og dets bevægelser har vi i Danmark tilsvarende oplysninger plus noget mere – nemlig, som tidligere nævnt, bl. a. oplysninger om befolkningens erhvervsfordeling.

Norge og Island er flere gange nævnt sammen med Danmark både i forbindelse med folketællingerne og med bestemmelser om kirkebogsføringerne. Når Island, Norge og Danmark således på befolkningsstatistikkens område har fælles rødder langt tilbage, beror det naturligvis på det statsretlige forhold, landene havde i den periode, der her er tale om: Danmark og Norge indtil 1814, og Danmark og Island endnu længere.

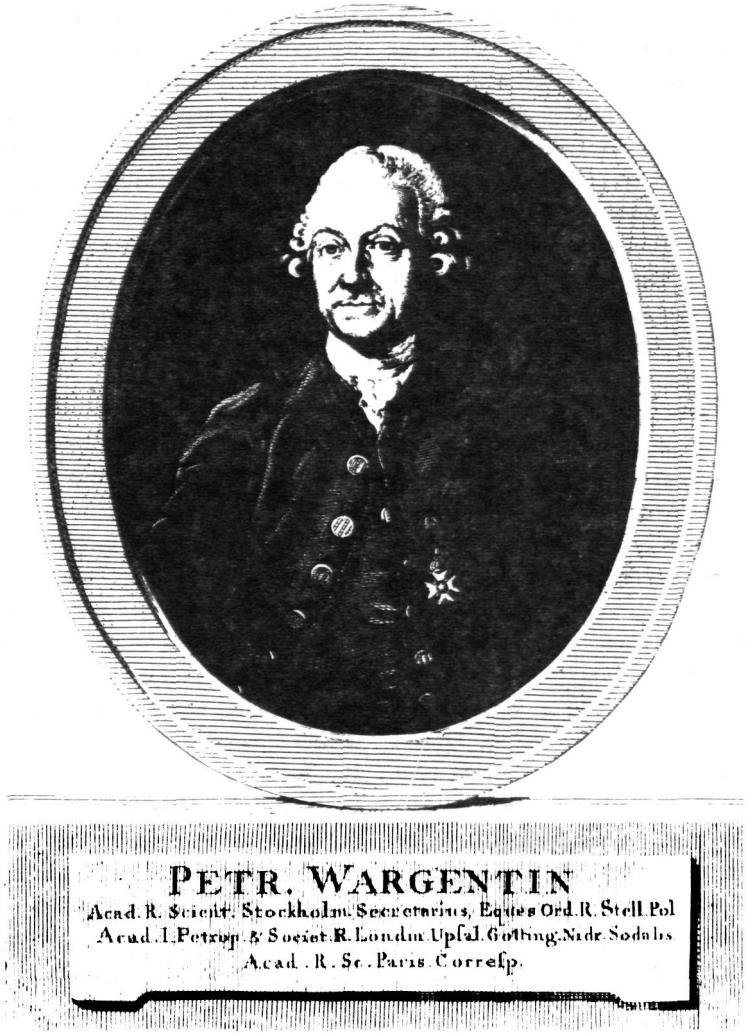
Én ting er ligheder og fælles tradition: men mon ikke der tillige er en del forskelle i statistiktradition og statistisk udvikling, som man i Norge og i Island kan pege på?

Nordiska och finländska tankar med anledning av 200-årsminnet av Pehr Wargentins död

av Göran Strengell[1]

Det är en stor glädje och ära för mig att få deltaga i detta möte där man högtidlighåller 200 års minnet av en stor svensk befolkningsstatistiker, Pehr Wargentin. Han har betytt mycket för Sverige, men även för Finland som ända till år 1809 eller i ca 600 år var förenat med Sverige. I dag har vi anledning att litet blicka bakåt och söka inspiration för våra arbetsuppgifter.

[1] Bitr. avdelningschef vid Statistikcentralen, Helsingfors (individstatistik).



Jag börjar med en varm hälsning från min chef, generaldirektören vid Statistikcentralen i Finland, professor Olavi Niitamo. Han sade sig vara mycket stolt över det faktum, att Pehr Wargentin hade rötter även i Finland; hans far och farfar var ju födda där, "Brain-drain" är således inte något nytt fenomen.

Sedan framför jag en hälsning från Nordiska utskottet för befolkningsstatistik, NUB, vars ordförande jag varit i tio års tid. Utskottet skulle egentligen ha hållit sitt möte i Oslo, men på grund av detta Wargentin-seminarium flyttades det hit till Stockholm. Och det var ju en god idé!

Detta utskott för befolkningsstatistik är

ett av det 15-tal statistiska utskott som de nordiska länderna för närvarande håller sig med. Detta utskott är också det äldsta av de nuvarande utskotten. Det grundades i juni 1952 i Köpenhamn och fyller inkommande sommar 31 år. En kort tid i jämförelse med det 200 års minne som nu firas, men dock innehållsrik. Utskottet grundades av dåvarande cheferna Petter Jakob Bjerve, Einar Cohn, Walter Lindberg, Klemens Tryggvason och Karin Kock.

Många kanske undrar vad de nordiska utskotten sysslar med. Beträffande utskottet för befolkningsstatistik så har det under de gångna 30 åren sysslat med ett otal olika frågor. En lista över dem är på samma gång en återspegling av den utveckling som ägt rum inom befolkningsstatistiken. Jag skall ge några exempel.

På 1950-talet behandlade utskottet befolkningsprognoser, avgränsningen av tätorter, statistik över födelser, dödsfall, äktenskap, klassificering av yrken, tabuleringsprogram för befolkningsrörelsen samt planeringen av 1960 års folk- och bostadsräkning. Många av dessa frågor skulle säkert ha intresserat även Pehr Wargentin.

På 1960-talet dominerades utskottets arbete av frågan om socio-ekonomisk ställning, yrkesklassificering, yrkesställning, nordiska flyttningsbetyg, sameräkning, labour-force-undersökningar, övergång till automatisk databehandling, inrättande och utnyttjande av centrala befolkningsregister jämte ibruktagande av personnummer. Det var i början av 1960-talet som frågan om "arkivstatistiska system" introducerades av Svein Nordbotten från Norge. Även planeringen av 1970 års folk- och bostadsräkning gav utskottet arbete.

På 1970-talet gick utvecklingen vidare i den riktning EDB samt den sociala utvecklingen förutsatte: man diskuterade inom utskottet ett allt mer omfattande utnyttjande av administrativa register inte bara inom folk- och bostadsräkningarna och den löpande befolkningsstatistiken. Det arkivstatistiska tänkandet hade trängt djupt in i vårt medvetande och präglat våra attityder. Ett nytt problem inom statistiken blir allt mera brännande: de papierlösa äktenskapen. År 1979 bryter man ut frågor rörande arbetskraftsstatistik som får ett eget utskott. Statistiken över arbetsmarknaden har blivit allt mera central. På tröskeln till 1980-talet talade man allt mera om longitudinella studier över 10–20 år vilka möjliggörs av datorer, relativ stabilitet i klassificeringar samt lång användning av personnummer.

Då man firar minnet av Pehr Wargentin, den svenska statistikens fader, kan man fråga sig, vem är då den finska statistikens fader? Har det funnits personer som betytt särdeles mycket för uppkomsten av statistiken i Finland? Det är lätt att säga vem som är den finska befolkningsstatistikens fader. Han hette Pehr Wargentin! Men strängt taget finns det nog flera faderskandidater.

Banden till Sverige brast år 1809 då Finland blev ett storfurstendöme under den ryske kejsaren och upphöjt till "nation". Men redan före 1809 fanns det personer i Finland som odlade den statistiska vetenskapen. Den mest kända och även en pionjär var biskopen i Åbo Carl Fredrik Mennander, samtida med Wargentin. År 1769 publicerade han på anmodan av Wargentin en uppsats om "Åbo stifts tilväxt i folkrikhet på 30 år". Han granskade även primärmaterialet kritiskt och ansåg bl.a. att en och annan präst inte alltid var så noggrann med folkbokföringen.

Under åren 1809–1865 sköttes Tabellkommissionens åligganden i Finland av en avdelning inom den kejserliga senaten för Finland, dvs. inom regeringen och närmast

som bisyssla av en person inom ekonomidepartementet. År 1812 antog senatens ekonomidepartement nya formulär för insamlande av uppgifter från församlingarna. Särskilda problem vållades av det faktum, att landet nu också hade en ortodox minoritet med egna församlingar. Ett sammandrag för tiden 1807–1812 färdigställdes år 1814 och publicerades delvis i Finlands Allmänna Tidningar. Försök att göra en fullständig statistisk beskrivning över hela Finland gjordes under denna tid av bl. a. Carl Christian Böcker (död 1841) och Rob. H. Rehbinder som år 1836 utgav "Underdåning berättelse angående Storfurstendömet Finlands tillstånd och förvaltning ifrån och med år 1826 intill närvarande tid". Berättelsen innehöll uppgifter om folkmängd, nativitet och mortalitet samt folkmängdens ökning och minskning. Ett namn värt att lägga på minnet är Gabriel Rein, som 1839 utgav Statistische Darstellung des Grossfürstenthums Finnland", som blev den första i tryck utgivna statistiska framställningen av Finland.

Två andra namn som bör nämnas är F. J. Rabbe och Paavo Tikkanen. Rabbe forskade i dödlighetsförhållanden och folktillväxt medan Tikkanen bl. a. hårt kritiserade bristen på ett ordentligt organiserat statistikväsen. På samma sätt som Wargentin påpekade Tikkanen vikten av att uppgifterna sammanställs och publiceras varvid de personer som inlevererar primäruppgifterna inspireras till ökad ansvarskänsla. Det var även på Paavo Tikkanens initiativ som tabeller i församlingarnas, domkapitlens och prosteriernas arkiv från 1749–1809 insamlades till senatens arkiv.

Den 4 oktober 1865 undertecknades det fundationsbrev, som betydde att Finland fick en statistisk byrå. Dess första chef blev ovannämnda professor, kanslirådet Gabriel Rein. Rein hade också varit medlem av den kommitté, som sedan 1861 förberett byråns grundande.

Statistikcentralen räknar år 1865 som sitt födelseår. Här följer ett utdrag ur fundationsbrevet:

> "Alexander den Andre ... Storfurste till Finland ...
> Till ... Gabriel Rein.
> Sedan Vår Senat för Finland ... hemställt att ... använda 16 000 mark årligen ... hafva Vi ... velat ... utse Eder att ... i egenskap av ledare öfvertaga merberörda arbeten ... Eder till kännedom och underdåning efterrättelse.
> Helsingfors den 4 Oktober 1865."

Tabellverket grundades av en konung den 3 Februari 1748, Statistikcentralen av en kejsare 117 år senare. Båda ämbetsverken och med dem hela Norden hyllar Pehr Wargentin och hans insatser som banbrytare inom statistiken.

Befolkningsstatistikens utveckling i Norge

av Gerd S. Lettenström[1]

1. Folketellinger

Alt fra 1660-årene finnes det bevart lister fra mantall, men dels gjaldt de ikke hele landet, og de tok ikke med kvinner unntatt de som var eiere av gård.

Den første egentlige folketellingen i kongeriket Danmark-Norge ble holdt i 1769. Det er uklart hvilke behov den var tenkt å dekke utover den almene nytte. Fra Rentekammeret i København fikk biskoper og stiftsbefalingsmenn brev med vedlagt et passende antall tabeller som skulle fordeles til magistrater og sogneprester. En tabell skulle vise folkemengdens fordeling etter kjønn og alder, den andre en fordeling på kjønn og yrke. De første resultater fra denne tellingen ble gitt i 1786 i Materialen zur Statistik der Dänischen Staaten.

I 1880-årene, dvs. 100 år senere, var det i Det statistiske byrå stor interesse for bearbeiding av de gamle mantall, og også tellingen fra 1769 skulle bearbeides og publiseres. Det er uklart hvorfor publiseringsplanene ble skrinlagt.

I år 1800 ble det så sendt ut kongelig befaling om å holde folketelling i Danmark, Norge og Island i 1801. Dette ble den første nominative tellingen. Også nå var det trolig den store interesse for befolkningen som lå bak. Resultatene fra denne tellingen ble gitt i avhandlinger av Pram og Thaarup og i tabeller vedkommende folketellingene i 1801, 1825 (Cristiania 1874).

De følgende fem folketellinger var summariske og lignet mer 1769-tellingen. Folketellingen i 1865 betydde et stort fremskritt da bl. a. selvtellingssystemet ble tatt i bruk. Oppgavene var også langt mer detaljerte enn ved tidligere tellinger.

2. Folkemengdens bevegelse

Helt fra de eldste tider har det i Norge, slik som i Sverige og Danmark, vært den geistlige embetsmann som har ført de offisielle registre over giftermål, fødsler og dødsfall.

Allerede i kirkeritualet av 1685 og i Christian V. norske lov av 1687 finnes påbud om at prestene skulle føre kirkebok over vigde, fødte og døde, men først i 1812 ble det fastsatt bestemte skjema for kirkebøkene i Danmark og Norge. Etter adskillelsen fra Danmark ble det gitt nye bestemmelser i Norge.

Det er uklart fra hvilket tidspunkt en begynte å bruke kirkebøkene som grunnlag for statistiske oppgaver. En skrivelse til biskopene fra 1720 omhandler innsending av påbudte oppgaver over fødte og døde, men først i 1735 fikk innberetningene fast form. Da ble biskopene i Norge og Danmark pålagt ved hvert års utgang å innhente oppgave over alle fødte og døde fra sogneprestene, og sende en utførlig beretning om dette til Commerce Collegium i København.

I 1775 ble det fastsatt et bestemt skjema for de årlige innberetninger til kollegiet over gifte, fødte og døde. Blant annet skulle døde menn og kvinner fordeles etter alder i 10-årige aldersklasser.

De første offisielle statistiske oppgaver over gifte, fødte og døde ble gitt fra 1801.

[1] Byråsjef ved Statistisk Sentralbyrå, Oslo.

For enkelte perioder før den tid finnes endel tabeller vedrørende folkemengdens bevegelse i Norge i publikasjoner som gjaldt de danske stater.

De statistiske tabeller i Norge ble utarbeidd av "tabellkontoret" under Finansdepartementet, og den første publikasjonen kom i 1839. Tabellkontoret fikk etterhvert en mer selvstendig stilling, og i 1876 gikk det over til eget direktorat under navn av Det Statistiske Sentralbyrå.

For tiårsperioden 1856–1865 ble statistikken over folkemengdens bevegelse utgitt som egen publikasjon (NOS C nr. 1). Denne publikasjonen betegner det første skritt i retning av en moderne behandling av de statistiske data, og den ble et mønster for de senere fremstillinger om folkemengdens bevegelse. Den ble utarbeidd under ledelse av byråsjef Anders Nicolai Kiær, senere direktør i Det Statistiske Sentralbyrå. I en omfattende innledning til tabellene foretok han en vitenskapelig analyse av resultatene. Publikasjonen gav sammenlignbare oppgaver over folketallet og folkemengdens bevegelse tilbake til 1801. Det var gitt beregninger vedkommende ekteskap og fødsler til belysning av giftermålshyppighet og fruktbarhet. Endelig må nevnes en mortalitetstabell for Norge for årene 1856–1865, den første som den gang ble offentliggjort i Norges Offisielle Statistikk. Resultatene var også sammenlignet med tidligere beregninger over dødeligheten og middellevetiden i forskjellige aldre i perioden 1821–1850, beregninger som var foretatt av statistikeren og samfunnsforskeren Eilert Sundt i hans arbeid Om dødeligheten i Norge, utgitt i 1855.

3. Eilert Sundt

Eilert Sundt ble født i 1817, akkurat 100 år etter Wargentin. Han var utdannet teolog og var universitetsstipendiat i kirkehistorie. Ved sitt arbeid som lærer i tukthuset kom han i kontakt med fantene (landstrykere) og fattet interesse for disse folk. Senere utvidet han sin interesse til de lavere klasser i sin alminnelighet. Han pekte på ønskeligheten av å undersøke de moralske og materielle forhold de lavere klasser levde under, for så å utbre opplysninger om dem.

Med stipendium foretok han mangeårige reiser rundt i hele landet. Frukten av hans studier ble en rekke verdifulle verker, som kom ut fra 1855 og utover. Fra 1855 har vi "Om giftermål i Norge" og "Om dødeligheten i Norge".

Sundt utviklet seg til en sosialstatistiker av høy rang. Mens han fra først av sto noe fjernt fra den statistiske metode og dannet seg et bilde av forholdene ved å spørre seg for hos folk, fikk han etterhvert øyene opp for at tallene gav det beste svar.

Han brukte vinterhalvårene til omfattende studier av topografisk-statistisk art. Han studerte inngående det danske og norske tabellverk, statistikk over Sverige og samlingsverk som inneholdt statistisk materiale for de danske provinser og danske stater. Han orienterte seg også i europeisk samfunnsvitenskapelig litteratur. Fattigproblemet førte til at Sundt alt i begynnelsen av 1850-årene kastet seg over studiet av europeisk vitalstatistikk. Dødelighetsproblemet hadde alt lenge engasjert europeiske forskere, og mange fremhevet at den gjennomsnittlige levealder var et viktig kriterium på folkets velferd og sivilisasjon.

I hans verk om dødeligheten foreligger dødelighetstabeller for tre ti-års perioder før 1850, han gir tall særskilt for de enkelte deler av landet, og han foretar sammenlikninger med andre land. I sammenlikningen med Danmark, Sverige og England nevner han også en tabell for Sverige utarbeidd av

astronomi-professor Selander for perioden 1830–35.

I det vitenskapelig sett betydeligste verket "Om giftermål" utviklet Sundt den teori som siden er blitt kalt Eilert Sundts lov.

Hans teori gjaldt variasjonene mellom fødselskullene med visse mellomrom. Et fåtallig fødselskull vil i sin tid gi en fåtallig ungdomsflokk, mens et stort fødselskull vil gi mange unge og flere vigsler etter en bestemt periode. Han fulgte variasjonene bakover i tid så langt han hadde materiale. Fra 1860-årene ble denne bevegelse mer uregelmessig som følge av utvandringen. Grunntanken i hans teori er imidlertid blitt bekreftet også i senere år.

Sundt så ikke på sine arbeider bare som vitenskapelige publikasjoner, men også som nyttig lesning for folket. Dette preget hans stil og tilrettelegging av stoffet. Han legger tydelig vinn på å skrive enkelt og lett forståelig.

4. Anders Nicolai Kiær

Kiær var født i 1838 og døde i 1919. At Statistisk Sentralbyrå ble opprettet i 1876, skyldtes i høy grad hans initiativ og arbeidsinnsats. Ikke minst på befolkningsstatistikken gjorde han en betydelig innsats. Hans oversikter over folkemengdens bevegelse 1856–1885 og resultatene fra folketellingen 1875 ble av grunnleggende betydning.

Ved Kiærs død i 1919 skrev den danske statistiker Harald Westergaard om ham at han var den skandinaviske statistikks nestor. Videre ble nevnt den store andel som Kiær hadde i utviklingen av norsk offisiell statistikk. Han omfattet befolkningsstatistikken med levende interesse, og folketellinger, folkemengdens bevegelse og dødelighetstabeller bærer tydelig vidnesbyrd om hans arbeidskraft og initiativ. Westergaard omtaler videre Kiærs innsats i det internasjonale arbeid, der han var med å grunnlegge Internasjonalt Statistisk Institutt. Han nevner også at det var Kiær som tok initiativ til et nærmere samarbeid mellom de tre skandinaviske lands offisielle statistiske byråer som begynte i 1889.

Det ble sagt om Kiær at han hadde noe felles med den bekjente norske sosialforsker Eilert Sundt. Han hadde, sier Westergaard, i hvert fall som denne, vidtgående samfunnsinteresser.

Simpel sammenligning af indkomstoplysninger[1]

af Mogens Nygaard Christoffersen[2]

1. Indledning

Spørgsmålsformulering spiller en særlig rolle, når det drejer sig om belastende eller kontroversielle emner. I denne forbindelse nævner mange forfattere indkomstspørgsmål som et eksempel på spørgsmål med dels en lav besvarelsesprocent, dels med en høj grad af unøjagtighed. Undersøgelserne heraf bygger oftest på en simpel sammenligning af to forskellige spørgeteknikker. Det traditionelle indkomstspørgsmål "Hvor stor var Deres skattepligtige indkomst sidste år?" bliver sammenlignet med resultaterne af et batteri spørgsmål, der kortlægger de samlede indkomstforhold. Selvangivelsen, der i realiteten er et detaljeret postspørgeskema, er et specialtilfælde af den sidste type spørgeteknik.

I en undersøgelse af Monroe G. Sirken m. fl. finder man, at den mediane indkomst ligger 14 pct. højere ved den detaljerede kortlægning end oplyst ved anvendelsen af tre brede spørgsmål.

Monroe G. Sirken konkluderer, at det er en mere pålidelig metode at spørge om tidligere erhvervsmæssig beskæftigelse end blot at stille nogle få indkomstspørgsmål. Det er værd at undersøge, om det forøgede besvær hermed står i rimeligt forhold til den forøgede præcision[3].

Michael E. Borus fandt ved sammenligning af det traditionelle indkomstspørgsmål med en detaljeret spørgeteknik, der kortlagde respondentens erhvervsaktivitet, at det traditionelle indkomstspørgsmål var unøjagtigt for de højere indtægtsgrupper[4].

William B. Locander konkluderer, at jo flere spørgsmål, der bliver stillet om specifikke indkomstkilder, jo mere præcis bliver opgørelsen over respondentens samlede indkomst. Det fremhæves herudover, at respondenterne er særlig påvirkelige af mindre ændringer af spørgeformen, når det drejer sig om besvarelse af belastende og kontroversielle emner[5].

2. Datamaterialet i forsøget

I undersøgelsen af den 20–69 årige befolknings leveforhold blev der i august–september 1976 interviewet en landsdækkende, repræsentativ stikprøve på godt 5000 personer[6]. Blandt de mange spørgsmål om leveforholdene blev også medtaget et om skattepligtig indkomst i 1975. Spørgsmålet

[1] Kilde: Socialforskningsinstituttets studie 46, Bjarne Hjorth Andersen og Mogens Nygaard Christoffersen: Om spørgeskemaer. København 1983.

[2] Forskningsassistent vid Socialforskningsinstituttet, København.

[3] Monroe, G. Sirken m. fl.: The Survey of Consumer Finances and the Census Quality Check. I National Bureau of Economic Research: An appraisal of 1950 Census Income Data. New York 1958, s. 127–168.

[4] Michael E. Borus: Response Error and Questioning Technique in Surveys of Earnings Information. Journal of the American Statistical Association, vol. 65, nr. 330, 1970, s. 556–575.

[5] William B. Locander & John P. Burton: The Effect of Question Form on Gathering Income Data by Telephone. Journal of Marketing Research, vol. 13, nr. 2, 1976, s. 189–192.

[6] Materialet er udførligt beskrevet i Erik Jørgen Hansen: The Distribution of the Living Conditions in Denmark. Socialforskningsinstituttets publikation 110, København 1982.

lød: "Hvor stor var Deres egen skattepligtige indkomst i *1975*?"

De interviewede blev i undersøgelsen præsenteret for et kort med 20 indkomstintervaller, hvorefter de blev bedt om at indplacere deres indkomst i et af disse 10 000 kroners intervaller.

Ved de traditionelle former for pålideligheds kontrol udnytter man en viden eller formodning om alternative datakilders pålidelighed, hvorefter datakilderne sammenholdes objekt for objekt.

Ofte vil man være i den situation, at der ikke findes alternative objektive datakilder, der kan anvendes til evaluering af interviewdatas gyldighed. Imidlertid findes der i Danmark udviklet et rimeligt pålideligt statistikregister for skattepligtige indkomster. Til brug for en kvantitativ evaluering af indkomstoplysningerne blev de relevante registeroplysninger udtrukket for de interviewede personer.

Skattevæsnets slutligningsopgørelse blev opgivet i 100 kronersintervaller.

Indkomstspørgsmålene synes at indeholde flere dimensioner. For det første det *socialt acceptable aspekt*, som både kan betyde, at de lavtlønnede overvurderer deres indkomst, og at personer med de højeste indkomster undervurderer deres indkomster. Dette aspekt kan også resultere i høje andele af ved ikke/vil ikke svare i ydergrupperne.

For det andet indeholder indkomstspørgsmålet et *vidensaspekt*, idet begrebet skattepligtig indkomst er et teknisk-administrativt, kompliceret begreb. Personer, der har fortrolighed og erfaring med sådanne, har lettere ved at besvare spørgsmålet. For visse persongrupper gælder det, at opgørelsen af deres indtægtsforhold er af særlig kompliceret natur. Specielt for de mindre selvstændige erhvervsdrivende er indkomstopgørelse i skatteteknisk forstand betydeligt vanskeligere at håndetere end for lønmodtagere.

Endelig indholder sådanne retrospektive indkomstspørgsmål en *hukommelsesmæssig faktor*, idet respondenterne bliver udspurgt i august–september 1976 om, hvad de skrev i deres selvangivelse et halvt år tidligere. Spørgsmålet er i vid udstrækning besvaret ud fra hukommelsen, hvilket ydermere vil formindske besvarelsernes præcision og samtidig forøge mulighederne for, at andre fejlkilder får en større indflydelse på besvarelsern.

3. Forsøgets resultater

Sammenligningen mellem interviewpersonernes egne og skattevæsnets oplysninger viste, at de 20–69 årige danskere overvurderede deres skattepligtige indkomst med 2,5 pct. i forhold til skattevæsnets slutligningsopgørelse. Når man alene henholder sig til gennemsnitsbetragtninger på landsplan, viser der sig at være en påfaldende god overensstemmelse mellem den simple spørgeform og det meget grundige, detaljerede selvangivelsesskema, der som bekendt tilmed er forbundet med et kontrolsystem og et sanktionssystem over for forsømmeligheder i udfyldelsen.

Imidlertid dækker disse gennemsnitsbetragtninger over store forskelle mellem forskellige befolkningsgrupper og indtægtsniveauer.

Det synes umiddelbart vanskeligt at give en beskrivelse af den noget diffuse sammenhæng mellem skattevæsnets ligningsopgørelse og undersøgelsens indkomstoplysninger (se fig. 1). Analysen viste imidlertid, at denne omfattende information lod sig reducere til et simpelt statistisk udtryk:

$Y = A \cdot X + B$, (den fuldt optrukne linje), hvor Y er den under interviewet oplyste

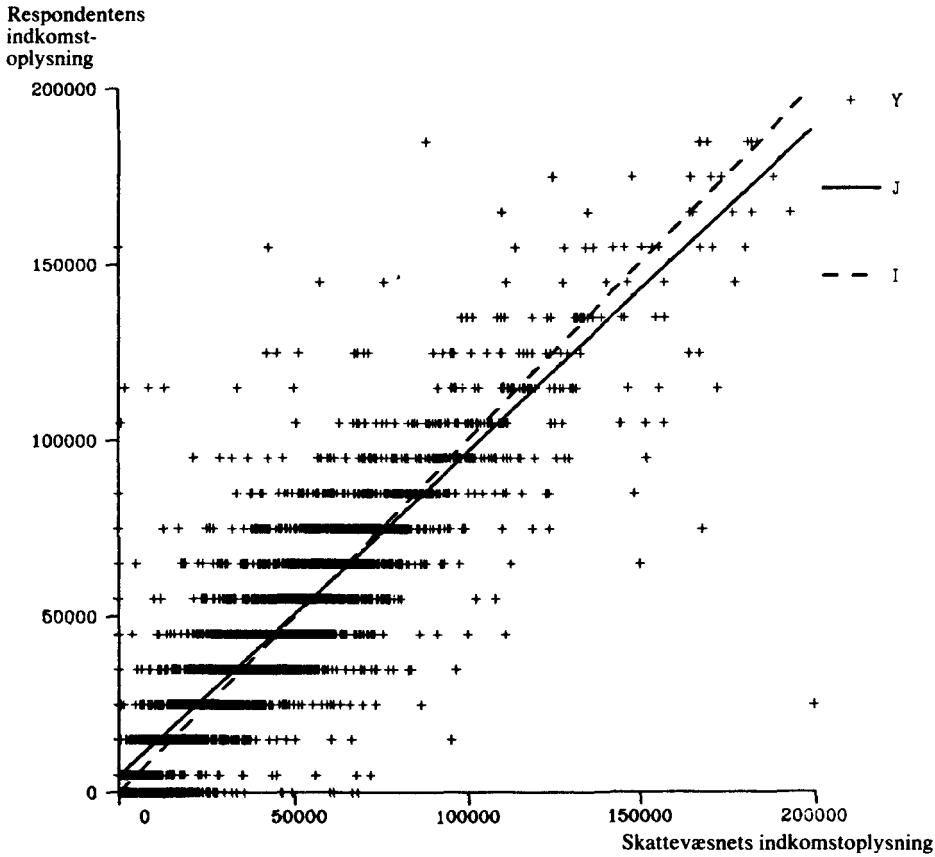


Fig. 1. Skattevæsnets slutligningsopgørelse, afbildet over for respondentens egne indkomstoplysninger

Værdierne er angivet i 1975-års kroner. Punkterne angiver skattevæsnets slutligningsopgørelse for indkomståret 1975 sammenholdt med personens egne indkomstoplysninger i levekårsundersøgelsen for alle personer, der havde indtægter under 200 000 kr. I alt er afbildet 4 901 punkter, svarende til 95 pct. af samtlige interviewede personer med oplyst indkomst. Den stiplede linje angiver identitetslinjen, hvor skattevæsnets oplysninger stemmer fuldstændig overens med respondenternes oplysninger. Den fuldt optrukne linje angiver den gennemsnitlige overensstemmelse mellem de to oplysninger. Den fuldt optrukne linje er fremkommet ved lineær regression. Skattevæsnets indkomstoplysninger er angivet i 100 kronors intervaller, mens respondentens indkomstoplysninger blev angivet på forevisningskort med 10 000 kr.-intervaller.

skattepligtige indkomst, og X er slutligningsindkomsten.

Analysen viser, at 80 pct. af variationerne i svarene på det simple indkomstspørgsmål lader sig forudsige ved hjælp af regressionslinjen ud fra den skattepligtige indkomst[1].

[1] Norman, H. Nie m. fl.: SPSS, Statistical Package for the Social Sciences. New York, 1975, kap. 18.4 og kap. 20.

Tab. 1. Den procentandel, der besvarer spørgsmålet om skattepligtig indkomst, fordelt efter erhvervsmæssig stilling og socialgruppe

	Socialgruppe				
	I+II	III	IV	V	Antal
	pct.				
Erhvervsmæssig stilling:					
Selvstændige	86	90	95	–	671
Funktionærer	94	97	97	–	1587
Arbejdere	–	–	99	96[1]	1362
Antal	456	890	1264	1010	3620

[1] Ufaglærte arbejdere.

For de indtægter, der lå under 200000 kr., og som udgjorde 95 pct. af samtlige respondenter, fandtes følgende sammenhæng mellem den angivne indkomst og slutligningsopgørelsen:

Opgiven indkomst = slutligningsindkomst · 0,92+4400[1].

[1] Den additive konstant er angivet i 1975-års kroner.

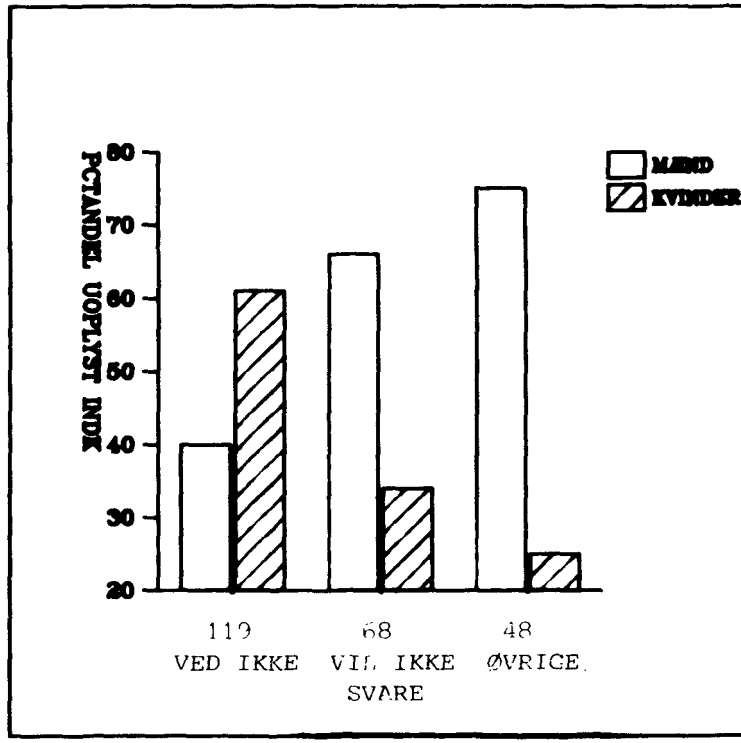


Fig. 2. Fordelingen af personer med uoplyst svar på indkomstspørgsmål ved levekårsundersøgelsen. Procentandele opgjort på mænd og kvinder og nægterbegrundelse

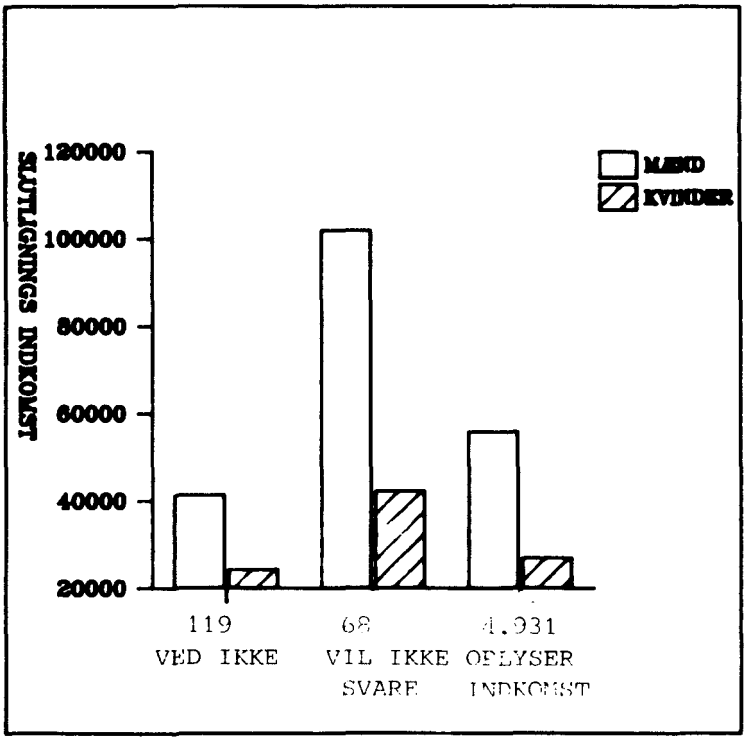


Fig. 3. Slutligningsindkomst for personer med oplyst og uoplyst indkomst ved levekårsundersøgelsen, fordelt på køn og nægterbegrundelse

Regressionsanalysen viser, at hældningskoefficienten er signifikant mindre end 1,0. Dette betyder, at der er en almindelig systematisk tendens til, at de allerlaveste indkomstgrupper svagt *overvurderer* deres indkomst, mens de, der ligger over gennemsnitsindkomsten, har en tilbøjelighed til at *undervurdere* deres indkomst, når de får stillet de traditionelle indkomstspørgsmål[1].

Mens de personer, der oplyser deres skattepligtige indkomst i levekårsundersøgelsen, gennemsnitligt overvurderer indkomsten ganske lidt, så er der samtidig en tendens til, at det er den med højere indtaegter, der *ikke* oplyser deres indkomst. Der var dog kun 5 pct. af samtlige adspurgte i levekårsundersøgelsen, der ikke ønskede eller ikke kunne oplyse deres skattepligtige indkomst.

Men der er en tydelig tendens til, at de, der ikke oplyser deres indkomst, i gennemsnit har en skattepligtig indkomst, der ligger en del over gennemsnittet for de andre. Gennemsnitsberegningerne skjuler store forskelle mellem forskellige beskæftigelsesgrupper. Mens 97 pct. af lønmodtagerne oplyser deres indkomst, er det kun 90 pct.

[1] Stikprøverne blev udtrukket i to trin med stratificeret udvalg, hvorved de statistiske beregninger er komplicerede. I nærværende analyse anvendtes chi-i-anden test, selv om denne forudsætter, at stikprøverne er udtrukket ved simpel tilfældig udvælgelse. En nærmere beskrivelse af omnibusundersøgelsernes stikprøveplan findes i Erik Jørgen Hansen (op. cit.).

af de selvstændige, der oplyser herom. De selvstændige erhvervsdrivende, der ikke oplyser deres skattepligtige indkomst i levekårsundersøgelsen, tjener i gennemsnit 80 pct. mere end de andre selvstændige.

Mens næsten alle unge oplyser om deres indtægtsforhold, er der en stigende tendens i de ældre aldersgrupper til ikke at ville eller kunne afgive oplysningen. Blandt de interviewede uden for erhverv er det kun pensionisterne, der er tilbageholdende med at oplyse om indkomstforholdene. De arbejdsløse, hjemmearbejdende, uddannelsessøgende oplyser gerne deres skattepligtige indkomster. Man kunne fortolke disse tendenser som en bekræftelse på, at besvarelsen af spørgsmålet om skattepligtig indkomst stiller store krav til hukommelsen og forudsætter et vist kendskab til begrebets teknisk-administrativt komplicerede natur.

I de ældre aldersklasser og hos pensionisterne finder man heller ikke nævneværdige forskelle mellem gennemsnitlige indkomster blandt dem, der oplyser, og dem, der ikke oplyser deres indkomstforhold i levekårsundersøgelsen.

Hvad angår kønsforskelle viser materialet, at der er forskellige svartyper hos dem, der ikke oplyser deres skattepligtige indkomst. Mens 2/3 af dem, der direkte udtrykker, at de ikke *vil* svare, er mænd, så er næsten 2/3 af dem, der svarer, at de ikke *kender* deres skattepligtige indkomst, kvinder.

Ved sammenligning af gennemsnitsindkomsten hos de respondenter, der henholdsvis oplyser og ikke oplyser deres indkomst, ses de største indkomstforskelle blandt de 30–50-årige mænd, der er funktionærer og selvstændige i socialgruppe I og II. De, der ikke oplyser deres indkomst, er yderligere karakteriseret ved fortrinsvis at bo i hovedstadens forstæder og have en højere uddannelse.

4. Konklusion

Undersøgelsen viste, at unøjagtighederne og de manglende oplysninger var meget forskelligt fordelt i de forskellige befolkningsgrupper.

Det er overvejende de højere statusgrupper med de store indkomster, der ikke oplyser deres indkomst. Hypotesen om, at spørgsmålet om den personlige skattepligtige indkomst også indeholder et element af social acceptabilitet, synes at bekræftes af nægteradfærden. De store indkomstforskelle mellem dem, der oplyser, og dem, der ikke oplyser deres indkomst, synes at hænge nøje sammen med personens eller familiens sociale status.

Personer, der oppebærer en relativt høj skattepligtig indkomst og derved afviger fra det normale, viser en mærkbar tilbøjelighed til at undlade at oplyse den skattepligtige indkomst.

Avgiftsprocenten i ATP

av docent Ann-Charlotte Ståhlberg[1]

0. Inledning

Kostnaderna för ATP ökar av flera skäl. Antalet ATP-pensionärer ökar i och med att allt fler tjänar in ATP-rätt. Allt fler kvinnor yrkesarbetar och tjänar in pensionsrätt. Varje ny årgång pensionärer har fler intjänade år än tidigare årgångar. Regeln att pensionen beräknas på de 15 bästa inkomståren får successivt genomslag. Förhållandet mellan den förvärvsarbetande delen av befolkningen och den del som får ATP förändras kraftigt. År 1982 går det fyra förvärvsarbetande på en ålders- eller förtidspensionär inom ATP. År 2030 är förhållandet två förvärvsarbetande per ålders- eller förtidspensionär inom ATP.

I denna artikel skall jag belysa hur stor vikt man kan lägga vid olika faktorer som bestämmer kostnaderna för ålderspensionen i ATP. I första hand är det de principiella förhållandena jag kommer att redogöra för, eftersom faktiska beräkningar i de flesta fall är uteslutna av kostnadsskäl.[2]

Låt oss till en början anta att varken befolkningsstorleken eller åldersfördelningen förändras. Då kan vi diskutera kostnader och kostnadsförändringar i pensionssystemet som inte har att göra med ändrade demografiska förhållanden. Låt oss också anta att den allmänna reallöneutvecklingen är noll för att skilja reallöneutvecklingens betydelse för pensionskostnaderna från andra kostnadsbestämmande faktorer. I ett renodlat fördelningssystem betalar den yngre generationen de gamlas pensioner och avgiften ett bestämt år kan bestämmas av följande relation:

$$q \times N \times \bar{y} = c \times R \times \bar{y} \text{ eller } q = c \times \frac{R}{N} \quad (1)$$

där q är avgiftsprocenten, N är antalet personer som betalar avgifter, R är antalet pensionärer, $\bar{y}$ är genomsnittsinkomsten och c är en faktor som bestämmer pensionsstorleken. I ATP är $c = 0{,}60$.

År 1979 får stå modell för vår steady-state population. Då fanns det 4,48 män mellan 16 och 64 års ålder på varje man som var 65 år och äldre. Det fanns 3,37 kvinnor mellan 16 och 64 år på varje kvinna som var 65 år och äldre och det fanns 3,85 personer (män och kvinnor) mellan 16 och 64 år på varje person (man och kvinna) som var äldre än 64 år.[3] Detta ger följande värden på avgiften q i ekvation (1):

$$\begin{aligned} q_{\text{män}} &= 13{,}4\ \%^{4} \\ q_{\text{kvinnor}} &= 17{,}8\ \% \\ q_{\text{totalt}} &= 15{,}6\ \% \end{aligned} \quad (2)$$

1. Vad 30-årsregeln och 15-årsregeln i ATP betyder för kostnaderna

Nu bestäms inte pensionen i ATP som ovan av genomsnittsinkomsten. För det första gäller att det normalt krävs 30 år med pensionsgrundande inkomster för att man

[1] Socialforskningsinstitutet vid Stockholms universitet.

[2] Denna artikel ingår i projektet "Det svenska socialförsäkringssystemet", som finansieras av Riksbankens jubileumsfond.

[3] Statistisk årsbok 1980.

[4] $q_{\text{man}} = 0{,}60 \times \frac{1}{4{,}48} = 13{,}4\ \%$.

skall få full pension. Som mest kan en person ha 49 år med pensionsgrundande inkomster. Men flera år än 30 påverkar inte pensionsstorleken. Däremot påverkas avgiftsunderlaget. Enligt arbetskraftsundersökningarna fanns 88 % av alla män och 74 % av alla kvinnor mellan 16 och 64 års ålder i arbetskraften år 1979. Av samtliga 16–64-åringar var 81 % i arbetskraften.[1]

Om vi antar att samtliga får ihop åtminstone 30 år med pensionsgrundande inkomster och att alla i arbetskraften också är sysselsatta, kommer under i övrigt samma förutsättningar som i ekvation (1) 30-årsregeln (regeln om att det normalt krävs 30 år för att få full pension) att innebära följande ändringar i pensionsavgiften:

Nu blir

$$\begin{aligned} q_{\text{män}} &= 15{,}2\ \%^{2} \\ q_{\text{kvinnor}} &= 24{,}1\ \% \\ q_{\text{totalt}} &= 19{,}2\ \% \end{aligned} \qquad (3)$$

För det andra gäller att det är de 15 bästa inkomståren som räknas. Hur mycket större genomsnittsinkomsten under de 15 bästa åren är jämfört med genomsnittsinkomsten under samtliga år finns det inga uppgifter på. Jag har grovt uppskattat skillnaden till ca 20 % i genomsnitt.[3] I så fall höjs avgiften i ekvation (3) med 20 % och

$$q_{\text{totalt}} = 23{,}1\ \%^{4}. \qquad (4)$$

Om vi dessutom tar hänsyn till att en allmän arbetslöshet påverkar avgiftsunderlaget, men inte pensionspoängen (ersättning från arbetslöshetskassa är pensionsgrundande men inte avgiftsgrundande) blir

$$q_{\text{totalt}} = 23{,}5\ \%^{5} \qquad (5)$$

i det fall vi antagit att arbetslösheten är 2 %.

Som vi ser höjer 15- och 30-årsreglerna under de givna förutsättningarna pensionskostnaderna med nästan 50 %. Det är klart att resultaten blir annorlunda om förutsättningarna ändras. Ju högre arbetskraftsdeltagande och ju jämnare livsinkomstprofiler, desto mindre betyder 15- och 30-årsreglerna. Till exempel skulle ett arbetskraftsdeltagande på 86 % i stället för 81 % (vilket är vad SCB räknar med i sin prognos för år 2000) betyda att 30-årsregeln höjer pensionskostnaden med 16 % i stället för 23 %.

2. Vad ändrade demografiska förhållanden betyder för pensionskostnaderna

I tab. 1 visas dels hur åldersstrukturen väntas se ut olika år fram till år 2025, dels pensionsavgiften de olika åren, när avgiften bestäms som i ekvation (1) och samma förutsättningar som i (1) gäller.

År 1990 är avgiften under de givna förutsättningarna 6 % högre än år 1979 som en följd av ändrade demografiska förhållanden. År 2005 är den 5 % lägre och år 2010 är den 4 % högre. Det är först därefter åldersförändringarna är stora nog för att ha en ordentlig effekt. År 2025 är t.ex. avgiften hela 23 %

[1] Information i prognosfrågor 1981:1.

[2] $q_{\text{man}} = 0{,}60 \times \frac{1}{4{,}48 \times 0{,}88} = 15{,}2\ \%$.

[3] Riksförsäkringsverket har uppgifter på medelpensionspoängen under perioden 1960-1982. För åren därefter finns ett antal prognoser över pensionspoängens utveckling. Jag har räknat ut att genomsnittsinkomsten de 15 bästa åren under perioden 1960–2009 är ca 15 % högre än genomsnittsinkomsten under hela perioden (reallöneutvecklingen antas vara 0 % åren 1983–1989 och därefter 1 % årligen). Detta ger sannolikt en underskattning av 15-årsregelns (regeln om att det är de 15 bästa inkomståren som bestämmer pensionens storlek) kostnader. En bättre uppskattning får man om man i stället följer den genomsnittliga inkomstutvecklingen för en kohort. För dem som är födda 1944 gäller för genomsnittet att de bästa 15 åren är drygt 20 % högre än hela genomsnittsinkomsten. För dem som är födda 10 år senare gäller ungefär detsamma. 20 % får därför gälla som en grov uppskattning.

[4] $q_{\text{totalt}} = 19{,}24 \times 1{,}20 = 23{,}1\ \%$.

[5] $q_{\text{totalt}} = 23{,}09 \times 1{,}02 = 23{,}5\ \%$.

Tab. 1. Befolkningen olika år[1] *samt pensionsavgiften olika år*

År	Antalet personer 16–64 år gamla på varje person som är 65 år och äldre	Pensionsavgiften (q) i procent
1979	3,85	15,6
1985	3,70	16,2
1990	3,63	16,5
1995	3,74	16,0
2000	3,95	15,2
2005	4,04	14,8
2010	3,70	16,2
2015	3,32	18,1
2020	3,19	18,8
2025	3,13	19,2

högre än år 1979. Enbart förskjutningar i åldersstrukturen mot flera gamla per person under 65 år har således inte den dramatiska effekt på kostnaderna som man kanske tror många gånger.

3. Vad förtidspensionerna betyder för kostnaderna

Utvecklingen av antalet förtidspensionärer spelar en stor roll för kostnadsutvecklingen i ATP. Fler förtidspensionärer innebär både att pensionsutbetalningarna ökar och att avgiftsunderlaget minskar. Antalet förtidspensionärer väntas öka. Man räknar med att de år 2010 kommer att vara dubbelt så många som år 1979.[2] Tar vi hänsyn till förtidspensionärerna får vi färre personer i åldrarna 16–64 år på varje pensionär (ålders- och förtidspensionär). Hur den relationen väntas se ut olika år framgår av tab. 2.

Om vi antar att förtidspensionen i genomsnitt är hälften så stor som ålderspensionen skulle q-värdena i tab. 1 komma att ändras till dem i tab. 2 när förtidspensionerna inkluderas. Det innebär att förtidspensionen under dessa förutsättningar svarar för ca 10-20 % av pensionskostnaderna.

4. Vad övergången till ett fullt utbyggt system betyder för pensionskostnaderna

ATP är ännu inte i full funktion, utan vissa övergångsbestämmelser gäller. De som redan hade uppnått pensionsåldern då ATP infördes kunde aldrig få någon pension från ATP. Det kunde bara de få som tjänat in pensionspoäng. Det innebär att ATP-pensionärerna var få i starten, men att antalet ökar till dess alla som pensioneras har haft möjlighet att tjäna in pensionsrätt. De första ATP-pensionärerna hann inte heller få ihop så många pensionsgrundande år, men varje ny årgång pensionärer har flera intjänade år än tidigare årgångar. Pensionerna blir med andra ord större.

För att få en bild av hur kostnaderna förändras under en successiv övergång till ett fullt utbyggt system, har jag räknat ut ATP-avgiften olika år i ett fördelningssystem. Samma renodlade modell som ligger bakom ekvationen (1) har använts: befolkningsstrukturen är oförändrad, den allmänna reallöneutvecklingen likaså. Genomsnittsindividen i varje kohort har då samma åldersinkomstprofil. Jag gör i det här fallet ingen

Tab. 2. Antalet personer 16—64 år gamla på varje ålders- och förtidspensionär samt pensionsavgiften olika år

År	Antalet personer 16–64 år gamla på varje ålders- och förtidspensionär	Pensionsavgiften (q) i procent
1979	3,24	17,3
1985	2,95	18,7
1990	2,87	19,2
1995	2,88	19,0
2000	2,93	18,4
2005	2,91	18,3
2010	2,68	19,9
2015	2,48	21,8
2020	2,40	22,6
2025	2,35	23,2

[1] Se Information i prognosfrågor 1980:6.
[2] Se ATP-prognos 1982 från riksförsäkringsverket.

Tab. 3. Avgiftsprocenten under ett övergångsskede

Antal år efter genomförandet	Avgiften i procent	Ökning i procent
10	3,4	
15	6,9	103
20	10,8	56
25	14,1	30
30	16,1	14
35	17,0	6
40	17,4	2
45	17,5	
50	17,6[1]	

skillnad på kvinnor och män. Åldersfördelningen är den som gäller för män. Den genomsnittliga inkomstutvecklingen är identisk med den genomsnittliga inkomstutvecklingen 1960–1982 för män födda 1944. Därefter har inkomsten skrivits upp med 1 % realt om året. Vad som under dessa förutsättningar händer med avgiftsprocenten framgår av tab. 3.

I exempel ovan har ATP-avgiften 45 år efter pensionssystemets införande stigit till 17,5 %. Takten i avgiftsökningarna är hög till att börja med, men avtar så småningom.

5. Slutsatser

Genom att förenkla verkliga förhållanden och hålla vissa kostnadsbestämmande faktorer konstanta, har jag kunnat illustrera effekten av varje sådan faktor för sig. De demografiska förändringarna, om man med dessa menar förskjutningar i åldersstrukturen mot relativt flera *äldre* människor, har inte den dramatiska effekt på kostnaderna som många gånger hävdas i den allmänna debatten. Inte förrän omkring år 2015 och därefter blir den demografiska effekten påtaglig. Dessförinnan är den mycket liten.

Som mest kan en person ha 49 år med pensionsgrundande inkomster. Men fler år än 30 påverkar inte pensionsstorleken (detta är 30-årsregeln). Däremot påverkas avgiftsunderlaget. Vidare är det inte genomsnittsinkomsten utan de 15 bäst betalda inkomståren som bestämmer hur stor pensionen blir (detta är 15-årsregeln). Ju jämnare individuella livsinkomstprofiler och högre arbetskraftsdeltagande, desto mindre betyder 15- och 30-årsreglerna. I ett system under uppbyggnad som vårt ATP får 15-årsregeln först successivt genomslag. I ett fullt utbyggt system skulle 15- och 30-årsreglerna uppskattningsvis svara för 25–30 % av ATP-kostnaderna. För att komma fram till detta har jag räknat med att genomsnittsinkomsten under de 15 bästa åren är 15–20 % högre än hela genomsnittsinkomsten och att arbetskraftsdeltagandet är drygt 85 % för 16–64-åringarna.

Utvecklingen av antalet förtidspensionärer spelar en rätt så stor roll för kostnadsutvecklingen i ATP. Fler förtidspensionärer innebär både att pensionsutbetalningarna ökar och att avgiftsunderlaget minskar. Antalet förtidspensionärer väntas vara dubbelt så många som år 1979 om 30 år och då uppskattningsvis svara för 10–15 % av pensionskostnaden. Hälften går kanske till förtidspensionerade av arbetsmarknadsskäl.

[1] Jag får ett något lägre värde här än värdet på $q_{\text{män}}$ i ekvation (4). Det beror på att jag inte har kunnat använda samma uppgifter över antalet yrkesverksamma i de båda fallen.

The Advantage of good Methods. —Considerations on the Application of Markov Models[1]

by Bo Rennermalm[2]

1. Introduction

Methods developed within traditional demography are concentrated on how to study and how to compare populations of various kinds. The techniques are applied to large as well as to small populations. The methods have been developed for aggregates of individuals and are based on a deterministic approach. This is a weakness, since totals and rates on the national or regional level often are strongly influenced by stochastic variation (Hoem, 1983). A population consists of individuals whose apparently stochastic behaviours contributes to the variation of the finite population characteristics.

This paper argues for methods based on a probabilistic formulation of individual behaviour. Experience shows that the methods are splendid when we study transitions between various demographic statuses. The methods are especially useful when we analyse dynamic behaviour, which changes with age or which depends on how long an individual has been in a state. The statistical theory provides us with a set of sharp tools of analysis. Furthermore we obtain more detailed results than by a conventional approach. Another advantage is that the models are non-parametric, since we do not make any special assumptions about underlying functions.

2. Good methods are practical

We cannot obtain knowledge and a deep understanding of a phenomenon without prior ideas of how it behaves. The central point in this statement is that we must formulate our hypotheses before we look at the facts. Our ideas or our hypotheses "should be invented to explain data and not derived from them" (Hempel, 1972). The purpose of the analyses and the computations, based on collected data, should be to investigate the veracity of our conceptualization of the phenomenon. Of course our hypotheses should be well-founded and supported by earlier work. It is difficult to accept explanations that contradict existing theories.

However we cannot be sure that there exists a generally accepted theory which explains the phenomenon. Competing or incompletely formulated theories confuse us. Today—in the day of computers—we often forget to construct a model of how the data are generated and to use it to make our hypotheses precise. Handy and standardized program products can mislead us to make a lot of computations and afterwards explain the phenomenon with the seemingly most interesting result in our hands. This course of action can lead to wrong conclusions, since the results probably often are

[1] This paper was presented 1982 in Kungälv Sweden at a Conference arranged by the Scandinavian Demographic Society. The paper is forthcoming in Scandinavian Population Studies No. 6.

[2] Statistics Sweden.

conditional upon the special outcome of the data, which may have no connection to the underlying process (Rennermalm, 1981).

It is a popular myth that we can obtain knowledge without thinking before we start with the computations. But we cannot expand our knowledge, not even with the supercomputers and superprograms of today or tomorrow, without any kind of prior theory and hypotheses. The theories and the models help us in a relevant way to collect, measure and evaluate facts. Otherwise, without any idea of the underlying process this work becomes superhuman, since everything must be collected, measured and assessed.

Thus, in general we can say that good theories and models are the most practical things that exist. Without them our acting and searching for knowledge become very unrational.

3. The necessary probability model

Typical demographic models on the individual level consist of a number of demographic statuses and events which inform us of the indivual changes of status. Some examples of demographic statuses are "married", "consensual union" and "parity 0". We will often use "state" as a synonym for status. We also assume that these states are clearly separated from each other and that they could be numbered. In reality it is difficult exactly to tell when an individual changes status. For example many cohabitations start gradually and it is first after a while that the individual can say "now I am cohabiting". I disregard this type of difficulties.

It is often convenient to illustrate the states and the transitions between them with boxes and arrows in the manner of Fig. 1 (from Hoem & Rennermalm, 1982). The boxes then denote various states and the arrows symbolize possible demographic events, i.e. transitions between states.

Every individual life history can be regarded as a realisation of an underlying stochastic process whose nature can be very complicated. Here, as in all knowledge searching activities, we use a simplified model of reality. We assume that every life history is a realisation of a time-inhomogenous Markov process or a semi-Markov process (Feller, 1971, Hoem, 1969, Hoem & Funck Jensen, 1982). The number of states is assumed to be finite and the transitions are allowed to occur at arbitrary times, i.e. our time parameter is continu-

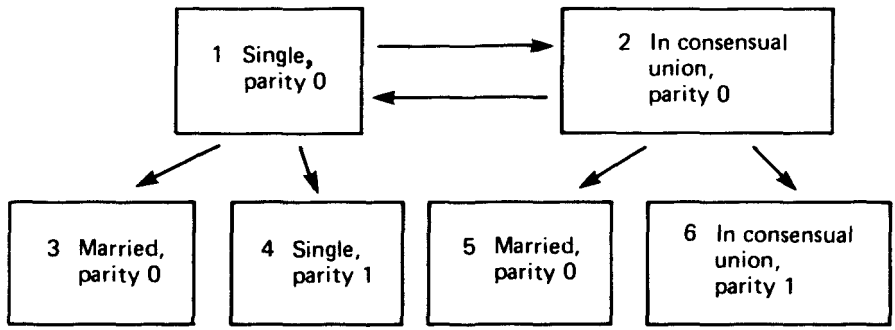


Fig. 1. Transition diagram of how women cohabit, marry and get their first child

ous. We also assume independence between different life histories.

The Markov property means that the probability for an individual to change status only depends on his/her latest state. Loosely speaking we say that the process has no memory since it only pays attention to where it is and not how it came there. The time-inhomogeneity property means that the probability to make a transition is time dependent, i.e. an individual who has belonged to a state for a long time, may show another behaviour than a new member in that state.

Some people may think that such a model does not describe reality, i.e. that the Markov property does not fit human behaviour. The Markov assumption certainly has its limitations in the same manner as every assumption in all other models has. But a good strategy is to start with the simple and handy instead of getting lost in complicated models. Experience shows that the Markov process is very useful. So, why carry coals to Newcastle?

In the following we need some notation. Consider an individual and let

$$I_j(x) = \begin{cases} 1 & \text{if the individual is in state } j \text{ at time } x, \\ 0 & \text{otherwise.} \end{cases}$$

$I_j(x)$ is an indicator variable that tell us the state of the individual at time x. The transition probability can be defined as

$$P_{ij}(x,y) = P(I_j(y) = 1 \mid I_i(x) = 1),$$

i.e. the probability that an individual belongs to state j at time y conditional that he/she was in state i at time x. Together the transition probabilities $P_{ij}(x,y)$ and the initial probability distributions $l_i(0) = P(I_i(0) = 1)$ determine the whole probability distribution. Thus all parameters of interest can be computed with this knowledge. But the transition probabilities are many times impractical as tools in the analyses.

Another much better parameter is the intensity, defined as

$$u_{ij}(x) = \lim_{y \downarrow x} P_{ij}(x,y)/(y-x) \text{ for } i \neq j.$$

The intensity gives the force or the power of the change of the probability per unit time. In a small time interval $(x, x+dx)$, $u_{ij}(x)dx$ is approximately equal to the probability that an individual moves from state i to j, given that he/she was in state i at time x. The intensity function $u_{ij}(x)$ determines the whole probability distribution in the same manner as the transition probabilities do. By a system of differential equations we can compute $P_{ij}(x,y)$ from $u_{ij}(x)$, or vice versa.

Sometimes the intensities are regarded as abstract entities which are difficult to understand. In this paper we shall make these central functions visible both in text and in figures.

4. The intensity—The sharp tool of analysis

To sum up there are two ways to study our phenomenon, either by means of the transition probabilities $P_{ij}(x,y)$ or by means of the intensities $u_{ij}(x)$. For analytical purpose $P_{ij}(x,y)$ are impractical as well as puzzling. They are impractical since they depend on two time variables instead of one. But more serious is the risk of being mislead when we compare $P_{ij}(x,y)$ from different groups of individuals. For example, if we look at Fig. 1 we can use some mathematics to show that the probability to start a cohabitation before age x is

$$\int_0^x \exp\left[-\int_0^s (u_{12}(t)+u_{13}(t)+u_{14}(t))\,dt\right] u_{12}(s)\,ds. \qquad (1)$$

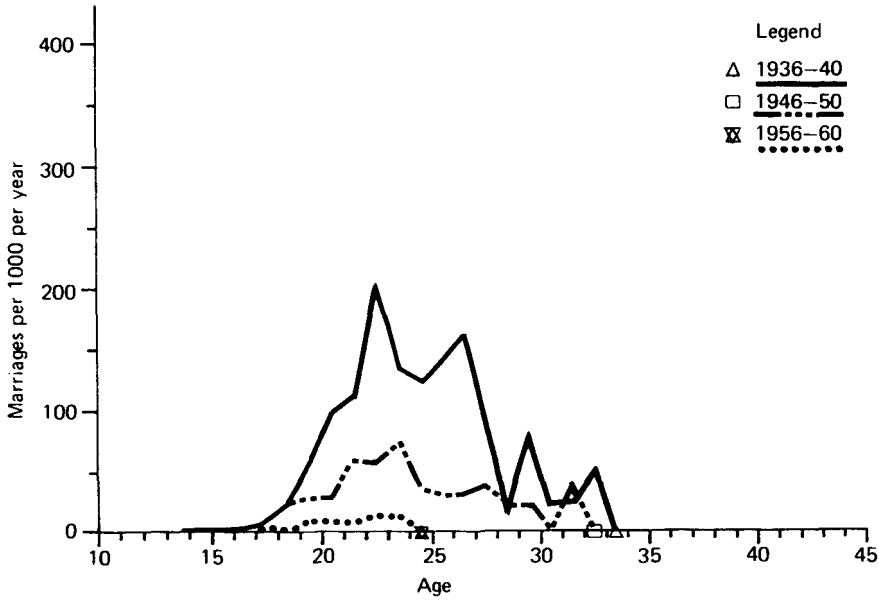


Fig. 2. Marriage intensities for single childless Swedish women born in 1936–40, 1946–50 and 1956–60 without prior cohabitation. From Hoem & Rennermalm (1982)

This probability is a function of three competing risks $u_{12}(x)$, $u_{13}(x)$ and $u_{14}(x)$. Thus a change or stability in this probability cannot be easily explained. We have to take all risks in consideration simultanously.

Assume e.g. that we compare an older and a younger cohort of women and we find that the probability to start a cohabitation has increased. This information alone tells us little, since the increase could be produced by a decreasing nuptiality intensity $u_{13}(x)$ while other intensities, including that for cohabitation i.e. $u_{12}(x)$, have been unchanged.

Thus there is a risk to get erroneous conclusions, since in models with competing risks, a change in a transition probability could point in the opposite direction of the underlying change of the intensities. Unfortunately analyses based only on transition probabilities are not unusual. The probability $P_{ij}(x,x+1)$ is often estimated by "the proportion of transitions between state i and j during time x to $x+1$".

The pure intensity functions $u_{ij}(x)$ do not suffer from this weakness. Different behaviours of groups are very distinctly mirrored by the intensity functions, provided the number of observations is large enough.

In Fig. 2 we show the estimated intensities $u_{13}(x)$ defined in our model in Fig. 1. The curves describe the nuptiality for Swedish women without previous cohabitation, born in 1936–40, 1946–50 and 1956–60. We can see very distinctly how this behaviour has changed. The intensities for the youngest cohort of women is about a tenth of the intensities of the oldest cohort, a tremendous decrease. Very few young Swedish women of today marry without prior cohabitation.

Since our youngest cohort of women was

no older than age 25 by the time of interview, their curve ends at this age. A real advantage with this method versus a traditional approach is that censoring is no problem any more. For every individual we only need to measure the time of the exposure. It does not matter if the time of exposure ends with a transition to a new state or ends by censoring.

The peak in Fig. 2 around age 23 of the intensity curve for women born in 1936–40 has almost disappered nowadays. This means that the young women of today do not have the same popular age at first marriage (without previous cohabitation) as young women had yesterday.

It is possible that we could have discovered this change by analysing the transition probabilities. But in that case we could never be sure about our conclusions. Now by analysing the intensities we can trust our findings.

In Fig. 3 we show the estimates of the intensities $u_{21}(t)$ in our model. Here we use "t" to indicate that the time parameter is duration and not age. The curves show the intensities of dissolution of childless cohabitations, i.e. the transition from state 2 to state 1. Disregarding the stochastic variations, which is caused by our small number of observations, we can see that women born in 1956–60 dissolve a consensual union to a greater extent than older cohorts of women. With traditional methods and so few observations we would never have discovered this type of change in behaviour.

Note that we are allowing women in state 2 to move back to state 1 (Fig. 1). With a

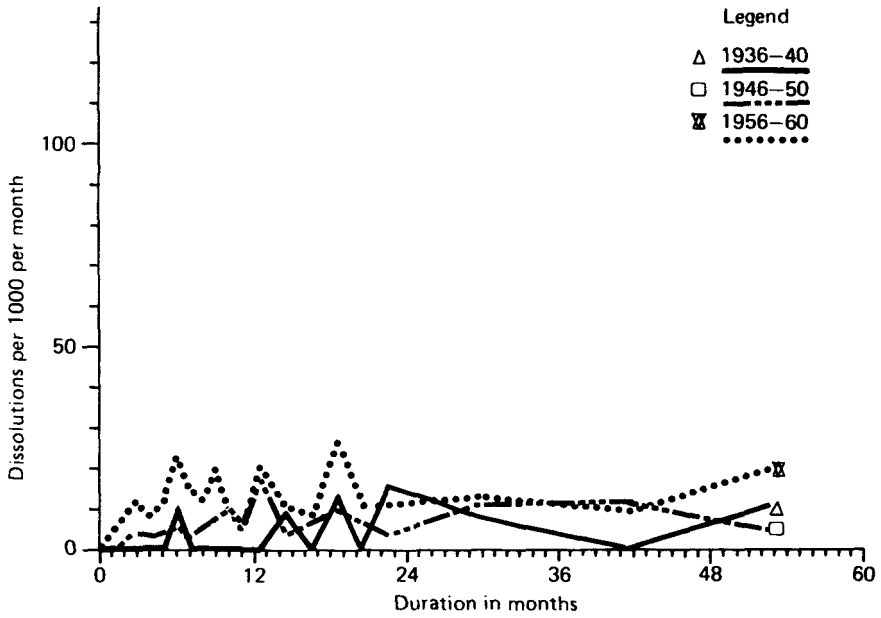


Fig. 3. Cohabitation dissolution intensities for childless Swedish women born in 1936–40, 1946–50 and 1956–60. From Hoem & Rennermalm (1982)

traditional approach this would have caused us real trouble. In fact, the model in Fig. 1 is a mixed time-inhomogenous age-dependent Markov process and a time-inhomogenous semi-Markov process. The process is semi-Markov since the intensities in state 2 depend on duration in that state.

Table 1. Decrements for childless consensual unions by duration of union. From Hoem & Rennermalm (1982)

Duration of consensual unions (in years)	Unions no longer intact and childless due to nonmarital childbearing, Per cent	
	Women born 1936–40	1956–60
1	12	8
3	16	27
5	16	31
8	16	40

The tremendous increase shown in Table 1 in the childbearing proportion can mislead us to adopt the explanation that fertility in childless consensual unions has increased. At least we could be willing to believe that the fertility differs between the oldest and the youngest cohort of women. This is wrong. In fact the underlying behaviour (intensities) has been unchanged for women to move from state 2 to state 6 in our model.

The rates shown in Fig. 4 give evidence for our statement about unchanged fertility in consensual unions. We use the word "rate" to indicate that for each cohort the intensities calculated for durations from 7 to 60 months have been combined into one single number. This can be done since the intensity curves are relative constant during this time interval. The tremendous decrease in marital rates for cohabiting women shown in Fig. 4 has in fact caused the above

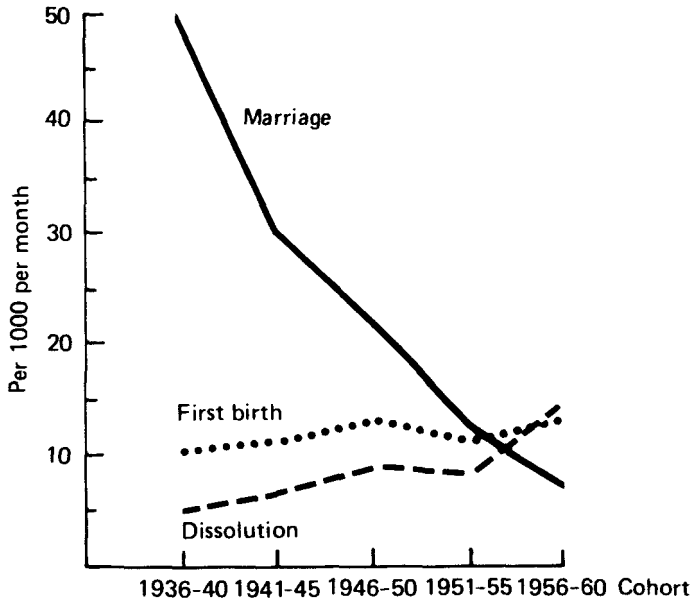


Fig. 4. Rates of marriages, first birth and union dissolution for cohabiting childless Swedish women born in 1936–1960. From Hoem & Rennermalm (1982)

mentioned increase in Table 1. The decrements in Table 1 are functions of the same type as specified in formula (1).

This striking example shows how easy it is to make wrong conclusions, when the analyses only are based on transition probabilities alone.

Of course our findings based on the curves in Fig. 2 and 3 are not unexpected. Most people would have guessed the decrease in marriage without previous cohabitation and the increase in dissolutions of consensual unions. But now we have established these findings in an exact numerical way. Furthermore ragged curves of the type in Fig. 2 and 3 serve as excellent tools for detecting changes and understanding the underlying individual behaviour.

To sum up, by applying a semi-Markov model we have learned something new about Swedish population trends: Consensual unions have not replaced marriage, since today a consensual union is not necessarily the start of a marriage as it was yesterday. Furthermore, nowadays consensual unions dissolve in higher extent, and cohabiting women have not adopted the higher fertility of their married sisters.

5. A rich description of the results

As mentioned earlier the intensity functions are the motor in the Markov process, i.e. something more valuable than merely a tool of analysis. By means of the intensities and some mathematics we can compute and estimate every characteristic of interest and in this way illustrate the effects of changes in the underlying behaviours. Thus the de-

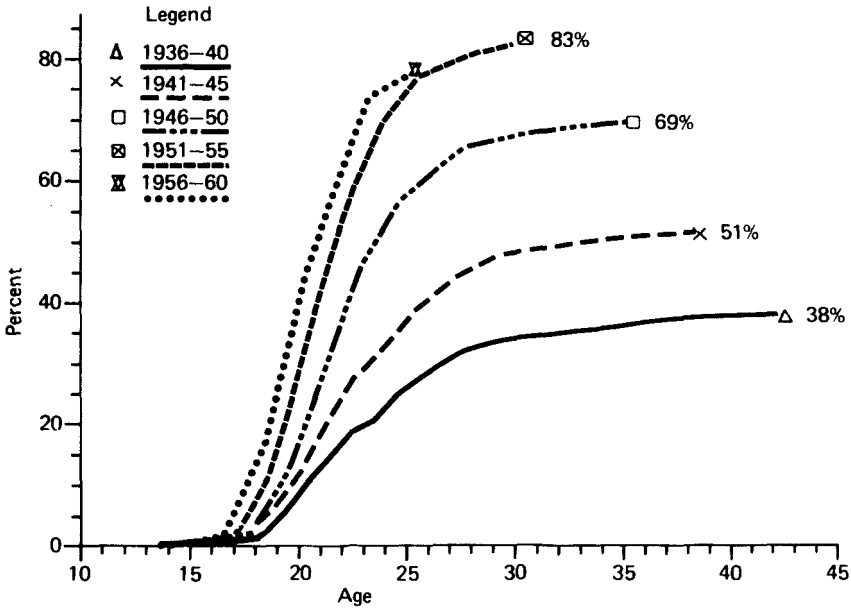


Fig. 5. Estimated proportions of women who start a cohabitation before age x. From Hoem & Rennermalm (1982)

scription of the results can be made very rich and realistic. It is realistic since our assumptions are rather few and natural.

An example on such an estimated characteristics are the curves in Fig. 5. They show the probability to start a cohabitation before age x. See the above formula (1). The curves are based on the estimated intensities. Note the surprisingly high proportion of consensual unions in our oldest cohort of women.

We give some further examples of corresponding characteristics which could be computed and estimated by the intensities:
– Other transition probabilities $P_{ij}(x,y)$ than the ones mentioned in the text.
– The expected time a woman in state 1 at age x will spend in this state up to any first departure over the subsequent n years.
– The probability of experiencing $n \geq 1$ consensual unions by age x and being single at that age.

Note that our inference is model-based, since we make our inference to the assumed underlying Markov process. We also regard our numerical results as estimates and not as fixed finite populations characteristics. Thus, we can use methods of mathematical statistics to test various kinds of hypotheses. Hoem & Funck Jensen (1982) give a complete discussion about these methods and the weakness of some ordinary demographic techniques.

6. Acknowledgements

I want to thank Jan. M. Hoem and Ingrid Lyberg for their helpful comments and suggestions to improve my paper.

7. References

Feller, W. (1971): An Introduction to Probability Theory and Its Applications, Vol. 2. New York, Wiley.

Hempel, C. (1972): Vetenskapsteori. Studentlitteratur. Lund. (In Swedish).

Hoem, J. M. (1969): Purged and partial Markov chains. Skandinavisk Aktuarietidskrift 52.

Hoem, J. M. (1983): Multistate mathematical demography should adopt the notions of event-history analysis. Dep. of Statist. University of Stockholm, Demographic Research Report No. 10.

Hoem, J. M. & Funck Jensen, U. (1982): Multistate life table methodology: A probabilistic critique. Copenhagen Univ. Lab. of Actuarial Math., Working Paper No. 38.

Hoem, J. M. & Rennermalm, B. (1982): Cohabitation, marriage, and first birth among never-married Swedish women in cohorts born 1936–1960. Dep. of Statistics, University of Stockholm, Demographic Research Report No. 8.

Rennermalm, B. (1981): Analys – vad är det? Statistisk tidskrift 1981: 1. (In Swedish).

Invandrare som brottsoffer[1]

av fil. kand. Tarja-Liisa Leiniö[2]

1. Inledning

Ett av huvudmålen i den svenska invandrings- och minoritetspolitiken är jämlikhet mellan invandrare och den infödda befolkningen. Målet innebär bl. a. ett löfte om ansträngningar för att skapa likvärdiga levnadsvillkor för båda grupperna. Invandrare garanteras lika behandling med den infödda befolkningen (SOU 1974: 41). Ett led i dessa ansträngningar var tillsättningen av en statlig diskrimineringsutredning. Den såg över brottsbalken och tryckfrihetsordningen avseende de delar som reglerar behandling av folkgrupper av viss ras, visst etniskt eller nationellt ursprung (SOU 1981: 38). Det i dessa bestämmelser definierade juridiska skyddet begränsar sig emellertid till att gälla *grupp* av personer.

En enskild *individ* skyddas juridiskt av brottsbalkens bestämmelser (BrB 5: 3 om skydd mot kränkande tillmäle, BrB 4: 5 om skydd mot bl. a. hot om brottslig gärning, BrB 4: 7 om ofredande) utan urskiljning av etniskt eller nationellt ursprung. Utredningen uppmärksammade det förhållandet att invandrare i dessa situationer har endast samma skydd som tillfaller varje enskild individ i befolkningen men de kan vara i behov av bättre skydd. Utredningen föreslog därför en förändring av BrB 5: 5 första stycket som innebär att brottet förolämpning kan åtalas av åklagare efter anmälan (i stället för att en enskild själv skall stämma gärningsmannen inför domstol), om åtal är påkallat ur allmän synpunkt (SOU 1981: 38, s. 73).

I denna uppsats skall belysas om invandrares skydd mot angrepp mot sin person avviker från det skydd den infödda befolkningen har. Innan man kan tala om diskriminering mot invandrare bör man göra en distinktion mellan oavsiktlig och avsiktlig gärning. Blir man ofredad i sin tjänsteutövning utan att det på något sätt är förknippat till vem man är som privatperson, kan vi inte tala om diskriminering. Händer ofredandet på grund av att offret är representant för en icke-svensk nationalitet föreligger diskriminering.

2. Syfte

Avsikten är att med hjälp av intervjufrågor om invandrare och svenskar under en viss tidsperiod blivit utsatta för olika typer av våldsbrott studera eventuella skillnader mellan de två grupperna. De frågor vi söker besvara är: Vilka skillnader föreligger mellan svenskar och olika invandrargrupper? I vilken utsträckning kan observerade skillnader återföras på variation i olika bakgrundsvariabler och därigenom förklaras eller tolkas? Föreligger det skillnader mellan olika invandrargruppers utsatthet för vissa typer av våldshändelser?

[1] Uppgifterna i artikeln grundar sig på data från SCBs undersökningar om levnadsförhållanden (ULF 1978, 1979 och 1980). Databearbetningen har bekostats av Brottsförebyggande rådet.

[2] Forskningsassistent vid Institutet för social forskning, Stockholms universitet.

3. Metodfrågor

3.1. Om att mäta utsatthet

De uppgifter som vi kommer att redovisa är hämtade ur intervjuer där man ställt frågor om den intervjuade under en ettårsperiod har utsatts för vissa typer av våldshandlingar. Resultaten av en sådan undersökning påverkas, förutom av de intervjuades faktiska utsatthet, av om de överhuvudtaget kommer ihåg det som har hänt under ifrågavarande period, av hur de uppfattar och tolkar det som hänt samt av om de väljer att berätta om det inträffade eller inte. S. k. teleskop-effekt, dvs. att man placerar en före tidsperiodens början inträffad händelse i den period som avses i intervjun, har visat sig vara betydande vid undersökningar, där händelser tillbaka i tiden har kartlagts (Reynolds, et al., 1977). I detta sammanhang bortser vi från denna effekt i resultaten.

Tidigare undersökningar har visat att s. k. offerundersökningar av intervjutyp är ett ganska dåligt mätinstrument när det gäller vissa typer av brott (Persson, 1977:6, 1977:7). Våldshändelser där gärningsmannen är känd, är en släkting eller en familjemedlem såsom vid hustrumisshandel, är kraftigt underrapporterade i sociologiska undersökningar. Vid den undersökning ur vilken vårt material har hämtats intervjuades intervjupersonens make eller maka eller den med intervjupersonen sammanboende också. I många fall var båda två närvarande vid själva utfrågningen. Det finns risk att eventuella våldshändelser inom familjen är underrapporterade i större utsträckning än vid intervjuer där endast en person i familjen intervjuas.

Vid primärgruppsvåld är skuldfrågan sällan entydig. Den intervjuades känsla av medskyldighet till det inträffade och skamkänslor gör att han eller hon inte gärna berättar om det. Vid en intervjusituation har den intervjuade heller inte klart för sig vilka frågor som kommer att ställas i fortsättningen. För att slippa att eventuellt redogöra för hur det gick till kan hon eller han välja att inte nämna det alls. Detta är naturligtvis allvarligt ur undersökningens tillförlitlighetssynpunkt, speciellt som det har visat sig att våldshändelser med allvarliga personskador med läkar- och sjukhusbesök som följd i övervägande fall är resultat av våld inom primärgruppen, oftast inom familjen (Lenke, 1978). De händelser som tillförlitligast belyses i intervjuundersökningar är angrepp som kommer utifrån den egna gruppen.

Intervjupersonen kan också ha en felaktig uppfattning om vad som har hänt. I allmänhet påverkar detta emellertid i riktning mot överrapportering, dvs. att man tolkar en olyckshändelse som en medveten brottslig gärning.

3.2. Är våldet vanligare bland invandrare än bland svenskar?

Mekanismer som leder till våld kan tänkas skilja sig åt mellan infödda och invandrare dels på grund av kulturella skillnader dels på grund av att gruppernas förankring i det svenska samhället är annorlunda. I brist på undersökningsresultat som kunde leda oss på spåren av sådana mekanismer och eventuella skillnader mellan grupperna får vi nöja oss med att antyda några faktorer som kan orsaka skillnader. I en i Finland genomförd offerundersökning visade det sig att våld och hot om våld där är betydligt vanligare än i Sverige eller i de övriga nordiska länderna (Nordiska Rådet 1983, Sveri, 1982). Uppbrottet från den invanda omgivningen, vare sig det är fråga om intern flyttning eller flyttning över landsgränserna är

påfrestande för individen. Aleksandra Ålund (1978) har i sina undersökningar spårat förekomst av spänningar inom immigrantfamiljerna, dels mellan makar, dels mellan föräldrar och barn. En källa till dessa spänningar är familjens möte med det svenska samhället. Invandringen medför för många inte bara en flyttning från ett land till ett annat utan samtidigt en flyttning från landsbygden till storstäder. Efter invandringen blir invandrarkvinnor, som tidigare varit ekonomiskt beroende av mannen, förvärvsarbetande och får därmed en självständigare roll även inom familjen. Barnen umgås med svenska ungdomar som har ett annat, friare beteende och oftast en mindre auktoritär uppfostran. Det är inte otänkbart att dessa förändringar i den invandrade familjens liv ibland kan leda till våldshandlingar.

Koncentrationen av enskilda invandrargrupper till ett eller ett fåtal bostadsområden kan leda till våld mellan representanter för dessa grupper och övriga på samma områden boende grupper. Risken är större, ju mera avvikande, i utseende och/eller beteende invandrargruppen är från de infödda. Ett sådant exempel är konflikterna i Södertälje mellan assyrier och svenska ungdomar.

Förutom att våld inom invandrargrupperna kan vara vanligare än bland svenskarna och att invandrare kan utsättas för våld från omgivningen i större utsträckning än svenskarna, påverkas resultaten också av eventuella skillnader mellan invandrare och svenskar när det gäller att rapportera om våldshändelser generellt i en intervjuundersökning. Man kan anta att personer som av tradition eller av yttre omständigheter såsom vid olika grader av påfrestningar som föranletts av flyttningen till ett främmande land, är vana vid och accepterar våldshandlingar har en större tolerans för sådant. Sådana fenomen som hämnd och självförsvar, vilket i vissa fall kan innebära våld, kan antas vara både accepterat och i verkligheten förekommande bland några invandrargrupper men är inte accepterade i det svenska samhället. Detta gäller i första hand primärgruppsvåld. Det är heller inte otänkbart att man skyddar medlemmar av den egna gruppen mot majoritetssamhället och att man i detta syfte inte vill avslöja ens allvarliga gärningar. När vi jämför olika nationalitetsgruppers utsatthet, kan våra resultat påverkas av gruppers olika tolerans för våld, av deras eventuellt skilda våldsuppfattningar samt av eventuella skillnader i att rapportera om våldshändelser.

4. Undersökningsmaterialet

Det empiriska material som här redovisas utgörs av statistiska centralbyråns (SCB) undersökningar av levnadsförhållanden för åren 1978, 1979 och 1980[1]. Sammanslagningen av tre års undersökningar föranleddes av att antalet personer med någon utländsk bakgrund för ett enskilt år uppgår till endast ca 1 000 personer. Ett sådant antal kan inte indelas i något större antal undergrupper. Finländare, den dominerande invandrargruppen i Sverige, utgör nästan hälften av personer med utländsk bakgrund.

Förfarandet med sammanslagningen har både för- och nackdelar. Främsta fördelen är det större antalet intervjuer. En nackdel är att vi inte riktigt vet vilken betydelse det har för resultaten att uppgifterna har sam-

[1] För ingående redovisning av undersökningarnas genomförande, urval, bortfall och innehåll, se Teknisk rapport avseende 1977 och 1978 års undersökningar. Appendix 4. Projektet har rapporterats i ett 30-tal undersökningsrapporter i serien Levnadsförhållanden, SCB.

lats in under tre år. Det finns dock ingen anledning att anta att sammanslagningen på något sätt skulle leda till mindre pålitliga mätningar, kanske tvärtom. I de fall uppgifter insamlades vid alla tre årens intervjuer, avser resultaten genomsnittet för de tre åren, i de fall frågorna ställdes under två av de tre åren, avses genomsnittet av dessa.

5. Utlänningsgrupper

Invandrargrupperna har urskiljts i materialet med hjälp av intervjufrågorna avseende föräldrarnas nationalitet. Frågorna löd: "Är någon av Era föräldrar utländsk medborgare eller har någon av dem varit det?... Vilken nationalitet? (för far resp. mor)." Endast de vars båda föräldrar är eller har varit utländska medborgare har inkluderats i invandrarkategorierna. För att kunna särskilja invandrargrupper med olika länderbakgrund är en ytterligare förutsättning att båda föräldrarna har samma nationalitetsbakgrund. I de fall föräldrarna kommer från två olika utländer skulle vi inte kunna hänföra intervjupersonen entydigt till någon speciell nationalitetsgrupp. Genom detta förfarande vill vi särskilja två grupper vars bakgrund *entydigt* skiljer sig från varandra, nämligen infödda svenskar och "rena" invandrare. Personer med två "lika" utländska föräldrar fördelar sig på olika grupper på det sätt som visas i nedanstående tablå.

Föräldrarnas medborgarskaps/ ursprungsland	Antal[1]	Andel
1 Finland	977	51,7
2 Danmark, Norge, Island	281	14,8
3 Väst- och Östtyskland, Österrike, Schweiz	192	10,1
4 Frankrike, Holland, Belgien, Storbritannien, Luxemburg, Eire	36	1,9
5 Estland, Lettland, Litauen	62	3,3
6 Sovjetunionen och Österuropa (utom Estland, Lettland, Litauen och Jugoslavien)	227	12,0
7 Jugoslavien	141	7,4
8 Grekland	58	3,1
9 Italien	22	1,2
10 Spanien, Portugal, Turkiet	52	2,7
11 USA, Kanada	20	1,0
12 Mexico, Mellanamerikanska staterna, Sydamerika	31	1,6
13 Australien, Nya Zeeland	2	0,1
14 Afrika	22	1,2
15 Asien	58	3,7
Totalt	1891	100,0

[1] Avser oviktade tal, dvs. faktiska antalet intervjuade utan korrigering för bortfall och för att olika familjekategorier t.ex. har olika sannolikheter att komma med.

Den här tillämpade definitionen av vem som är invandrare är naturligtvis diskutabel. Den vanligaste avgränsningen är uppgiften om medborgarskap. Medborgarskapsbytet kan emellertid ses som en i första hand juridisk handling som *i sig* inte behöver ändra vare sig individens eget beteende eller hur han eller hon blir behandlad i vardagslag. Vistelsetiden i Sverige fram till den tidpunkt då man söker svenskt medborgarskap varierar stort både mellan och inom olika nationalitetsgrupper. Politiska flyktingar i allmänhet söker svenskt medborgarskap så fort det är möjligt. För arbetskraftsinvandrare förhåller det sig annorlunda. Av de år 1968 invandrade finländarna t.ex. hade fram till 1980 års utgång endast ca 18 procent bytt medborgarskap medan resten fortfarande hade sitt finländska medborgarskap. Av samma års jugoslaviska invandrare hade endast 14 procent blivit svenska medborgare fram till samma tidpunkt. Däremot hade ca 60 procent av 1968 års polska invandrare fått

svenskt medborgarskap fram till 1980 års utgång (SOS, 1980).

5.1. Andra generationens invandrare

Att definiera gruppen andra generationens invandrare är, om möjligt, ännu svårare än att bestämma vem som är invandrare. Någon vedertagen praxis existerar inte. Bland olika mer eller mindre genomförbara alternativ har vi valt att utgå från den vid SCB utnyttjade definitionen "i Sverige födda svenska medborgare med någon eller båda föräldrarna födda i utlandet". Dock har vi särskiljt dem med en svensk och en utländsk förälder från dem med två utländska föräldrar. Det verkar omotiverat att sammanföra barn till två utlänningar som kommit till Sverige av ekonomiska eller av politiska skäl och som kanske tänker att återvända till sitt hemland till samma grupp med barn till en svensk och en utlänning som flyttat till Sverige för att förenas med sin "kärlek". Förankringen till det svenska samhället är betydligt starkare i det senare än i det förra fallet (se vidare, Reinans, 1982, Invandrare och Minoriteter, 1983).

Gruppen andra generationens invandrare enligt definitionen ovan består till 85 procent av personer med endast en utländsk förälder. Den vanligaste föräldrakombinationen är svensk far och utländsk mor vilket gällde för 362 personer i vårt material. 133 av dessa har finländsk mor. 21 personer har två utländska föräldrar som kommer från två olika länder, utanför Sverige.

Möjligen är det av betydelse om det är fadern eller modern som kommer från utlandet. Så skulle kunna vara fallet om det t. ex. visade sig att svenska kvinnor gifter sig med utländska män av en viss nationalitet medan svenska män gifter sig med utländska kvinnor av en viss annan nationalitet och att dessa utländer skiljer sig från varandra betydligt i fråga om beteende, kultur osv. Så skulle också kunna vara fallet om det fanns tydliga skillnader, t. ex. i fråga om utbildning och yrke, mellan svenska män som gifter sig med utländskor och svenska kvinnor som gifter sig med utländska män. Det finns dock inga genomförda studier som skulle belysa dessa frågor.

En annan tänkbar indelning vore att beakta ålder vid invandring i stället för födelseland och medborgarskap. Oberoende av medborgarskap skulle t. ex. de som fått hela sin utbildning i Sverige kunna hänföras till andra generationens invandrare.

5.2. Överlappning av grupperna

På grund av de definitioner som här används är grupperna delvis överlappande. Definitionen "två utländska föräldrar från samma ursprungsland" inkluderar även andra generationens invandrare med två utländska föräldrar i de fall föräldrarna kommer från samma land. De 21 personer med två utländska föräldrar av olika nationalitet som tillhör andra generationens invandrare är däremot uteslutna ur analysen där invandrare har delats in i olika nationalitetsgrupper. Det finns ingen logisk nationalitet för dessa personer. Dessa 21 ingår i analysen där andra generationens invandrare diskuteras.

Denna överlappning medför att nationalitetsgruppernas koncentration till de allra yngsta åldersgrupperna ökar något. Följden blir att ålderseffekten, vilket i sig är viktigt ur utsatthetssynpunkt, förstärks i resultaten. Vi kommer dock att ta hänsyn till skillnader i gruppernas åldersfördelning vid redovisningen av resultaten. De grupper som kommer att särredovisas i det följande framgår ur nedanstående tablå.

Övriga grupper är antingen för små eller för heterogena för att särredovisas.

Kategori	Antal[1]
Infödda svenskar	22 568
Alla invandrare med två "lika" föräldrar	2 100
Därav från:	
Finland	916
Övriga Norden	272
Väst- och Österskland, Österrike, Schweiz	181
Estland, Lettland, Litauen, Sovjetunionen och övriga Österuropa (utom Jugoslavien)	301
Jugoslavien	140
Andra generationens invandrare	
Med en svensk, en utländsk förälder	692
Med två utländska föräldrar	129

[1] Viktade tal.

6. Resultatredovisning

Undersökningsresultaten redovisas efter de frågor som formulerades inledningsvis under punkt 2. De är: Vilka skillnader i utsatthet för våld föreligger mellan svenskar och olika invandrargrupper? I vilken utsträckning kan observerade skillnader återföras på variation i olika bakgrundsvariabler och därigenom förklaras eller tolkas? Föreligger skillnader mellan olika invandrargrupper?

7. Utsatthet för våld eller hot om våld

De frågor som ställdes till de intervjuade angående om de hade blivit utsatta för någon våldshändelse under det senaste året var inte exakt desamma för de tre undersökningsåren 1978, 1979 och 1980. I samband med 1978 års undersökning om levnadsförhållanden genomfördes en noggrannare kartläggning av utsatthet (SCB, 1981). Utfrågningen på trygghetskomponenten inleddes åren 1978 och 1979: "Har Ni själv under det senaste året, alltså sedan ..., varit utsatt för någon eller några av följande händelser ...

1. Våld som ledde till synliga märken eller kroppsskada?
2. (Har Ni därutöver blivit utsatt för) Våld som *inte* ledde till synliga märken eller kroppsskada?
3. (Har Ni därutöver blivit utsatt för) Hot eller hotelser om våld som var farliga eller så allvarliga att Ni blev rädd?"

För varje uppräknad händelse enligt de tre frågorna fortsatte 1978 års intervju med frågor om händelseplats, om man anmält det inträffade till polisen, om det krävde besök hos läkare, tandläkare eller sjuksköterska, om man blev sjukskriven eller om följden blev att man inte kunde sköta sina vanliga sysslor. Dessa frågor ställdes inte 1979.

År 1980 hade den första frågan delats upp på 1) Våld som ledde till sådana skador att de krävde besök hos läkare, tandläkare, eller sjuksköterska, 2) (Har Ni därutöver blivit utsatt för) Våld som ledde till synliga märken eller kroppsskada men som ej krävde besök hos läkare, tandläkare eller sjuksköterska? De tre övriga frågorna var desamma som 1978 och 1979. Intervjusituationen var med andra ord något annorlunda vid de tre årens undersökningar.

7.1. Skillnader mellan svenskar och invandrare

5,3 procent[1] av infödda svenskar och 8,3 procent av invandrare har utsatts för minst en våldshändelse under en ettårsperiod. Av de olika nationalitetsgrupperna har finländska, de med baltisk, sovjetisk eller annat österuropeiskt ursprung samt jugoslaviska invandrare utsatts i större utsträckning (andelarna utsatta är i ovannämnd ordning 8,8, 7,6 och 11,4%) medan övriga nordiska

[1] Resultaten i tabellform kan rekvireras från författaren.

och de med tysk, österrikisk eller schweizisk bakgrund har lägre andelar (5,5 resp. 6,6%). 9,4 procent av andra generationens invandrare med en och 20,9 procent av dem med två utländska föräldrar har utsatts för minst en våldshändelse.

Jämfört med risken att utsättas för något våld under en ettårsperiod i de olika grupperna ser risken att utsättas för två eller fler våldshändelser bland de utsatta helt annorlunda ut. Bland utsatta infödda svenskar är sådan risk drygt 40 procent och bland alla invandrare nära 50 procent.[1] Bland andra generationens invandrare med en svensk, en utländsk förälder ligger risken att utsättas för två eller fler våldshändelser bland de utsatta på ca 45 procent[2] och bland dem med två utländska föräldrar på 30 procent.[3]

Andelen invandrare som utsatts för våld med synliga märken eller kroppsskada är dubbelt så stor som motsvarande andel bland svenskarna (3,4 resp. 1,7%). Även de lindrigare våldsformerna har oftare rapporterats av invandrargruppen. De mest utsatta grupperna är den finländska, den jugoslaviska och de båda andragenerationsgrupperna. Övriga nordbor, vars andel utsatta för något våld var densamma som de infödda andel, har en större andel som utsatts för den allvarligaste våldstypen, våld med kroppsskada (3,3%).

Utsatthet kan också mätas med medeltal rapporterade våldshändelser, dels i hela gruppen, dels bland dem som utsatts minst en gång. Bland alla invandrare är det framför allt våld utan märken och hotelser om våld som några individer har utsatts för upprepade gånger. Medeltalet per utsatt för våld med synliga märken eller kroppsskada däremot är av samma storleksordning bland såväl infödda svenskar som bland invandrare. Den andelsmässigt mest utsatta gruppen, andragenerationens invandrare med två utländska föräldrar, har de lägsta medeltalen av våld med eller utan kroppsskada. I denna grupp har med andra ord proportionerligt fler utsatts för enstaka våldshändelser.

7.2. Kan skillnaderna återföras på vissa bakgrundsvariabler?

Risken att utsättas för olika typer av våld är ojämnt fördelad så att män drabbas oftare än kvinnor. De yngsta männen är den mest utsatta gruppen. Ensamstående drabbas oftare än samboende. Våldet är vanligare i Stockholm och i Göteborg än annorstädes i landet. Även boendeform och yrke samvarierar med utsattheten för våld. Våldet är vanligare bland de höghusboende än t.ex. villaboende och är vanligare bland arbetare än bland tjänstemän. I det följande skall vi studera om de ovan redovisade skillnaderna i utsatthet mellan infödda och invandrare kan förklaras när hänsyn tas till eventuella skillnader i gruppernas sammansättning vad avser dessa egenskaper (se Appendix).

Vid *könsspecificeringen* visar det sig att skillnaden mellan dels svenska och invandrade kvinnor dels mellan svenska kvinnor och andra generationens kvinnor med en utländsk förälder krymper. Skillnaden i utsatthet mellan svenska män och svenska kvinnor är 2,5 procent men är ca 5,0 procent mellan könen bland alla invandrare. Denna skillnad är drygt 6 procent i andragenerationsgruppen med en och 20 procent i den med två utländska föräldrar. Det är med andra ord framför allt invandrade män

[1] 869 svenskar har utsatts för en, 534 för minst två våldshändelser. Bland invandrare är motsvarande antal 86 resp. 88.
[2] 34 personer i denna grupp har utsatts för en och 30 för minst två våldshändelser.
[3] 19 personer i gruppen har rapporterat endast en och 9 minst två händelser.

som står för skillnaden i *andelen utsatta* mellan svenskar och invandrare.

Infödda män och kvinnor och invandrade män uppvisar i genomsnitt lika medeltal våldshändelser per utsatt (2,35–2,46) medan invandrarkvinnor erfarit i *medeltal* en hel våldshändelse fler (3,51) än de övriga grupperna. Bland andragenerationsgrupperna har män med en utländsk förälder ungefär samma medeltal (2,52) som de infödda medan kvinnornas medeltal i denna grupp är 0,5 färre händelser (2,05). Andragenerationens män med två utländska föräldrar har i genomsnitt rapporterat nästan en hel färre händelse (1,59) än de infödda medan gruppens kvinnor har ungefär samma medeltal (2,33) som de infödda. I den sistnämnda invandrargruppen och bland alla invandrare är förhållandet likadant mellan könen i medeltal rapporterade händelser bland de utsatta medan andragenerationsgruppens (två utländska föräldrar) medeltal ligger för båda könen en hel händelse lägre än bland alla invandrare.

Upprepade våldshändelser med synliga märken eller kroppsskada som följd är vanligare bland kvinnor än bland män i grupperna infödda svenskar och alla invandrare medan det motsatta förhållandet gäller för andragenerationen med en utländsk förälder.

Åldersmässigt är invandrargruppen i sin helhet yngre än infödda svenskar. Den dominerande åldersgruppen bland de förstnämnda är 25–44-åringar. Åldrarna 65–74 är kraftigt underrepresenterade bland invandrarna. Invandrarna är med andra ord koncentrerade till de generellt mer utsatta åldersgrupperna.

När hänsyn tas till ålder visar det sig att det faktiskt bara är i den allra yngsta åldersgruppen där invandrarna *inte* är mer utsatta än de infödda. I de två äldsta grupperna kvarstår skillnaden på 3-procentsnivån och till och med ökar något i åldersgruppen 25–44 år. Hälften av invandrarna tillhör denna sistnämnda åldersgrupp.

De båda andragenerationsgrupperna är rena ungdomsgrupper. 3/4 är yngre än 45 år och 40 procent är under 25 år.

Invandrares fördelning i olika *familjetyper* skiljer sig från motsvarande fördelning bland de infödda. Medan andelarna ensamstående resp. samboende är exakt lika i båda grupperna är andelen utan hemmavarande barn lägre bland invandrare än bland svenskarna. Såväl de ensamstående som de samboende invandrare som inte har hemmavarande barn står för merparten av skillnaden i utsatthet för olika våldstyper. Skillnaden i utsatthet krymper mellan infödda och invandrare som har barn, oberoende av om de är ensamstående eller samboende.

Bland andragenerationsgrupperna är de samboende kraftigt underrepresenterade. Var femte ensamstående med barn och var sjätte ensamstående utan barn av dem med en utländsk förälder har drabbats av någon typ av våld. Var fjärde ensamstående utan barn i andragenerationsgruppen med två utländska föräldrar har likaså utsatts för våld.

Invandrares *bostadsort* är oftare än de inföddas någon av storstäderna. Koncentrationen till Stockholm och Göteborg, där våldet generellt är vanligast, är tydlig bland invandrare. Men utsattheten för de i dessa två storstäder bosatta invandrarna är klart större än för de på samma orter boende infödda svenskarna.

Den ortstypsindelning (de tre storstäderna var för sig och övriga landet) som vi har tillämpat här är rätt primitiv. Men även om vi skulle redovisat ortstyperna med största möjliga precision skulle vi knappast kunnat fånga in skillnader i boendemiljöer, skillnader som kan vara av betydelse i detta

sammanhang. Dessa skillnader kan bättre beskrivas med hjälp av uppgifter om boendeform eller *hustyp*, dvs. om man bor i villa eller radhus, i relativt små (1–3 våningar) flerfamiljshus eller om man i stället bor i stora flerfamiljshus (4 eller fler våningar).

När vi tar hänsyn till gruppernas boendeform kan vi konstatera att skillnaderna i utsatthet för våld mellan de infödda och invandrare i det närmaste försvinner för radhusboende och för boende i en- eller tvåvånings flerfamiljshus. Skillnaderna mellan grupperna krymper med en tredjedel (från 3 till 2 procent) när bostadsformen är antingen villa eller 3-vånings flerfamiljshus och kvarstår som oförändrad (3 procents skillnad) för kategorin 4 eller fler våningars flerfamiljshus. Andragenerationens utsatthetsmönster tycks inte ha någon logisk samvariation med deras bostadsform. Antagligen har det att göra med deras starka koncentration till de yngsta åldersgrupperna och att övervägande andel av dem är ensamstående.

7.3. Sammanfattning

Generellt har invandrare utsatts för olika typer av våldshändelser i större utsträckning än infödda svenskar. Skillnaden försvinner inte när hänsyn tas till skillnader i gruppernas sammansättning avseende kön, ålder, familjeförhållanden, bostadsort och bostadsform. Genom att kontrollera dessa egenskaper förtydligas dock utsatthetsbilden, dvs. vi isolerar i någon mån grupper med överrisk att utsättas för olika former av våld.

Bland alla invandrare är det i första hand män, personer i åldrarna 25–44 år, personer utan hemmavarande barn oberoende av om de är ensamstående eller samboende, personer som är bosatta i Stockholm eller i Göteborg och personer vars bostad återfinns i flerfamiljshus med fyra eller fler våningar som har de *högsta andelarna* utsatta. Man bör dock hålla i minnet att vi inte har tagit hänsyn till dessa egenskaper *samtidigt* utan var för sig.

Bland andragenerationen med en utländsk förälder är dels män, dels ensamstående, speciellt ensamstående med barn, grupper med stora överrisker att ha utsatts för någon form av våld. I andragenerationsgruppen med två utländska föräldrar står de yngsta och de ensamstående utan barn för den tydligaste överrepresentationen för att ha drabbats.

Medan *medeltalet* rapporterade våldshändelser av olika typer var av samma storleksordning bland dels de utsatta infödda (oberoende av kön) dels bland utsatta invandrarmän, visar resultaten att de utsatta invandrarkvinnorna i genomsnitt har rapporterat klart större antal våldshändelser. Denna skillnad kan bero på två saker: Det kan betyda att invandrarkvinnor *faktiskt* är *mer utsatta*. Eller det kan betyda att de *rapporterar fler* inträffade *händelser* än vad svenska kvinnor gör.

Frågan om våldets händelseplats ingick i 1978 års intervju men har inte bearbetats. Samma fråga ställdes i en liknande intervju 1981[1], där finländska och jugoslaviska invandrare ingår. Enligt den undersökningen är den vanligaste händelseplatsen den egna bostaden för det våld svenska och jugoslaviska kvinnor rapporterar. Bland finländska

[1] Socialforskningsinstitutets levnadsnivåundersökning 1981. Generellt visar denna undersökning lägre andelar utsatta bland såväl jugoslaver (6,7 %) som bland finländare (3,9 %). Orsaken till dessa skillnader kan vara att våldet minskat sedan 1978–1980. Mer sannolikt är skillnaden dock resultat av att frågorna inte är exakt lika i de två undersökningarna. Levnadsnivåundersökningen kan på grund av sin ytligare frågeformulering ha ett högre mörktal, dvs. större underrapportering.

kvinnor är händelseplatsen oftast tunnelbana, tåg, buss eller -station eller någon annan offentlig plats (gata, torg osv.). Detta tyder på att det kan finnas skillnader mellan olika kvinnogrupper i fråga om vilken typ av våld det är fråga om. Finländska kvinnor skulle med andra ord drabbas av "gatufridsvåld" medan det bland jugoslaviska och svenska kvinnor skulle röra sig om primärgruppsvåld.

Den alternativa förklaringen att invandrarkvinnor skulle rapportera fler våldshändelser än svenska kvinnor kan tänkas bero på t.ex. en annorlunda uppfattning av intervjusituationen. Invandrarkvinnor kan uppfatta intervjuandet anonymare än vad svenska kvinnor gör.

7.4. Skillnader i utsatthet mellan invandrare med olika etnisk bakgrund

Andelen utsatta för någon typ av våld var högst bland jugoslaviska (11,4%) och bland finländska (8,8%) invandrare. Andelen utsatta var lägre bland de baltisk-sovjetiska (7,6%), de tyska, österrikiska och schweiziska (6,6%) och övriga nordiska invandrare (5,5%). Intensitetstalen, dvs. medeltalet, för något våld bland de utsatta visar att kvinnorna i grupperna finländare, tyskar, österrikare och schweizare samt i gruppen baltisk-sovjetiska invandrare har klart högre medeltal än män generellt och kvinnorna i de två övriga grupperna. Högsta medeltalen per utsatta för våld med synliga märken återfinns dock bland övriga nordiska män samt bland kvinnorna i den tyska gruppen.

Invandrargruppernas sammansättning på egenskaperna ålder, familjetyp, boendeform och bostadsort skiljer sig, i vissa fall rätt betydligt. T.ex. är andelen 15–24-åringar något större bland finländare än de övriga grupperna. Andelen 45–64-åringar å sin sida är klart lägst bland finländare och de jugoslaviska invandrarna. Åldersmässigt tillhör dessa två sistnämnda nationaliteter de generellt mest utsatta grupperna.

Andelen ensamstående är störst bland nordbor medan över hälften av jugoslaverna är samboende med barn. Bland de tyska, österrikiska, schweiziska och de baltisk-sovjetiska invandrarna dominerar samboende antingen med eller utan barn. Ur familjetypssynpunkt är alltså jugoslaverna oftare grupper med minst och nordborna grupper med störst sannolikhet att utsättas för olika typer av våldshändelser.

Övriga nordbor och tyska invandrare bor oftare (ca 50%) i villor eller radhus än de övriga invandrargrupperna. Denna andel är lägst (ca 20%) bland jugoslaverna. Koncentrationen till flerfamiljshus med minst fyra våningar är tydligast bland jugoslaver samt bland invandrare från de baltiska länderna, Sovjetunionen och övriga Österuropa. Det är framför allt den sistnämnda gruppen samt jugoslaver och finländare som ur bostadsformssynpunkt kan förväntas ha högre utsatthet.

Våldet är vanligast i Stockholm och i Göteborg. Av de olika invandrargrupperna har finländare inte den högsta andelen storstadsbor men de storstadsboende bland dem är oftast bosatta i Storstockholmsområdet. Endast var tionde övrig nordbo är bosatt i detta område men var femte av dem bor i Göteborgsområdet. Jugoslaver har ungefär lika andelar boende i vart och ett av de tre storstäderna. Ur ortstypssynpunkt är alla invandrare relativa högriskgrupper. Koncentrationen till storstäderna i allmänhet och till Stockholm och Göteborg i synnerhet ökar sannolikheten bland dem att utsättas för olika typer av våldshandlingar.

Sammanfattningsvis är de faktorer som samvarierar med större utsatthet för olika typer av våldshändelser bland finländare;

många unga, ensamstående personer bosatta i Storstockholmsområdet. Bland jugoslaverna är det åldersfördelningen och storstadsboendet, som i och för sig är karakteristiskt för alla invandargrupper, som medför större risker att utsättas för våld.

7.5. Individuella skillnader i utsatthet

I det föregående har vi diskuterat skillnader i utsatthet för olika former av våldsbrott mellan infödda svenskar och invandrare som *grupper*. Vi kunde konstatera att invandrare är mer utsatta än svenskar och att dessa skillnader inte försvann när hänsyn togs till skillnader i gruppernas sammansättning med avseende på kön, ålder, familjesituation, boendeform eller bostadsort. För att se om individuella skillnader i utsatthet förklaras av samma bakgrundsvariabler (med tillägg av en mera detaljerad familjecykelkategorisering (se någon av ULF-rapporterna för detaljer) och av yrkestillhörighet), genomfördes en regressionsanalys på *individnivå*. Beroende variabel är medelantalet rapporterade våldshändelser, dels de tre våldsformerna tillsammantagna, dels var och en av våldstyperna för sig.

Det summerade genomsnittliga antalet våldshändelser av någon form som rapporterats av infödda svenskar var 0,13. Motsvarande medeltal bland invandrarna var 0,23. Skillnaden är statistiskt säkerställd. När man tar hänsyn till köns- och åldersammansättningen sjunker skillnaden i medeltal våldshändelser något, nämligen till 0,10. Ytterligare reduktion åstadkoms inte av att man tagit hänsyn till familjesituation. När man utöver kön och ålder också kontrollerar för boendeform resp. bostadsort erhålles en reduktion av skillnaden i utsatthet i det förra fallet till 0,07 och i det senare till 0,08. När man kontrollerar för kön, ålder och yrke återfinns en skillnad på 0,09 mellan en genomsnittlig svensk och en invandrare.

Medeltalet rapporterade våldshändelser med synliga märken eller kroppsskada som följd var bland svenskar 0,03 och bland invandrare signifikant högre eller 0,06. Med köns- och åldersskillnader konstanthållna kvarstår skillnaden oförändrad men krymper något när man dessutom kontrollerar för familjetyp eller familjecykel (till 0,026 resp. 0,025). Ytterligare reduktion uppnås vid hänsyn tagen till boendeform (ner till 0,020) och ortstyp (0,022) resp. yrke (0,023) utöver kön och ålder.

Vid våld utan synliga märken upprepas ovanstående mönster. Skillnaden reduceras något när man tar hänsyn till kön, ålder och familjesituation (från 0,02 till 0,017) medan den största reduktionen åstadkommes vid kontroll av boendeform (till 0,011) och ortstyp (till 0,012) utöver kön och ålder. Att ta hänsyn till yrkestillhörighet utöver kön och ålder bidrar ytterst lite jämfört med när endast kön och ålder var konstanthållna (0,016 jfrt med 0,017).

Det genomsnittliga antalet hotelser som har rapporterats av infödda svenskar är 0,06 jämfört med invandrares medeltal på 0,13. Även denna skillnad är statistiskt säkerställd. Denna skillnad kvarstår oförändrad vid kontroll av kön, ålder och familjetyp samtidigt. En reduktion på 0,01 av skillnaden uppnås när man tar hänsyn till å ena sidan familjecykel och å andra sidan yrkestillhörighet med samtidig kontroll av kön och ålder. Kön, ålder och hustyp resp. kön, ålder och ortstyp medför en reduktion från 0,06 till 0,04 i medeltal rapporterade hotelser om våld mellan en genomsnittlig svensk och en invandrare.

Sammanfattningsvis kan sägas att skillnaden i medelantal våldshändelser mellan en

svensk och en invandrare som bäst reduceras till hälften (vid summerat medelantal våldshändelser) när hänsyn tas till kön, ålder och hustyp samtidigt. Vid varje våldstyp ger denna kontroll den största reduktionen av skillnaden.

En motsvarande regressionsanalys där infödda svenskar och andragenerationens invandrare utgör populationen visar det sig att egenskapen att endast ha en utländsk förälder har en signifikant effekt i individens benägenhet att utsättas för ett i genomsnitt större antal våldshändelser än egenskapen att vara infödd. Att ha två utländska föräldrar däremot medför ingen signifikant skillnad jämfört med infödda svenskar i utsatthetshänseende. Köns- och åldersstandardiseringen medför att effekten av etnisk bakgrund (en resp. två utländska föräldrar) inte längre har en signifikant effekt på utsatthetsbenägenheten avseende medeltal våldshändelser. Det visar sig med andra ord att skillnader i köns- och åldersfördelningen också förklarar skillnaderna i utsatthet när den studeras på individnivån.

7.6. Utsatthetsmönstret bland infödda och bland invandrare

Grupper med överrisk att utsättas för olika typer av våldsbrott, mätt i medeltal rapporterade händelser, är män, 15–24-åringar, ensamstående, höghusboende, arbetare och boende i Stockholm eller i Göteborg. Detta gäller framför allt bland infödda svenskar.

Bland invandrare är bilden något annorlunda. För det första visar det sig inte finnas någon signifikant skillnad i medelantalet våldshändelser mellan män och kvinnor. För det andra återfinns inte motsvarande överrisk för 15–24 åringar bland invandrare som bland de infödda. I andragenerationsgruppen kunde ingen signifikant ålderseffekt mätas. För det tredje gav inte yrkestillhörigheten bland invandrare något utslag i form av högre eller lägre utsatthetsbenägenhet såsom var fallet bland infödda.

Avsaknaden av ålderseffekt bland andragenerationen och av yrkeseffekt bland alla invandrare har troligen samma förklaring. Den skulle vara avsaknaden av variation i fråga om åldersspridning i den förra gruppen och av yrkesspridning i den senare gruppen.

Ur nationalitetssynpunkt blir vår slutsats att etnisk bakgrund i sig (dvs. vilken enskild nationalitet man representerar) inte har något förklaringsvärde för ens utsatthet för olika antal våldshändelser utöver det förklaringsvärde som ges av egenskapen att vara invandrare i stället för att vara infödd svensk.

8. Oro för våld

Nästan hälften av våldshändelserna äger rum på "gata, torg eller annan allmän plats" (28%) på "restaurang, dansställe, folkpark eller annat nöjesställe" (14%) eller på "tåg, buss, tunnelbana eller -station" (6%) (SCB, 1981). En enskild individ kan förebygga att utsättas för denna typ av våld genom att begränsa sin vistelse på sådana ställen eller genom att undvika dessa platser på speciellt "farliga" tidpunkter, sena kvällar och helger.

I undersökningen ställs frågan: "Har det på senare tid hänt att Ni avstått från att ge Er ut på kvällen av rädsla att bli överfallen, rånad eller på annat sätt ofredad?" Resultaten av svarsfördelningen visar att en betydande andel personer ofta eller ibland faktiskt avstår från att gå ut på kvällen av rädsla för att något skall hända dem. Man kan påstå att rädslan är överdriven. Ca 5 procent av befolkningen har faktiskt råkat ut för våld och allt våld har ingalunda ägt rum utanför hemmets väggar medan 12,5

Tab. 1. Gruppernas fördelning efter bakgrundsvariablerna kön, ålder, familjetyp, hustyp och bostadsregion. Antal och andel

	Infödda		Alla invandrare		Finland		Övriga Norden		Tyskland, Österrike (1)		Baltiska länder, Sovjetunionen (2)		Jugoslavien		Andragenerationen med			
															en utl. förälder		två utl. föräldrar	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Män	11 105	49,2	978	47	386	42	102	38	86	47	146	48	79	56	372	54	74	57
Kvinnor	11 463	50,8	1 121	53	530	58	170	62	95	53	155	52	61	44	320	46	55	43
15–24 år	3 510	15,8	313	15	179	20	30	11	16	9	23	8	22	16	274	40	51	39
25–44 år	8 277	37,2	1 041	50	460	50	101	37	71	39	127	42	97	69	249	36	47	36
45–64 år	7 124	32,0	630	30	241	26	124	46	75	41	120	40	21	15	133	19	16	12
65–74 år	3 316	14,9	116	5	36	4	17	6	19	11	32	10	–	–	35	5	15	12
Ensamstående utan barn	6 994	31,0	563	27	266	29	73	27	33	18	75	25	34	24	323	47	77	60
Ensamstående med barn	641	2,8	102	5	58	6	15	5	9	5	9	3	3	2	28	4	4	3
Samboende utan barn	7 536	33,4	575	27	219	24	82	30	71	39	113	37	25	18	150	22	28	22
Samboende med barn	7 406	32,8	859	41	374	41	103	38	68	38	105	35	78	56	192	28	20	15
Villa/radhusboende	12 900	56,5	734	35	305	34	144	53	91	50	106	36	28	19	344	50	57	44
1–3 vån. flerfam.hus	6 834	29,9	774	37	378	42	82	30	55	30	93	32	74	49	213	31	47	36
4 vån./högre flerfamiljshus	3 099	13,6	575	27	222	24	46	17	35	19	92	32	48	32	129	19	25	19
Storstockholm	3 515	15,6	566	30	251	27	29	11	53	29	85	28	22	16	182	26	43	33
Storgöteborg	1 785	7,9	245	12	69	8	57	21	24	13	43	14	24	17	75	11	18	14
Stormalmö	1 222	5,4	137	6	10	1	19	7	8	4	45	15	25	18	39	6	6	5
Övriga landet	16 049	71,1	1 151	55	586	64	168	62	97	54	129	43	68	49	395	57	62	48

[1] Schweiz.
[2] Övriga Österuropa.

procent av befolkningen säger sig ofta eller ibland ha avstått från att gå ut på kvällen på grund av rädsla. Det visar sig också att kvinnor oftare än män känner sådan oro och därför avstår från att gå ut medan förhållandet mellan könen i fråga om faktisk utsatthet är det motsatta (se också, Lenke, 1982). Resultaten tyder på att sådan oro är mera utspridd bland invandrare än bland svenskar oberoende av kön.

En annan sida av upplevd oro för eventuella våldshändelser belyses med frågor om man under senare tid har "känt sig orolig för våld/hot om våld i det egna bostadsområdet" eller "utanför det egna bostadsområdet". Bland såväl svenskar som bland invandrare är rädslan för våld vanligare när det gäller ställen utanför det egna bostadsområdet än i de egna bostadskvarteren. Återigen visar det sig att invandrare oftare än svenskar känner sådan oro, oberoende av om frågan gällde det egna bostadsområdet eller utanför det.

Eventuellt betecknande för bostadsortens kvalitet är att endast 6 procent av de infödda känner sig oroliga för våld i de egna bostadskvarteren jämfört med ca 12 procent av invandrarna. Relativt sett är skillnaden mindre mellan grupperna avseende andelen som sagt sig ha känt oro för våld utanför det egna bostadsområdet (15 procent av infödda och 20 procent av invandrare).

9. Sammanfattning

I artikeln diskuteras utsatthet för våld som har lett till synliga märken eller kroppsskada, för våld utan sådana synliga påföljder samt för hot eller hotelser om våld som var farliga eller så allvarliga att man blev rädd. Vi jämför svenskar och invandrare, dels olika grupper där invandrares båda föräldrar har eller har haft samma utländska medborgarskap, dels den s.k. andragenerationens invandrare fördelade på dem med en svensk och en utländsk förälder och dem med två utländska föräldrar oberoende av om föräldrarna kommer från samma eller från två olika länder utanför Sverige.

Resultaten som avser åren 1978–1980 genomförda intervjuundersökningar, visar att invandrargruppen i sin helhet genomgående har utsatts för våld i större utsträckning än den infödda befolkningen. Skillnaderna kvarstår, dock på en lägre nivå, även efter att man har tagit hänsyn till skillnader mellan grupperna vad avser sådana egenskaper som har befunnits samvariera med utsattheten. De egenskaper vi syftar på är kön, ålder, vilken familjetyp man tillhör, vilken bostadsform man bor i och var någonstans i landet man bor samt ens yrkestillhörighet.

Dessa egenskaper samvarierar med utsattheten bland såväl svenskar som bland invandrare. Det föreligger emellertid tre undantag som tyder på att utsattheten varierar i viss mån annorlunda bland svenskar resp. bland invandrare. Det första är att invandrarkvinnor *inte* har en signifikant lägre sannolikhet att utsättas för olika våldshändelser än invandrarmän. Det andra är att medan utsattheten bland de infödda är störst i den yngsta åldersgruppen är den bland invandrare högre bland 25–44-åringar än bland 15–24-åringar. Det tredje är att är man invandrare, spelar yrkestillhörigheten ingen signifikant roll för den genomsnittliga risken att bli utsatt. Det beror tydligen på att invandrarnas utsatthetsnivå dels är relativt hög, dels varierar mycket mellan individerna, dels att det inte finns någon speciellt stor variation i invandrares yrkesstruktur till skillnad från svenskarnas fördelning efter yrke. Antalsmässigt domineras invandrargruppen av finländare, vilka

till ca tre fjärdedelar är arbetare, den generellt mest utsatta yrkesgruppen i hela befolkningen. Om tillräcklig yrkesspridning fanns bland invandrare är det möjligt att deras utsatthet skulle ha samma mönster som den bland svenskarna. Fullt möjligt är emellertid att det även i det fallet skulle visa sig att invandrarna var mera utsatta.

I brist på motsvarande spridning i fråga om ålder bland andragenerationens invandrare ger ålder ingen direkt effekt på utsattheten i denna grupp. Frånvaron av de äldre, generellt mindre utsatta, medför dels att det inte finns tillräckligt stora jämförbara åldersgrupper, dels att utsattheten för hela gruppen blir större.

Nationalitetstillhörigheten bland invandrare gav ingen signifikant förklaring till skillnader i utsattheten för olika våldsbrott. De skillander som finns mellan olika nationalitetsgrupper förklaras främst av skillnader i gruppernas sammansättning efter ålder, familjetyp, bostadsform och bostadsort.

Huruvida det är fråga om diskriminering, interna motsättningar inom grupperna, konflikter inom familjerna, förekomsten av andra sociala problem eller skillnader i levnadssätt mellan invandrare och svenskar som orsakar invandrares större utsatthet låter sig inte uttalas med hjälp av det material vi här har kunnat bearbeta. I en brittisk studie där man också ställde frågor om hur ofta man gick ut på kvällarna och i vilket syfte, fann man en stark korrelation mellan frekvensen att gå ut och att bli utsatt för våldshändelser. (Sparks, et al. 1977). Det är inte osannolikt att invandrare har en livsstil som innebär att de utsätter sig själva för större risker genom att vistas ute på kvällarna och på platser där våldet är vanligast.

Oron för våld och hotelser om våld är mera utspridd bland invandrare än bland svenskar. Denna oro kan sägas vara högst befogad med tanke på den större erfarenheten av våldshändelser som invandrarna har. Såväl bland svenskar som invandrare är emellertid oron mera utspridd än den faktiska risken att usättas för olika typer av våldshändelser.

10. Referenser

Invandrare och Minoriteter, 1983: 2: "Vilka är andragenerationens invandrare?" Sven A. Reinans vs. Jan Ekberg.

Lenke, L. (1978): "Risken för våldsbrott i Sverige – i komparativ belysning", Stencil, Institutionen för Sociologi och Kriminologi, Stockholms universitet, 1978.

Lenke, L. (1982): "Gatuvåldet i Stockholm – En kriminologisk analys av våldsbrottsutvecklingen i Stockholm" Stencil, Stockholms universitet 1982.

Nordiska Rådet och Nordiska statistiska sekretariatet, 1983: Levnadsnivå och ojämlikhet i Norden. En komparativ analys av de nordiska levnadsnivåundersökningarna. I Finland angav 18%, i Danmark 6%, i Norge 5% och i Sverige 9% att de hade utsatts för våld.

Persson, L. G. W. (1977: 6): "Brottsstatistikens tillförlitlighet" i Statistisk tidskrift 1977: 6.

Persson, L. G. W. (1977: 7): Offer för tillgrepp, skadegörelse och våld – en redovisning av 1974 års offerundersökningar, SCB, Promemorior, 1977: 7.

Reinans, S. A. (1982): "Om den andra generationen" i (red), Hamberg, E. och Hammar, T.: Invandringen och framtiden, Publica 1982.

Reynolds, P. D. & Blyth, D. A. (1974): "Sources of Variation affecting the Relationship Between Police and Survey-based Estimates of Crime Rates" i Drap-

kin, I. & Viano, E. (eds.): Victimology (vol. 1–5). Lexington 1974.

SCB, (1981): Offer för vålds- och egendomsbrott, Rapport nr 24, SCB, Stockholm 1981.

SOU 1974:41; Invandrarutredningen.

SOS, 1980; Befolkningsförändringar, del 3, 1980. Tabell 7.6. Utlänningar som erhållit svenskt medborgarskap efter invandringsår och medborgarskap vid invandringen i promille av resp. immigrantkohort 1968–1980.

SOU 1981:38; Hets mot folkgrupp.

Sparks, R. F., Genn, H. G., Dodd, D. J. 1977: Surveying Victims. A study of the Measurement of Criminal Victimization, Avon 1977.

Sveri, K. (1982): "Comparative Analyses of Crime by Means of Victim Surveys – The Scandinavian Experience", i Schneider, H. J. (red), The Victim in International Perspective, 1982.

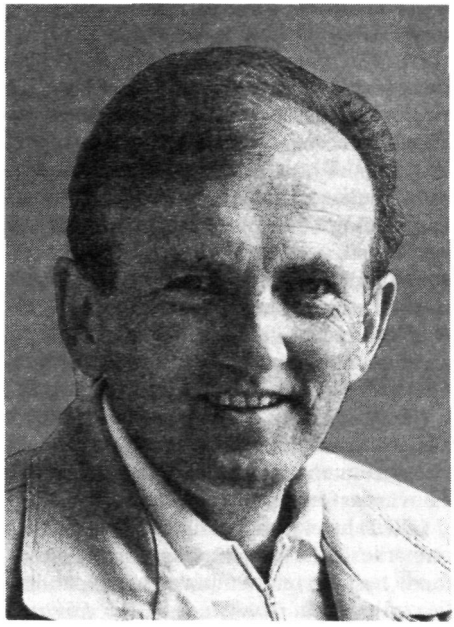
Ålund, A. (1978): Migration och sociala förändringsprocesser, Umeå universitet, Sociologiska inst. Nr. 44, 1978.

Meddelanden

SCB får ny chef



Lennart Nilsson



Sten Johansson

Generaldirektören i statistiska centralbyrån, Lennart Nilsson, har fr. o. m. den 1 april 1983 avgått från sin tjänst, för att tillträda befattningen som generaldirektör och chef för riksskatteverket. Han har varit SCBs chef sedan 1 januari 1979. Till ny chef har regeringen utsett professor Sten Johansson. Han tillträder sin befattning den 1 augusti 1983.

Född 1939 i Övermorjärv, Norrbottens län, studerade Sten Johansson i Uppsala och blev fil.kand. 1963 (historia, sociologi, ryska) och fil.lic. 1967. Åren 1967–1971 medverkade han i låginkomstutredningen och erhöll år 1972 en professur i välfärdsforskning vid Institutet för social forskning (SOFI) knutet till Stockholms universitet. 1982 – i samband med regeringsskiftet – blev han sakkunnig åt statsrådet Ingvar Carlsson i statsrådsberedningen.

Sten Johanssons kontakter med SCB har under åren varit många och långa. Han har vid ett flertal tillfällen medverkat i denna tidskrift med artiklar och han har varit ansvarig för levnadsnivåundersökningarna 1968, 1974 och 1981, som han gjorde i samarbete med SCB.

Europeiska statistiska möten

Miljöstatistiska indikatorer (ECE)

I Europeiska statistikerkonferensens serie av informella möten kring olika miljöstatistiska frågor sammanträdde representanter för ett tiotal länder i Genève den 7–9 mars 1983 för en förberedande diskussion om miljöindikatorer. Från Sverige deltog avdelningschef Knut Medin, SCB.

Inom ECE har ett arbete på miljöindikatorer bedrivits under ledning av *Senior Advisers to ECE Governments on Environmental Problems*. En *task force* med Jugoslavien som *lead country* och med Sverige bland deltagarländerna har tagit fram en uppsättning om 120 indikatorer av vilka 40 ansetts vara särskilt angelägna. Med utgångspunkt härifrån sker för närvarande en försöksverksamhet. *Senior Advisers* har emellertid samtidigt överlämnat frågan till andra ECE-organ, däribland statistikerkonferensen.

Inom ramen för det österuropeiska ekonomiska samarbetet har i bortåt tio år funnits ett system av miljöindikatorer. Detta system är nu föremål för översyn.

OECD har sedan åtskilliga år arbetat med ett närbesläktat projekt, där man för närvarande tar fram andra utgåvan av en publikation med titeln *The State of the Environment in OECD Member Countries*.

Diskussionerna vid mötet baserades i stor utsträckning på ett dokument som redogjorde för de tre nämnda projekten. I ett särskilt dokument behandlades sambandet med sociala indikatorer.

I Kanada har diskussionen om miljöstatistikens utformning utgått från den *stress* på miljön som orsakats av föroreningar etc. Vid mötet framlades ett dokument med titeln *Indicators of Ecosystems under Stress*. Ansatsen är intressant bl. a. därför att den så renodlat placerar biologiska aspekter i centrum.

Avsikten med detta möte var att det skulle förbereda ett möte under första halvåret 1984 med hela arbetsgruppen för miljöstatistik. Som ämnen för detta möte föreslogs (1) behovet och användningen av miljöindikatorer inklusive sambandet med miljöstatistiken som helhet, (2) miljöområden som bör täckas av ett system med miljöindikatorer, (3) kvaliteten på miljöindikatorerna särskilt med sikte på internationella jämförelser, (4) indikatorer för ekosystem under stress, (5) sammansatta miljöindikatorer.

Knut Medin

OECD-möten

Expertmöte om FoU-statistik

Expertgruppen för forskningsstatistik (FoU-statistik) höll möte i Paris 24–25 februari 1983. Från Sverige deltog Tage Berglund, SCB.

Ett förberedande möte med de nordiska ländernas representanter hölls 23 februari i den svenska OECD-delegationens lokaler. Vid detta möte gjordes en avstämning av de nordiska ländernas synpunkter och en fördelning av arbetsuppgifter inför expertgruppsmötet.

Den första diskussionspunkten avsåg FoU-statistiken inom högskolesektorn. Kvaliteten på denna statistik är inte särskilt tillfredsställande för flertalet medlemsländer.

En allmän kvalitetshöjning är därför önskvärd. Både nationella och internationella ansträngningar behövs. Detta har påpekats bl. a. från de nordiska ländernas sida både i expertgruppen och i den politiska kommitté (Committee for Scientific and Technical Policy, CSTP) under vilken expertgruppen sorterar. Arbetet med FoU-statistiken för detta område kommer därför att få en högre prioritet under de närmaste åren.

Huvuddiskussionen avsåg den allmänna inriktningen av det framtida arbetet inom detta område, nationellt och inom OECD-sekretariatet.

I en kort historisk tillbakablick konstaterades att FoU-statistiken från början var inriktad på en beskrivning av resursinsatserna (input indicators) för FoU-verksamheten. Ett relativt tillfredsställande FoU-statistiskt system, med det ovan nämnda undantaget för högskoleområdet, hade utvecklats. I en analys av konsumentbehov

m. m. som gjordes 1977–78 av en särskild arbetsgrupp tillsatt av CSTP (Mullin Group) konstaterades ett behov även av informationer som avsåg resultat eller effekter (output indicators) av FoU-verksamheten. En serie av arbetsgruppsmöten och seminarier har därefter genomförts där pågående arbeten med olika resultatindikatorer i olika länder har redovisats och diskuterats.

Det är nu dags för ett ställningstagande till vilken balans man skall ha i fördelningen av de statistiska resurserna mellan insatsindikatorer och resultatindikatorer. De arbeten avseende resultatindikatorer som hade gjorts eller pågick i olika länder var av tillfällig karaktär, ofta forskningsprojekt vid universitet eller forskningsinstitut. Om en allmän och kontinuerlig dataförsörjning skulle komma till stånd krävdes nationella ställningstaganden på politisk nivå med tilldelande av ansvar och resurser för verksamheten.

Det konstaterades att det var önskvärt med mera löpande verksamhet avseende de mest användbara resultatindikatorerna och att konsumenternas behov tillgodosågs bäst vid en "harmonisk" balans mellan insats- och resultatindikatorer.

Om denna nya målsättning för det statistiska arbetet skulle antas ansågs det lämpligt att expertgruppens mandat vidgades och medlemsländernas representation sågs över. Eventuellt borde gruppens namn ändras till "Expert Group on Science Indicators".

Vid mötet diskuterades även vissa metodfrågor, bl. a. problem med att avgränsa försvarsforskning mot civil forskning i olika medlemsländer med betydande försvarsforskning.

Information lämnades vidare om utvecklingen av OECDs databank avseende FoU-statistiken. Genom införandet av ett nytt system (OSIRIS) hoppades man få möjligheter att enkelt och snabbt besvara frågor från olika medlemsländer, förutom att man för sitt eget arbete fick ett kraftfullt instrument för lagring, underhåll och hantering av FoU-indikatorer.

Tage Berglund

Utbildnings- och sysselsättningsåtgärder för ungdomar i 16–19-årsåldern

OECDs sekretariat för "Social affairs, manpower and education" har inför ett ministerrådsmöte våren 1983 fått i uppdrag att sammanställa statistik som rör utbildningsdeltagande och sysselsättningsförhållanden för ungdomar i 16–19-årsåldern i ett urval OECD-länder. Genom utbildningsdepartementet blev jag tillfrågad om jag kunde sammanställa aktuella uppgifter för Sveriges del och presentera dem vid en konferens i Paris 6–8 oktober 1982.

Man hade inbjudit representanter från Italien, Frankrike, Holland, Västtyskland, Storbritannien, Japan och Sverige. Från de fyra förstnämnda länderna samt från Sverige presenterades papers som diskuterades vid konferensen. De två huvudpunkterna i diskussionen blev materialens jämförbarhet samt likartade trender i de olika länderna. Japans och Storbritanniens representanter bidrog med muntliga redogörelser för förhållandena i resp. land.

Nomenklaturdiskussionen visade sig mycket befogad, då de redovisningskategorier som användes i rapporterna ofta stod för skilda saker trots att gemensamma beteckningar används. "Technical education" på gymnasial nivå innefattar t. ex. för Italiens del både ekonomiska och tekniska inriktningar. Kursplanerna på resp. område innehåller dessutom väsentliga mått av humaniora, närmast jämförbart med humanistiska linjen i Sverige (utom språkprogrammet).

I samtliga fem länder som lämnat skriftliga rapporter kunde en gemensam trend mot ökat intresse för yrkesutbildning märkas bland ungdomarna. Trenden tog sig dock olika uttryck i de olika länderna. Ungdomar i Västtyskland visade ett fortsatt starkt intresse för teoretiska studier på gymnasial nivå, men ett vikande intresse för universitetsstudier. I stället fortsatte en starkt ökande andel av ungdomarna som fullföljde teoretiska gymnasiestudier med lärlingsutbildning. I Italien har den "studieavbrottstradition" som finns i landet markerats ytterligare genom övergång till yrkesutbildning bland avbrytare på teoretis-

ka linjer. Det traditionella lärlingssystemet i Italien växer också på bekostnad av yrkesutbildningen i skolmässiga former.

Ungdomsarbetslösheten visar ett likartat utvecklingsmönster men på skilda nivåer för de olika länderna. Japan utgör ett undantag både när det gäller ungdomsarbetslöshet och intresse för yrkesutbildning. Japanska ungdomar fortsätter i väsentligt större utsträckning än ungdomar från övriga representerade länder med teoretiska studier både på gymnasial och eftergymnasial nivå. Ungdomsarbetslösheten visar någon tendens till ökning, men ligger ändå på en nästan försumbart låg nivå.

Den som är intresserad av att ta del av de skriftliga rapporterna kan vända sig till Mats Myrberg, prognosinstitutet, SCB.

Mats Myrberg

Det nordiska chefsstatistikermötet 1982

Det 30e nordiska chefsstatistikermötet hölls den 9–11 augusti 1982 i Oslo. I mötet deltog 24 personer från de nordiska statistiska centralbyråerna samt Nordiska statistiska sekretariatet. Den för mötet utarbetade dokumentationen, diskussionsreferaten samt de nordiska statistiska utskottens verksamhetsrapporter 1979–1982 och samarbetsprogram för 1982–1985 ingår i mötesrapporten, som kommer att publiceras i Nordisk statistisk skriftserie. I det följande ges därför endast en summarisk redogörelse för mötet.

1. Dagordning

Mötets dagordning omfattade bl. a. följande punkter:

- Utbildning av personal i och utanför centralbyråerna.
- Lovmodelarbejdet i Danmark.
- Den tekniska utvecklingen och dess konsekvenser för statistikproduktionen.
- Databaser för statistikproduktion – möjligheter och problem för de nordiska centralbyråerna och den nordiska statistiken.
- Användaranpassad statistikproduktion.
- Videregivelse eller salg af elementargruppeoplysninger med henblik på videre bearbejdning hos modtager.
- Det nordiska statistiska samarbetets framtida utformning.
- De nordiska statistiska utskotten.
- Det internationella statistiska samarbetet.

2. Utbildning av personalen i och utanför centralbyråerna

Diskussionsunderlaget till denna punkt hade utarbetats av Statistisk Sentralbyrå (SSB) i Norge.

Verksamheten vid en modern statistisk centralbyrå är mångsidig. Detta återspeglas naturligtvis i arbetskraftens sammansättning:

- Handläggare med och utan akademisk utbildning.
- ADB-personal med och utan utbildning vid universitet eller högskola.
- Kontorspersonal.

En väl genomförd rekryteringspolitik på alla nivåer är naturligtvis en första förutsättning för att centralbyrån skall få den personal som behövs. Dessutom bör de anställda erbjudas utbildning de tidigare ej haft möjlighet att gå igenom samt utbildning för såväl fördjupning som vidareutveckling av tidigare förvärvade kunskaper.

Delar av den interna utbildningen vid centralbyråerna bör vara obligatorisk, medan andra delar av utbildningsprogrammet kan framstå som ett erbjudande till de grupper till vilka de riktar sig. Bland de obligatoriska kurserna kommer naturligtvis utbildningen av den personal som skall vara verksam inom centralbyråns ADB-avdelning.

Vid en utvärdering av utbildningsproblemen bör man speciellt fästa uppmärksamheten på den största personalgruppen vid centralbyrån – kontorspersonalen – som vanligtvis inte har någon utbildning utöver grundskola eller gymnasium. Goda allmänkunskaper är naturligtvis viktiga men inte tillräckliga för att kunna utföra de speciella arbetsuppgifter som de blir ålagda. Merparten av de specialkunskaper som krävs tillägnas genom det dagliga arbetet. En obligatorisk kurs med tonvikten lagd på praktisk

statistik vid sidan av information från personalavdelningen och kunskaper om tekniska hjälpmedel synes vara på sin plats för att i någon grad hjälpa denna personalkategori ur den fackliga isolation arbetet annars lätt för dem in i. Liknande kurser får troligen störst genomslagskraft om de berörda personerna kallas in efter att ha arbetat 6–12 månader vid centralbyrån. Också akademiker behöver undervisning i form av obligatoriska kurser. Denna personalkategori har ofta stora teoretiska kunskaper medan de däremot inte sällan är dåligt insatta i hur en statistisk centralbyrå arbetar.

Kontorspersonalen har under årens lopp samlat på sig ett avsevärt praktiskt statistiskt kunnande. Med utvidgade teoretiska kunskaper kommer många av dessa medarbetare att kunna utföra väsentligt mer krävande arbetsuppgifter. Den teoretiska utbildning det här är fråga om bör erbjudas de personer som önskar en sådan utbildning. För att kurser av detta slag skall uppnå önskad effekt bör de sträcka sig över flera hundra undervisningstimmar. Undervisning bör ges i elementär statistik, ekonomisk teori, demografi och sociologi samt ADB och administration. I Norge har man i dag 20 års erfarenheter av denna kurstyp. Erfarenheterna av utbildningen är så goda, sett ur såväl personalens som SSBs synpunkt, att dessa kurser ofta betraktas som flaggskeppet i utbildningsverksamheten.

Under diskussionen konstaterades att erfarenheter av och problem i samband med personalutbildningen vid de statistiska centralbyråerna uppvisar många gemensamma drag. Kontakter mellan de utbildningsansvariga vid resp. centralbyrå borde därför vara ett lämpligt forum för ett mer ingående informationsutbyte. Chefsstatistikerna tillstyrkte i princip att de utbildningsansvariga träffas någon gång under verksamhetsperioden 1982–1985. I alla de nordiska länderna har utbildningsverksamheten uppvisat en stark tillväxt under senare år. Ett nordiskt samarbete i dessa frågor borde öka möjligheterna för ytterligare förbättringar av den interna utbildningen.

3. Lovmodelarbejdet i Danmark

Från dansk sida hade ett diskussionsunderlag utarbetats för denna punkt på dagordningen. År 1978 påbörjades arbetet med att utveckla en modell för konsekvensberäkningar för olika lagförslag eller lagjusteringar. Dessa beräkningar skall utföras snabbt och kräver tillgång till en mängd statistiska uppgifter och samtidigt skall samspelet mellan olika lagregler kunna redovisas.

Modellen är en central ADB-modell, som finns i Danmarks Statistik och användarna – för närvarande 7 ministerier och styrelser – har tillgång till denna via terminal. Modellen består i huvudsak av tre komponenter:

- ett kartotek med gällande lagar, främst på skatte- och socialutgiftsområdena. I denna formaliserade modell av lagregler ingår inte lagstiftningens parametrar dvs. inkomstgränser och procenter. Av praktiska orsaker har dessa placerats i en s. k. lagdatabank;
- modellbefolkningar, som är det statistiska underlaget för beräkningarna;
- beräknings- och analysdelen, som möjliggör utförandet av önskade räkneoperationer o. d.

Modellen är i själva verket en ram för konsekvensberäkningar. Konsumenterna bestämmer själva parametrarnas värden och kan med hjälp av modellen beräkna t. ex. vilka följder ändringar i vissa skatter och bidrag får för olika låginkomstfamiljer.

Modellen kan tillsvidare endast ge svar på frågor om de omedelbara fördelnings- och budgetmässiga följderna. De indirekta konsekvenserna ingår inte i modellen, men supplerande beräkningar kan utföras med hjälp av andra modeller, t. ex. med Danmarks Statistiks ekonomiska modell ADAM.

Den danska lovmodellen har väckt stort intresse i de övriga nordiska länderna. I Norge har mindre modeller utarbetats för motsvarande ändamål och i Finland och Sverige föreligger planer på att utarbeta liknande rammodeller. Mötesdeltagarna diskuterade ansvarsförhållande, uppdateringsproblem och användningsmöjligheter. Det rådde delade meningar om vilka institutio-

ner som skall ha rätt att utnyttja modellerna. I Danmark är användarkretsen begränsad på grund av det detaljerade innehållet. Man önskar skydda materialet mot eventuell bakvägsidentifiering och dessutom är antaganden om den framtida utvecklingen och påtänkt lagstiftning som ingår i modellen hemliga.

4. Den tekniska utvecklingens konsekvens för statistikproduktionen

Från finländsk sida presenterades ett dokument om den informationstekniska utvecklingen under de senaste tio åren. Utvecklingen på både hardware- och softwareområdena går mot ökad decentralisering, ökade möjligheter att använda ADB-media utan allt för stora krav på förkunskaper och ett ökat intresse bland konsumenterna att utnyttja ADB.

Betraktar man den tekniska utvecklingens konsekvenser för utvecklingen i de statistiska centralbyråernas externa miljö kan man konstatera att antalet informationskällor märkbart ökar och att den direkta statistikinsamlingen därigenom kommer att minska. Samtidigt måste man öka resurserna för koordinering och standardisering av data både inom den offentliga och den privata sektorn.

Beträffande konsekvenserna för de statistiska centralbyråernas interna verksamhet kan sägas att möjligheterna att decentralisera databehandlingen ökat. Därigenom kan produktionsprocessen omorganiseras och många arbetsmoment och delprocesser kan förläggas till ämnesenheterna. Personalen vid ämnesenheterna får utvidgade arbetsuppgifter samtidigt som den nya tekniken övertar en stor del av rutinarbetet.

Genom decentraliseringen minskar kommunikationsbehovet mellan enheterna och antalet fel på grund av kommunikationsmissförstånd förväntas minska på sikt. Dessutom minskar flaskhalsarna i processerna och arbetet kan ses som en helhet dvs. fel som beror på en splittrad arbetsprocess minskar.

Den nya tekniken ger också nya möjligheter till användning av äldre data och sammankoppling av data från olika källor och processen kommer att bli ännu billigare och enklare än idag.

Den tekniska omvälvningen ställer emellertid stora krav på data-administrationen. De statistiska centralbyråernas datalager måste ses som en helhet och allt mer uppmärksamhet bör riktas på att göra denna helhet enhetlig. En viktig förutsättning för processen är utbildning av personalen. En stor del av utbildningen kan ske i form av inlärning genom arbete, men dessutom behövs kurser om bl. a. kopplingsmöjligheter och om externa datasystems innehåll och kvalitet.

Ytterligare betonades att utvecklingen bör styras av centralledningen och helst ske stegvis.

Mötesdeltagarna var eniga om att utvecklingen går mot ökad decentralisering av data- och informationsbehandlingen inom samtliga nordiska centralbyråer, men att man får gå försiktigt tillväga så man inte bortrationaliserar de goda sidorna hos en centraliserad organisation. Bl. a. bör genomförandet av standardiserade systemlösningar och program styras centralt och en central övervakning av de decentraliserade systemen kan också bli nödvändig.

5. Databaser för statistikproduktion — möjligheter och problem för de nordiska centralbyråerna och den nordiska statistiken

Inför diskussionerna kring databaser för statistikproduktion hade ett dokument utarbetats vid statistiska centralbyrån (SCB) i Sverige. Inom flera nationella statistiska centralbyråer diskuteras vilken funktion databaser skall ha vid distribution av statistik. Vid SCB pågår en orientering mot en mer marknadsnära roll. Det innebär att intresset mer än tidigare riktas mot att föra ut statistik till de olika användarkategorierna på ett sätt som är anpassat till deras behov. Man siktar alltmer till att söka tillgodose det samlade statistikbehovet och utgår då inte enbart från den egna producerade statistiken. Man söker också presentera statistiken på ett sådant sätt att det gagnar den enskilde användaren.

En allt viktigare del av statistikbehoven gäller internationell statistik och däribland

även statistik om de nordiska länderna. Via ett vidgat samarbete skulle man kunna tillgodose behoven av datadistribuerad statistik från de nordiska länderna. I första hand skulle ett sådant samarbete baseras på utbyte av statistik i maskinläsbar form. I andra hand skulle det kunna vidgas dels kring ett tekniskt samarbete kring databasuppbyggnaden dels till att söka bygga upp databaser med rimligt jämförbar nordisk statistik. Samarbetet skulle också kunna avse distributionen inom Norden av statistik från internationella organ såsom OECD, IMF, FN med flera. Speciellt inom det privata näringslivet betraktas Norden ofta som en marknad. En viktig del av försörjningen med internationell statistik gäller därför i själva verket statistik från Norden. För att tillgodose denna efterfrågan skulle man därför vilja se dels de olika ländernas centralbyråer som förmedlare av varandras nationella statistik dels öka mängden av nordisk statistik – i meningen statistik där man så långt möjligt sökt skapa jämförbarhet mellan de olika länderna.

Under diskussionen berördes bl. a. frågan om marknadsföring av redan existerande och planerade tidsseriedatabaser. Problem som centralbyråerna börjat ägna allt större uppmärksamhet. Eftersom statistikbyråerna inte längre har en monopolställning på distribution av statistik gäller frågan om marknadsföringen skall ligga hos centralbyråerna eller dataserviceföretag. Centralbyråernas uppgift att handha statistikproduktionen kommer att kvarstå också i framtiden. Men att kräva monopol på marknadsföringen av offentlig statistik borde inte vara möjligt.

Den ökade efterfrågan på internationell och däribland nordisk statistik för bl. a. konjunkturbedömningar och marknadsundersökningar kan göra det svårt för centralbyråerna att tillgodose alla önskemål. Ett närmare nordiskt samarbete i databasfrågor skulle kunna lösa en del av dessa problem. Genom ett utbyte av statistik i maskinläsbar form skulle man kunna tillgodose efterfrågan på statistik från de nordiska länderna. Arbetet bör i första hand inriktas på att finna gemensamma lösningar på enklare praktiska problem när det gäller utbytet. Möjligheterna till mer sofistikerade problemlösningar som syftar till att koppla samman databaser vid de nordiska statistikbyråerna kan komma att diskuteras på längre sikt.

6. Frågor kring statistikens presentation. – Användaranpassad statistikpresentation

Statistiska centralbyrån (SCB) i Sverige hade utarbetat diskussionsunderlaget till denna punkt på dagordningen. Frågan om hur vi distribuerar och presenterar den statistik som produceras har de senaste åren alltmer kommit i förgrunden inom SCB. Bakom detta ligger en förändrad syn som lite tillspetsat kan uttryckas så att centralbyrån bör utvecklas från ett tabellverk till ett informationsverk.

Användaranpassning innebär att ta hänsyn till att statistikanvändarna är olika. De har:

- olika behov av statistik
- olika gott om tid för att ta del av statistik
- olika intresse för statistik
- olika förutsättningar för att förstå statistik.

Det finns flera vägar att gå fram för att åstadkomma en användaranpassning. Hur man väljer att göra och hur långt man går i anpassning till olika användares behov är en fråga om vad som är praktiskt och ekonomiskt möjligt.

En första åtgärd för att användaranpassa publiceringen kan vara att sprida olika mängd statistik till olika användarkategorier. En väg att gå kan vara att minska mängden detaljuppgifter i statistikredovisningen så att innehållet i publikationerna närmar sig den redovisningsnivå som svarar mot behovet hos tjänstemän i förvaltningen, allmänutredare m. fl.

En andra åtgärd för att användaranpassa publiceringen, genom att ta hänsyn till olika användares tidsresurser, förutsättningar att tillgodogöra sig statistik samt intresse för statistik, kan vara att presentera statistiken med olika grad av vidareförädling, dvs. öka inslaget av analys vid presentationen.

En tredje åtgärd skulle kunna vara att öka problemområdesorienteringen. Ett sätt

skulle då vara att satsa på problemområdesorienterade publikationer som ett komplement till den nuvarande produktionsorienterade redovisningen.

Ett fjärde sätt, slutligen, skulle vara att i presentationen bygga på de erfarenheter från masskommunikationsområdet som visar hur information skall vara uppbyggd för att väcka intresse och för att bibehålla läsarens intresse.

Det framhölls under diskussionen att försöken med att få en komplett bild av användarnas statistikbehov har visat sig vara en i det närmaste omöjlig uppgift. Ur centralbyråernas synpunkt sett borde det därför vara lämpligast att pröva sig fram till en publiceringsform som kan accepteras av de flesta. Trots utvecklingen av datatekniken kommer kraven på en genomtänkt publicering att finnas för lång tid framåt. Databaserna kommer visserligen att öka i betydelse men det dröjer ännu lång tid innan allmänheten kan tillgodogöra sig den information som finns och kommer att finnas tillgänglig i databaser. Publikationer som har karaktären av baspublikationer, t. ex. folk- och bostadsräkningar och utrikeshandelsstatistik, är publikationer av bestående värde och kan inte ersättas av spridning via databaser.

7. Videregivelse eller salg af elementargruppeoplysninger

Många statistikkonsumenter har behov av att få tillgång till datamaterial som de snabbt och flexibelt kan använda i olika analyser. För att tillmötesgå dessa önskemål kan de statistiska centralbyråerna etablera databaser eller ge uppgifterna i maskinläsbar form. Uppgifterna bör vara mycket detaljerade för att motsvara ändamålet. De kan ges i form av "modelldata" dvs. varje post gäller ett avidentifierat statistiskt objekt (en person, familj, ett företag osv.).

De statistiska centralbyråerna är restriktiva med att utlämna detaljerat material till forskare och forskningsinstitutioner och vissa åtgärder måste vidtagas för att förhindra röjande:

a) Mottagarkretsen begränsas och materialet betraktas som förtroligt. Material som eventuellt skall publiceras bör kontrolleras av centralbyrån.
b) Datainnehållet begränsas. Avgörs från fall till fall beroende på ändamålet och forskarens "marknadskännedom".
c) Införande av "störningar" i dataposterna. Dessa ändringar förändrar bilden av de enskilda dataposterna, men påverkar inte det statistiska resultatet.

I diskussionen rådde det enighet om att de statistiska centralbyråerna bör vara restriktiva med att utlämna primärmaterial o. d. och att man inte önskar få någon låst situation genom förbud. Möjligheten att efter prövning överlåta dylikt material till forskare och forskningsinstitutioner för godtagbara ändamål ville man fortfarande ha kvar. Detta trots skärpta integritetsregler som pålagts av de registerkontrollerande myndigheterna eller av centralbyrån själv. Det konstaterades att samarbetet med de registerkontrollerande myndigheterna i Danmark, Norge och Sverige i dag är tillfredsställande sedan den offentliga integritetsdebatten dämpats samtidigt som kravet på mer detaljerade uppgifter ökat från vissa konsumentgrupper.

Ytterligare jämfördes ländernas regler för och övriga praktiska detaljer vid utlämnande av primärmaterial. Det nämndes också i detta sammanhang att Europarådet håller på med att behandla en rekommendation om datamoral inom forskning och statistik. Rekommendationen omfattar rätt vittgående regler och tar också upp problemet med indirekt röjande. Om rekommendationen skall tillämpas kan den få oacceptabla följder för statistikproduktionen, t. ex. om varje enskild tabell skall tas upp till integritetsprövning och speciellt för kontroll av eventuella möjligheter till indirekt röjande.

8. Det nordiska statistiska samarbetets framtida utformning

Under denna punkt presenterades från Nordiska statistiska sekretariatets sida ett dokument om utvecklingen av det nordiska statistiska samarbetet och förslag till dess framtida utformning med hänsyn till det kärva ekonomiska läget.

Chefsstatistikermötet framhöll värdet av

samarbetet och ställde sig positiv till att utvidga samarbetet till nya ämnesområden. Det restriktiva budgetläget i de nordiska länderna kräver dock att det totala antalet möten inte får överstiga antalet under verksamhetsperioden 1979–1982. För att skapa utrymme för ett samarbete på flera områden enades man om huvudprincipen att utskotten bör hålla ett möte vartannat år, men att fler möten kan hållas om så anses vara nödvändigt efter godkännande av senare chefsstatistikermöten. Dessutom betonades att projektverksamheten bör minska i omfattning.

Chefsstatistikermötet behandlade även de överlappningsproblem som kan uppstå i utskott vars verksamhetsområde är av övergripande karaktär. Det beslöts att den nuvarande utskottsstrukturen inte skall förändras för de existerande utskotten, men att de synpunkter som framförts skulle beaktas vid mötets genomgång av utskottens programförslag.

Beträffande de besparingsförslag som framförts av utskotten ansågs de flesta vara av sådan karaktär att de kan prövas.

I frågan om upprättandet av permanenta utskott för finans- resp. kreditmarknadsstatistik fattades inte något beslut vid detta möte utan frågan sköts på framtiden.

Utbytet av information per korrespondens mellan utskottsmedlemmarna har inte alltid fungerat tillfredsställande och det uppdrogs åt Nordiska statistiska sekretariatet att tillse att nödvändig information sprids till utskottsmedlemmarna och andra utskott.

Chefsstatistikerna beslöt att utskottens rapporteringsskyldighet till de kortare chefsstatistikermötena bör begränsas. Framförallt gäller detta utskott som inte hållit möte under det gångna verksamhetsåret eller inte avser föreslå ändringar i det arbetsprogram som fastslagits för treårsperioden.

9. Nordiska statistiska utskott i chefsstatistikermötets regi

Chefsstastikermötet behandlade de nordiska statistiska utskottens verksamhetsrapporter för 1979–1982 och förslag till samarbetsprogram för 1982–1985. Det beslöts att två nya utskott skulle upprättas; ett för inkomst- och förmögenhetsstatistik och ett annat för databas- och andra databehandlingsfrågor. Vidare fastlades att det Nordiska utskottet för ett system för socio-demografisk statistik (NUSD) skall byta namn till Nordiska utskottet för socio-demografisk statistik (NUSD).

Huvudpunkterna i de enskilda utskottens arbetsprogram är följande:

9.1. Nationalräkenskaper (NUNA)

1. Allmän orientering om utvecklingen på nationalräkenskapsområdet.
2. Institutionella sektorer (intäkts- och kapitalräkenskaper).
3. Byggnads- och anläggningsbranschen.
4. Varuhandel och avanser (särskild orientering).
5. Frågor i förbindelse med revisionen av SNA.
6. Kapitalstorleksberäkningar.
7. Kvartalsvisa nationalräkenskaper.
8. Sysselsättning och lön (särskild orientering).

9.2. Utrikeshandelsstatistik och varunomenklaturfrågor (NUVA)

1. Fortsatt behandling av olika frågor inom utrikeshandelsstatistikens område såsom dess omfattning och avgränsning, tidslokalisering och internationella jämförbarhet. Speciellt bör de områden genomgås där de nordiska länderna för närvarande ej synes följa FNs nya rekommendationer från januari 1981.
2. Informationsutväxling, särskilt orientering från dansk sida om pågående arbeten i den Europeiska Gemenskapen (EG).
3. Utarbetande av ett förslag till utformningen av en dokumentation av eventuella gemensamma nordiska varuuppdelningar under varupostnivån i HS (dvs. under den sex-ställiga nivån).
4. Granskning av FNs förslag till reviderad SITC (SITC, rev 3). Utarbetande av svar på eventuell remiss från FN av förslaget.
5. Utarbetande av nycklar mellan den nu

gällande nordiska varuförteckningen och HS.
6. Utredning av vilka arbetsuppgifter med HS som skall åvila NUVA sedan HS börjat tillämpas, t.ex. frågor angående nya varuuppdelningar i den internationellt fastställda nomenklaturen eller på gemensam nordisk nivå.

9.3. Jordbruksstatistik (NUJO)
1. Trädgårdsstatistik.
2. Statistik rörande företag och brukare inom deltidsjordbruket.
3. Skogsstatistik.
4. Publikationsstruktur.
5. Statistikens kvalitet.
6. Frågor i anknytning till World Agricultural Census 1990.
7. Övriga frågor, härunder statistik över hästar, pälsdjursavel m.m.

9.4. Näringsstatistik (NUN)
1. Nyckeln mellan de nordiska bransch- och varunomenklaturerna (avslutning och uppföljning av arbetet).
2. Enhetsproblematiken i näringsstatistiken (bl.a. hjälpverksamhet inom koncern/företag/arbetsställe).
3. Uppföljning av utvecklingen av nordiska databaser (innehållsmässiga aspekter).
4. Fördjupat informationsutbyte beträffande:
   a) Byggnads- och bostadsstatistik.
   b) Energistatistik.
   c) Konjunkturstatistik.
   d) Statistik som belyser näringslivsstrukturen.
   e) Administrativt material i näringsstatistiken.

9.5. Samfärdselstatistik (NUSA)
1. Vägtrafikolycksstatistik.
2. Bilregisterfrågor.
3. Ekonomisk statistik.
4. Sjöfartsstatistik (in- och utrikes).
5. Översikt över producenter av statistik inom samfärdsel och turism.
6. Information om nationella program inom samfärdselstatistikens område.

9.6. Socio-demografisk statistik (NUSD)
1. Standarder för definitioner och klassificeringar.
   a) Familje- och hushållsbegreppen.
   b) Socio-ekonomiska grupperingar.
   c) Åldersklassificeringar.
   d) Uppföljning av frågor som NUSD begärt att andra utskott skall behandla.
2. Kombinerat utnyttjande av socio-demografiska data.
   a) Sociala relationer.
   b) Offentlig verksamhet, förbindelser mellan SNA och SSDS (med utbildning som praktiskt exempel).
   c) Utveckling av en databasorienterad modell för belysning av fördelningen av inkomster och transfereringar samt simulering av fördelningseffekter vid förändringar i transfereringssystemet.
   d) Bortfallsproblem i ett integrerat personstatistiskt system.
   e) Sambandet mellan hushållsbudgetundersökningar och nationalräkenskaper.
   f) Statistical matching.
3. Utvecklingen av den statistiska sociala rapporteringen.

Utskottet kommer att arrangera ett seminarium gemensamt med NUUL under 1984.

9.7. Befolkningsstatistik (NUB)
1. Den löpande befolkningsstatistiken (aktuella komponenter, t.ex. migration, samboende utan vigsel, fruktsamhet).
2. Yrke-dödlighetsundersökningar.
3. Longitudinella undersökningar på basis av FoB 1960–1980.
4. Uppföljning av FoB 1980.

En särskild expertgrupp för behandling av punkt 2 kommer att sammanträda i anslutning till utskottets första möte i perioden.

9.8. Arbetsmarknadsstatistik (NUA)
1. Uppföljning av rapporten från NAUTs seminarium om arbetsmarknadsstatistik i november 1980.
2. Behandling och uppföljning av projektet "Registerbaserad sysselsättningsstatistik i Norden".

3. Revision av Nordisk yrkesklassificering (NYK) samt ILOs förslag till revision av ISCO.
4. Arbetslöshetsstatistik.
5. Arbetstidsfrågor.
6. Arbetsförhållanden och arbetsmiljö.
7. Internationell arbetsmarknadsstatistik.

9.9. Utbildnings- och kulturstatistik (NUUK)

Utbildningsstatistik

1. Befolkningens utbildningsnivå.
2. Lärarstatistik.
3. Registrering av användarbehov och behovs- och användaranpassning av utbildningsstatistiken inkl. problem vid användningen av administrativa register.
4. Longitudinella studier.
5. Utbildningsindikatorer och sociala indikatorer på utbildningsområdet.
6. Vuxenutbildning och specialundervisning.

Kulturstatistik

7. Individorienterad kulturstatistik.
8. Nationella kulturstatistiska publikationer.
9. Offentliga och privata utgifter till kulturändamål.
10. Andra kulturområden.
    a) Bibliotheksstatistik.
    b) Kulturindustri.
    c) Internationellt arbete med kulturstatistik.

9.10. Kriminalstatistik (NUK)

1. Nordisk jämförande brottsstatistik.
2. Standard för klassificering av brott och straffrättsliga reaktioner.
3. Offerstudier (dold brottslighet).
4. Brottsskadestatistik.
5. Återfallsstatistik.

Chefsstatistikermötet upplöste huvudutskottet bestående av representanter för justitiedepartementen, kriminologisk forskning och de statistiska centralbyråerna. Utskottet består i fortsättningen enbart av representanter för de statistiska centralbyråerna.

9.11. Socialstatistik (NUS)

1. En genomgång av den redovisning som förekommer i socialstatistiken och jämförelse med motsvarande redovisning i nationalräkenskaperna.
2. Systematisk genomgång i de nordiska länderna av den redovisning som förekommer i NOSOSKOs publikation för dokumentation av källor för olika uppgifter samt av principer för aggregering och för klassificering av förekommande poster.
3. En genomgång av redovisningen som sker till NOSOSKOs publikation med beaktande av SNAs grupperingsprinciper och särskilt då SNA grupp 5.
4. Följa utvecklingen inom området för sociala indikatorer, som faller inom utskottets verksamhetsfält.
5. Utbyte av erfarenheter ifråga om utvecklingen av socialstatistik.
6. Ytterligare frågor som kan aktualiseras i samband med uppgiftsinsamlingen till och presentationen av nordisk socialstatistik i NOSOSKOs publikationer.

9.12. Miljöstatistik (NUM)

1. Seminarium om vattenstatistik under våren 1983 (arrangerat gemensamt av NUM och Arbetsgruppen för miljödata).
2. Geografiska områdesbegrepp.
3. Avfallsklassificering.
4. Information om utvecklingen av miljöstatistik i de nordiska länderna inklusive uppföljning av arealklassificeringsprojektet.
5. Naturresursräkenskaper.
6. Erfarenheterna av tidsserieredovisning.
7. Utnyttjandet av industrins näringsgrensindelning i miljöstatistiken.
8. Eventuell uppföljning av seminariet om vattenstatistik.

9.13. Prognoser (NUP)

1. Prognosseminarum kommer att hållas under år 1984.

9.14. Undersökningar om levnadsvillkor (NUUL)

1. Rapportutkast från expertgrupp 1 – Samnordisk rapport om levnadsförhållanden.

2. Orientering om pågående metodologiskt arbete gällande förändringsskattningar, datakvalitet och felmodeller i samband med levnadsnivåundersökningar.
3. NUSDs önskemål om att NUUL behandlar socio-ekonomisk indelning och hälsoindikatorer.
4. Information om levnadsnivåundersökningar, tidsanvändningsstudier och socialstatistiska översiktspublikationer.
5. Inventering och sammanställning av aktuella välfärdsfrågor i länderna som kan behandlas inom levnadsnivåundersökningarna och utgöra underlag för framtida nationella publikationer.
6. Föreslagna åtgärder från expertgrupp 1 för att åstadkomma ökad jämförbarhet och samordning mellan levnadsnivåundersökningarna.
7. Metodologiska frågor vad gäller förändringsskattningar, datakvalitet och felmodeller i samband med levnadsnivåundersökningar.
8. Ett gemensamt seminarium med NUSD år 1984.

9.15. Kontaktutskottet för statistiska metodproblem i statistikproduktionen (KOMET)

1. Metodproblem i miljöstatistiken.
2. Utvalda metodproblem vid registerbaserade räkningar.
3. Utnyttjandet av data från register och urval.
4. Estimering för små områden.
5. Statistical matching.
6. Metoder för analys och presentation av upprepade undersökningar.
7. Uppföljning av arbetet om "kvalitetsredovisning".
8. Information om metodstatistiskt arbete i de nordiska länderna.

Utskottets verksamhet kommer att koncentreras till frågor som gäller matematisk-statistiska metoder i statistikproduktionen.

9.16. Kontaktmöten

Chefsstatistikermötet godkände att följande kontaktmöten får arrangeras under verksamhetsperioden:

1. Kontaktmöte om biblioteksverksamheten vid de statistiska centralbyråerna.
2. Kontaktmöte om informationsverksamheten vid de statistiska centralbyråerna.
3. Kontaktmöte om utbildningsverksamheten vid de statistiska centralbyråerna.

10. Projektverksamhet

Chefsstatistikermötet behandlade rapporter om följande projekt:

1. LIST STAT – samnordisk databas för statistikmaterial. Projektet finansieras av NORDINFO.
2. Registerbaserad sysselsättningsstatistik i Norden. Projektet finansieras av Nordiska ministerrådet. Medlemmar av NUA ingår i projektets styrgrupp.
3. Samnordisk rapport om levnadsförhållanden. Projektet utföres av NUULs expertgrupp 1.
4. Förstudie avseende kultur- och mediestatistik i Norden. Projektet avslutat 1981.
5. Revision av Nordisk yrkesklassificering (NYK).
6. Nordisk arealklassificering. Projektet genomfördes i regi av NUM och rapport har publicerats som nr. 29 i serien Tekniske rapporter.

11. Annat statistiskt samarbete i Norden

För denna punkt på dagordningen förelåg rapporter över verksamheten 1979–1982 i

- Nordiska socialstatistikkommittén (NOSOSKO).
- Nordiska medicinalstatistikkommittén (NOMESKO).
- Nordiska arbetsmarknadsutskottets (NAUT) utskott för arbetsmarknadsforskning.
- Arbetsgruppen för miljödata (AMD).
- Nordiska ämbetsmannakommittén för regionalpolitik (NÄRP)/Basprojektet.
- Nordiska samarbetsdelegationen för samhällsforskning (NOS-S)/Nordisk databas för regionala tidsserier (NDRT).
- Nordiska samarbetsorganisationen för teknisk-vetenskaplig forskning (NORDFORSK).

12. Sekretariatets publikationsverksamhet

Följande publikationer har utgivits av Nor-

diska statistiska sekretariatet under 1981/82:

12.1. Nordisk utredningsserie
Nordisk statistisk årsbok 1981, NU 1981: 15, Stockholm 1982.

12.2. Nordisk statistisk skriftserie
Nr. 38. Social security in the Nordic countries 1978. Stockholm 1981 (engelsk utgåva).

Ovanstående publikation har utarbetats av Nordiska socialstatistikkommittén (NOSOSKO) och finansierats av de nordiska ländernas socialministerier.

12.3. Tekniske rapporter
Nr. 28. Nordiska seminariet om kriminalstatistik, Utö Kursgård, Sverige, 14–17 september 1981. Köpenhamn 1982.

12.4. Nyt fra Nordisk statistisk sekretariat
Ett nummer har utkommit i juni 1982.

13. Nordiskt deltagande i internationella möten

Chefsstatistikermötet behandlade en översikt över de statistiska centralbyråernas deltagande i ECEs och OECDs möten under den gångna treårsperioden. Rigsstatistiker N. V. Skak-Nielsen, Danmarks Statistik, redogjorde för det statistiska samarbetet inom EG och från finländsk sida redogjordes för det statistiska samarbetet mellan Rådet för ömsesidigt ekonomiskt bistånd (SEV) och Finland.

14. Avtal mellan statistikbiblioteken

Chefsstatistikerna godkände en överenskommelse mellan statistikbiblioteken i Danmark, Finland, Norge och Sverige om viss ansvarsfördelning vid anskaffning och förvaring av statistiskt material.

15. Nästa större chefsstatistikermöte

Det bestämdes att det 31a nordiska chefsstatistikermötet skall hållas i Sverige år 1985.

Lars Finell *Harry de Shàrengrad*

IIASAs seminarium om informationssystem för integrerad regional utveckling

Den 15–17 december 1982 anordnade International Institute for Applied Systems Analyse (IIASA) ett seminarium om informationssystem för integrerad regional utveckling. Från Sverige deltog representanter från Stockholms läns landsting, Stockholms kommun och undertecknad. I övrigt kom deltagarna från IIASAs medlemsländer. Under seminariet diskuterades möjligheter att nå en bättre anpassning av skilda informationssystem till planerarnas behov. Speciellt diskuterades problemen med att många informationssystem är sektorsanknutna och att det är svårt att skapa helhetsbilder.

Som underlag för seminariet förelåg rapporter från ett urval av medlemsländerna. Den svenska hade författats av Ola Nygren från statistikutredningen och undertecknad. Vid presentationen av rapporterna lades speciell vikt vid problemen med att göra prognoser och beskrivningar av konsekvenser av olika politiska åtgärder. En fråga som restes var hur systemanalyser skall utvecklas för att ge större flexibilitet i prognosarbetet. Många planerare och forskare önskar tillgång till mikrodatabaser med vars hjälp de själva kan göra bearbetningar och prognoser. Inga länder har färdiga lösningar på detta, även om man i många länder nu bygger upp databaser för statistiskt bruk.

Stor uppmärksamhet ägnades åt det regionalstatistiska system som presenterades av den amerikanske delegaten. Han pekade på problemen med att de regionala statistiska serierna inte är koordinerade och menade att det borde vara möjligt att arbeta fram någon form av motsvarighet till nationalräkenskaperna. Likartade tankar fördes fram av andra deltagare, som menade att avsaknaden av en systemansats var besvärande.

De olika inläggen kommer att presenteras i en bok från IIASA.

Gösta Guteland

Statistical Computing Project

Statistical Computing Project är ett projekt som startats med stöd av UNDP under ECE. Inom projektet finns fyra arbetsgrupper.[1]

Projektets styrkommitté och tre av de fyra undergrupperna höll möten i Aten under veckan 15–19 november 1982, styrkommitténs möte den 15–17 november, och de övriga – RAPID Joint Group, Data Editing Joint Group och METIS Joint Group – den 18–19 november. METIS-gruppens möte bevakades ej av Sverige.

1. Styrkommittén

Statistical Computing Projects styrkommitté höll sitt möte den 15–17 november. Ordförande för mötet var Christer Arvas och Sveriges representant var Märit Leringe bägge SCB. Utöver Sverige var 13 europeiska länder representerade. Frankrike deltog som observatör.

Projektet kommer att pågå t.o.m. 1984, vilket innebär en förlängning med ett år. Information om detta gavs av projektets samordnare, som samtidigt underströk vikten av att utvecklingsländer kunde tillgodogöra sig de resultat som kommer ut av det.

De fyra undergrupperna rapporterade om det arbete som utförts under året och redogjorde för sina planer för 1983.

För "Table Generation Group" aviserade Danmark att beslut fattats om utveckling av INTERTAB, ett nytt programpaket för tabellproduktion.

"Data Editing Group" med Ungern som ledande land rapporterade om arbetet med granskningsprogrammet AERO och med specifikationen av GSPS, "General Survey Processing System". Gruppen föreslogs att hitta en organisatorisk lösning för att på bästa sätt kunna fortsätta inom båda dessa huvudområden.

För METIS-gruppen, som ägnar sig åt metadatafrågor, rapporterade Tjeckoslovakien om årets aktiviteter. Flera deltagare ansåg att gruppen inte skulle ägna sig åt programutveckling utan fastmer åt att specificera de funktioner som krävs för hantering av metadata. Resultatet av arbetet kommer att presenteras i en handbok.

Sverige gav en resumé av RAPID-gruppens aktiviteter. Bl.a. informerades om det programutvecklingsarbete som påbörjats av en liten arbetsgrupp med deltagare från Ungern, DDR och Sverige. Styrkommittén påpekade att någon jämförelse mellan olika databassystem för statistikbearbetning bör göras.

För nästa period valdes mr T. Walzak från Polen som ny ordförande. Kommittén uttryckte sin stora uppskattning för Christer Arvas' insats som ordförande under projektets två första år.

2. RAPID-möte

I mötet som leddes av Märit Leringe deltog representanter för Ungern, DDR, Italien och Bulgarien. Större delen av första dagen ägnades åt att ge närmare information om det arbete som pågick till länder utanför gruppen. Applikationer från DDR, Ungern och Sverige som utnyttjar RAPID beskrevs och "RAPID Guideline" delades ut. Flera länder visade mycket stort intresse för det programmeringssamarbete som inletts. En beskrivning av arbetsmetoderna och utvärdering av resultatet såg många fram emot, inte minst de som berördes av tabellgruppens satsning.

Italien presenterade sina planer för användningen av RAPID. Från och med 1983 är de intresserade av att deltaga i arbetet med gemensam programutveckling.

DDR rapporterade om läget med RAPID på EC-datorer. Då Bulgarien planerar att installera RAPID på EC höll sig diskussionerna länge kring Bulgariens speciella problem.

Till sist planerades det kommande årets möten inkl. de planerings- och testsessioner som krävs för den gemensamma programutvecklingen.

3. Data Editing Joint Group

I mötet som leddes av mrs Katalin Szasz deltog representanter från 9 länder samt mr Ernö Stauder, SCPs samordnare. Från Sve-

[1] Se Statistisk tidskrift 1982: 5, sid. 387.

rige deltog Christer Arvas och Leopold Granquist, SCB.

Gruppens arbetssätt och nuvarande inriktning redovisas i rapporten från det föregående mötet i Madrid, 14–16 april 1982.[1] En arbetsuppgift som tillkommer är att ett av projektets slutresultat skall bli ett slags handbok i "editing methodology", vilken blir en redigerad sammanställning av de uppsatser som presenteras vid mötena under projektets livstid.

Mötet i Atén hade följande ämnen:

Datakvalitetsproblem; rapporter om programmet DIA och den interaktiva versionen av programmet AERO; generellt system för statistikproduktion; jämförande utvärdering av generella granskningsprogram; interaktiv granskning; det fortsatta arbetet inkl. handlingsplan fram till nästa möte.

Sverige presenterade en uppsats om kvalitet vid maskinell granskning byggd på erfarenheter från SCB och en evalvering av användningen av CANEDIT vid granskningen av 1976 års folk- och bostadsräkning i Kanada. Enligt uppsatsen är dagens generella granskningsprogram alltför inriktat på att få in så mycket kontrollrutiner som möjligt i stället för att fokuseras på att upptäcka förekommande systematiska fel. Detta diskuterades livligt och det beslöts att gruppen ska fortsätta med detta ämne med Sverige som ansvarig samordnare av bidrag från deltagarländerna.

Spanien gav en rapport över sitt arbete med att utveckla DIA till att omfatta deterministiska imputeringar avseende systematiska fel inom ramen för den s.k. Fellegi-Holtmetodologin. I Jugoslavien utvecklar man det ungerska programmet AERO till ett interaktivt program. Erfarenheterna rapporterades.

England samordnar och ansvarar för ämnet "Generalized Survey Processing System" (GSPS). En uppsats redovisades om innehåll och funktioner i ett sådant system. Det diskuterades hur man ska gå vidare. Diskussionerna utmynnade i beslut om att arbeta efter två ansatser "top down" resp. "bottom up". I top down-ansatsen ska ett användar "interface" definieras (Frankrike ansvarigt) och i bottom up-ansatsen fortsätter England sitt arbete med checklista och studier av existerande programvaror.

Ungern presenterade ett förslag på vad en dokumentation av ett generellt granskningsprogram bör innehålla för att tillåta jämförande studier av granskningsprogram. Programmet AERO redovisades i enlighet med förslagets riktlinjer.

En uppsats om interaktiv granskning lades fram av Ungern. Det beslöts att papperet skulle vara projektets slutdokument i ämnet.

Slutligen diskuterades bl.a. spridningen av AERO och programpunkterna för nästa möte som kommer att äga rum i Titchfield 5–8 april 1983.

Märit Leringe *Christer Arvas*
Leopold Granquist

[1] Statistisk tidskrift 1982: 5 sid. 386–387.

SCBs yttrande över rapport 2 från UHÄs datareferensgrupp "Förstärkning av den administrativt inriktade utbildningen inom dataområdet"

ADB-tekniken är sedan lång tid det dominerande tekniska hjälpmedlet i SCBs verksamhet, där framställning av statistisk information är den centrala uppgiften. Den önskvärda kompetensprofilen på SCBs handläggande personal kan allmänt sett sägas vara goda kunskaper i statistik, ADB och något eller några olika ämnesområden, t.ex. ekonomi eller sociologi. SCB har dessutom stort behov av kvalificerade experter på olika områden och då inte minst inom ADB-området.

Under senare år har SCB nästan enbart anställt högskoleutbildade personer på tjänster där ADB-innehållet dominerat (programmerare och systemerare). Nyanställda programmerare får genomgå en ca 10 veckor lång SCB-anpassad utbildning i ADB, som i sig alltid torde vara nödvändig men också anpassad till den grundutbildning som ges vid högskolan. Inom SCB ges sedan en fortlöpande vidareutbildning.

Vid behov av nyrekrytering av personer med krav på kvalificerade ADB-kunskaper är de som genomgått i första hand den systemvetenskapliga linjen av stort intresse för SCB.

En viktig del i allt samhälleligt förändringsarbete bör vara en analys av tillgängligt *statistiskt informationsunderlag*. Ett problem när det gäller överväganden om högskoleutbildningen på ADB-området är den i dag i viss utsträckning ofullständiga högskolestatistiken. SCB har t.ex. resultatstal på linjenivå men ej med uppdelning på kurser och ämnen. Till en del förklaras bristerna av att ADB-området är ganska nytt och mycket snabbt föränderligt. Det har därför inte hunnit byggas upp något omfattande statistiskt bakgrundsmaterial, som kan ligga till grund för beslut om förändringar. SCB vill ändå här peka på betydelsen av ett bra statistiskt underlag för uppföljningen av förändringar av olika högskoleutbildningars innehåll. UHÄ och högskolan måste beakta detta i sin verksamhet.

På uppdrag av datadelegationen tog SCB för ungefär ett år sedan fram en rapport (En beskrivning av ADB-personalen, 1982-01-31), som visade på läget för *ADB-utbildade personer* kring 1980. I denna rapport konstateras bl.a. att det då fanns ca 17 000 yrkesverksamma programmerare/systemmän. Denna yrkesgrupp kan sägas karakteriseras av att jämfört med andra yrkesverksamma vara yngre och mer välutbildad och dessutom starkt koncentrerad till Stockholms-området (hälften av de yrkesverksamma).

Prognosinstitutet vid SCB genomför regelbundet en s.k. arbetskraftsbarometer. I denna urvalsundersökning ingår sedan lång tid utbildningsgruppen programmerare/systemerare. Ett antal arbetsgivare tillfrågas om tillgången på sökande till lediga tjänster, rekryteringsläget och förändringen av antalet anställda de närmaste tre åren. I årets undersökning, som nyss färdigställts, kan man bl.a. konstatera att arbetsgivarna uppger brist på sökande med praktik men balans vad avser nyexaminerade. Antalet programmerare som efterfrågades motsvarar 12 % av dem som nu har anställning. Antal anställda programmerare/systemmän förväntades öka med ca 7 % per år de närmaste tre åren.

SCB uppskattar att nyrekryteringsbehovet för närvarande är i storleksordning 1 500 per år från de utbildningsvägar på ADB-området som nu finns. Mot bakgrund av vad som har framkommit i SCBs undersökning, ligger det vidare nära till att anta att bristen skulle kunna täckas på ett bättre sätt, om det fanns fler att rekrytera med systemvetenskaplig utbildning. Det finns inte heller någon anledning att tro att detta nyrekryteringsbehov på sikt kommer att vara mindre än nu.

SCB anser att *datareferensgruppens slutsatser* i stort verkar rimliga. Detta dock sagt med hänsyn till att vi inte tagit del av gruppens första rapport och att utredningsmaterialet i vissa avseenden enligt vår bedömning inte är tillräckligt genomarbetat. Speciellt vill SCB ge sin syn på följande punkter:

– SCB kan hålla med om att systemvetenskapliga linjen för närvarande inte bör byggas ut från tre till fyra år. På sikt bör dock en sådan utbyggnad kunna ske (se också nedan). Redan nu bör man dock pröva möjligheterna till översyn av linjeinnehållets omfattning och sammansättning.

– SCB anser att en ökad antagning till linjen bör prioriteras högt (jfr ovan om utbildningsprognoser). Referensgruppens föreslagna utbyggnadstakt verkar dock väl hög med hänsyn till bl.a. den begränsade tillgången på kvalificerade lärare. Hur detta i praktiken skall lösas framgår inte av datareferensgruppens skrivningar. En förstärkning på utrustningssidan är säkert önskvärd men kan på inget sätt ersätta den nödvändiga tillgången på kvalificerade lärare.

– SCB delar helt gruppens uppfattning att systemvetenskapliga linjen bör utformas utifrån en samhällsvetenskaplig grundsyn. Vad detta i praktiken betyder framgår dock inte av föreliggande rapport. Hur kommer det faktiska linjeinnehållet att se ut? Enligt vår bedömning kommer den samhällsvetenskapliga grundsynen att bli svår att klara inom en treårig utbildningslinje. En successiv utbyggnad bör därför redan nu förberedas. Den systemvetenskapliga linjen bör

enligt SCB ge de grundläggande ADB-tekniska kunskaperna. Linjen bör dock också innehålla utbildningsinslag som belyser datoriseringens samhällskonsekvenser.

– SCB vill speciellt betona betydelsen av nödvändiga och tillräckliga inslag av statistikutbildning i systemvetenskapliga linjen. Det är också av betydelse att linjen ger möjligheter till fortsatta kvalificerade statistikstudier. Det finns t. ex. ett behov av utbildning för samhällsekonomiskt utredningsarbete. Viss samordning med utbildningslinjerna inom AES-sektorn kan vara önskvärd. SCB instämmer här i arbetsgruppens skrivningar i avsnitt 5.2.6.

– SCB anser inte att ett system med etappavgångar bör finnas på en treårig utbildningslinje.

– SCB anser liksom referensgruppen att satsningar på vidareutbildning och forskning är av stor betydelse. Detta gäller inte minst områden som berör datoriseringens samhällskonsekvenser i vid bemärkelse. Det finns dessutom en stor grupp yrkesverksamma utan längre teoretisk utbildning på ADB-området.

SCBs yttrande över länsdemokratikommitténs betänkande "Vidgad länsdemokrati" (SOU 1982: 24)

SCB saknar anledning att medverka i den grundläggande principdiskussionen om vidgad länsdemokrati på annat sätt än att här framhålla, att statistisk information om samhällets utveckling och tillstånd utgör en viktig förutsättning för att säkra det demokratiska inflytandet. SCB begränsar således sitt yttrande till de delar av betänkandet som berör samhällets informationsförsörjning. Länsdemokratikommitténs förslag i betänkandet kan sammanfattas enligt följande:

- Förstärkt samordning på regional nivå – länsstyrelsens möjligheter att samordna de statliga sektorverksamheterna i länet förstärks.
- Vidgad länsdemokrati genom landstingen – landstingens inflytande över vissa länsorgan stärks.
- Ändringar i administrativa förfaranden – i huvudsak förändringar i arbetsformerna för länsstyrelsens förtroendemannastyrelse.
- Årsredogörelse för utvecklingen m. m. i länet – länsstyrelsen utarbetar en årsredogörelse över statlig och annan verksamhet i länet som överlämnas till regeringen, länets kommuner och landstinget.

I anslutning till den pågående decentraliseringsprocessen betonas, bl. a. i detta betänkande, vikten av samordning. Det innebär att det ställs stora och nya krav på samspelet mellan länsmyndigheter, landsting och primärkommuner inom ramen för vad som generellt kallas samhällsplanering. *På den kommunala nivån* har kraven på samspel mellan de kommunala förvaltningarna vuxit sig allt starkare och bl. a. resulterat i behov att samordna den statistiska belysningen av kommunen och dess delar. Mer eller mindre formaliserat har detta resulterat i grundsynen, att så kallade *gemensamma planeringsförutsättningar* bör ligga till grund för planeringen.

SCB menar att kommitténs förslag om förstärkt samordning på regional nivå ställer krav på ett mer enhetligt underlag också för olika *regionala organs* verksamhet. Vidare måste planeringen *på regional och kommunal nivå* bygga på i vissa avseenden gemensamma förutsättningar. SCB har intrycket av att denna grundsyn delas av allt fler aktörer i samhällsplaneringen i takt med att kraven på samordning av olika samhällsinsatser blir större. Här bör också beaktas att i de kommande årens kärva ekonomiska klimat kan behovet av uppföljning av olika åtgärder förväntas öka. Denna *uppföljningsverksamhet* underlättas om informationsunderlaget möjliggör exempelvis jämförelser mellan olika åtgärders effekter i olika regioner.

Under den senaste tioårsperioden har teknisk och annan utveckling medverkat till att olika former av registerhantering har öppnat nya möjligheter till decentraliserad

produktion av statistik och information över huvud taget. Möjligheterna att på ett relativt enkelt och billigt sätt ta fram information har således ökat. Samtidigt ökar dock *risken för* att den allt större mängden information skapar ett *informationskaos*.

Ett mer enhetligt informationsunderlag kan uppnås genom att en viss beskrivning av samhället på regional nivå ges en *särskild auktorisation*. SCB har goda möjligheter att svara för en stor del av informationsunderlaget för denna beskrivning genom det väl fungerande löpande samarbetet mellan representanter för regionala och lokala organ och SCB. Vidare möjliggör *SCBs ökade satsning på olika statistiska databaser* en ökad och mer behovsanpassad spridning av statistik direkt till användarna via terminal. Redan idag distribueras en stor del av SCBs statistik på detta sätt till regionala och lokala intressenter. Här bör beaktas att statsmakternas ökade krav på avgiftsfinansiering av statlig verksamhet bl. a. har medfört att ett stort antal statistikanvändares kostnader för den statistiska information som SCB tillhandahåller har ökat. Erforderliga *ekonomiska resurser* behöver därför ställas till de regionala organens förfogande för att möjliggöra framtagandet av för resp. län relevant information.

SCB har som central förvaltningsmyndighet för den statliga statistikproduktionen ett särskilt ansvar för att utvecklingen inte leder till att onödigt dubbelarbete, bristande samordning och oekonomiska lösningar i samhällets informationsförsörjning. SCB kommer att i samarbete med berörda intressenter fortsätta arbetet för att undvika en sådan situation.

SCB vill således även i anslutning till länsdemokratikommitténs slutbetänkande peka på vikten av att informationsförsörjningen fungerar på ett utifrån demokratiska aspekter tillfredsställande sätt på alla regionala nivåer och att därvid jämförbarhets- och samordningsaspekterna beaktas. Verket vill slutligen framhålla betydelsen av att sambandet mellan inflytande och information fungerar och att informationsförsörjningen är en viktig del i en decentraliserings- och demokratiseringsprocess.

SCBs yttrande över betänkandet "Information om arbetsmiljörisker (SOU 1982: 30)

1. Inledande synpunkter

SCB delar helt utredningens uppfattning vad gäller behovet av en samlad översiktlig information om arbetsmiljön. Utredningen framhåller flera fördelar med en sådan information av statistisk art. Dessa har under en längre tid varit avgörande för SCBs utvecklingsarbete rörande en officiell arbetsmiljöstatistik.

SCB har i likhet med utredningen funnit att bristen på en samlad arbetsmiljöinformation blivit alltmer besvärande. SCB har därför påbörjat ett utvecklingsarbete inom detta område. Det är vidare positivt att utredningen föreslår att medel skall tilldelas SCB för fortsatt arbete med att ta fram officiell statistik som belyser förhållandena i arbetslivet. SCB har i sitt utvecklingsarbete inom detta område bl. a. utprovat olika metoder för att inhämta information om arbetsmiljön och hoppas inom kort kunna lägga fram vissa förslag till en löpande datainsamling.

Vidare framhåller utredningen behovet av samordnade insatser för att säkra möjligheterna att sammanställa och erhålla data ur olika register med information om samband mellan arbete och hälsa. Lika och entydiga definitioner och indelningar bör därvid eftersträvas i de skilda registren. Även SCB har som statistikproducerande myndighet erfarenheter av behovet av statistisk samordning för information som skall användas inom ett samhällsområde. SCB delar därför utredningens bedömningar att denna fråga särskilt skall uppmärksammas. I enlighet med utredningens förslag bör arbetarskyddsstyrelsen i samråd med bl. a. SCB allmänt verka för en samordning av den registrerade arbetsmiljöinformationen.

2. SCBs utvecklingsarbete på arbetsmiljöområdet

SCB har sedan några år tillbaka bedrivit ett arbete med att utreda behov av och möjligheter att utveckla en officiell statistik om

arbetsmiljön. I detta arbete har ingått ett metodologiskt utvecklingsarbete, som syftar till att ta fram metoder för statistisk datainsamling inom arbetsmiljöområdet. De metodologiska svårigheter som utredningen nämner utgör också enligt SCBs bedömningar besvärliga problem som man på ett eller annat sätt måste komma förbi för att med tillfredsställande kvalitet kunna redovisa beskrivande arbetsmiljöstatistik.

I SCBs metodarbete har bl. a. studerats möjligheten att i statistiken utnyttja den information om arbetsmiljön som kan hämtas från enskilda arbetstagare, via intervju- och enkätundersökningar. Sådana undersökningar riktade till olika befolkningsgrupper används idag i statistikarbetet i många länder. Kvalitetsproblemen är dock många och de är världen över föremål för ett omfattande utredningsarbete.

För att studera kvaliteten och tillförlitligheten när det gäller arbetsmiljöinformation har SCB mycket detaljerat kartlagt arbetsförhållandena för ett urval arbetstagare. Detta har gjorts med hjälp av tekniska arbetsmiljömätningar, expertutfrågningar och observationer. Till de personer vars förhållanden kartlagts, har sedan ställts frågor om exakt samma förhållanden. Olika formulär och utfrågningsmetoder har använts och tillförlitligheten har prövats när det gäller en rad områden, från beskrivningar av konkreta fysiska förhållanden till frågor om psykosociala förhållanden. De hittills framkomna resultaten pekar på att det i princip är möjligt att via intervju- och enkätfrågor få tillförlitliga resultat. Metoden tycks dock i första hand fungera bra inom områden om vilka arbetstagarna har direkta egna erfarenheter. Frågeställningarna måste vidare vara enkla och klara och efterfråga *i första hand beskrivningar och inte bedömningar.*[1]

Resultaten från dessa metodstudier kommer att avrapporteras under innevarande budgetår. Samtidigt kommer vissa förslag till löpande datainsamling att framläggas.

[1] Vissa preliminära resultat finns redovisade i en bilaga till yttrandet, vilken ej publiceras här.

Resultaten och erfarenheterna som framkommit i ovan beskrivna metodarbete bör enligt SCBs mening ligga till grund dels för viss ny datainsamling, dels för närmare utredning rörande metoder för databearbetning och redovisning av bearbetningsresultat.

Utredningen föreslår att SCB bl. a. skall få i uppdrag att utreda behovet av metodutveckling för uppskattning av risker i arbetsmiljön och för att kontinuerligt följa arbetsmiljöns "riskprofil" samt att lägga fram förslag om hur behovet skall tillgodoses. SCB ser positivt på ett sådant uppdrag och är beredd att genomföra arbetet vilket bör kunna bedrivas i samarbete med olika intressenter och parter inom arbetsmiljöområdet. Enligt utredningen skulle arbetarskyddsfonden kunna bidra med medel till att finansiera det föreslagna metodutvecklingsarbetet.

3. Arbetsmiljöinformation i ett vidare sammanhang

Det bör uppmärksammas att arbetsmiljöinformation i vissa fall måste samordnas med annan information för att ge en fullständig och korrekt bild av verkligheten. Exempelvis finns samverkande negativa hälsoeffekter av svaveldioxid och fasta partiklar i luft. Exponering för den ena föroreningen kan ske i arbetsmiljö och för den andra i yttre miljö. Utredningen har såvitt SCB kunnat finna ägnat mycket ringa uppmärksamhet åt problem av detta slag. För SCB, som bl. a. skall verka för samordning mellan statlig och annan statistikproduktion, förefaller detta vara en påfallande brist. I detta remissvar kan endast en aspekt på denna fråga behandlas, nämligen behovet av samordning av information i statistisk form om arbetsmiljön och om den yttre miljön.

Att inordna arbetsmiljöinformation i ett vidare sammanhang torde vara relevant på alla olika områden där utredningen kommer med förslag (kap. 9–15). Främst kanske det dock ännu gäller bevakning av arbetsmiljörisker och bevakning av kunskapsutvecklingen. Exempelvis synes de epidemiologiska bevakningssystemen – vars betydel-

se även SCB vill understryka – vara lämpliga för registrering av geografisk anknytning, vilket i efterhand kan utnyttjas för sammanställningar av yttre miljöfaktorer och arbetsmiljödata.

Utvecklingsarbete pågår inom SCB beträffande sådan statistik över den yttre miljön som är av betydelse i detta sammanhang, t.ex. avseende s.k. belastningar. Som en första början har SCB på OECDs begäran tagit fram enkla uppgifter om hur stor andel av Sveriges befolkning som bor i områden med vissa höga genomsnittliga svaveldioxidhalter i (utomhus-)luften. Lokala uppgifter av detta slag kan tas fram för vissa tätorter. Redovisningsformer och datakvalitet i dessa sammanhang diskuteras för närvarande.

Samordningen mellan arbetsmiljö och yttre miljö bör inte bara gälla samverkande faktorer. En annan viktig samordning är den som krävs för att beräkna den totala exponeringen för en viss faktor. I vissa arbetsmiljöer finns t.ex. höga halter av koloxid – men samma personer utsätts också för koloxid vid resor till och från arbetet, och ibland i hemmet, i synnerhet om man där använder ved eller gas. Mätningar av total exponering för koloxid företas nu i USA. Inom SCBs allmänna utvecklingsarbete på luftkvalitetsstatistik diskuteras även dessa frågor.

En fråga av något annorlunda karaktär gäller transporter och annan hantering och användning av giftiga varor och ämnen. Transporter kan ses som ett problem både för dem som utför transporterna och för den omgivande befolkningen och miljön. Jordbrukets användning av bekämpningsmedel kan innebära både hälsorisker för jordbrukarna och miljöproblem. Informationsbehovet sammanfaller delvis. Även de kartläggningar som utredningen föreslår på detta område bör genomföras med medverkan från SCB.

4. Utvecklingen av SCBs dödsorsaks- och dödsfallsregister

Utredningen understryker vikten av att register med information som kan belysa riskerna i arbetslivet finns allmänt tillgängliga. Register som nämns i detta sammanhang är bl.a. SCBs dödsorsaks- och dödsfallsregister. Det senare registret gör det möjligt att kombinera uppgifter om yrke, näringsgren m.m. från folk- och bostadsräkningarna med uppgifter om dödsorsaker, dödsålder, hemort etc. från dödsorsaksstatistiken så att information kring sambanden mellan dödlighet, dödsorsaker och miljöfaktorer kan erhållas.

Användningen av dödsorsaksstatistiken har ökat mycket kraftigt under senare år. Uppgifter om dödsorsaker används alltmer direkt inom den medicinska vetenskapen samt i hälso- och sjukvårdsplaneringen. Den ökade användningen av dödsorsaksstatistiken har medfört att nya krav ställts på statistikens innehåll och kvalitet. Inom SCB pågår därför arbeten med att förbättra dödsorsaksstatistiken i olika avseenden. Speciellt har åtgärder vidtagits för att förbättra datakvaliteten.

Den ökande regionala bevakningen av befolkningens hälsotillstånd har även medfört att det ställts ytterligare krav på den regionala redovisningen av dödsorsaksstatistiken. Mot denna bakgrund föreslår SCB i sin anslagsframställning för 1983/84 att verket får bygga upp ett system för bevakning av dödlighetsutvecklingen och dödligheten för olika dödsorsaker inom olika regionala områden. En löpande bevakning av dödligheten möjliggör även att vissa arbetsolyckor med dödlig utgång kan uppmärksammas inom olika geografiska områden. Enligt SCBs uppfattning kan ett bevakningsprogram för dödligheten ses som ett värdefullt komplement till de regionala epidemiologiska bevakningssystem, som föreslagits komma till stånd och som utredningen understrukit behovet av.

SCBs dödsfallsregister syftar till att ge underlag för orienterande studier av överdödlighet inom olika yrken, socio-ekonomiska grupper, näringsgrenar m.m. Dödsorsaksstatistiken och folkräkningsmaterialen utgör basmaterial för dödsfallsregistret. SCB har tidigare lagt upp ett register för 1960-talet byggt på uppgifter från folk- och bostadsräkningen 1960 och dödsorsaksstatistiken för åren 1961–1970. Uppläggningen

av detta register finns redovisad i rapporten Dödsfallsregister 1961–1970, nr 1981:5 i serien Promemorior från SCB.

För att få ett register som är mer aktuellt och som täcker längre observationsperioder pågår vid SCB arbeten med att också skapa dödsfallsregister med uppgifter från folk- och bostadsräkningen 1970 och uppgifter från dödsorsaksstatistiken för åren 1971–1980. Avsikten är att sedan årligen komplettera dödsfallsregistret med uppgifter om dödlighet och dödsorsaker. Registret kan betraktas som ett komplement till andra register av liknande karaktär, t. ex. socialstyrelsens cancer- och miljöregister. SCB planerar att presentera de första resultaten från dödsfallsregistret 1971–1980 under innevarande budgetår.

5. Arbetsskadestatistikens innehåll och användning

SCB vill i sin egenskap av medproducent av den officiella arbetsskadestatistiken särskilt ta upp de problem som rör denna statistik. Mot bakgrund av det stora intresset för och efterfrågan på uppgifter om arbetsskador anser SCB i likhet med utredningen att denna statistik kvalitativt sett bör kunna förbättras.

SCB delar utredningens uppfattning att inrapporteringen av arbetsskadeuppgifterna från arbetsskadeanmälan till informationssystemet om arbetsskador (ISA) ej fungerar helt tillfredsställande. För att uppnå erforderlig kvalitet i arbetsskadestatistiken krävs nu betydande arbetsinsatser i fråga om komplettering och upprättning av ofullständigt lämnade uppgifter på ingående blankettmaterial. Åtgärder bör därför vidtas för att höja kvaliteten i redovisningen till ISA.

Utredningen föreslår att särskilda föreskrifter, som säkerställer att uppgifterna om arbetsskador kan utnyttjas för statistiken, skall utarbetas. Enligt SCBs uppfattning är det angeläget att sådana föreskrifter kommer till stånd och att de sammanställs i samråd med SCB. De föreslagna föreskrifterna bör kunna bidra till att kvaliteten på indata förbättras. Det är därvid väsentligt att även uppgifterna om företaget och arbetsstället omfattas av dessa föreskrifter.

Utöver uppgifterna från arbetsskadeanmälan inhämtas uppgifter om näringsgren och arbetade timmar till ISA via en samkörning med det centrala företagsregistret (CFR). Uppgifterna om arbetade timmar används i arbetsskadestatistiken för en beräkning av relativa arbetsskadefrekvenser. CFR baseras på information från bl. a. andra myndigheters register. Arbetstidsuppgiften inhämtas från riksförsäkringsverkets (RFVs) register över arbetsgivare. Fr. o. m. 1982 har denna uppgift utgått från RFVs register.

Slopandet av arbetstidsuppgifterna i RFVs material medför problem även för SCB eftersom detta mått används för storleksklassificering av företag och arbetsställen i CFR. Inom SCB pågår därför ett utredningsarbete som syftar till att beräkna ett nytt storleksmått baserat på kontrolluppgifterna (KU) från det s. k. RS-systemet (rationalisering av skatteadministrationen). Det nya måttet skall i första hand användas för klassificeringar i samband med urvalsdragningar men det skulle eventuellt även kunna användas för arbetsskadestatistikens behov. Måttet skall om möjligt avspegla arbetsvolymen i företaget dvs. motsvara det tidigare använda timbaserade måttet. I utredningsarbetet kommer även att beräknas jämförelser mellan timmåttet och det nya KU-baserade måttet så att eventuella systematiska skillnader kan belysas.

Vidare föreslår utredningen att verksamheten omkring ISA skall ordnas så att arbetsskadematerialet bättre kan utnyttjas för olika ändamål utöver de periodiska publikationerna som tas fram inom ramen för den officiella statistiken. SCB vill i detta sammanhang peka på de möjligheter verket har att medverka även i arbetet med framställning av de mer speciella uttagen av arbetsskadestatistik. Detta kan ske inom ramen för verkets uppdragsverksamhet.

Under innevarande år har inom SCB ett nytt ADB-system utvecklats för framställning av den officiella arbetsskadestatistiken. Detta system har getts en flexibel utformning och är nästan helt uppbyggt kring

olika standardprogram. Då ändringar i detta system är enkla att genomföra kan det med fördel även utnyttjas för uttag som syftar till att belysa särskilda frågeställningar inom arbetsmiljöområdet, t.ex. för uttag av uppgifter för en särskild bransch, yrkeskategori eller för en viss typ av skada.

6. Avslutande kommentarer

SCB har i sitt utvecklingsarbete på arbetsmiljöområdet funnit – i likhet med utredningen – att behovet av en officiell arbetsmiljöstatistik är stort. Myndigheter, partsorganisationen, samarbetsorgan och fackliga organisationer på arbetslivets och arbetsmiljöns område är intresserade av att följa arbetsmiljöns kvalitet och utveckling. SCB har därför i sin anslagsframställning för 1982/83 föreslagit att verket får fortsätta med utvecklingen av en officiell arbetsmiljöstatistik så att en samlad översiktlig statistik om arbetsmiljön kan presenteras. Förslaget avslogs emellertid, varför SCB på nytt har anmält arbetsmiljöstatistiken i sin anslagsframställning för 1983/84.

Mot denna bakgrund vill SCB särskilt understryka utredningens bedömning och förslag att SCB skall tilldelas nya medel, utöver vad som ryms inom befintliga budgetramar, för framställning av en officiell arbetsmiljöstatistik. Det är därvid väsentligt att denna statistik – utöver sammanställningar av redan befintlig statistik – även grundar sig på viss ny datainsamling. De årliga kostnaderna för denna verksamhet har av utredningen beräknats till 500 000 kr i 1982 års priser. SCB vill ansluta sig till denna kostnadsberäkning.

Vidare är SCB beredd att åtaga sig ett uppdrag om att närmare utreda behovet av metodutveckling för uppskattning av risker i arbetsmiljön samt att lägga fram förslag om hur behovet skall tillgodoses. Enligt SCBs uppfattning bör detta arbete bedrivas utifrån de resultat och erfarenheter som framkommit i SCBs metodutvecklingsarbete på arbetsmiljöområdet samt i samarbete med intressenter och partsorganisationer inom detta område.

Arbetsmiljöinformationen bör i vissa fall inordnas i ett vidare sammanhang för att en fullständig och korrekt bild skall kunna ges av verkligheten. Bidragande till detta är existensen av faktorer som i samverkan kan ge negativa hälsoeffekter, där exponering för den ena föroreningen kan ske i arbetsmiljö och för den andra i yttre miljö. En samordning av den statistiska informationen om arbetsmiljön och om den yttre miljön är därför i flera fall önskvärd. SCB anser att större uppmärksamhet bör ägnas dessa frågor inom de olika områden där utredningen lägger fram förslag.

SCB vill avslutningsvis framföra att verket delar utredningens uppfattning att åtgärder bör vidtas för att säkerställa informationen i arbetsskadestatistiken. De förslag som utredningen kommer med i detta sammanhang bör genomföras i samråd med SCB.

Studiebesök i Danmark och Norge med anledning av SCBs fördelningsmodellprojekt

Inom SCB har ett särskilt fördelningsmodellprojekt startats. Ett viktigt syfte med detta är att utveckla modeller för konsekvensberäkningar vid ändringar av skatte- och bidragsregler, subventioner m.m. Undertecknade, som tillsammans med Staffan Wahlström (projektansvarig) arbetar med SCBs fördelningsmodellprojekt, besökte 2–3 december 1982 såväl Danmark som Norge för att i resp. land studera motsvarande modeller.

I Danmark studerades den s.k. lovmodellen. Denna har skapats för att tillgodose det ökade intresset för studium av fördelningseffekter. Den gynnas av att man i Danmark har byggt ut de personnummerorienterade dataregistren i mycket stor omfattning. Beslut om att upprätta lovmodellen fattades av den danska regeringen under 1980. Det ansågs vara en fördel att arbeta med ett centraliserat system som kunde användas av flera intressenter.

På ekonomidepartementet finns ett särskilt sekretariat som svarar för den löpande driften. Övergripande frågor behandlas i ett särskilt lovmodellråd. Danmarks Statistik

har en nyckelroll i arbetet som innebär ansvar för anskaffande av de s. k. modellbefolkningarna och datamässig uppläggning. Budgeten för 1983 uppgick till något mer än tre milj. danska kronor.

De användare som finns av modellen kan själva göra egna bearbetningar direkt via terminal. Detta kan t. ex. gälla ändringar i skattesatser, avdrags-, bidragsregler. Vid mer komplicerade beräkningar måste dock sekretariatet kontaktas. Det finns en handbok som har utarbetats för användarna. Dessutom har man en särskild grundkurs på två dagar.

Lovmodellen består av tre delar nämligen:

1. Modellbefolkningarna
2. Lovkartotek
3. Lovdatabank

Det finns många olika modellbefolkningar. Den största modellbefolkningen består av 138 000 personer fördelade på 90 000 familjer. För varje familj finns 14 variabler och för varje person 80 variabler. För att skapa denna stora modellbefolkning har fyra olika register sammanslagits. Varje användare kan anpassa lovmodellens filer till en egen modellbefolkning om så önskas. Denna möjlighet har dock endast utnyttjats sparsamt.

I lovkartoteket finns information om systembeskrivningar, beräkningsrutiner och analysmetoder. Av särskilt intresse i systemet är de moduler som tar hand om framskrivning, anpassning, kalkylering och analys. Det finns således möjlighet att korrigera beräkningarna med hänsyn till ändringar i befolkningssammansättning, aktuella inkomstförhållanden m. m.

Lovdatabanken innehåller lagregler, vissa makrobelopp och tidsseriedata. Gränsdragning mot lovkartoteket är flytande.

Hela systemet är uppbyggt med utgångspunkt i SAS (ett standardprogram för statistisk analys).

Några exempel på typiska analyser som genomförts är följande:

1. Analys av marginaleffekter. Den sammanlagda marginaleffekten av skatter och inkomstberoende transfereringar.
2. Utveckling av den reala disponibla inkomsten för arbetar- och tjänstemannahushåll med 2 vuxna och 2 barn.
3. Analys av förändringar i skattesystemet, t. ex. effekten av ändringar i det standardavdrag som får göras med 5 % av lönesumman, dock högst 3 200 Dkr.

Den danska lovmodellen används flitigt. Främst av olika departement.

I Norge är situationen annorlunda. Modellerna har här skapats för att lösa vissa bestämda uppgifter. De har vuxit fram successivt sedan slutet på 60-talet. Till skillnad från Danmark är varje modell specialprogrammerad.

Modellerna handhas av en särskild skatteforskningsgrupp som finns vid Statistisk Sentralbyrå i Oslo. Fördelen med detta ansågs vara att man därigenom lättare än utomstående kunde utnyttja tillgänglig statistik och påverka denna.

Det som modellerna kan beräkna är först och främst effekter av ändringar i personliga skatter och arbetsgivaravgifter, transfereringar, avgifter och subsidier på konsumtionsvaror samt andra ändringar som har konsekvenser på prisnivån och det relativa priset på dessa varor och tjänster.

I likhet med den danska lovmodellen är huvudregeln att modellerna endast kan behandla de direkta effekterna. Det finns emellertid ett undantag och det gäller den s. k. INSIDENS-modellen, där en metod är inbyggd som gör det möjligt att ta hänsyn till att hushåll kan förändra sin konsumtion som en konsekvens av ändring i de relativa priserna.

I korthet finns följande modeller:

1. KFS. Den har som utgångspunkt typhushåll. Det som används för att typbestämma hushållen är inkomstbegreppet som varieras efter önskemål.
2. SKATT. Denna modell används för att kunna beräkna taxerad inkomst och därmed sammanhängande skatteinkomster.
3. LOTTE. Modellens grundmaterial är deklarationsblanketten. Den används t. ex. vid analys av ändringar i skattebestämmelser.
4. INSIDENS. Med hjälp av denna modell kan t. ex. ändringar i subventionssystem studeras.

Uppdragsgivare är i första hand Finansdepartementet och de politiska partigrupperna i Stortinget. Socialdepartementet och Forbrukardepartementet, intresseorganisationer, dagspress m. m. var andra användare.

Till skillnad från Danmark genomfördes alla bearbetningarna på modellerna av medarbetare i skatteforskningsgruppen.

Det var värdefullt att på detta tidiga stadium under uppläggning av SCBs fördelningsmodellprojekt få ta del av de erfarenheter som fanns i Danmark och Norge på detta område.

Bengt Eklind *Rolf Johansson*
Göran Jostin

Det finsk-ungersk-svenska samarbetet kring miljöstatistiska metodfrågor

I mars 1981 överenskom cheferna för de statistiska centralorganen i Finland, Ungern och Sverige om gemensamma insatser kring olika metodfrågor inom naturmiljöstatistiken. För ändamålet tillsattes en arbetsgrupp med två representanter från vardera landet.[1]

[1] Se Statistisk tidskrift 1981: 3, sid 236.

Arbetsgruppen höll sitt andra sammanträde den 28–30 april 1983 i Helsingfors. Från Sverige deltog avdelningschef Knut Medin och avdelningsdirektör Thomas Polfeldt, SCB. Förutom informationsutbyte kring aktuella utvecklingsprojekt m. m. behandlades två huvudämnen:

1. Statistisk beskrivning av luftkvaliteten.
2. Bedömning och redovisning av miljöstatistikens kvalitet (inkl. läsanvisningar i miljöstatistiska publikationer)

Till mötet hade vart och ett av de tre länderna tagit fram ett större dokument om det ena av dessa ämnen och ett mindre dokument om det andra ämnet.

Det kunde konstateras att väsentliga insatser gjorts för utveckling av statistiken över luftkvaliteten. Som ett resultat av mötet kommer en gemensam metodpublikation om luftkvalitetsstatistik att utges hösten 1983. Den innehåller bearbetade versioner av de tre vid mötet framlagda rapporterna.

Nästa möte hålls i Stockholm våren 1985. Även vid detta tillfälle kommer två huvudämnen upp till behandling, nämligen (1) avfallsklassificering och avfallsstatistik och (2) tidsserieanalys inom miljöstatistiken. Alla tre länderna åtog sig att utarbeta rapporter i båda ämnena.

I enlighet med det avtal, som gäller för samarbetet, övergick ordförandeskapet i arbetsgruppen efter detta möte från Sverige till Finland.

Knut Medin

Summary in English

P. W. Wargentin, 1717–1783. "The Father of Swedish Statistics"

This year, 200 years have passed since the death of Pehr Wilhelm Wargentin, called the father of Swedish statistics. In commemoration, Statistics Sweden has published a memorial volume (No. 15 in the series Selections) with seven essays by Wargentin, drawn from the 1754, 1755 and 1766 records of the Royal Swedish Academy of Sciences. Also included is a newly written essay by Dr Erland Hofsten, in which he presents Pehr Wargentin as an individual and a demographer and describes Wargentin's part in the establishment of the Swedish population statistics and the conditions leading to the creation of a central statistical office.

Statistics Sweden also arranged a statistics seminar on 19th January, 1983, at which this publication was presented by Dr Hofsten. Professor Erik Bylund (University of Umeå) spoke about "Population Statistics in a Historical-Geographical Perspective", and Professor Sune Åkerman (also Umeå) about "Population Statistics in Historical-Demographical Research".

On the same day as this seminar was held, the Nordic Committee on Population Statistics was meeting in Stockholm, and the four representatives of the other Nordic countries made short addresses at the seminar about the population statistics in their respective countries. Three of these speeches have been reproduced in this issue of Statistical Review (though without English summaries). They are: "Befolkningsstatistikkens udvikling i Danmark indtil 1849" (The development of the population statistics in Denmark before 1849) by Mrs Lene Skotte, Statistics Denmark, Copenhagen; "Nordiska och finländska tankar med anledning av 200-årsminnet av Pehr Wargentins död" (Some Nordic and Finnish reflections on the occasion of the 200-year anniversary of the death of Pehr Wargentin) by Mr Göran Strengell, Central Statistical Office, Helsinki; and "Befolkningsstatistikens utveckling i Norge" (The development of population statistics in Norway) by Mrs Gerd Lettenström, Central Bureau of Statistics, Oslo.

Information on income obtained from an interview survey compared with information given to taxation authorities

by M. Christoffersen

1. Introduction

Question wording plays a special role when embarrassing or controversial topics are dealt with. In this connection, many authors mention income questions as an example of questions with both a low response rate and a high degree of inaccuracy. Responses to the traditional income question: "What was your taxable income last year?" are compared with the Danish income tax return.

2. Survey Data

The survey of the living conditions of the Danish population aged 20–69 conducted in August–September 1976 comprised a nationally representative sample of well over 5000 persons. Among the variety of questions on living conditions one dealt with the respondents' taxable income in 1975. The question read: "What was your own taxable income in *1975*?"

Those interviewed were shown a card

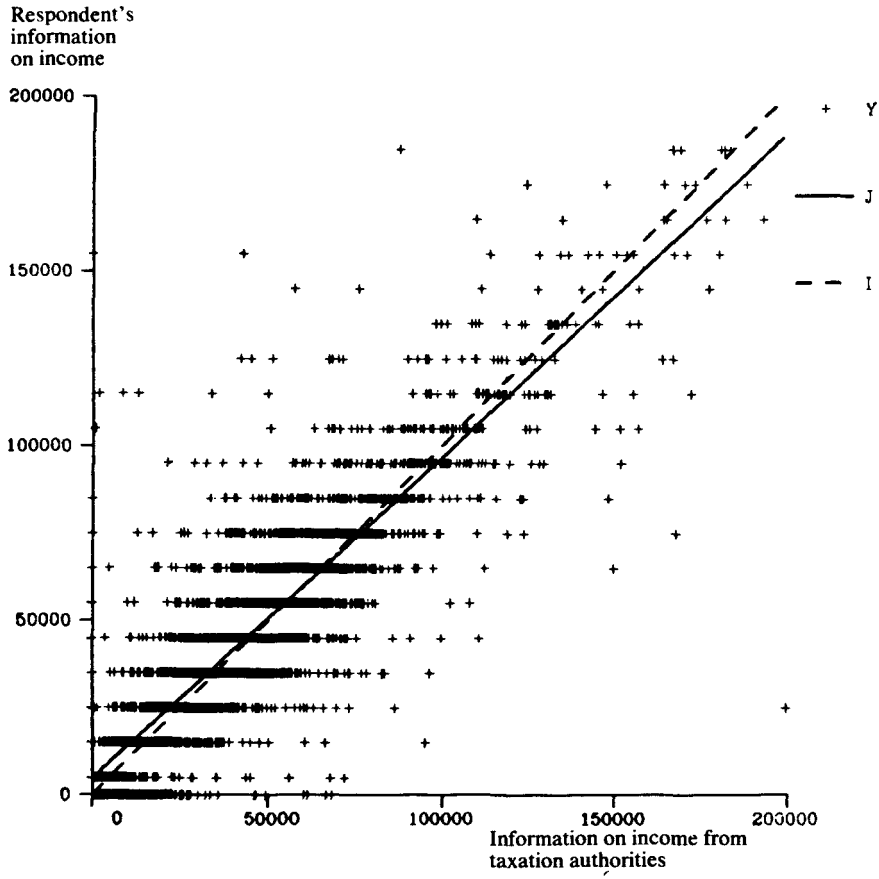


Fig. 1. The final tax assessment plotted against the respondent's own information on income

The dotted line shows the identity line where there is complete agreement between the information given by respondents and by taxation authorities. The solid line is fitted using linear regression techniques.

listing 20 income intervals after which they were asked to indicate which of the DKr. 10000-intervals applied to them.

The final assessments of the taxation authorities were stated in intervals of DKr. 100.

The taxation authorities' information on income is stated in intervals of DKr. 100 whereas the respondent's own information is given on the basis of prompt cards with DKr. 10000-intervals.

3. Survey Results

A comparison between the respondents' own information and that of the taxation authorities showed that, in relation to the final assessments of the taxation authorities, 20–69-year-old Danes overstate their taxable income by 2.5 per cent.

Table 1. Percentage of respondents who answered the question on taxable income, distributed according to occupational status and social group

	Social Group				
	I+II	III	IV	V	Number
	per cent				
Occupational status:					
Self-employed	86	90	95	–	671
Salaried employees	94	97	97	–	1 581
Workers	–	–	99	96[1]	1 362
Number	456	890	1 264	1 010	3 620

[1] Unskilled workers.

The analysis showed that this more detailed information could be reduced to a simple linear regression:

$y=a+bx$ (the solid line),

where y is the taxable income disclosed during the interview and x is the income according to the final assessment.

The analysis shows that 80 per cent of the variations in the answers to the simple income question can be predicted by means of the regression line based on taxable income.

The regression analysis shows that the slope is significantly less than 1.0, which implies a systematic trend of a slight *overstatement* of incomes in the very lowest income groups while those with above average incomes tend to *understate* their incomes when asked the traditional income questions.

While the respondents in the survey of living conditions who disclose their taxable income generally overstate this slightly, there is at the same time a tendency for those in the higher income brackets *not* to state their income. However, only 5 per cent of all the persons interviewed in the survey of living conditions were unwilling or unable to state their taxable income.

While practically all young people disclose their income, there is a tendency for the older age groups to be unwilling or unable to give such information. Among the economically inactive persons interviewed, only pensioners were reluctant to provide information on their financial circumstances. The unemployed, persons not employed outside the home and students willingly revealed their taxable income. These tendencies could be interpreted as confirming that the answering of the question on taxable income makes heavy demands on memory and assumes certain knowledge of the complicated nature of the concept seen from a technical-administrative point of view.

As far as sex differences are concerned, the material shows that different types of response are found among those not disclosing their taxable income. Whereas two thirds of those *refusing* to answer are males, almost two thirds of those stating that they *do not know* their taxable income are females.

ATP Contributions

by A. Ch. Ståhlberg

Changes in the relative size of the contributions paid by the employers to the Swedish Supplementary Pension Scheme (the ATP scheme) depend upon several factors, e.g. demographic changes, changes in the number of early-retirement pensioners, and the accumulation process. They are also a consequence of some rules of the scheme, where the 30-year rule stipulates that 30 years of income are sufficient for obtaining a full pension and the 15-year rule that the

size of the pension is calculated from the best 15 income years. The article presents an analysis of the effects of each factor on the pension costs.

If demographic changes mean shifts in the population structure towards a relatively greater number of elderly persons, these changes will for a long time have but a minor influence on the pension costs, and not until the year 2015 will any noticeable demographic effects start to appear. The implications of a greater number of early-retirement pensioners are both an increase of the pension payments and a decrease in the number of persons on which the contributions are calculated. In another 30 years, the number of early-retirement pensioners is expected to be double that in 1979 and to account for some 10–15 per cent of the pension costs. Half of these costs are likely to pertain to persons whose early-retirement pensioning is due to shortage of work.

The smoother the individual life income patterns and the higher the employment ratio, the smaller the effects of the 30-year and 15-year rules. In a scheme still under construction as the Swedish ATP scheme, the 15-year rule will only gradually take full effect. In a fully extended scheme, the two rules will account for 25–30 per cent of the pension costs.

Immigrants as Victims of Crime

by T.-L. Leiniö

The rate of exposure to "violence resulting in visible marks or bodily harm", "violence without visible marks" and "threat(s) sufficiently serious or dangerous to scare the victim" is significantly lower among native Swedes than among immigrants. In the article, immigrants are defined as persons both of whose parents are 'or have been' nationals of a foreign country. Even second-generation immigrants, those with either one or two immigrant parents, are more frequently exposed to various forms of violence than native Swedes. *Immigrant males* account for the greatest difference between Swedes and immigrants with respect to the *percentage of persons exposed to violence or threats*, while *immigrant females* have reported a significantly *greater mean number of acts of violence* than either immigrant males or native Swedes of either sex.

The difference between Swedes and immigrants remains even after adjustments have been made for differences in the composition of the groups in respect to sex, age, family situation, type of housing and urban/rural domicile. The type of housing is a more important factor than the other variables in explaining the difference. The age distribution of the second-generation groups, with their heavy concentration in the youngest age classes, explains the difference in exposure rate between these groups and the native Swedes. Which particular nationality an immigrant belongs to seems to be of no importance; it is the very condition of being an immigrant instead of a native Swede that raises the exposure rate.

Litteratur

Bokanmälningar

Pehr Wargentin — den svenska statistikens fader. En minnesskrift. Statistiska centralbyrån, Stockholm 1983. Urval 15. 180 s. Kr. 75:–.

Sverige äger en i det närmaste obruten serie av befolkningsstatistiska data ända sedan mitten av 1700-talet. År 1749 inrättades nämligen det s. k. Tabellverket som kom att bli grunden för det svenska befolkningsstatistiska systemet. Den person som kom att betyda mest för Tabellverkets verksamhet under dess första år var astronomen och statistikern Pehr Wargentin född 1717 och död 1783. Med anledning av att det i år är 200 år sedan Pehr Wargentin dog har statistiska centralbyrån gett ut en minnesskrift som innehåller 7 originaluppsatser ur Kongl. Svenska Vetenskapsakademiens Handlingar. I boken finns dessutom en introduktion om Pehr Wargentin och om Tabellverkets historia av Erland Hofsten.

För att förstå den roll befolkningsstatistiken kom att spela och anledningen till att den fick en så god start, måste man känna till något om det ideologiska klimatet vid denna tid. Sverige var sedan några år tillbaka besegrat i ett långvarigt krig och man måste bygga upp landet igen. Under inflytande av merkantilistiska tankegångar höll man på att inventera de resurser som landet hade. Till sådana räknades inte bara naturaförekomster utan också befolkningen. Det fanns därför en anledning till att skaffa en noggrannare bild av befolkningens förändringar. Man hade bl. a. en föreställning om att det förelåg en stark befolkningsminskning på grund av en omfattande utvandring, särskilt från jordbruksbygder.

Under en följd av år hade man diskuterat möjligheterna att samla in uppgifter om befolkningen. En förutsättning för ett sådant arbete var givetvis den redan i slutet av 1600-talet införda kyrkobokföringen. Att befolkningsstatistiken fick en så välorganiserad ram redan från början berodde på en stor framsynthet hos de personer som tog aktiv del i arbetet med planeringen. Man tog också lärdom av erfarenheter från andra länder så att deras misstag kunde undvikas. En betydelsefull roll spelade förmodligen Pehr Elvius som var matematiker och astronom. Redan 1744 publicerade han en uppsats om barnantalet i Uppsala med utgångspunkt i data om födda åren 1694–1743. En annan betydelsefull insats gjordes av Jakob Albrecht von Lantingshausen, som senare blev överståthållare i Stockholm. När han efter mångårig tjänst i utlandet återkom till Sverige skrev han ett memorial om nyttan av befolkningsundersökningar utifrån erfarenheter från England och Holland. Lantingshausens memorial blev underlag för ett definitivt förslag som stadfästes av Konungen 1748 och vilket innebar att Tabellverket skulle träda i funktion den 1 januari 1749.

Pehr Elvius fick inte leva för att se Tabellverket i verksamhet någon längre tid. Redan sommaren 1749 avled han hastigt under en resa till södra Sverige. Till efterträdare utsåg man den 32-årige Pehr Wargentin, också han astronom från Uppsala.

Fram till sin död författade Pehr Wargentin en rad skrifter inom befolkningsstatistiken. Dessa publicerades i Kongl. Svenska Vetenskapsakademiens Handlingar och behandlade så vitt skilda ämnen som säsongväxlingar i antal födda och döda, folkutflyttningen ur Sverige, befolkningstillväxten i Sverige och i länen m. m. Han publicerade också några uppsatser på andra håll med befolkningsstatistiskt innehåll.

I den skrift som nu getts ut av statistiska centralbyrån finns förutom beskrivningen av Wargentins verksamhet som demograf

och Tabellverkets första år också i faksimil några av hans tidigaste uppsatser. Det är en serie av artiklar med titeln "Anmärkningar om nyttan af årliga Förtekningar på födda och döda i et Land", som utkom i Vetenskapsakademiens kvartalshäften 1754 och 1755. Med finns också en artikel om dödligheten: "Mortaliteten i Sverige, i anledning av Tabell-Verket", vilken utkom första kvartalet 1776.

Dessa artiklar har väl i dag närmast ett historiskt kuriosaintresse men belyser ändå, på ett ganska intressant sätt, tankar hos en vetenskapsman från senare delen av 1700-talet, som idag ändå förefaller rätt moderna. Bland uppsatserna i serien "Om nyttan af förtekningar på födda och döda ..." är den femte särskilt intressant eftersom Wargentin i den diskuterar olika dödsorsaker. Statistiken var väl ganska ofullständig och beskrivningen av dödsorsakerna lämnade väl också en del övrigt att önska. Wargentin själv var kritisk mot klassificeringen och gör en del allmänna reflexioner kring detta. Han försöker sig också på en jämförelse av dödligheten i städer och på landsbygd vilket följande citat visar:

"Det förtjänar visserligen eftertanka, at Hetsig feber och i synnerhet Slag, Apoplexia, hafva et så grufveligt öfvervälde i Stockholm, mer än på Landet och i andra stora Städer. Hvad annat kan dertil vara orsaken, än oordenteligare lefnad, öfverflöd i mat och dryck, samt i synnerhet et omåtteligt bruk af utländska starka liqueurer?"

Den sista uppsatsen som behandlar dödligheten i Sverige finns tidigare publicerad av Livförsäkringsbolaget Thule i anledning av den 9:e Internationella Aktuariekongressen 1930. Den innehåller en dödlighetstabell vilken var den första i sitt slag avseende en hel nations befolkning. (Däremot är det inte en egentlig livslängdstabell i modern bemärkelse.)

För den som är intresserad av den svenska befolkningsstatistiken och dess historia samt av statistiska centralbyråns föregångare Tabellverket ger boken om Pehr Wargentin mycket av intresse. Den spänner över en tidig period i verksamheten medan entusiasmen var stor och innan förfallet trädde in i slutet av 1700-talet. Faksimiltrycket från Vetenskapsakademiens Handlingar är kanske lite svårläst, men om man ger sig tid att läsa igenom dem finner man mycket av intresse som inte bara är kuriosa. Genom att studera dem kan man få en intressant bild av dåtidens verklighet och en bakgrund till det vi utläser ur de historiska tabeller vi idag har tillgång till.

Ingvar Holmberg

Lärare för livet? Lärarnas arbetsmarknad. Tendenser 1960–81 med utblick mot 1995. Information i prognosfrågor 1982:6. SCB 1982. 93 s. Kr. 30:–.

Syftet med prognosinstitutets rapport "Lärare för livet?" är enligt förordet "att ge en lättöverskådlig beskrivning av uppbyggnaden av lärarprognoserna" samt "ge en redogörelse för hur tillgången på lärare påverkas av hur många av de lärarutbildade som väljer läraryrket och hur många som lämnar det". Denna redogörelse är lite mer komplicerad än vad man vid första reflektionen tror att det rör sig om.

I sina första kapitel redovisar rapporten en hel del statistik från skolområdet – "landets största arbetsplats". Man tar upp en del fakta om lärarutbildningen och förvånas över att det faktiskt inte är särskilt många antagningsplatser på lärarutbildningarna med tanke på att det är en mycket stor yrkeskår – över 140000 personer. Teckningslärarlinjen har 96 antagningsplatser, lågstadielärarlinjen 336 och matematik- och naturvetenskapliga ämneslärarlinjen 290 platser.

När man läst igenom vilka faktorer som påverkar behovet av lärare och faktorer som påverkar tillgången börjar man undra om det verkligen går att göra en trovärdig sammanvägning för att få fram en prognos. Förordets ambition att ge en lättöverskådlig beskrivning av lärarprognosernas uppbyggnad har man enligt min mening inte lyckats med. Delvis beror det på att det faktiskt *är* komplicerat men också på att man inte i ett eget avsnitt gjort en beskrivning av en generell prognosmodell. Vidare saknas en hel del definitioner av centrala begrepp – vad

betyder exempelvis "lärarbenägenhet", något om lärares benägenhet att arbeta som lärare förmodligen, men hur detta beräknas ges inget svar på. Det hade också varit bra att få en definition på vilka lärare som finns i lärarregistret vilket bl. a. skulle kunna förklara nytillskotten i lärarkåren i åldern 50–60 år. Delvis beror dessa nytillskott på att lärare som varit verksamma i annat yrke eller hos annan arbetsgivare tas bort i lärarregistret för att sedan uppstå som nytillskott när de återvänder till skolan. Man frågar sig om det inte vore angeläget att man skapar ett komplett lärarregister där alla lärare – oavsett om de är yrkesverksamma eller inte – finns registrerade.

Avslutningsvis ger rapporten ett antal förklaringar till att prognoser inte alltid "slår in", det kan vara brist på statistik, ändrade bestämmelser på undervisningsområdet eller andra faktorer som inte går att kalkylera med.

Jag tycker nog inte att man ska ha så orimliga krav på prognoser som utgör ett led i en planeringsprocess. Syftet skall inte vara att prognosera läget eller ens det mest sannolika läget ett visst år utan prognosen bör redovisa läget i ett antal alternativ där man laborerar med exempelvis antalet antagningsplatser, klasstorlekar etc. Sedan är det politikerna som med prognosen som ett av underlagen skall ta ansvaret för utvecklingen.

Björn Steneberg

Bergman, L. R. och Wärneryd, B., Om datainsamling i surveyundersökningar. Vilken metod är bäst och vad får vi egentligen veta? Statistiska centralbyrån. Liber/Allmänna förlaget. 1982. 241 s. Kr. 125:–.

För att få information om tillståndet och utvecklingen inom olika samhällssektorer i vårt land genomföres ofta s. k. "surveyundersökningar", dvs. undersökningar som innebär insamling av data om och direkt från människor tillhörande olika befolkningsgrupper. Studierna görs ofta i form av frågeundersökningar gällande representativa urval av befolkningen i ett land eller annat område.

Surveymetodiken förbättrades och tillämpades i allt större omfattning under 1950- och 1960-talen. Skälen till den ökade användningen var flera bland vilka kan nämnas t. ex. undersökarens önskan att själv få bestämma vilken typ av information han/hon vill ha (vilket ju inte vanligen är möjligt vid utnyttjande av sekundärdata), undersökarens önskan att rikta undersökningen mot någon viss specifik typ av population, som kanske inte är urskiljbar bland tillgängliga sekundärdata samt ett ökat intresse för att tillförsäkra olika befolkningsskikt och -kategorier möjligheter att framföra uppfattningar och önskemål i olika samhällsfrågor. Surveymetodiken har ibland uppfattats som ett medel för ett ökat medborgarinflytande på beslutsfattandet i frågor av betydelse för medborgarna i samhället. Metoden har ofta värderats positivt därför att den vanligen ger användaren en känsla av att han vet vad han gör, när han väljer att studera ett problem med en surveyundersökning.

Problemen med att genomföra surveyundersökningar har otvivelaktigt ökat under 1970-talet, som exempel kan nämnas svårigheten att övertyga utvalda undersökningspersoner om att de bör deltaga i undersökningen (den s. k. bortfallsproblematiken har alltså vuxit), svårigheten att få undersökningspersonerna att godtaga relevansen i vissa av formulärets olika frågor och frågeområden samt de starkt ökade styckekostnaderna per genomförd utfrågning. En person eller organisation som funderar på att söka belysa ett problem med hjälp av en surveyundersökning befinner sig alltså i en besvärlig beslutssituation. Skall man våga satsa den ofta ansenliga summa som en dylik undersökning vanligen kostar för att få den önskade informationen. Och kommer man i så fall att faktiskt få veta vad man ville veta? Detta beror i hög grad på kvaliteten i den genomförda studien.

Vid en forskarkonferens i statistisk metodik 1979, anordnad av SCB, kretsade diskussionen till stor del kring kvaliteten i studier av förändringar och differenser. I diskussionen av denna problematik efterlyste jag en sammanställning över de erfarenhe-

ter som under årens lopp erhållits av SCBs utredningsinstitut, vad gäller felbenägenheten vid användning av olika variabler och mått i olika typer av undersökningar av olika slag av populationer.

Det var därför med tillfredsställelse som jag mottog upplysningen att Lars R. Bergman och Bo Wärneryd nu hade samlat ihop sina mångåriga erfarenheter från genomförandet av stora undersökningar i en bok med titeln: "Om datainsamling i surveyundersökningar" med den intresseväckande undertiteln: "Vilken metod är bäst och vad får vi egentligen veta?"

I en tidigare bok av Langlet-Wärneryd redovisades erfarenheter av blankett- och frågekonstruktion och i den nu aktuella boken har ett vidare ämnesområde behandlats, nämligen hela datainsamlingsprocessen. Boken innehåller 9 kapitel, där det första – "Begrepp och perspektiv" – drar upp gränserna för bokens innehåll. Det betonas att diskussionen kommer att gälla huvudförfarandena vid datainsamling och kretsa främst kring problem vid intervjuer och postenkäter samt huvudsakligen gälla standardiserade förfaranden. Tre olika betraktelsesätt på kvalitetsproblematiken presenteras varefter författarna exemplifierar en statistisk kvalitetssyn som innebär att kvaliteten knyts till resultatens (de statistiska tabellernas) överensstämmelse med de "sanna" resultaten. Två typer av kvalitetsegenskaper anges – *relevansfaktorer* och *feltyper*. Relevansfaktorerna gäller statistikens innehåll och aktualitet och feltyperna gäller systematiska fel (skevhet eller bias) och slumpmässiga fel (precision).

Därefter presenteras en psykometrisk kvalitetssyn där begreppen reliabilitet och validitet är centrala. Reliabilitetsbegreppet anknyter till det statistiska precisionsbegreppet men reliabilitetsbegreppet gäller inte slutresultatets tillförlitlighet utan främst tillförlitligheten i en enskild mätning. Validitetsbegreppet ansluter sig närmast till relevans- och skevhetsbegreppen i den statistiska synen på kvaliteten. Uttalanden om kvalitet kan dels vara absoluta (dvs. hur pass väl ett visst förfarande fungerar), dels vara relativa (dvs. vilket förfarande är bäst). De kan också avse hur pass nära en total kvalitetstäckning ett visst förfarande är samt variera beträffande i vilken grad de är kvantitativt grundade.

I kap. 2 behandlas de olika leden i en statistisk undersökning och exemplifieringar ges av olika datainsamlingsmetoder. Som illustration av en postenkätundersökning har valts SCBs Omnibus, telefonintervjuer illustreras med hjälp av Arbetskraftsundersökningen (AKU), besöksintervjuer med hjälp av Undersökningen av levnadsförhållanden (ULF) och bokföringsmetoden illustreras med hjälp av Hushållsbudgetundersökningen. Den som är intresserad av syftet, uppläggningen och av vilka som undersöks i dessa stora återkommande studier kan här på ett översiktligt sätt erhålla dylik information.

Kap. 3 behandlar mätsituationen vid olika datainsamlingsmetoder. Först presenteras en modell över mätsituationen, där de centrala delarna utgörs av två "feedback" processer, en beträffande uppgiftslämnarens sätt att tillgodogöra sig frågorna, den andra gällande hans svar. Därefter diskuteras mätsituationen vid en postenkätundersökning. Det – jämfört med en intervjuundersökning – normalt sett högre bortfallet vid en postenkät anges bero dels av att det är svårare för en person att komma sig för att svara på en postenkät än vid en intervju samt på att det är "socialt" svårare att avvisa en intervjuare än att avstå från att skicka in en enkät. Därefter behandlas mätsituationen vid en intervjuundersökning. Här framhålls dess fördelar i möjligheten till feedbackprocesser såväl i frågor som svar.

Vidare påpekas den uppenbara risken för intervjuareffekter i detta sammanhang, dvs. att intervjuaren påverkar den svarande så att fel kan uppstå. Ett vanligt sådant fel är att intervjuaren inte följer givna instruktioner.

Som exempel nämns att i en undersökning ställdes så mycket som en tredjedel av frågorna ej precis i enlighet med formuleringen och det var *de erfarna* intervjuarna som svarade för flest avvikelser. En annan undersökning visade att endast 37% av intervjuarbeteendena var normal fråge-

ställningsaktivitet medan 31% gällde "frågefeedback" (frågeupprepning) och 26% "svarsfeedback" (uppmuntrande kommentarer). Trots att många feedbackprocesser är mycket kortvariga är det troligt att de kan ha stor betydelse för hur ett svar kommer att se ut. Dylika processer måste ändå tillåtas, eftersom intervjuns genomförande grundar sig på att ett dylikt socialt samspel kan utvecklas mellan intervjuare och svarande. Feedbackprocesserna måste dock kontrolleras så långt det är möjligt t. ex. genom noggrann träning av intervjuarna. En jämförelse görs också av undersökningssituationen vid telefon- resp. besöksintervjuer. Effekterna av skillnader i "avstånd" mellan intervjuare och svarande vid dessa två förfaranden diskuteras. När man söker hålla allt annat konstant finner man endast små skillnader i resultaten vid användning av dessa metoder.

Ett avsnitt behandlar intervjuareffekter och åtgärder för att reducera dylika. Kapitlet avslutas med ett försök att avgöra vilken datainsamlingsmetod som förordas av uppgiftslämnarna. Något klart besked kan författarna ej ge.

Kap. 4 gäller uppgiftslämnarrollen och inställningen till surveyundersökningar. Först behandlas kontakten med uppgiftslämnaren och orsaker till vägran. En sammanställning ges av åtgärder som kan verka motivationshöjande på deltagandet och en redovisning görs av resultaten från två studier av vägrans orsaker. Bland resultaten kan nämnas att det på ett allmänt plan synes finnas en utbredd brist på förtroende för undersökningar främst vad gäller sekretess och syfte. Individuellt fanns det en rad samverkande faktorer till deltagande/ej deltagande.

Sedan behandlas undersökningens ämne och målpopulation. Det hävdas att undersökningens relevans för målpopulationen synes vara den näst viktigaste faktorn för bortfallsnivån vid postenkäter. Viktigast är "antalet kontaktförsök". Språkförståelseproblematiken nämns och man påpekar att språkförståelsen är mindre god för en hel del uppgiftslämnare även bland svenskar. Kapitlet avslutas med ett avsnitt om skyddet av den personliga integriteten. Man konstaterar att mätningar av tilltron till att personuppgifter behandlas på ett för individen betryggande sätt visar att många är misstrogna i detta avseende. Samtidigt synes det svårt att mäta effekterna av ett mer eller mindre långt gående uppgiftslämnarskydd.

I kap. 5 diskuteras val av metod med hänsyn till informationsproblemet. En rad problemområden behandlas. Kapitlet upptar knappt 60 sidor och är synnerligen innehållsrikt med rader av exemplifieringar samt viktiga tips och påpekanden av betydelse för den som står i begrepp att beställa eller själv genomföra en surveyundersökning.

Den slutsats som redovisas är att några dramatiska skillnader mellan olika datainsamlingsmetoder oftast inte kan iakttagas och att de skillnader som trots allt förekommer sannolikt till stor del kan förhindras om man vid utformning av frågor och formulär söker anpassa sig till de specifika omständigheter som olika mätsituationer kan innebära. De aspekter på informationsproblemet som behandlats i kapitlet sammanfattas i tre grupper:

1. Grad av teknisk kommunikationssvårighet.
2. Grad av kommunikationssvårighet med avseende på känslighet.
3. Omfånget.

I fråga om 1 skulle besöksintervjun få ett plus framför övriga diskuterade datainsamlingsmetoder. I fråga om 2 skulle postenkät och besöksintervju (med speciellt inlagt skyddsförfarande vara fördelaktigast) och i fråga om 3 skulle besöksintervju få plus framför övriga förfaranden.

I kap. 6 behandlas kostnadsrelationer och problem vid s. k. metodmix. De huvudsakliga kostnadsparametrarna i en datainsamling presenteras och även den skattade kostnadsrelationen mellan olika datainsamlingsstrategier för en typundersökning. De rörliga kostnaderna för olika slag av undersökningar anges ungefärligen och skillnaderna är dramatiska mellan en postenkät och en besöksintervjuundersökning.

Metodmix, dvs. användning av flera olika metoder t. ex. för att driva ned bortfallet vid

frågeundersökningar, kan naturligtvis få effekter på svarskvaliteten. Studier av detta problem har författarna ej lyckats hitta, men man misstänker att försök att minska bortfallet kan leda till att data av tvivelaktig kvalitet kommer att läggas till undersökningsmaterialet.

I kap. 7 jämföres metoder med avseende på några speciella problem. Det gäller sådant som frågeordning och kontextproblem, standardisering av förfaranden, graden av ställningstagande till en fråga samt partiellt bortfall.

Kap. 8 tar upp förfaranden vid kvalitetsbelysning av datainsamlingen. Ett avsnitt behandlar möjligheterna att följa hur data genereras och vilka proceduravvikelser som förekommer. Exempel på genomförandet av dylika kvalitetsbelysningar vid SCB ges. Vidare behandlas replikering och återintervju med exemplifieringar på sådana kvalitetskontroller inom ramen för AKU 1978. Till sist ges exempel på ett valideringsförfarande tillämpat på frågor om arbetsmiljö ingående i ULF.

Det nionde och sista kapitlet i boken är en sammanfattning av vad man kommit fram till i de tidigare kapitlen. Sammanfattningen görs i form av en rad korta påståenden om de tre huvudförfaranden för datainsamling som granskats i de föregående avsnitten. Meningen är sannolikt att söka underlätta besvarandet av frågan i bokens huvudtitel. Något entydigt svar kan naturligt nog ej ges utan detta är beroende av frågeställningar, egenskaper hos undersökningspopulationen samt av krav på mätsituationen. Sammanställningen torde dock kunna vara till god hjälp vid val av lämplig undersökningsplan och vid utvärdering av undersökningsrapporter för läsare som ej är väl förtrogna med surveymetodik.

De båda författarna har producerat en lättläst och innehållsrik bok, späckad med praktiska illustrationer av användning av surveymetoder och därmed sammanhängande problem. Boken torde ha stort intresse för såväl utredare och beställare av utredningar som för lärare i ämnen där utrednings- och undersökningsmetodik ingår som viktiga kursmoment.

Per Kempe

Brändström, A. & Sundin, J. (ed), Tradition and Transition. Report No. 2 from the Demographic Data Base, University of Umeå. Kungälv 1981. 255 s.

"Tradition and Transition" består av sex uppsatser och har underrubriken "Studies in microdemography and social change". Den sista i raden av uppsatser är en beskrivning av den demografiska databasen.

Volymens titel sägs vilja framhålla (den analytiska) betydelsen av faktorer som har att göra med kultur, normer och traditioner vid vårt studium av det förflutna. Sanningen kommer man ofta inte fram till på ett enda sätt utan det gäller att kombinera det som står till buds. – Eller som det står i förordet: "The best results will probably be found when 'hard' and 'soft' information both support one's theories".

Den demografiska databasens organisation och sätt att fungera har tidigare presenterats i Statistisk tidskrift (1978:6). Som framgår av Christina Danells uppsats blev databasen en särskild och permanent enhet vid Umeå Universitet år 1978. De första sju församlingarna som togs in i databasen var spridda över hela landet. Det var Fleninge (Skåne), Locknevi (Småland), Gullholmen (Bohuslän), Trosa (Södermanland), Svinnegarn (Uppland), Tuna (Medelpad) och Nedertorneå (Norrbotten). Sedan några år tillbaka pågår registreringen av Sundsvallsregionen och församlingarna Lögdö, Selånger, Tynderö och Attmar börjar nu enligt uppgift bli fullt färdiga. Vid mitten av 1980-talet är förhoppningsvis hela regionen registrerad och arbetet med Linköpingsområdet har kommit igång. Det arbetet beräknas pågå in på 1990-talet.

Det har sålunda hittills inte funnits så mycket att hämta från databasen. Men bättre och bättre blir det. Huvudsaken är i alla fall att arbetet fortsätter och att insatserna inte minskar. I varje fall är det endast en av de fem ämnesinriktade uppsatserna i "Tradition and Transition" som nästan helt bygger på uppgifter från databasen. Uppsatsen heter "Infant mortality in a changing society. The effects of child care in a Swedish parish 1820–1894". Författare är Anders

Brändström och Jan Sundin, forskare resp. chef vid databasen.

Med hänsyn till den svenska folkbokföringens beskaffenhet har databasen kommit att avse tiden från omkring 1820 till 1895. Det är för övrigt under denna period som de stora förändringarna på det befolkningsmässiga området började inträffa i vårt land. Dessa förändringar har både initierat och initierats av en mängd yttre faktorer inom områden som teknik, medicin, ekonomi, kommunikation, kultur. Studium av 1800-talet är en enormt spännande sysselsättning för den som är intresserad av mänsklig och social utveckling. Avsikten med Brändströms – Sundins studie är att bidra till kunskapen om hur och varför spädbarnsdödligheten minskade i en svensk församling under den demografiska transitionens tidigare skeden. Spädbarnsdödligheten är inte enbart intressant som isolerad företeelse utan den är en jämförelsevis god indikator på hur det över huvud taget ser ut i samhället. Detta framhålls också i uppsatsen.

Liksom andra demografiska komponenter uppvisar både dödligheten i allmänhet och den spädbarnsspecifika en regional variation. Omkring år 1860 hade Jämtlands län lika låg dödlighet som i hela landet år 1940. Perioden 1860–1866 var spädbarnsdödligheten i Jämtlands län 92 promille, i Västernorrlands län 174 promille och i Stockholms stad 271 promille!

Omkring mitten av 1700-talet uppgick spädbarnsdödligheten (dvs. antal döda under första levnadsåret per 1 000 levande födda) till mer än 200 promille. I dag är talet omkring 7. När man betänker att det är fråga om medeltal så förstår man att variationerna regionalt eller mellan olika befolkningsgrupper måste ha varit betydande för ett par hundra år sedan.

Det material Brändström – Sundin använder sig av utgörs dels av statistik på kontraktsnivå avseende åren 1806 och 1855, dels utgörs det av bearbetningar av databasens församlingar Fleninge, Trosa, Svinnegarn och Nedertorneå. I nämnd ordning redovisas studier av det regionala mönstret, sociala skillnader, åldersberoendet under första levnadsåret, biometrisk analys, säsongvariationer över kalenderåret. Ett studium som jag möjligen saknar är dödlighet med hänsyn till barnets ordningsnummer (paritet). Hur mycket man skulle ha fått ut av det i ett "högdödlighetssamhälle" är kanske osäkert, men variabeln ordningsnummer torde i ett sammanhang som detta vara mera viktig än exempelvis säsongvariation över året.

När man går igenom den statistiska delen i en undersökning som denna får man fram en mängd upplysningar och indikationer. Men tyvärr finns det ju inte kvantitativa uppgifter om annat än vissa grundläggande variabler. Bakomliggande orsaker får ofta prövas utifrån befintliga ämneskunskaper, t. ex. i det här fallet om hur det svenska samhället i stort och smått fungerade för ett eller annat århundrade sedan. Den verkligt intresseväckande informationen ges i slutavsnitten, där författarna försöker tolka erhållna resultat med hjälp av bl. a. dödsorsaksstatistik och rapporter från provinsialläkare om barnens skötsel.

När man tänker på hur människorna levde – eller tvingades leva – förr i tiden, och vilken kunskapsnivå de hade om hur kroppen fungerar, skulle man litet drastiskt kunna säga att det inte är så märkligt att dödligheten var hög utan snarare förvånande att den inte var ännu högre. För att hålla sig till spädbarnsdödligheten synes man helt allmänt kunna fastslå att det var de bristande kunskaperna om hygien och möjligheterna att "praktisera hygien" som gjorde att många barn avled strax efter födelsen. Att den nyfödda måste ha modersmjölk förstod inte alla, inte heller att ett spädbarns mage inte tål vanlig mat eller komjölk och framför allt inte om födan serveras sur eller ur en oren nappflaska (motsvarande).

De citat ur läkarrapporter som ges i uppsatsen säger kort och rakt på sak mera om våra förfäders liv och leverne än många konventionella folklivsskildringar. (Enligt vissa uppgifter skall för övrigt under de senaste åren ett minskat "breast-feeding" ha orsakat en uppgång av spädbarnsdödligheten i en del u-länder.)

Men det var inte bara hos allmogen som

vetskapen om hygienens betydelse ofta var obefintlig. (Det bör dock framhållas att de indelta soldaterna bibringades förhållandevis goda kunskaper om hur de skulle sköta kroppen.) Här kan bara erinras om den kamp den ungerske läkaren och obstetrikprofessorn Ignaz Philipp Semmelweis fick föra vid mitten av 1800-talet för att få antiseptiken accepterad vid förlossningar. I korthet gick Semmelweis' upptäckt ut på att läkarkandidaterna på den förlossningsklinik i Wien där han tjänstgjorde tillförde kvinnorna som de vårdade barnsängsfeber på grund av att de blivande läkarna ofta gick direkt från obduktionsbordet till förlossningssängen. Genom att införa större renlighet vid förlossningarna kunde Semmelweis nedbringa dödligheten hos barnsängskvinnorna från tio till tre procent. – Men motståndet var hårt från de medicinska auktoriteterna.

Men döden uppträder i många former. Vid en genomgång av dödböckerna för en bohuslänsk församling avseende 1860-talet upptäckte jag att två mödrar samma år och månad under sömnen råkat kväva sina barn till döds.

En annan uppsats med nära demografisk anknytning är Sune Åkermans (professor i historia) "How did the great decline in fertility start?" I uppsatsen beskrivs en intervjuundersökning som genomfördes i Rätan i Jämtlands län år 1979. I den introduktion som tillställdes de för intervju uttagna personerna (i huvudsak sådana som bildade familj under 1920–1930-talen) framhålls betydelsen av att få nedtecknat upplevelser och minnen, eftersom kommande generationer "annars knappast kan föreställa sig hur det var att vara barn när seklet var ungt eller att komma ut på arbetsmarknaden eller att bli hembiträde eller husmor på 1920-talet". Frågeformuläret var ganska omfattande och anknöt till, förutom vissa karakteristika för intervjupersonen, barndom och närmaste släktingar, livet som vuxen och livet som pensionär. Att göra en samhällsvetenskaplig undersökning med hjälp av retrospektiva intervjuer är inte särskilt enkelt i något avseende. Detta framgår av den mycket noggranna beskrivningen av undersökningens uppläggning och genomförande, vilken är värd att beaktas av dem som sysslar med liknande. Slutsatsen blir emellertid att Rätan-studien visar att retrospektiva intervjuer med vissa givna förutsättningar går att genomföra.

"Control, punishment and reconciliation" är en studie av "parish justice" i Sverige före 1850. Författare är Jan Sundin. Artikeln presenterar det decentraliserade rättsväsende som fanns inom kyrkoadministrationen och som var av särskild betydelse under 1600–1700-talen. Läsaren får här en ingående beskrivning av den lokala rättsskipningen inkl. bakgrunden till förhållanden och utveckling. Det var tydligen inte bara Konungen som var ansvarig inför Gud allena utan även husbönderna verkar i stor utsträckning ha haft det så ställt i disciplinära frågor gentemot familj och anställda.

De två huvudsakliga institutionerna inom församlingen var kyrkostämman (beslutsorgan) och kyrkorådet (styrelse och förvaltning). Kyrkodisciplinen övervakades av kyrkorådet. Till sitt förfogande hade stämma och råd en krets av lokala polismän ("sexmän"), vilka var ansvariga för att myndigheternas beslut verkställdes. Vid mitten av 1800-talet tillkom sockennämnder som en följd av den under årens lopp utvecklade byråkratin. Dessa övertog uppgifter från kyrkoråden, i huvudsak de som hade att göra med lag och ordning. Bland förseelser som behandlades av församlingarna kan nämnas frånvaro från husförhören, svärjande, vanhelgande av gudstjänstdagen. Under det sistnämnda sorterade en varierande flora av brott. Här kan framhållas dryckenskap, kortspel, dans, högljuddhet, skvallrande. Föräldrar som inte såg till att barnen kom till kyrkan på söndagarna och husbönder som inte gav sina tjänare tillfälle att besöka gudstjänsten skulle också straffas. De straff som utdömdes var för vuxna i allmänhet böter eller stocken. Prästen kunde neka en felande rätten att få nattvarden. Mest grymt synes dock barnen ha behandlats. Kroppslig aga skulle utdelas i närvaro av sexman eller av kyrkorådet utsett vittne. Men de kunde även bli dömda till stocken. Barn fick alltså

i fastbundet tillstånd och kristet nit bli begabbade av församlingsborna på söndagarna. Förhoppningsvis dömdes inte barn till gapstocken särskilt ofta, men nog har det förekommit.

I slutet av uppsatsen försöker författaren att genom olika källstudier få en uppfattning om vad som verkligen hände. Resorna i det förflutna som läsaren får följa med på går genom Bygdeå 1726–1759, Stigsjö 1684–1739, Bergsjö 1750–1850 och Harmånger 1750–1850. I allmänhet verkar det vara spriten som ligger bakom de händelser som fått kyrkan att ingripa. Domarna inskränktes ofta till en varning och många gånger torde den tilltalade inte ha inställt sig så att ärendet kunnat avgöras.

Möjligen var det så att de kyrkliga domarna och bestraffningarna mildrades under 1700-talet jämfört med vad som hände tidigare. Regionala olikheter kan ju också ha förekommit. Det mest makabra som inträffat i Sverige på kyrkotuktens område torde väl ändå vara det som ärkebiskop Abraham Angermannus iscensatte under sin resa i Linköpings stift år 1596. Med svenska mått mätt bör händelserna ha varit i stort sett obeskrivliga. Risinstitutionen upplevde sin höjdpunkt, bakdantare och skvallerbärare hade högkonjunktur och de dömande behövde inte alltför länge fundera på rätt och orätt, lögn och sanning.

De två återstående uppsatserna i volymen är rubricerade "Mortality and morbidity in Swedish iron foundries 1750–1875" samt "Swedish blacksmiths in the nineteenth century". Båda är författade av Jan Sundin och Lars-Göran Tedebrand (professor i historia). Det skulle föra för långt att här försöka referera eller kommentera också dessa. Vare nog sagt att de är lärorika och fyllda av fakta. Särskilt skall kanske den senare artikeln framhållas på grund av den okonventionella och medryckande uppläggningen, genom vilken man får följa och jämföra en individuell och en statistisk livscykel under perioden 1835–1893.

Som slutord skall bara framhållas att det måste vara ett enormt arbete som ligger bakom "Tradition and Transition". Beskrivningarna är noggranna och dokumenterade. Läsaren bör för övrigt inte glömma att studera notapparaten, i vilken många tips om intressanta referenser kan hämtas.

Rune Tryggveson

Hofsten, E., Indien. Om samhället och befolkningsproblemen. Norstedts 1982. 79 s. Kr. 48:–.

Erland Hofsten är en av våra mer flitiga resenärer till olika, för oss nordbor, mer eller mindre exotiska platser i världen. Dessa resor brukar i allmänhet ge avkastning i form av en eller flera artiklar och i vissa fall mer omfattande skrifter. I början av 1981 tillbringade Hofsten en period på 3 månader som gästprofessor vid University of Delhi. De intryck som han då fick, tillsammans med erfarenheter från tidigare besök i Indien, har nu samlats och presenteras i en bok med titeln: "Indien. Om samhället och befolkningsproblemen".

Här i Norden betraktar vi väl gärna Indien som ett enhetligt land, stort till ytan och med en mycket talrik befolkning. Tillgänglig statistik kompletterar bilden med uppgifterna att befolkningen uppgår till närmare 700 miljoner och att ytan är 3 à 4 gånger större än Sveriges (3,3 miljoner km^2 och inte 5,5 som det står i boken).

Frågan är om indiern i allmänhet uppfattar sitt land som en enhetlig nation. Den misstanke vi tidigare haft att så inte skulle vara fallet bekräftas av Erland Hofsten. Den indiska halvön är en subkontinent med 3 nationalstater. Indien är den största av dessa och består i sin tur av 22 delstater samt 9 självstyrande territorier. Varje delstat har en egen lagstiftande församling och ett eget parlament. Skillnaden mellan dessa delstater är mycket stora i politiskt, ekonomiskt, socialt och kulturellt avseende. Kanske t. o. m. större än mellan de tre nationalstaterna och det är kanske inte så underligt eftersom några av delstaterna är lika stora som de största av de europeiska länderna.

Vår allmänna kännedom om Indien och förhållandena där är väl ganska begränsad och ändå är det ett urgammalt kulturland.

Indiens tidigare historia har vi kanske hört talas om i förbigående när vi läst om Alexander den Stores erövringståg i Asien eller om mongolernas härjningar under vår medeltid. Mer detaljerad kunskap har vi kanske från den tid då engelsmännen fick herraväldet över den indiska halvön en bit in på 1700-talet eller från mitten av 1800-talet när Indien blev en formell kronkoloni till England. Hofsten tecknar de stora dragen av Indiens fyratusenåriga historia, från omkring 2000 år före vår tideräknings början fram till våra dagar.

Dagens Indien vet vi kanske mer om, men det är ändå svårt att föreställa sig de stora skillnader som råder mellan olika grupper i det indiska samhället. Hofsten ägnar stort utrymme åt att beskriva samhället och förutsättningarna för ekonomisk och social utveckling. I boken ställs den betydelsefulla frågan om barnbegränsning kan vara en lösning på Indiens problem.

Det är väl få länder där frågan om familjeplanering diskuterats så lidelsefullt som i Indien. Hofsten påpekar att tvångssteriliseringarna under 1970-talet inte förändrade Indiens befolkningstillväxt i någon nämnvärd grad. Han menar också att det inte är befolkningstillväxten som sådan som är det största hindret för utvecklingen utan det är andra krafter som verkar.

Utbildning är en faktor som ofta åberopas när det gäller förutsättningar för förändring i fruktsamheten. Det finns talrika undersökningar i Indien som visar att fruktsamheten är lägre bland familjer där föräldrarna har någon form av utbildning. Men att från detta dra slutsatsen att det nyligen introducerade programmet för skolutbildning skulle leda till någon väsentlig sänkning av fruktsamheten kan vara riskabelt. Så länge den stora massan av befolkningen anser sig ha glädje och nytta av att föda många barn kommer de att fortsätta med det även om deras utbildning höjs.

Det kan också finnas särskilda anledningar till att man vill bibehålla en stark befolkningstillväxt. Ett exempel som nyligen fått en fruktansvärd aktualitet, är konflikten i Assam, dvs. den delstat som gränsar till Bangladesh. Under åren efter Indiens delning och Bangladesh's självständighet lämnade många människor Bangladesh på grund av sin hinduiska religion. De flesta flyttade till Assam där nu inemot hälften av befolkningen består av invandrade bengaler. Detta är orsaken till de svåra motsättningar som råder mellan assameser och bengaler. De förra uppfattar bengalerna som inkräktare vilka är på väg att ta över deras land. När det så beslutades att bengalerna skulle få rösträtt ledde det till den blodiga konflikt som vi kunnat läsa om i tidningarna. Att i en sådan miljö propagera för barnbegränsning, när det för assameserna gäller att bibehålla sin överlägsenhet i antal, förefaller inte vara särskilt lämpligt. Det föreligger liknande potentiella konfliktsituationer även i andra delar av Indien, t. ex. när det gäller förhållandet mellan hinduer och muslimer.

I en avslutande del av boken tar Hofsten upp sambandet mellan sociala förändringar och folkökning ur ett mer generellt perspektiv. Med exempel från Irland och några andra europeiska länder försöker Hofsten dra några slutsatser om den möjliga utvecklingen i Indien.

I Irland ledde ekonomiska åtgärder inom jordbruket till en mycket kraftig folkökning som sedan följdes av en lika kraftig folkminskning genom svält och emigration. Även på andra håll i Europa fick man en stark folkökning om än inte lika snabb som på Irland. Där avlastades det värsta trycket genom emigration. Efter hand som industrialiseringen kom i gång kunde också folkökningen sugas upp av en växande industri.

Frågan är om några av dessa erfarenheter kan äga tillämpning på Indien. Hofsten diskuterar migration och folkmängdstillväxt, familjeplanering etc. och drar slutsatsen att ingen av dessa faktorer egentligen kan leda till en ändring i förhållandena på längre sikt för Indiens del. Det är helt andra åtgärder som behöver vidtagas: "En ny social ordning är nödvändig och det enda sättet att åstadkomma en sådan tycks vara genom det tryck som folkökningen bland de nu undertryckta åstadkommer".

Det är en intressant bok om ett intressant land som Hofsten presenterar. Frågorna

hopar sig om hur man här ska komma tillrätta med sina stora problem. Det är givetvis inte lätt att sammanfatta ett så rikt stoff på ett ganska begränsat utrymme som en bok på cirka 75 sidor utgör; jag tycker att Hofsten har lyckats bra med den uppgiften. Det gamla citatet från den tyske 1700-talsskalden Claudius gäller i hög grad Erland Hofsten: "Wenn jemand eine Reise tut, so kann er was erzählen".

Ingvar Holmberg

Hofsten, E., Demografins grunder (3e upplagan, Studentlitteratur. Lund 1982. 191 s. Kr. 110:−.

Demografins grunder utkom första gången 1969 och presenterades då ganska utförligt (Statistisk tidskrift 1970:6). Boken var en av de första läroböcker på svenska som försökte ge en så fullständig bild av demografins alla olika aspekter som möjligt. I och med den andra upplagan som kom några år senare gjordes en grundlig omarbetning. Den tredje upplagan som nu föreligger överensstämmer i stort sett med den föregående.

Den väsentligaste förändringen är en uppdatering av de statistiska uppgifterna där så varit möjligt. Dessutom har det tillkommit lite om folkökning och familjeplanering i Indien – förvisso ett resultat av Hofstens senaste besök i detta land.

Bokens målsättning är fortfarande att vara en introduktion i demografi både för dem som tänker specialisera sig i ämnet och för dem som behöver demografi som hjälpvetenskap.

Som en introduktion försvarar boken väl sin plats. Ett minimum av matematik används för att beskriva en rad viktiga demografiska fenomen. Lite mer matematiskt inriktade härledningar har förvisats till ett Appendix.

Under senare år har det skett en ganska snabb utveckling på den demografiska teorins område. Detta gäller framför allt metoder för att framställa multipla dekrementtabeller som utgör grunden för beräkning av t.ex. livslängdstabeller för arbetskraften och för civilståndsförändringar. Förhoppningsvis kommer en översikt av detta nya, mycket viktiga område att finnas med i nästa upplaga av boken.

Ingvar Holmberg

Rolski, T., Stationary Random Processes Associated with Point Processes, Lecture Notes in Statistics, Vol. *5*, Springer-Verlag, New York 1981, ISBN 0-387-90575-8. vi + 139 s. $11:40.

Teorin om stokastiska processer handlar om slumpmässiga förlopp i tiden. För oss statistiker på SCB kan stokastiska processer bl.a. ge modeller för analys av tidsserier.

Ämnet för den här boken är en ganska ny teori som bygger på följande ansats. Man har samtidigt att göra med dels en stokastisk process $X(t)$, alltså en slumpmässig storhet som också är en funktion av tiden, dels en s.k. punktprocess $N(t)$. En punktprocess här är ett speciellt slags stokastisk process, som är lika med noll för alla tidpunkter utom för vissa tidpunkter Y_0, Y_1, Y_2, ..., där den är lika med ett. Man kan säga att det är fråga om ett slumpmässigt val av vissa tidpunkter Y_0, Y_1, Y_2, ... Alternativt kan $N(t)$ få vara en s.k. "märkt" punktprocess, vilket är ett vidare begrepp än punktprocess. I en märkt punktprocess finns viss slumpberoende information knuten till varje vald tidpunkt. Ett sådant här par av processer (X, N) kallas, onödigt tungt och mystifierande tycker jag, för en "stokastisk process associerad med en märkt punktprocess". Detta måste naturligtvis förkortas, och på engelska blir det då "r.p. à m.p.p." Kunde man inte säga "märkpunkterad process"?

Ett fält för tillämpning, som behandlas i boken, finns inom teorin för köer. I köteorin uppfattas en kö av kunder som ett slumpberoende förlopp. Ett sätt att beskriva en kö är följande. Låt $V(t)$ ange hur lång tid det skulle ta att få slut på den kö som eventuellt finns vid tidpunkten t, om det inte kom några nya kunder efter t. Denna process $V(t)$ kallas för köns virtuella väntetidsprocess. Låt vidare $N(t)$ vara en punkt-

process som anger de tidpunkter då det kommer en ny kund. Om man vill hålla reda på olika kategorier av kunder så kan $N(t)$ få vara en märkt punktprocess som även anger kategori för resp. kund. Nu blir paret (V, N) en "r.p. à m.p.p." som på ett väsentligt sätt beskriver hur kön utvecklas.

I fokus för bokens framställning står resultat av vissa allmänna egenskaper hos sådana här par av processer (X, N). En central fråga torde något populariserat kunna formuleras så här: Jämför å ena sidan hur förloppet kan väntas bete sig kort efter starten, å andra sidan hur det kan väntas bete sig på lång sikt. Vilka samband finns mellan sannolikhetsfördelningarna?

Här skall inte gås in på detaljer eller strikta formuleringar, men idéerna kan något antydas. En process sägs vara stabil om fördelningen för medelvärden över långa tidsintervall av processen tenderar att likna en viss fördelning. Den sistnämnda fördelningen blir då stationär dvs. oföränderlig över tiden. Centrala begrepp är också Palm-fördelning och Palm-version. Man säger att (X, N) har Palm-fördelning, och är en Palm-version, om sannolikhetsfördelningen för (X, N) uppfyller ett visst invariansvillkor med avseende på tidsförskjutningar av längden Y_1. En egenskap hos Palm-versioner är att de så att säga har "synkroniserad start" vid tidpunkten 0, i det att $P(N(0) = 1) = 1$. Om en Palm-version är given så finns en motsvarande stationär fördelning som gäller för dess genomsnittliga beteende över långa tidsintervall.

I boken ges en formel som uttrycker denna stationära fördelning med hjälp av den givna Palm-fördelningen. Likaså finns en formel som återskapar Palm-fördelningen ur den motsvarande stationära fördelningen. Vidare ges en del andra formler som uttrycker relationer mellan de båda fördelningarna. Dessa relationer avser bl. a. sannolikheter rörande tiden 0, alltså av formeln $P(X(0) \in E)$, där E är en mängd.

Resultat av det sistnämnda slaget är kanske de som i princip kunde vara de mest intressanta för praktiska beräkningar. Tyvärr är dock formlerna åtminstone delvis ganska komplicerade, med bl. a. upprepade summeringar. I boken ges emellertid även en del följdsatser till relationerna.

Man kan alltså konstatera att Palm-fördelning kommer in som ett naturligt villkor i framställningen. Resultaten medger synbarligen ganska allmängiltiga tillämpningar inom bl. a. köteori (den s. k. G/G/s-modellen för köer).

För att läsa boken bör man nog vara något så när insatt i modern teori för stokastiska processer. Det kan dock noteras att den högre teorin egentligen håller sig i bakgrunden. Själva idéerna, både i resultat och bevis, rör sig väsentligen på en ganska elementär nivå. Boken har förvisso pedagogiska förtjänster. Teori och exempel, mest från köteori, varvas boken igenom. Först behandlas fallet med diskret tidsparameter, sedan det analoga men svårare – och väl intressantare – fallet med kontinuerlig tidsparameter. Men boken har förstås karaktär av matematisk monografi, och den är varken någon populärframställning eller någon egentlig lärobok.

Kan dessa r.p. à m.p.p. ha något intresse för oss som arbetar på SCB? Det är nog inte alls uteslutet att de kan få det. Det talas ofta om att framgent administrativa register kommer att få en allt större roll som primärmaterial till offentlig statistik. Detta kommer då även att ändra kraven på de statistiska metoderna. En av de faktorer som kan vara av vikt är att sådana register ofta baseras på rapporter om händelser som inträffar. Detta i sin tur kan göra att processmodeller av olika slag blir relevanta på ett nytt sätt.

Martin Ribe

Utkom från trycket den 31 augusti 1983.

STATISTISK TIDSKRIFT

UTGIVEN AV STATISTISKA CENTRALBYRÅN

Ansvarig utgivare: Sten Johansson. Redaktör: Nils Welander

Adr: Karlavägen 100. Postadress: 115 81 Stockholm

Telefoner: 08/14 05 60, ankn. 4260 (redaktören personligen)
och 4265 (Helen Lidén)

Prenumeration genom bokhandel eller hos Liber distribution, Prenumerationsorder, 162 89 Stockholm

Utkommer med 5 häften om året

Pris ca 105: – per år eller ca 20: – per häfte (inkl. moms)

Tredje följden Årg. 21 Nr 4 1983

Innehåll

Artiklar

Hans Ylander:
Ett par kartografiska metoder i statistikproduktionen vilka bygger på SCBs tätorts-avgränsningar 269

Yngve Fritzell:
Yrkesfördelningen 1753–1805 enligt Tabellverket: de särskilda städerna 279

Iiris Niemi:
Systematic Bias in Hours Worked? 326

Meddelanden

Europeiska statistiska möten:
Metodfrågor i miljöstatistiken (ECE) 331
Nordiska statistiska utskott:
Utskottet för befolkningsstatistik 331
Utskottet för miljöstatistik 334
Seminarium om vattendata 335
Utskottet för utbildnings- och kulturstatistik 336
OECD-möten:
Möte om miljöstatistik 338
Contact 83 339
SCBs yttrande över underhållsfondsutredningens slutbetänkande "Återinvestera i bostäder" (SOU 1982: 65) 339
Skördestatistiska nämnden 340
Personalnotis 340

Summary in English 341

Litteratur

Nyare svensk och utländsk litteratur 344
Bokanmälningar:
Hofsten, E., Befolkningslära – demografi för lekmän (Ingvar Holmberg) 347
Sachs, L., Applied Statistics. A handbook of techniques (Jörgen Dalén) 347

STATISTICAL REVIEW

PUBLISHED BY STATISTICS SWEDEN, STOCKHOLM, SWEDEN

Five issues a year

Third series. Vol. 21 Nr 4 1983

Contents

Articles

Hans Ylander:
Two cartographic methods based on the delimitation of urban areas at Statistics Sweden 269, 341

Yngve Fritzell:
The Swedish Town Population in the 18th Century; 68 Towns by Occupation and Social Status 279, 341

Iiris Niemi:
Systematic Bias in Hours Worked? 326

Notes

European statistical meetings:
Methodological problems in environmental statistics (ECE) 331
Inter-nordic statistical committées:
Committee on demographic statistics 331
Committee on environmental statistics 334
Seminar on water information and water statistics 335
Committee on educational and cultural statistics 336
OECD-meetings:
Meeting on environmental statistics 338
Contact-83 339
Statistics Sweden: Comments on the final report "To Reinvest in Housing" (SOU 1982:65) 339
The Harvest Statistics Board 340
Personnel note 340

Summary in English 341

Literature

Recent Swedish and foreign literature 344
Book reviews:
Hofsten, E., Befolkningslära – demografi för lekmän (Ingvar Holmberg) 347
Sachs, L., Applied Statistics. A handbook of techniques (Jörgen Dalén) 347

Ett par kartografiska metoder i statistikproduktionen vilka bygger på SCBs tätortsavgränsningar

av avdelningsdirektör Hans Ylander

O. Bakgrund

Förhållandet mellan kartor och statistik är dubbelriktat. Det innebär att statistikresultat kan presenteras i kartform, samtidigt som kartor kan vara ett utgångsmaterial för statistikproduktion, dvs. kartan kan användas som statistikkälla.

I det förra fallet talar man om *tematiska kartor*. Ett särskilt tema som exempelvis befolkning, temperatur eller industriproduktion presenteras i sin rumsliga spridning i kartform. En sådan kartpresentation kan ge särskilda fördelar i form av överblick av ett stort material samt möjligheter att se spridningsmönster och samband med andra element i landskapet. En tematisk karta inrymmer alltså både deskriptiva och analytiska användningsmöjligheter.

I det senare fallet där kartan används som datakälla, eller tillsammans med annat material som underlag för statistikproduktion, måste man givetvis utgå från en verklig karta som den topografiska eller ekonomiska kartan. Ur en sådan karta, som ju gör anspråk på att beskriva en stor del av naturlandskapet och odlingslandskapet jämte bebyggelsen, kan man lyfta ut ett element som man är intresserad av och redovisa dess yta eller antal i form av statistik.

Det är närmast den senare typen av kartografisk metodik som här skall beskrivas i form av utnyttjande av den ekonomiska kartan för datafångst och statistikredovisning. Det rör sig om två ganska osofistikerade metoder, nämligen dels en arealredovisning av tätortsexpansion på jordbruksmark, dels en inventering av allmänt tillgänglig mark för rörligt friluftsliv kring våra större tätorter.

Båda metoderna har först prövats i försöksverksamhet vid enheten för naturresurser och miljö vid SCB och sedan börjat tillämpas i mer reguljär statistikproduktion. Resultaten kommer bl. a. till användning i arbetet med den fysiska riksplaneringen.

Gemensamt för de båda metoderna är också att de utgår från de tätortsavgränsningar på kartor som SCB har genomfört från 1960 års folkräkning och fram till FoB 80. Då dessa tätortsavgränsningar dessutom i sig är en tillämpning av kartografisk metodik i statistikproduktionen kan det finnas skäl att starta med en översikt av tätortsavgränsning och tätortsredovisning både före och efter 1960 års folkräkning.

1. SCBs redovisning av städer/landsbygd och tätorter/glesbygd

Urbaniseringen i Sverige, dvs. städernas tillväxt och landsbygdens avfolkning, har inte bara försiggått under decennierna efter andra världskriget som man i förstone tänker på, utan den har pågått i över 100 år.

År 1880 bodde ca 15 % av befolkningen i städer och köpingar. Nu är förhållandet omkastat så att 17 % bor i glesbygd och resten, 83 %, i tätorter. Denna väldiga befolkningsomflyttning, som nu har avstannat, speglar

givetvis landets omvandling från agrarsamhälle till industristat.

Den officiella statistiken redovisar under 1800-talet befolkningen fördelad på *städer* och *landsbygd*. Mot slutet av århundradet tillkommer ett antal köpingar. Med städer och köpingar avses då och ända fram till 1950 *hela* det administrativa området. Landsbygden utgörs av landskommuner, även om dessa innefattar "stadsliknande samhällen".

Under industrialiseringens framväxt uppstod en rad nya sådana samhällsbildningar som industriorter och stationssamhällen som ej var städer. Detta gjorde att uppdelningen på städer och landsbygd blev ett alltmer ofullständigt mått på den fortskridande urbaniseringen.

Från folkräkningen 1900 meddelas för första gången vissa uppgifter om tätorter. Det rör sig dock bara om administrativa tätorter vid folkräkningarna 1900 och 1910, dvs. städer, köpingar och municipalsamhällen.

Försök hade också gjorts 1900 och 1910 att lämna uppgifter även för agglomerationer utan egen administration som de nya industriorterna och stationssamhällena. Men statistiken över dessa orter blev först 1920 tillräckligt fullständig för att kunna publiceras i folkräkningen.

Man finner alltså vid statistisk redovisning av demografiska, ekonomiska eller sociala data att man sedan gammalt skiljer på stad och landsbygd med hänsyn till olikheterna i näringslivets inriktning. Parallellt med denna redovisning växer under 1900-talet tätort-glesbygdsredovisningen fram. SCB redovisar administrativa och icke-administrativa tätorter.

Städernas administrativa områden motsvarade ända fram till 1940-talet ganska bra den faktiska utbredningen av stadsbebyggelsen. Men under 1940- och 1950-talen inkorporerades vidsträckta områden av renodlad landsbygdskaraktär med städerna.

Detta accentuerades ytterligare av de båda kommunindelningsreformerna 1952 och 1962. Alla kommuner blev större. Slutligen upplöstes 1971 helt skillnaden mellan landskommuner och städer i administrativ bemärkelse.

Man kan alltså notera under 1900-talets förlopp å ena sidan framväxten av stora industriorter på landsbygden utan administrativ status och å andra sidan städer som kommer att omfatta både en stadskropp, glesbygdsområden och andra mindre orter. En redovisning på tätort-glesbygd har då tillkommit i folkräkningssammanhang för att undvika de här anförda olägenheterna.

Till skillnad från stad och köping är tätorten ett renodlat bebyggelse- och befolkningsgeografiskt begrepp som är helt oberoende av de administrativa gränserna. En tätort avgränsas med hänsyn till bebyggelsens täthet och befolkningens storlek.

Definitionerna har förändrats något med tiden, men från 1960 års folkräkning fastslogs följande definition för de nordiska länderna:

"Som tätbebyggt område räknas alla hussamlingar med minst 200 invånare såvida avståndet mellan husen normalt inte överstiger 200 meter."

Till detta kommer ett markanvändningskriterium. Urban markanvändning ingår i tätortsområdet och kan ibland binda ihop olika bebyggelsegrupper.

Fram till och med 1950 års folkräkning var det pastorsämbetet i församlingen som levererade råmaterialet till tätortsstatistiken genom avprickningar i församlingsböckerna, som 1950 ersattes av mantalslängder som grundmaterial. För att underlätta arbetet följde man förteckningar över

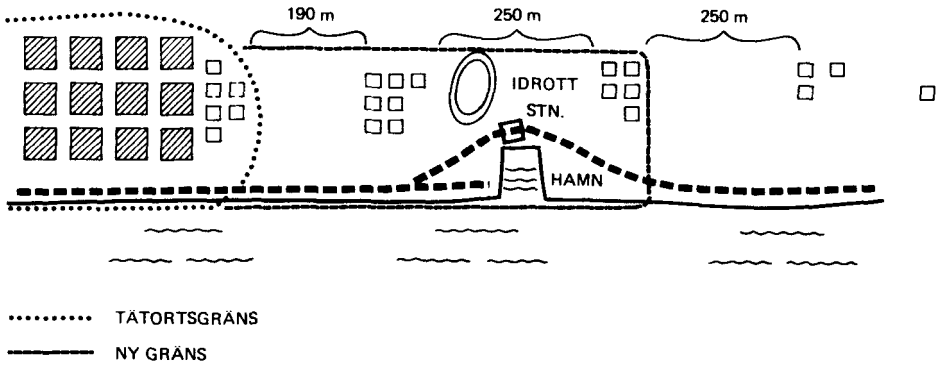


Fig. 1. Exempel på tätortsavgränsning. Den prickade linjen avgränsar en del av bebyggelsen i en tätort. Efter 5 år förnyas tätortsavgränsningen varvid området inom den finstreckade linjen, där det tillkommit ett par nya bebyggelseområden, fogas till tätorten.

Den bortre husgruppen binds samman med orten via "urban markanvändning". De försök med s.k. automatisk tätortsavgränsning som hittills har utförts, missar i detta fall både den urbana markanvändningen och den bortre husgruppen på 5 hus. I större orter ökar dessa fel ytterligare.

fastigheter som tidigare tillhört tätorten.

Men det ligger i sakens natur att SCBs instruktioner för arbetet kunde tolkas olika i skilda församlingar. Genom att man inte använde några kartor vid avprickningen i längderna kunde man inte se om avprickningen av tätortsbefolkningen motsvarade tätortens verkliga sammanhängande bebyggelse.

Efter en mångårig kritik främst från geograferna som pekat på dessa brister, beslöt SCB att till 1960 års folkräkning införa *två* genomgripande förbättringar av tätortsstatistiken.

Den *ena* var att man genomgående införde kartor och flygbilder i avgränsningsarbetet. Det *andra* var att man engagerade geografisk expertis för att utföra själva avgränsningsarbetet på kartor för hela landet. Detta skulle garantera en större enhetlighet och jämförbarhet i tätortsbegreppet än vad pastorsämbetena hade åstadkommit.

Alla orter som kunde tänkas fylla kravet på tätort enligt 1960 års definition avgränsades nu för första gången på kartor.

Den ekonomiska kartan i skala 1:10 000 var då, liksom nu, det bästa kartmaterialet genom sin detaljerade skala och sina fastighetsbeteckningar, som ger kopplingsmöjligheter till fastighets- och befolkningsregister. Ett problem var att den ekonomiska kartan 1960 bara täckte drygt en tredjedel av landet. I övriga delar fick man använda äldre lantmäterikartor, kommunala stadskartor och flygbilder i avgränsningsarbetet. Sedan dess har kartmaterialet successivt förbättrats.

Det centrala avgränsningsarbetet av rikets närmare 2000 tätorter genomfördes under drygt ett års tid. Därefter fick alla

avgränsningar länsvis granskas och fastställas vid särskilda länssammanträden. I dessa deltog överlantmätare, länsarkitekt, länsplanerare, vissa kommuner samt SCBs geograf.

Länsmötena förde det goda med sig att tätortsgränserna blev accepterade av de närmast berörda länsmyndigheterna. Samtidigt bidrog dessa med sin goda lokalkännedom till att vissa justeringar och aktualiseringar av gränserna kunde tillfogas på kartorna. Vidare kunde vissa namnfrågor diskuteras och fastställas. Den fortsatta bearbetningen skedde inom SCB. Tätortskartorna finns nu både i SCBs arkiv och på alla länsstyrelser.

Även vid folk- och bostadsräkningarna efter 1960 har samma avgränsningsmetod på kartor tillämpats och samma granskning och fastställande har gjorts i samråd med länsmyndigheterna.

Målsättningen har varit att utgå från 1960 års grundläggande arbete och aktualisera tätortsgränserna där så behövs med hänsyn till nytillkommen bebyggelse. Därmed har det varit möjligt att åstadkomma först en jämförbarhet i rummet och därefter successivt även en jämförbarhet i tiden.

2. Tätortsexpansion på jordbruksmark 1975–1980

SCBs tätortsstatistik är grundvalen för studier av den svenska urbaniseringen. Det faktum att SCB nu under en 20-årsperiod har avgränsat tätorter på kartor har också varit en förutsättning för redovisningen av *tätortsarealer* och förändringar i dessa arealer. Man kommer då över från den demografiska belysningen av urbaniseringsprocessen till rena markanvändningsfrågor. Denna sida av tätortsexpansionen har varit av stort intresse för arbetet med den fysiska riksplaneringen och dess inriktning på hushållning med mark och vatten. I 1977 års jordbrukspolitiska beslut fastslås bl.a. att den brukningsvärda åkerjorden i landet skall utnyttjas för jordbruksproduktion. Detta innebär stor återhållsamhet med tätortsutbyggnad på åkermark. Den här beskrivna metoden och arealstatistiken kan ge ett underlag att bedöma om intentionerna i 1977 års beslut har förverkligats.

Det enhetliga kartmaterial och de enhetliga tätortsavgränsningar som numera finns har varit utgångspunkter i framtagandet vid SCB av metoden att mäta tätortsutbyggnad på jordbruksmark. Den bygger på uppslaget att använda de transparenta kartor med 1975 och 1980 års tätortsgränser som nu finns vid varje överlantmätarmyndighet tillsammans med den tryckta ekonomiska kartan i färg. Detta är alltså en ny kombination av kartor i samma skala, som tidigare ej kommit till användning. I folkräkningssammanhang används de transparenta bladen med tätortsgränser rutinmässigt tillsammans med den ajourförda s.k. fastighetskartan för inkodning av nytillkomna fastigheter i tillväxtzonen mellan 1975 och 1980 års tätortsgräns.

Genom att byta ut den ajourförda fastighetskartan mot det ej ajourförda färgtrycket av ekonomiska kartan och lägga ett transparent blad med 1975 och 1980 års tätortsgränser över denna karta, kan man se om tätortsexpansionen under denna 5-årsperiod har skett på skog eller åkermark. Denna kartkombination ger med andra ord en möjlighet att konstatera vilken markanvändning som fanns i tillväxtbältet mellan 1975 och 1980 års tätortsgräns *innan* denna mark togs i anspråk av tätortsutbyggnaden.

I den mån utbyggnaden har ägt rum på jordbruksmark (= åker och kultiverad betesmark) planimetermäter man denna areal med en vanlig planimeter, eller ännu

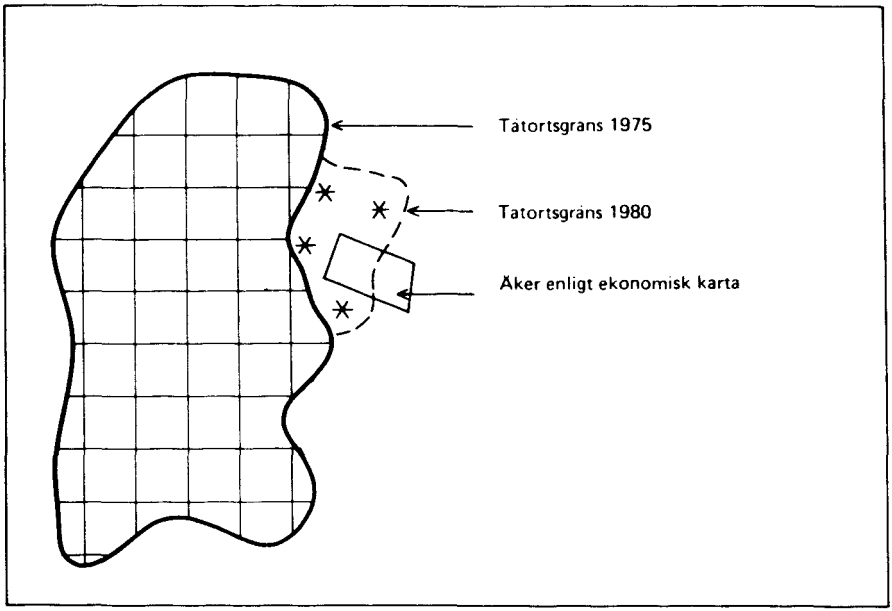


Fig. 2. Den rutade ytan är en tätort enligt 1975 års avgränsning med heldragen linje. Den streckade linjen innefattar den utvidgning av orten som utförts vid 1980 års tätortsavgränsning, och som bygger på ajourförd fastighetskarta.

Den något äldre, tryckta, ekonomiska kartan visar att området innanför den streckade linjen består av skog och åker. Den del av åkern som faller inom 1980 års gräns, arealmäts och redovisas som urbaniserad jordbruksmark.

hellre den snabbare elektroniska planimetern. Resultaten redovisas för varje enskild tätort och länsvis.

Jag har själv testat metoden i en försöksverksamhet på Västmanlands län. Då metoden visade sig fungera helt tillfredsställande och till en rimlig kostnad, beslöt SCB våren 1981 att tillämpa den i en *landsomfattande* inventering. Detta skedde i form av ett uppdrag till rikets överlantmätarmyndigheter. Dessa levererade uppgifterna under sommaren 1981.

För att säkerställa den statistiska kvaliteten i mätresultaten genomförde vi också en stickprovskontroll på samtliga berörda län. I några län hade man av olika skäl i viss mån underskattat de urbaniserade arealerna. Genom stickprovskontrollen kunde dessa uppgifter justeras. Totalt för landet innebar kontrollen en uppjustering på ca 10% i förhållande till de inlevererade värdena.

Ytterligare några kvalitetsfrågor bör påpekas. Rent allmänt kan sägas att kvaliteten är god i den bemärkelsen att arealuppgifterna grundar sig på ett över riket enhetligt kartmaterial och på tätortsavgränsningar som är samstämmiga för hela landet.

Ett problem är dock att den i övrigt förträffliga ekonomiska kartan av naturliga

skäl hänför sig till en lång produktionsperiod. Det innebär att kartmaterialet i några län är ganska föråldrat, ca 25–30 år gammalt. Kartan kan där redovisa åkermark som kan vara nedlagd, skogsplanterad eller kan ha fått en annan användning vid undersökningsperiodens början 1975.

Västmanlandsstudien visade på en metod att med hjälp av flygbilder kompensera för denna felkälla, som i detta län uppgick till 5%. Problemet kan vara större i de län som har äldre ekonomisk karta än Västmanlands län (1962–1963), och mindre i de län där kartan utarbetats senare.

Resultaten av inventeringen har publicerats i Statistiska meddelanden J 1982: 10. Här skall bara i all korthet omtalas att totalt för riket 10 400 hektar jordbruksmark togs i anspråk för tätortsutbyggnad under perioden 1975–1980.

Statens planverk och lantbruksstyrelsen har bidragit till finansieringen av undersökningen, då den har intresse för arbetet med den fysiska riksplaneringen.

3. Tätortsexpansion på jordbruksmark 1960–1980

Intresset för den ovan beskrivna inventeringen visade sig bli stort i både massmedia och bland en miljömedveten allmänhet. Inom SCB beslöt man då att söka belysa utvecklingen även under tidigare 5-årsperioder, eftersom det fanns anledning att misstänka att utbyggnaden på åkermark haft en betydligt större omfattning före 1975, då tätortsutbyggnaden allmänt var mycket större än under den senaste perioden.

Då tätortsavgränsningar på kartor, som inledningsvis har påpekats, finns från 1960 är det fullt möjligt att genomföra en sådan undersökning och därmed få möjlighet att följa utvecklingen under en längre tidsperiod. Detta förutsätter emellertid att den ekonomiska kartan har förelegat under hela perioden. I en sammanhängande östsvensk region omfattande ett 10-tal län är detta fallet.

Vid SCB har man därför startat med arealmätningar enligt den beskrivna metoden även för perioderna 1960–1965, 1965–1970 och 1970–1975 i några av de östsvenska länen. Tekniken innebär att man också måste överföra äldre tätortsgränser till den transparenta tätortskartan, vilken sedan placeras över den ekonomiska kartans färgtryck för att möjliggöra arealmätningar av ianspråktagen åkermark under de olika 5-årsperioderna.

De första resultaten omfattande Uppsala län, Södermanlands län och Östergötlands län publicerades hösten 1982 (MA-projektet. Rapport nr 8).

4. Inventering av "Allmänt tillgänglig mark"

En likartad kartografisk metod som den nyss beskrivna har jag också utvecklat och prövat i en begränsad försöksverksamhet innan den har satts i mer reguljär produktion. Avsikten har varit att söka tillgodose de behov som finns hos främst statens naturvårdsverk (SNV) att få den för allmänheten tillgängliga marken i någon form statistiskt belyst. Denna mark kan betraktas som en naturresurs. För att kunna hävda den allemansrättsliga marken gentemot exploaterande intressen behövs kunskap om denna resurs.

Det kan givetvis inte bli fråga om en totalinventering, och en sådan skulle knappast heller vara av intresse för denna fråga. Vad det gäller är nämligen *stadsbornas* (= tätortsbornas) möjligheter att komma ut i skog och mark på fritiden. Det gäller alltså alla dem som ger sig ut på promenader, cykelturer, badutflykter, orienteringstävlingar och liknande i städernas närmaste omgiv-

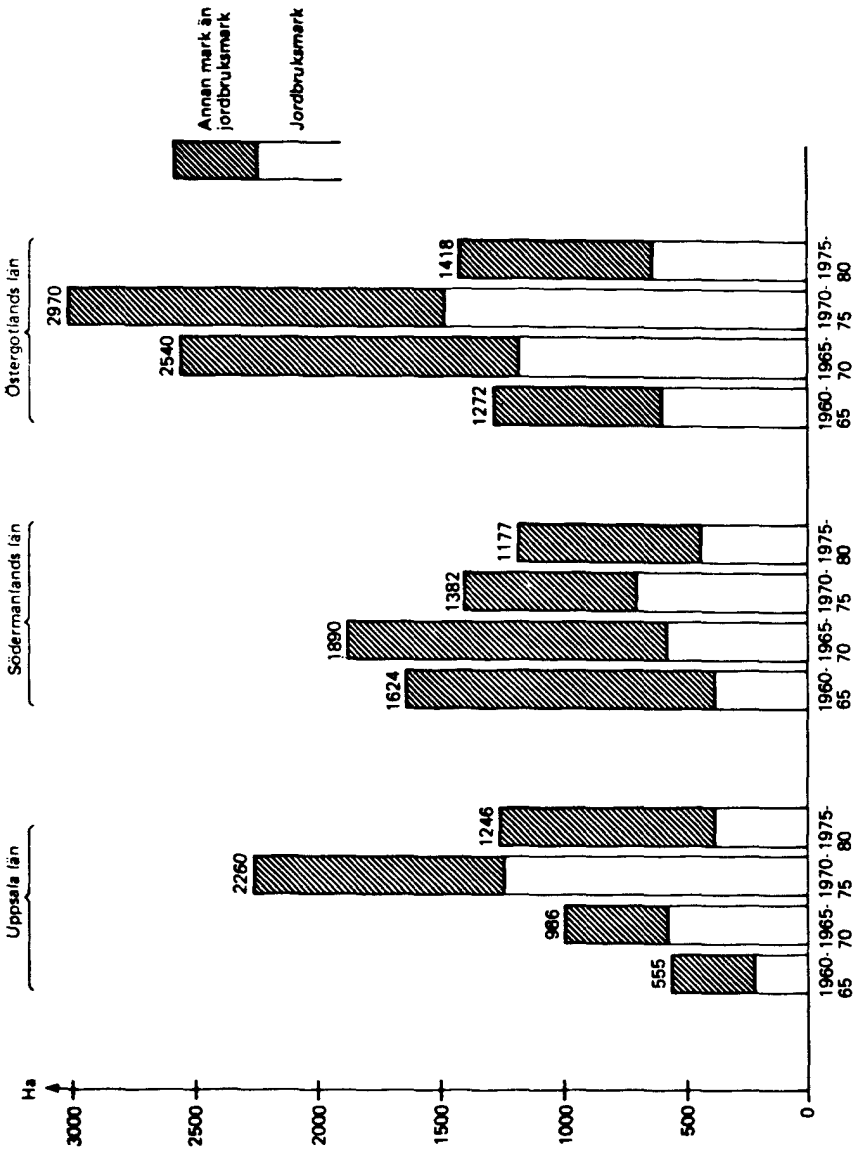


Fig. 3. Tätortstillväxten 1960–65, 1965–70, 1970–75 och 1975–80 fördelad på jordbruksmark och annan mark än jordbruksmark i Uppsala län (C), Södermanlands län (D) och Östergötlands län (E)

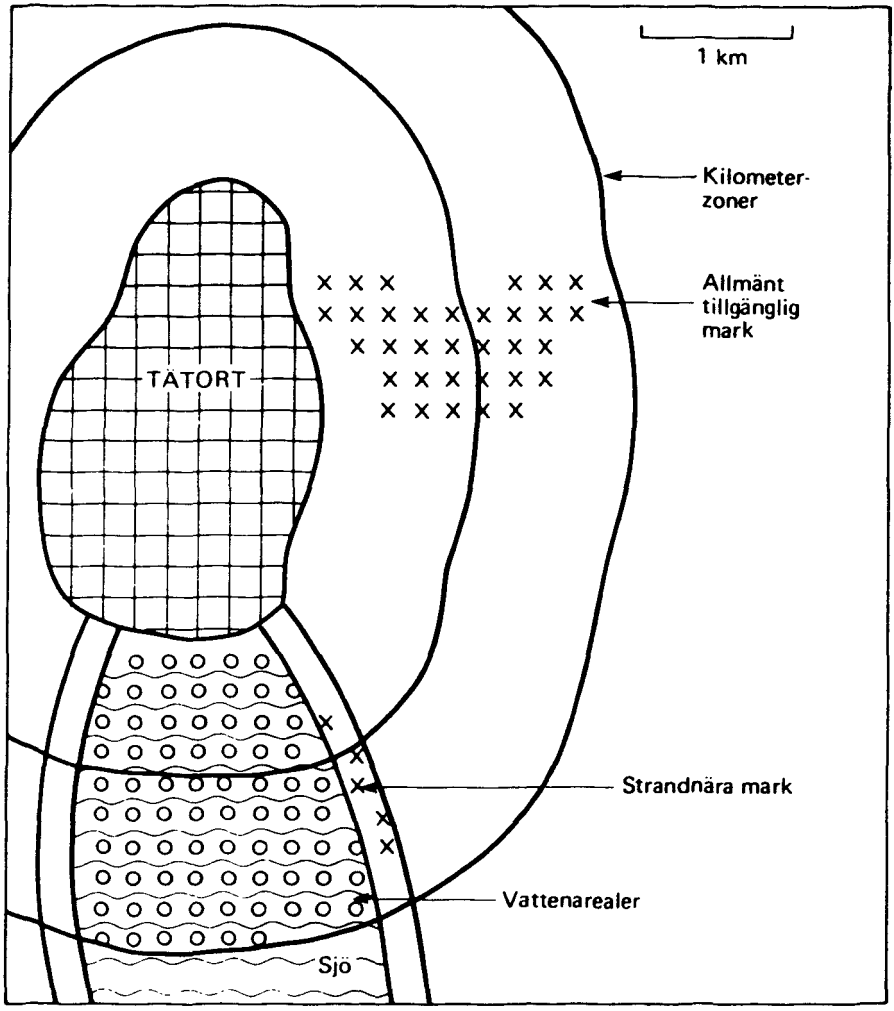


Fig. 4. Principskiss av den kartografiska inventeringsmetoden

Det rutade området är en tätort. Utanför denna har här två kilometerzoner ritats. Inom dessa representerar varje kryss, som ligger över den ekonomiska kartans gröna områden, fyra hektar allmänt tillgänglig mark. Motsvarande vattenarealer markeras av små cirklar. Arealerna summeras och redovisas i tabeller.

ningar. Dessa möjligheter växlar starkt, beroende på om staden är belägen på jordbruksslätterna eller i en annan, mer omväxlande naturtyp.

Till saken hör också att konkurrensen om marken kan vara stor i tätorternas när-

Tab. 1. Allmänt tillgängliga arealer mark och vatten i kilometerzoner runt Västerås, Enköping och Uppsala 1980, arealer i hektar

Tätort	Tätortens		Allmänt tillgängliga arealer				Allmänt tillgänglig mark per 1 000 inv.	
	befolkning	totalareal	Strandnära mark	Annan mark	Vatten	Totalt	Totalt	Därav strandnära
Västerås	97 463	4 651						
Zon 1			92	1 420	716	2 236	16	1
Zon 1–2			228	2 956	1 460	4 680	33	2
Zon 1–3			296	4 848	2 156	7 348	53	3
Enköping	18 577	934						
Zon 1			–	728	–	728	38	–
Zon 1–2			–	1 692	–	1 692	89	–
Zon 1–3			8	3 200	–	3 208	169	–
Uppsala	101 967	4 520						
Zon 1			152	1 192	320	1 664	13	1
Zon 1–2			184	2 872	520	3 576	30	2
Zon 1–3			196	4 940	760	5 896	50	2

zoner. 1977 års jordbrukspolitiska beslut om stor återhållsamhet med jordbruksmarken vid tätortsutbyggnad kan leda till större tryck på skogsmarken nära tätorten.

Begränsningen av inventeringen till de större tätorternas närzoner ger en möjlighet att utesluta hela glesbygden samt mindre tätorter. De människor som bor i sådana miljöer kan ju sägas ha praktiskt taget obegränsad tillgång till den fria naturen. Denna begränsning ger också en ingång till den kartografiska metoden i form av SCBs tätortsavgränsningar som ett utgångsmaterial.

Metoden går i korthet ut på att ett antal kilometerzoner ritas ut parallellt med tätortsgränsen på en transparent millimeterrutad film. Detta plastmaterial placeras sedan över den ekonomiska kartan i färg. Det ger möjlighet att se vilken form av markanvändning som finns i närzonerna runt tätorten, om det är jordbruksmark, någon bebyggelse eller annan mark. Vi har valt att definiera den allmänt tillgängliga marken som liktydig med den mark som är grön på den ekonomiska kartan (skogar, berg, myr- och hedmarker). Definitionen innebär en viss förenkling av verkligheten men träffar ändå ganska nära det begrepp det här är fråga om. Det är den mark som är allmänt tillgänglig under hela året.

Själva arealmätningen av de ytor som inventeras kan utföras på olika sätt: med planimeter eller med tillämpning av "punktmetoden", som är en areell sampling. Vi har valt den senare mätmetoden i denna inventering. Arbetsmaterialet i form av en millimeterindelad plast ger en enkel möjlighet att välja lämplig punkttäthet och kontrollera slumpfelen. En inventering av markanvändningen i punkter med två centimeters avstånd i ett regelbundet rutnät ger 4 hektar "infångad" mark per observation. Dessa summeras och redovisas i kilometerzoner utanför tätortsgränsen.

Förutom den allmänt tillgängliga mark som faller inom tätorternas närzoner, har det också ansetts vara av intresse att få en redovisning av de eventuella vattenarealer som ingår i dessa zoner. Skälet till detta är att vattenarealer under alla årstider har ett

värde för rekreations- och fritidsändamål. En begränsning har dock gjorts till insjöarealer.

En kategori mark som anses särskilt attraktiv för friluftslivet är den strandnära marken. Vi har därför avskilt denna mark som en särskild klass i ett bälte inom 200 meter från strandlinjen.

Metoden är medvetet enkel i sin tillämpning bl. a. för att garantera stor tolkningssäkerhet i de inventerade arealerna. Detta kan sägas vara fallet, särskilt i jämförelse med en del andra tillämpningar av fjärranalys som sett dagens ljus under senare år. En viss felkälla kan dock ligga i att kartmaterialet är av växlande ålder och aktualitet i olika län. Detta förhållande kan dock balanseras dels med att tätortsavgränsningarna är likåldriga och aktuella (1980 års gränser), dels med att en fritidshusinventering i kartform får komplettera den ekonomiska kartans bebyggelseinformation i varje län.

Yrkesfördelningen 1753–1805 enligt Tabellverket: de särskilda städerna

av fil. lic. Yngve Fritzell

Förkortningar:

M = Den svenska befolkningsstatistikens 200-åriga bestånd. Minnesskrift. Stockholm 1949. (Statistiska meddelanden. Ser. A. Bd 6: 3)
MTL = Mantalslängd
TK = Tabellkommissionen
SCB = Statistiska centralbyrån
GMR = Generalmönsterrullor
15< = 15 år och över

1. Inledning

I den fjärde artikeln[1] av min tillbakablick *över stånd och yrken har jag inte kunnat* täcka hela perioden, utan behandlar blott två år av 1700-talet och för jämförelse yrkesutövarna år 1805. Jag ger inte heller rikssiffror utan går ned till de enskilda städerna med deras yrkesgrupper. En brist är att inte alla städer har kunnat tas med, blott 68 av totalt 83. I fråga om de övriga 15 saknas uppgifter för 1751 och 1772, och de finns inte heller för något närliggande år. De enstaka data som föreligger i stadshistoriker är olikartade och ofta utan yrkesfördelning.

Redan kort efter tabellverkets tillkomst föreslogs att särskilda blanketter skulle användas för städerna. Det blev faktiskt också K. M:ts beslut år 1756, men verkställandet sköts upp till dess de blanketter som sändes ut på en gång för 25 år till kyrkorna, hade tagit slut. När därefter formulären ändrades för andra perioden (1775–1800), kunde inga stadstabeller göras, "emedan Commissionen redan haft mer arbete än som kunnat medhinnas" (Nicander 1800).

Tabellverkets upprinnelse har i vår tid utförligt skildrats främst av Hjelt och Arosenius. Här skall endast redogöras kort för dess organisation i början. Som det första statistiska ämbetsverket i världen utformades det av riksdagen under medverkan av Vetenskapsakademien. Det stadfästes år 1748. På två tryckta blanketter fyllde prästerna i församlingsvisa tabeller dels över befolkningsrörelsen, dels över folkmängden. De senare återges här på sid. 285–288 i faksimil för första och andra perioderna. Att märka är att "tredje tabellen" består av tre olika sådana, en över kön och ålder i kombination, en över kön och civilstånd och en över kön samt stånd och yrken (= "ståndstabellen"). I städerna medverkade magistraten vid ifyllandet. För de tre första åren samlades båda blanketterna in årligen, men därefter behövde folkmängdstabellen bara göras upp vart tredje år intill 1775 och sedan vart femte.

Insamlingen av materialet var delad i flera steg. Först gick tabellerna till prosten, som gjorde ett sammandrag på likalydande blankett. Prosteritabellerna sändes till konsistoriet, som författade en summatabell för stiftet. De län som helt eller delvis hörde till

[1] Föregående artiklar "Yrkesfördelningen 1840–1955 enligt Tabellverket jämfört med andra källor" (Statistisk tidskrift 1973: 5, sid. 358 ff.), "Yrkesfördelningen 1825–1835 enligt Tabellverket och andra källor" (Statistisk tidskrift 1976: 3, sid. 181 ff.) samt "Yrkesfördelningen 1805–1820 enligt Tabellverket och andra källor jämte sammandrag intill 1855" (Statistisk tidskrift 1980: 1, sid. 5 ff.).

stiftet fick därifrån sina särskilda sammandrag. Och slutligen var det länsstyrelsernas tur att göra sina tabeller och skicka dem till tabellverket som gjorde ett "generalsammandrag".

En så stark decentralisation visade sig olämplig. Det blev svårt med kontrollen att alla församlingar hade kommit med i summatabellerna. Svårare blev det, emedan rikets indelningar var högst invecklade. Somliga församlingar var "sönderstyckade på två, tre, ända till fyra prosterier – – – län och stift", och "för vidlyftighetens skull kunde inga folkmängdstabeller upprättas för varje sockendel" (Nicander 1812, cit. Lext, sid. 242).

Brister blev det förvisso i de tidiga befolkningsdata. Man måste skilja på luckor, som berodde på att områden hade uteslutits ur tabellverket, och på luckor som härrörde av att ståndstabellen tog upp bara en del av folket, ehuru ålderstabell och civilståndstabell hade fullständiga tal. I den här undersökningen är det första slaget av luckor redan nämnt: de 15 felande städerna (de har förkommit men ingår i rikssiffrorna i M). Halv lucka har några städer (10), de redovisar bara för ett av tabellåren, 1772 eller ett än senare år. För Ystad omfattar tabellen blott en av de två församlingarna. Ofullständig ståndstabell men fullständiga andra data är ingen ovanlig typ, men stånd och yrken har då kompletterats av mig på sätt som har beskrivits i varje fall.

2. Städerna i det tidiga tabellverket

Det kom att dröja till 1795, innan TK kunde räkna fram städernas slutsummor av folkmängden. Då hade man just gjort den ändringen att man till TK infordrade alla prosteritabeller. Men i saknad av stadstabeller måste TK för de flesta städer beräkna andelen av för stad och land gemensamma yrken. Det har också varit författarens problem (p. 4). Märk att inga ståndS- eller yrkesdata *publicerades* den gången. Därför har TKs tal för 1795 inte mycket att ge mig.

Meningen med den här undersökningen är främst att ge material för fortsatt forskning i de särskilda städernas historia. I andra hand bör den kunna lämna acceptabla rikssiffror för stadsyrken eller grupper därav och belysa deras utveckling. Tal för ett tidigt år var naturligtvis önskvärda. År 1751 har valts, fastän det är ytterligare tio städer som saknar data för det året. Det andra valda året, 1772, ibland ersatt av närliggande år, ger således bättre täckning. För båda åren gällde samma formulär, så att jämförelser bör bli hyggliga.

Men andra perioden är bara i undantagsfall använd, nämligen där inga siffror finns för slutet av den första. Att andra perioden i regel inte har behandlats har flera orsaker. Den tiden skulle personer med dubbla yrken redovisas dubbelt och därmed ståndstabellens slutsumma bli större än folktalet. Vidare splittrades vart stånd eller yrke i grupper efter civilstånd, (gifta, förut gifta, ogifta och barn <15) varvid ogifta 15< inte blev jämförliga med ungdomsgrupperna under första perioden. Slutligen var de flesta yrkesrubrikerna helt olika mot den tidigare periodens formulär och länstabellerna hade slopats, det togs bara in stiftstabeller till TK. Man har kallat 1770- och 1780-talen tabellverkets nedgångsperiod.

Arosenius tycks mena (sid. 41) att redan under första perioden ståndstabellen innehöll dubbelräkningar, därför att yrkestitlarna inte heller då uteslöt varandra. Men det är inte min erfarenhet vad gäller städerna. Redan 1751 stämmer ålders- och ståndstabellerna eller skiljer sig bara på enstaka individer i 70% av städerna. De andra

fallen fördelar sig lika på över- och underskott.

För jämförelser framåt i tiden har 1805 års data försökts. Från det året är dock familjerna inte längre sammanhållna som tidigare utan hopförda i stora yrkesgrupper. Det gör att blott de vuxna männen kan jämföras yrke för yrke. Man får i alla fall för yrkesutövarna en föreställning om förändringarna under ett halvt sekel – liksom för totalfolkmängden.

3. Källor till jämförelse

En stor och viktig källa till jämförelse är mantalslängden (MTL), en statsskattelängd som finns från 1600-talet. Dessa längder upprättades årligen av "mantalskommissarier". Hushållens huvudpersoner skulle infinna sig till mantalsskrivning. Prästernas medverkan var under den nu aktuella tiden begränsad.

Mantalslängderna hade inga tryckta formulär och fick därigenom rätt skiftande uppställning. Åldersbestämmelser fanns givetvis. Vid den tid det gäller var undre gränsen för skattskyldighet 15 och den övre i regel 63 år. Ännu vid 1700-talets mitt räckte det om man redovisade husbönderna med namn och deras skattskyldiga familjemedlemmar och tjänare med antal och eventuellt namn. Först 1765 kom föreskrift att även de mantalsfria skulle föras in i längden. Därtill kom att det tog tid innan bestämmelsen hade trängt in överallt. De befriade delades vanligen i tre grupper: underåriga, överåriga och "av andra orsaker befriade" (mest fattiga och sjuka men därtill adeln och dess husfolk samt soldater för egna personer men ej hustrur).

MTL, med en rad per familj, hade kolumner där folket fördes in med 1, 2 etc. personer. Sju kolumner torde ha varit det minsta före 1765: männer, hustrur, söner, döttrar, drängar, pigor och inhysesfolk utom summor. Efter sagda år kom till minst tre kolumner för befriade jämte deras summa och total.

När det gäller slutsummorna, är således jämförbarheten beroende av om de befriades antal har förts in i längden eller inte. En hel del har sagts om det i noter till särskilda städer men det skall också beröras i texten (p. 5).

Annat är det att kontrollera speciella yrken eller grupper, då måste hela längden studeras. I städerna är det regel att yrket anges för husbonden – men det är ändå många undantag, t. ex. "borgare" kan beteckna långa rader av invånare i somliga städer.

4. Stads- och landsförsamlingar

Förr var det mycket vanligt, att en stad och en landskommun utgjorde en församling (eller två men ändå) med gemensam kyrkobokföring och därmed redovisades i en folkmängdstabell. Som redan sagts har det blivit ett stort problem i min studie. Ytterligheterna kan lätt visas: Filipstad, som i sin förening med Färnebo landskommun har endast 1/6 av totala folkmängden, och Visby med 4 000 invånare emot 150 i den tillhörande Visby landskommun.

Man får ett begrepp om problemet när man hör att nära hälften av mitt material räknat i städer har varit av det kombinerade slaget, 29 städer av 68. Flertalet var mindre städer, men det stora Falun och de relativt stora Nyköping, Växjö och Örebro var förenade med betydliga landskommuner och måste brytas ut. Vilka de övriga är visas av noter märkta med asterisk.

För varje kombinerad stads- och landstabell var det uppenbart att två mantalslängder fanns till jämförelse, vilkas slutsummor borde tillhopa överensstämma med

ståndstabellens summor. Men för de särskilda yrkena blev en prövning mot MTL nödvändig. Det gällde att i grupper gemensamma för land och stad komma åt de personer som tillhörde staden. Problemet var störst där de båda delarna var ungefär jämnstarka.

Ståndspersonerna, i och för sig en diffus grupp, kunde t. ex. antas tillhöra staden till en större del än den som gällde befolkningen i övrigt. Men en del ståndspersoner (adeln kanske) borde övervägande tillhöra landet. Jag försökte ibland kollationera ståndspersonerna i MTL med siffrorna i ståndstabellen. Så kunde inte ske systematiskt, det skulle ha tagit för lång tid och ändå bli osäkert. Även om man kunde placera huvudpersonerna på rätt sida om stadsgränsen, så måste man ge flertalet sällskap av hustrur och ett lämpligt antal barn och ungdomar. Dessa senare grupper har därför inte något större värde i städer med bortopererad landsdel.

Gruppen fattiga, lytta m. fl. har i de kombinerade ståndstabellerna i allmänhet lagts ut på land och stad efter folkmängden.

För två norrlandsstäder kunde separata data göras endast genom utdrag ur hela prosteritabeller. Dessa behövdes för att undvika alltför stora luckor i norr, och man måste ändå ersätta 1772 med något av de sista åren av århundradet.

5. Mantalslängderna

Det har skrivits mycket om mantalslängderna, inte minst om graden av fullständighet i deras folktal. Inte så mycket av detta har handlat om en direkt konfrontation av MTL och husförhörslängderna i de enskilda *församlingarna*. Vid summering för större områden kommer alltid överskott och underskott att till en del uppväga varandra. När Lext för olika år i slutet av 1700-talet jämför de båda källorna, räknar han med länssiffror och får då bara en gång av nio större avvikelser än 2%. De flesta år har MTL högre tal än tabellverket – även här för 1772, se nedan.

Det finns ett material som gäller århundradets mitt och samtliga svenska städers befolkning.[1] Uppenbarligen är det grundat på mantalslängderna för 1747. Men som ovan sagts var MTL då ännu ofullständig på det sätt att endast de skattskyldiga var införda där. För de befriade har materialet för den skull bara – som Almqvist säger – högst ungefärliga tal som i vissa fall inte är rimliga. Själv finner jag det oklart om de "av andra orsaker" befriade är inkluderade i folktalen. Jag har här inskränkt mig till att undersöka, huruvida mina stadsbefolkningar för 1751 över- eller understiger summorna i 1747 års material. Det visar sig att det förra fallet alldeles dominerar: överskott i ståndstabellen i 47 fall. Resten består av några fall där båda talen är lika men mest av städer där jag har luckor. Mellantiden av 4 år kan omöjligt ha nämnvärd inverkan på siffrorna. Hur mycket som beror på att befriade "av andra orsaker" utelämnats har inte undersökts systematiskt.

Bättre syntes det vara att jämföra mitt stadsmaterial för 1772 och däromkring med MTL för så närliggande år som möjligt. Riksarkivets bestånd har anlitats, men där arkivbanden är skadade är de ofta inte tillgängliga. Någon generell undersökning rörande städerna känner jag inte utöver den av 1747. Resultatet har nu blivit tab. 1 A. Undantagna är Stockholm och Göteborg som ansetts grundligt kontrollerade ändå. Av kolumnen för MTL framgår vilka städer som angav enbart mantalet, dvs. de skattskyldiga, och vilka som upptog även de be-

[1] Almqvist, D., Tillståndet i Sveriges städer 1747.

Tab. 1 A. Tabellverket och mantalslängden i 66 städer omkring 1772
Anm. År för tabellverket är 1772 där ej annat anges. Stockholm och Göteborg uteslutna.

År för Tab. v.	MTL		Tabell-verket	MTL o. beräk-nat mantals-fria	Differens +	Differens −	% +	% −	Anmärkn.
	71	Sigtuna	517	529	12		2		
	81	Norrtälje	660	719	59		9		
	73	Vaxholm	758	824	66		9		
73	74	Uppsala	3926	2768 989		169		4	Akademien
	74	Enköping	1152	1158	6		1		
75–80	80	Nyköping	2456	2525	69		3		
	75	Mariefred	428	406		22		5	
	75	Strängnäs	1139	1031		108		10	
	73	Torshälla	608	579		29		5	
75	75	Trosa	499	512	13		3		
	71	Linköping	2301	2349	48		2		
	71	Norrköping	7721	8145	424		6		
	71	Söderköping	913	900		13		1	
	71	Skänninge	687	712	25		4		
	71	Vadstena	1408	1407		1		0	
	73	Jönköping	2860	1497 1214		149		5	
	72	Eksjö	883	336 393		154		17	
	73	Gränna	460	229 181		50		11	
90	84	Växjö	1322	763 550		9		1	
	77	Kalmar	3465	3554	89		2		
	77	Vimmerby	878	933	55		6		
	77	Västervik	2388	2510	122		5		
80	80	Visby	4156	4370	214		5		
66	66	Karlskrona	10200	10028		172		2	
	75	Karlshamn	2534	1683 1181	330		13		
73	75	Sölvesborg	373	248 178	53		14		
	86	Simrishamn	623	589		34		5	
	95	Ängelholm	584	563		21		4	
73–80	71	Malmö	3796	2439 1689	332		9		
66	71	Lund	1763	905 946	88		5		
80	76	Helsingborg	1308	1381	73		6		
	71	Ystad	1797 567	2338		26		1	S:ta Maria S:t Per
	71	Skanör/ Falsterbo	579	601	22		4		
73	84	Halmstad	1099	1208	109		10		
	84	Laholm	865	853		12		1	
	84	Falkenberg	527	602	75		14		
63–72	84	Varberg	1440	1491	51		3		
	84	Kungsbacka	329	423	94		29		Hög ökning 10 år?
	72	Kungälv	1032	1047	15		1		Se noterna
	72	Marstrand	932	987	55		6		
	72	Strömstad	481	620	139		29		Se noterna
	74	Borås	1995	1846		149		7	

(forts.)

Tab. 1 A (forts). Tabellverket och mantalslängden i 66 städer omkring 1772

År för Tab. v.	MTL		Tabell-verket	MTL o. beräk-nat mantals-fria	Differens +	Differens −	% +	% −	Anmärkn.
	74	Alingsås	1 092	848		244		22	Se noterna
	74	Åmål	732	677		55		8	
	74	Ulricehamn	462	410		52		11	
	73	Mariestad	895	875		20		2	
	72	Skara	846	786		60		7	
	73	Skövde	521	418		103		20	
	73	Karlstad	1 501	1 079	103		7		
				525					
75	73	Kristinehamn	1 174	634		36		3	
				504					
	73	Filipstad	476	358	37		8		
				155					
73	73	Örebro	2 366	1 418		158		7	
				790					
	72	Nora	644	406		40		6	
				198					
69	73	Västerås	2 543	2 550	7		0		
69	73	Köping	1 319	1 369	50		4		
	73	Sala	1 805	1 314	165		9		
				656					
	72	Falun	5 221	3 013	92		2		
				2 300					
	73	Hedemora	996	720	83		8		
				359					
	73	Säter	539	539		0		0	
	73	Gävle	3 757	2 716	514		14		
				1 555					
	73	Söderhamn	1 464	872		4		0	
				588					
	73	Hudiksvall	1 092	707	30		3		
				415					
	72	Härnösand	1 470	1 094	258		18		
				634					
95	94	Umeå	936	564		21		2	
				351					
80	94	Luleå	730	370		96		13	
				264					
90	94	Piteå	820	470		8		1	
				342					
		Summa	107 810	109 772	3 977	2 015			
					+1 962				

TAB. III.

NÖDIG UNDERRÅTTELSE

ÅR 1751 Om Mastrand Pl:s FÖRSAMLINGS *Invånares* Tilväxt *uti* Göteborgs STIFT, *genom hälsosamma inrätningar, eller* Aftagande *genom* Sjukdomar, Hunger *och* Vanskörsel, *samt brist på nyttige näringsfång, befunnen efter följande fördelning af* Ålder, Kön *och* Giftermål, *samt* Stånd *och* Omständigheter.

ÅLDER.

	MAN-KÖN.	QVIN-KÖN.
Yngre		
a. *åo* 1. *år* -	6	7
Emellan		
b. 1. *och* 3. -	8	12
c. 3. *och* 9. -	2	15
d. 5. *och* 10. -	17	21
e. 10. *och* 15.	17	25
f. 15. *och* 20.	11	30
g. 20. *och* 25.	13	25
h. 25. *och* 30.	16	29
i. 30. *och* 35.	16	16
k. 35. *och* 40.	5	11
l. 40. *och* 45.	12	15
m. 45. *och* 50.	8	15
n. 50. *och* 55.	14	19
o. 55. *och* 60.	5	17
p. 60. *och* 65.	8	15
q. 65. *och* 70.	2	14
r. 70. *och* 75.	3	9
s. 75. *och* 80.	1	1
ff. 80. *och* 85.		1
t. 85. *och* 90.	1	1
u. *öfver* 90. *år.*		
Summa eff. Könen	172	295
Summa	470	

STÅND.

	MAN-KÖN.	QVIN-KÖN.
1. a. *Ridderskap och Adel* - - - - -		
b. *Ungdom öfver* 15. *år.* - - - - -		
c. *Barn under* 15. *år* - - - - -		
2. a. *Prästerskap, Academie- och Scholæ-Stater*	4	4
b. *Ungdom öfver* 15. *år* - - - - -		
c. *Barn under* 15. *år* - - - - -	2	3
3. a. *Civila - Personer och deras underlydare* -	7	9
b. *Crono-Betjänter, Fogdar, Länsmän &c.*		
c. *Ungdom öfver* 15. *år* - - - - -		6
d. *Barn under* 15. *år* - - - - -	6	7
4. a. *Stadsens beskyddare Betjänter* - -		
b. *Dito Lagmän och Tjenstefolk öfver* 25. *år* - - - - - - - -		
c. *Dito under* 25. *år* - - - - -		
5. a. *Magistrat i Städerne* - - - - -	5	6
b. *Grefvar* - - - - - - - -		
c. *Rådmän* - - - - - - - -	4	3
d. *Mäklare och Stad-Rådmän* - - -		
e. *Borgerskapet och Ungdom öfver* 15. *år* -	5	12
f. *Barn under* 15. *år* - - - - -	5	9
6. a. *Fabriqueurer* - - - - - - -		
b. *Ungdom och Arbetare öfver* 15. *år* -		
c. *Barn under* 15. *år* - - - - -		
7. a. *Ämbets- och Handtwärks-folk* - -	11	12
b. *Gesäller* - - - - - - - -		
c. *Lärogåssar öfver* 15. *år* - - - -		
d. *Barn under* 15. *år* - - - - -		
8. *Ringare Borgerskap, Förswars- och Förderskarlar* - - - - - - - -	37	53
9. *Rådtens, Staters, Kyrko- och Stads-Betjänter*	5	6
10. a. *Fristat tahfast folk i Städerne* - -		2
b. *Dito Bröstligt, doch ej Hospitals-Hjon*	1	15
11. *Resande och Främlingar* - - - -		
12. *Skeppare och Sjöfarit folk* - - - -	12	10
13. a. *Handtwärkares och ringare Borgerskapets, samt Betjänters Barn, öfver* 15 *år* -	14	29
b. *Dito under* 15 *år* - - - - - -	34	57
14. a. *Samtelige Borgerskapets och Betjänters Tjenstefolk, öfver* 25 *år* - - - -		30
b. *Dito under* 25 *år* - - - - - -	2	4
15. *Större och mindre Seminarier på Landet*		
16. a. *Tilgifte med utsäde* - - - - -		
b. *Tilgifte utan utsäde* - - - - -		
17. a. *Fristat tahfast- och Gatuhus-folk på Landet*		
b. *Dito Bröstligt, doch ej Almoss-Hjon* -		
18. a. *Gårdningsmän och Sahm-Handtwärkare*		
b. *Andre Handtwärkare på Landet* -		
19. *Rytt:re, Dragoner, Soldater och Båtsmän*	3	3
20. *Strandsittiare och fiskerit frälse folk, doch ej Hemmansbrukare* - - - -	4	5
21. a. *Handtwärkare vid Bergwärken* - -		
b. *Brukisfolk dito* - - - - - -		
22. *Mälnare* - - - - - - - -	1	1
23. a. *Ogift Almoges Barn och Tjenstefolk, öfver* 15. *år* - - - - -		
b. *Dito under* 15. *år.* - - - - -	5	4
24. *Fattige Präst-Änkor* - - - - -		
25. *Gemenskt afflidne Krigsfolk utom Krigsmanskap* - - - - - - - -	4	2
26. a. *Wärkeligen intagne i Hospitaler* - -	1	6
b. *Wärkeligen intagne i Fattighus* - -		
27. a. *Blindige utom Hospitaler till, Fattighus*	1	1
b. *Gubar och urfvinige utan Hospital* -		
28. *Bröstlige af fallande fot och sinnessvaga sinnen* - - - - - - -		
29. *Lägrade Qwinfolk* - - - - -		2
30. *Rasp- och Spinhusfolk* - - - -		
31. *Lifstagar på Fästningar och Slott* -		
32. *Främmande Religions-Förwanter* - -		

GIFTERMÅL.

	MAN-KÖN.	QVIN-KÖN.
A. *Gifte Folk* -	85	85
B. *Änklingar och Änkor* - -	6	48
C. *Ogifte öfver* 15. *år* - -	31	80
D. *Barn under* 15. *år* - -	50	85
Summa eff. Könen	172	298
Summa	470	

OMSTÄNDIGHETER.

I. Hushåll *eller* Matlag *i Städerne, undantagne the Gemene af Krigsfolk, alla utfattige och eländige* - - - - -	100
II. Caffehus *i Städerne* - - - - -	
III. Källare *i Städerne* - - - - -	1
IV. Krogar *i Städerne* - - - - -	17
V. Hushåll, Matlag *eller* Räkstal *på Landet, undantagne the Gemene af Krigsfolk, alla utfattige och eländige* - - -	
VI. Gästgifverier *på Landet* - - - -	
VII. Krogar *på Landet* - - - - -	

BRUK.

I. Alla theſſe Summor inſlåtjas ej här med ſiffror af PASTOR, för än vid årets ſlut, ſedan the förut af honom blifvit adderade efter hoſgående Methodiſke Tabell.

II. PROBSTEN öfver Contraſtet, eller HÄRADS-SKRIFVAREN, adderar på thenna ſtadje Tabell alla theſſe ſin Probſterics Förſamlingars ſkrifvilte Siffre-Summor, uti hvarje ruta af Probſte-Tabellen; Utaf hvilken et ark inſändes årligen, i Januarii Månads ſlut, til Conſiſtorium, til et annat lika lydande blifver qvar hos Probſten, inhäftat uti theſs Tabell-Bok för Probſteriet, ſom ihnnas til Succeſſoren efter hans död.

III. CONSISTORIUM adderar ſåven, ruta från ruta, alla Probſtarnes Rute-Siffror på thenne ſin Tabell, ſom årligen, vid Februarii Månads ſlut, ſkall inſändes til alla i Stiftet varande Landshöfdingar, öfver the under theras Län boſtrande Kyrkor, ſedan Conſiſtorium aſtragit en lika lydande Tabell til ſin underrättelſe.

IV. LANDSHÖFDINGE-ÄMBETET förſörer härmed på lika ſätt, ſamt årligen inſänder ſin Tabell öfver hela Länet til Wälb. CANTELLIE-COLLEGIUM; men alla Kyrkornes årlige Tabeller inhäftas på en gång, ſåſom andre Curor, Kyrkliſva hos Paſtor, och inventeras vid Kyrkoråſe.

STOCKHOLM, Tryckt hos LARS SALVIUS

TAB. III.

År 1790. *FÖRTEKNING efter Ålder, Kön och Giftermål på alla i Wexiö Stads- och Lands-Församling varande Personer, jemte några Hushålls-omständigheter.*

FOLKHOPENS FÖRDELNING EFTER ÅLDREN.

ÅLDREN.	MANKÖN.	QUINKÖN.	ÅLDREN.	MANKÖN.	QUINKÖN.
a. *Yngre än 1 år*	28.	18.	n. 45 *och* 50.	61.	69.
b. *Emn 1 och 3 år*	60.	79.	o. 50 *och* 55.	75.	49.
c. 3 *och* 5.	45.	48.	p. 55 *och* 60.	75.	50.
d. 5 *och* 10.	118.	105.	q. 60 *och* 65.	19.	41.
e. 10 *och* 15.	127.	106.	r. 65 *och* 70.	94.	47.
f. 15 *och* 20.	95.	94.	s. 70 *och* 75.	27.	19.
g. 20 *och* 25.	91.	195.	t. 75 *och* 80.	5.	12.
h. 25 *och* 30.	81.	116.	u. 80 *och* 85.	4.	5.
i. 30 *och* 35.	68.	129.	x. 85 *och* 90.	1	1.
k. 35 *och* 40.	81.	102.	y. *öfver* 90.		
l. 40 *och* 45.	77.	77.	*Summa efter Kön.*	1074	1941.

Summa 3015.

GIFTERMÅL.	MANKÖN.	QVINKÖN.
A. *Gift Folk.*	760.	760.
B. *Änkling och Änkor.*	35.	136.
C. *Ogifta öfver 15 år.*	705.	429.
D. *Barn under 15 år.*	774.	716.
Summa efter Könen.	1074.	1941.
Summa Summarum.	27	15.

OMSTÄNDIGHETER.	
I. *Hushåll eller Matlag i Städerne undantagne gemene af krigs-folket samt alle utfattige och eländige.*	179.
II. *Caffehus i Städerne.*	1.
III. *Källare i Städerne.*	1.
IV. *Krogar i Städerne.*	15.
V. *Hushåll, Matlag, eller Röktal på landet, undantagne de gemene af krigs-folket samt alla utfattige och eländige.*	177.
VI. *Gästgifvarier på Landet.*	
VII. *Krogar på Landet.*	

ANMÄRKNINGAR.

I:mo Tab, III och IV författas på lika sätt, som tilförene Tab. III; Och som desse tvänne Tabeller, hädanefter icke oftare författas, än hvart femte år, så at de, två gånger i hvart decennio komma at upprättas; Så, på det dessas decennier måtte inträffa med decennierne i den allmänna tidräkningen; Så komma dessa tvänne Tabeller at för första gången upprättas år 1775, andra gången år 1780 och så vidare.

II:do För hvar Titel i Tabell IV, anföras Hustrur och Barn, Gift och Ogift Folk &c., som höra til samma Titel; Hvarvid, då förekomma skulle, at en och samma Person hörer til flera Titlar, såsom at en Ämbetsman är tillika Bruksidkare, en Bonde är tillika Fiskare och så vidare, så uptagas sådana Personer vid hvar Titel, til hvilka deras göromål höra, så at en och samma Person, kan flere gånger varda i denne Tabell inräknad, alt efter mångfaldigheten af des göromål, hälft Summorne i denne Tabell icke komma at sammanläggas til en generell Summa.

IV.

Personer, efter deras Stånd, Vilkor och andra lefnads Omständigheter.

A.	Gifte. Man-Kön.	Gifte. Quin-Kön.	Änklingar och Änkor. Man-Kön.	Änklingar och Änkor. Quin-Kön.	Ogifte öfver 15 år. Man-Kön.	Ogifte öfver 15 år. Quin-Kön.	Barn under 15 år. Man-Kön.	Barn under 15 år. Quin-Kön.
a. Kronones i tienst varande högre och lägre Civile Ämbetsmän i Städerne utom Clereciet.	15.	15.	2	11	6.	8.	18	26.
b. Dito Militaire Ämbetsmän i Städerne.	1	1.	—	—	2	—	—	1.
c. Magistrats-personer och Städernes hederligare Betiening såsom Secreterare, Notarier, Kamererare, Kämnärer, Assistenter, &c.	7.	7.	1	14	—	7	5.	2
d. Konstnärer som icke äro Borgare.	—	—	—	—	—	—	—	—
e. Studenter; Diäknar och Scholæ-Gossar i Städerne och på Landet.	—	—	—	—	7.	—	19	—
f. Rätters, Staters och öfrige publique Värkens i Städerne och på Landet ringare Betiente, såsom Vaktmästare, Pedeller, Klockare, Kammardrängar &c.	11.	11	1.	7.	4.	2.	12	6.
ff. Under-Officerare, Volontairer, Dragoner, Soldater, Båtsmän och Militaire-Reserve-karlar i Städerne.	5.	5.	—	1.	4.	—	6.	2
g. Kronones i tienst varande högre och lägre Civile Ämbetsmän på Landsbygden, utom Clereciet.	6.	6.	2	—	2.	2	4.	5.
h. Militaire Ämbetsmän på Landsbygden.	—	—	—	—	2	—	—	—
i. Bruksidkare ehvad slags Bergsbruk de drifva.	—	—	—	—	—	—	—	—
k. Hos private Personer i tienst varande hederligare Betiente i Städerne och på Landet, såsom Directeurer, Kamererare, Inspectorer, Gårdsfogdar, Bokhållare &c.	—	—	—	—	—	—	—	—
l. Ringare Betiente vid Bergslagerne, Bruks- samt Grufve-folk, eller alla ringare Personer som vid Bergvärken hafva at syssla.	—	—	—	—	—	—	—	—
m. Jägare, Skyttar, Skogsvaktare, Kronans och Privates.	—	—	—	—	—	—	—	—
n. Hushållerskor, Hus- och Kammar-Jungfrur, Hofmästare, Kammartienare, Laquaier, Kuskar, Löpare &c. i Städer och på Landet.	—	—	2	8.	11.	2	—	—
o. Handtvärkare och Gärningsmän på Landet, utan åtskilnad.	5.	3.	1.	2	1	1.	2	—
p. Fiskare och Skärkarlar.	1.	1	—	—	2	1	1.	—
q. Saltpetersiudare, Pärlefiskare, Dykerie-betiente, Lotsar, Färgkarlar, Sågare, Mölnare, Kolare och Trägårdsmästare &c. på Landsbygden.	4.	4.	1	—	1	—	2	2
r. Under-Officerare, Volontairer, Ryttare, Dragoner, Soldater, Båtsmän, Vargärningsmän och Militair Reserve-karlar på Landsbygden.	17.	17	—	—	7.	2	20.	14.
B.								
a. Lytte och bräckliga, af alla Stånd i Städer och på Landet, såsom galna, blinda, döfva, dumba, som til arbete äro oskickeliga, eller af ålder vanföra.	4.	4	5.	15.	2	17.	1	1
b. Lytta och bräckliga i Städer och på Landet, samt gamle Personer, som kunna göra något arbete.	5.	3	—	11.	3.	11.	—	—
c. Lytta och bräckliga, samt fattiga och nödlidande i Städer och på Landet, som äro intagne i Hospitaler och fattighus.	—	—	3.	17.	6.	11.	1.	2.
d. Rasp- och Spinhusfolk, samt fångar på Slott och Fästningar.	—	—	—	—	—	—	—	—
C.								
På de fem sista åren til och från Stiften, Städerne, Socknarne, och Probsterierne. — a. Inflyttande Främlingar.	1	1	—	—	1.	—	—	—
b. Utflyttande Främlingar.	—	—	—	—	1.	—	—	—

År 1790. TAB.

FÖRTEKNING på de i Wexio Stad och Socken varande

	Gifte.		Anklingar och Ankor.		Ogifte öfver 15 år.		Barn under 15 år.	
1:mo.	Man-Kön.	Quin-Kön.	Man-Kön.	Quin-Kön.	Man-Kön.	Quin-Kön.	Man-Kön.	Quin-Kön.
1:mo. Samtelige Ridderſkapet och Adelen utan åtſkilnad af Grefl. Friherliga och Adeliga värdigheten.	2	2	1	6	4	5	6	7
2:do. Ridderſkapet och Adelen ſom drifver Bergsbruk, Handel, Fabriquer eller i allmänhet Ståndsmanna näringar.	1	1	1	—	—	—	—	—
3:tio. Ridderſkapet och Adelen, ſom idkar Landtbruk eller i allmänhet Landtmanna näringar.	1	1	1	—	—	—	—	—
2:do.								
1:mo. Präſterſkapet i Ståderne.	1	1	1	5	2	1	4	7
2:do. Präſterſkapet på Landet.	—	—	—	—	—	—	—	—
3:tio. Conſiſtoriernes Ledamöter, ſamt Ämbetsmän vid Scholarne.	10	10	1	7	10	9	7	4
3:tio.								
1:o. Borgerſkapet ſom handlar i Groſs.	6	6	—	2	2	6	9	10
2:o. Minut-handlande Borgerſkapet.	6	6	1	2	2	12	6	7
3:o. Borgerſkapets öfrige handlande Ledamöter.	24	24	6	11	4	14	28	18
4:o. Fabriques Idkare i allmänhet.	62	62	1	2	12	20	71	61
5:o. Af ſig komne Ledamöter af Borgerſkapet.	4	4	—	—	49	—	—	—
6:o. Skrå- Ämbets- och Handtwärksmän.	—	—	—	—	51	—	14	—
7:o. Geſäller i Ståderne.	7	7	—	1	—	—	1	2
8:o. Lärogoſſar i Ståderne.	2	2	—	1	1	—	—	2
9:o. Ämbetsmän och Ståndsperſoner, ſom idka Borgare-näring.					4	—	—	—
10:o. Privilegierade Handtwärkare i Ståderne.	—	—	—	—	—	—	—	—
11:o. Föregående Handlandes, Fabriqueurers och Ståndsperſoners Contoirs- Bod- Fabriques- och allmänhet all deras Handels Betiening.	—	—	—	—	—	—	—	—
12:o. Spinneri-Idkare och dervid Arbetande på Landsbygd.	—	—	—	—	—	—	—	—
13:o. Skeppare och Siömän i Ståderne.	9	9	—	2	—	4	4	4
14:o. Förſvars och Fördelskarlar i Ståderne.	12	12	1	5	—	18	18	22
15:o. Friſkt och Arbetsfördt Inhyſes-folk, eller frie Arbetare i Ståderne.	2	2	—	—	74	155	1	9
16:o. Samtelige yngre och äldre Tienſtefolket och Legohionen i Ståderne.								
4:to.								
1:o. Bönder utan åtſkilnad af Hemmanens natur, och ehvad de bruka egen eller andras jord.	52	52	7	7	2	—	52	50
2:o. Bönder ſom jämte Landtbruk idka andra näringar, ſåſom Fiſke, Handtwärk, Stenhuggerie &c.	—	—	—	—	1	—	—	—
3:o. Ståndsperſoner på Landsbygden utom Adelen och Clericiet, ſom idka Landtbruk.	—	—	1	—	—	1	—	—
4:o. Nybyggare, ehuru ſtor eller liten jord de bebo.	—	—	—	—	—	—	—	—
5:o. Torpare, af hvad beſkaffenhet torpen vara må.	48	48	1		9	6	78	77
6:o. Arbetslöſe Backſtugu-Inhyſes-folk och Gatuhusmän &c. eller Frie-Arbetare å Landsbygden, ſom hafva egit bo.	47	47	7	35	8	32	27	27
7:o. Tienſtehionen hos menige Allmogen, hos Torpare och Inhyſes-folk, ſamt hos Ståndsperſoner af alla på Landet boende Ständ.	2	2	1	—	67	72	1	—
8:vo. Gammalt Husbondfolk på Landsbygden, ſom på ålderdomen ſitter hos ſina barn eller andra.	5	5	1	1	—	—	—	—

friade i längden. Den förra gruppen har nämligen två tal för varje stad: det angivna mantalet överst och mitt beräknade tillägg för de fria därunder, dessa städer är 24. De andra 42 har bara en siffra för MTL eftersom längden innehöll både skattskyldiga och andra. Tilläggen för de mantalsfria har gjorts på följande sätt. Antalet barn <15 har hämtats från ståndstabellen. De överåriga och de "av andra orsaker" befriade (se p. 3) har gemensamt skattats till samma andel som dessa kategorier har i de städer där de författningsenligt redovisas, nämligen 41% av *barnantalet*. Tillägget är således schematiskt bestämt. Men jag fann att den intressanta differensen mellan ståndstabell och MTL härigenom minskades i tretton fall, ökades i sex och blev nästan oförändrad i två fall.

Som förut sagts bör jämförelse inte göras mellan slutsummorna för gruppen. Avvikelser uppåt och nedåt har samma intresse och bör adderas. Slutsummorna divergerar med blott 1,8% av tabellverkets summa. Numeriska värdet av differenserna är däremot 5,6% eller nära 6000. Med två tredjedelar därav har mantalslängderna haft överskott i förhållande till tabellverket, eljest underskott.

Om man söker den inverkan som de schematiska tilläggen för överåriga och "andra orsaker" har, så finner man i städer med mina tillägg och ca 39000 invånare en avvikelse netto av 3,5% och brutto (numeriska värdet) av 7,3%. Den andra gruppen – städer med full uppgift om befrielser – har de motsvarande värdena 0,9% netto och 4,6% brutto. Den skillnaden i precision är inte mer än naturlig.

6. Kalkyl för felande städer

För att grovt beräkna folkmängden 1751 i de 15 städer som saknar "tredje tabellen" både då och 1772 har jag gått två vägar och skall jämföra resultaten. Städerna är Södertälje, Östhammar, Öregrund, Eskilstuna, Kristianstad, Landskrona, Uddevalla, Vänersborg, Falköping, Hjo, Lidköping, Lindesberg, Askersund, Arboga och Sundsvall. I det ena fallet har jag antagit att folkökningen i medeltal för första och andra perioderna var densamma som för de "kända" städerna. Enligt tab. 1 B och 1 C är den procentuella ökningen 12,0% och 10,5% respektive, motsvarande 23,8% för hela tiden. Folktalet i de 15 städerna var år 1805 tillsammans 22687. Vid en härpå grundad tillbakaräkning kommer man till en folkmängd omkring 1751 av 20000. Härvid har gjorts en detaljundersökning för Eskilstuna. Skälet var att dess år 1771 grundade "fristad" med skattefrihet för järnmanufaktur m.m. utvecklades mycket raskt och gjorde folkökningen särdeles hög under andra perioden.

Den andra metoden att fylla luckan för 1751 utgår från mantalslängdernas slutsummor. Som förut visats innehåller MTL denna tid enbart de skattskyldiga. Summa summarum är 8823. Nu antar jag att de befriade är proportionellt lika många som i de kända städerna. Proportionen visar sig mycket stabil, spridningen är liten, utom att några småstäder har höga tal för de mantalsfria. Materialet för 1772 – frånräknat Stockholm och Göteborg – visar medeltalet 66% för kvoten mantalsfria/skattskyldiga. Jag antar vidare att även 1751 den kvoten var nära lika. Ett visst stöd har jag av Almqvists data för 1747 (p. 5), ehuru de befriade där förmodas vara högst ungefärligt skattade. Motsvarande kvot hos Almqvist – med avdrag utom av de båda stora jämväl av Falun, som anger ett orimligt högt barnantal – utgör 70%.

Den första kalkylen, där tillväxten an-

tages ligga på städernas medelnivå, torde bli särskilt osäker. Bland växande städer märks några på västkusten (sillfisket) och Ulricehamn (knallehandeln) men likaså stora Gävle och Norrköping, bland krympande finns Borås och Falun (se noterna). Men i kalkyl 2 där man utgår från mantalet är stadens större eller mindre tillväxt inbakad i metoden. I valet mellan de två kalkylerna måste jag föredra den andra och sätter kvoten vid 70%. Den ger folktalet 15000 för 1751.

En viktig studie av förhållandet mellan mantalsskriven och total folkmängd har gjorts av D. Hannerberg. Han finner för Örebro-Värmlands län 1751 den kvoten vara 47,5%. Han jämför alltså de skattskyldiga med hela folktalet, men jag de befriade med de skattskyldiga. Att märka är att städerna har knappt 4% av länets folktal mot 10% i riket. Mitt resultat omräknat till Hannerbergs kvot blir 59% och tyder således på att förhållandevis väsentligt färre mantalsfria fanns i städerna än på landet. Mesta delen av skillnaden förklaras helt enkelt av det högre barnantalet på landsbygden. Ett stickprov i samma län gällande 1815 antyder att ca två tredjedelar av underskottet i stad av mantalsfria beror på den mindre barnprocenten.

Folktalet för 1751 skall också beräknas för de tio städer som har data för blott 1772, omkring, och 1805. De är Eksjö, Växjö, Visby, Malmö, Helsingborg, Varberg, Kristinehamn, Umeå, Piteå och Luleå. Här ger kalkyl 2 ca 13100. Jag har förut sagt att MTL inte alltid har varit tillgänglig just för mina tabellår. Sådana avvikelser gäller likaväl för några av de tio. Men jag har prövat ett par fall och funnit att tidsförskjutningarna har helt liten inverkan på folktalet.

Det återstår att för de 15 kalkylera folktalen för det mellersta tabellåret 1772. Nio av dem har fullständig redovisning i MTL med "min" kvot = 87. (Kristianstads data har måst hämtas ur 1786 års MTL som visar en militärstad med en på tre tillhörande krigsstaten och kvoten 123!) För de övriga, små utom Landskrona och Sundsvall, sätter jag kvoten till 70. Summan av de 15 blir för det mellersta året 20000.

Den totala svenska stadsbefolkningen skulle enligt mina beräkningar vid tabellverkets början ha legat på omkring 181000 (152900 + 15000 + 13100). Det är 10,0% av hela folket och 0,2% mer än vid den förut tidigaste räkningen år 1800. Städernas andel ligger år 1772 ännu något högre eller vid 207900 = 10,2% (187900 + 20000). Den nivån nåddes på nytt först under 1850-talet. Eftersom jag har sökt komplettera städerna för luckor, är det naturligt att talen jämförs med de rikssiffror som Sundbärg avsevärt höjde (införda i Historisk statistik). Dock har 1805 års tal åter sänkts litet men det inverkar föga här (det gäller pommernarmén, se min artikel i Stat. tidskr. 1980:1; tab. 3.2, not 2).

Städernas folkökning utgör intill 1772 26000 eller 14,9% och därefter till 1805 24700 eller 11,9%. Vad angår den avsaktade folkökningen efter 1772 faller den helt på Stockholm. Den dominerande huvudstaden har slagit av på takten med bara 6% ökning 1772–1805 emot 13,4% för övriga städer. I nästa punkt sägs något om konjunkturerna under de båda perioderna.

7. Stånd och yrken

Läsaren bör beakta att prästen fick en ny uppgift när han ålades att göra statistiska tabeller över sin hjord. Och den svåraste av dem var kanske ståndstabellen. Se faksimil (sid. 285–288) och nycklar (sid. 297). Just en ståndsperson skulle definieras – med hustru, barn och tjänare. I vissa yrken och i

en samlingsgrupp skulle ungdom över 15 år klassas för sig, ingen övre ålder angiven. Fattiga, lytta och bräckliga skulle hållas särskilda, oberoende av yrke. Man måste räkna med många misstag. Och de flesta felen lär inte kunna påvisas och rättas.

Inte desto mindre har jag varit angelägen att ändra uppgifterna blott när jag har haft mycket goda skäl, och att låta de övriga stå för "tabellförfattarnas" räkning.

De flesta av Sveriges städer var under 1700-talet väldigt små. Skulle de beskrivas var för sig kunde man för den skull knappast ge samma detaljer som primäruppgifterna hade. Den ståndS- och yrkesindelning, som jag har använt, är likväl inte mycket sammandragen i förhållande till församlingstabellens. Som exempel må nämnas att den talrikaste yrkesgruppen, "ämbets- och hantverksfolk", inte i församlingstabellen var uppdelad på de skilda ämbetena utan endast på mästare, gesäller och lärlingar. Däremot var ståndspersonerna i primärtabellen delade på adel, präster, kronobetjänter etc. och andra ståndspersoner, men är här hoplagda till en grupp, delad bara efter kön och ålder. Termen betjänt brukades för både högre och lägre personal.

Under period 1 finns ingen rubrik för militärbefälet, varken officerarna eller underofficerarna. Under period 2 räknas underofficerarna till de meniga. Dit har jag fört dem också för den första perioden, när tillägg har måst göras för luckor. När officerarna – givetvis ståndspersoner – först, under period 2, får en plats på blanketten, kallas de militära ämbetsmän. Ståndens betjäning med två grupper av vuxna är hos mig en grupp. F.d. soldater ingår här bland de aktiva.

Handelsnäringens fyra vuxna grupper har likaledes slagits ihop. Det kan synas egendomligt, men magistraten är faktiskt på första periodens blankett en undergrupp under handeln. Jag har likväl fört borgmästaren till ståndet.

Ovan har nämnts ungdomen. Barnen <15 har grupperats liknande, dock att några barn kan ha dolts bland fattiga och lytta. Sist i varje stadstabell har alla barn förenats till en pojkgrupp och en flickgrupp.

Första tabellåret är mycket enhetligt att redovisa talen för 1751 eller närmaste år. Bara Göteborg är nämnvärt undantag med 1757 års tal för större delen av staden, därtill kommer Köping med samma årtal. Det andra tabellåret 1772 företer däremot avvikelser med annat årtal (utöver 1773) i en fjärdedel av städerna. Blott en sjättedel av de 68 städerna angav dock år, då andra periodens formulär gällde. Siffrorna blev mera osäkra efter omräkning att gälla första perioden. Men i folktal räknat blir bara Stockholms avvikelse av betydelse, kanske är dess siffra (1769) lägre än den rätta med 1 000 personer. Utan större fel kan alltså mina summor för städerna som helhet anses representera de valda tabellåren.

Utvecklingen under *första perioden* kan avläsas i tab. 1 B på sid. 292, och omfattar de 58 städer som har uppgifter för både begynnelse- och slutår. Förändringarna uttrycks i absoluta tal och i procent.

Mankönet har totalt ökats med 13,3% och kvinnkönet med 11,0%, och till följd härav har könsproportionen minskats från 1 194 till 1 163/1 000. Talen är som alltid i städerna högre än rikssiffrorna (1 127 och 1 097 resp.) men liksom de i sjunkande.

Gustav Sundbärg ger i sin bearbetning av tabellverket en överblick över konjunkturväxlingarna. Tiden 1747–60 syns som frihetstidens mest lysande period, dock med tydlig nedgång mot slutet. Åren 1761–65 är åtminstone medelmåttiga, men 1766–75

Tab. 1 B. Stånd och yrken i 58 städer 1751–1772

	1751		1772		Förändring		%	
	Män	Kvinnor	Män	Kvinnor	Män	Kvinnor	M	Kv
1–2 Ståndspersoner/ungd.	4515	4675	5030	5155	+ 515	+ 480	+11	+10
3 Barn <15	2034	2038	2077	2036	+ 43	− 2	+ 2	0
4 Handelsmän, rådmän	2142	2203	2455	2447	+ 313	+ 244	+16	+12
5 Bodsvenner/ungd.	1559	877	1693	816	+ 134	− 61	+ 9	− 7
6 Barn <15	1479	1478	1271	1300	− 208	− 178	−14	−12
7 Fabrikörer	353	308	549	539	+ 196	+ 231	+55	+75
8 Arbetare/ungdom	4003	6225	4129	4864	+ 126	−1361	+ 3	−22
9 Barn <15	1641	1565	1854	1866	+ 213	+ 301	+13	+19
10 Soldater o. båtsmän	6877	4905	6629	3799	− 248	−1106	− 4	−23
11 Skeppare o. sjömän	2021	2041	2327	2719	+ 306	+ 678	+15	+33
12 Hantverkare	4754	4970	5485	5986	+ 731	+1016	+15	+20
13 Gesäller	3766	651	4040	1043	+ 274	+ 392	+ 7	+60
14 Lärlingar 15<	2929	2	3580	8	+ 651	+ 6	+22	.
15 Fria arbetare m. fl.	7610	11497	7710	10936	+ 100	− 561	+ 1	− 5
16 Publ. staters betj.	2104	2231	3135	3399	+1031	+1168	+49	+52
17 Ungdom 15<	1538	4149	2155	5163	+ 617	+1014	+40	+24
18 Barn <15	14074	13942	16483	16917	+2409	+2975	+17	+21
19 Ståndens betj. 15<	2247	4068	2568	5076	+ 321	+1008	+14	+25
20 Andras tj.-folk 15<	2818	9967	3838	12239	+1020	+2272	+36	+23
21 Tjänstefolk <15	712	1299	1151	1767	+ 439	+ 468	+62	+36
22 Fattiga, lytta m. fl.	766	3872	1039	4035	+ 273	+ 163	+36	+ 4
23 Summa	69942	82963	79198	92110	+9256	+9147	+13,3	+11,0
	152905		171308					+12,0
Barn <15	19940	20322	22836	23886	+2896	+3564	+14,5	+17,5
	40262		46722					

Tab. 1 C. Vuxna män enligt stånd och yrken samt totalfolkmängden i 68 städer 1772–1805

	1772 (ca)	1805	Förändring	%
1–2 Ståndspersoner/ungdom[1]	5593	6585	+ 992	+19
4 Handelsmän, rådmän	2814	2994	+ 180	+ 6
5 Bodsvenner/ungdom[1]	1949	2276	+ 327	+17
7 Fabrikörer	583	529	− 54	− 9
8 Arbetare/ungdom[1]	4176	5081	+ 905	+21
10 Soldater och båtsmän	7342	8426	+ 1084	+15
11 Skeppare och sjömän	2753	3567	+ 814	+30
12 Hantverkare	6228	8055	+ 1827	+29
13 Gesäller	4433	5663	+ 1230	+28
14 Lärlingar 15<	4037	7162	+ 3125	+78
15 Fria arbetare m. fl.	8069	7264	− 805	−10
16 Publika staters betj.	3287	2649	− 638	−19
19 Ståndens betjäning 15<	2653[2]	7223	+ 732	+11
20 Andras tjänstefolk 15<	3838			
22 Fattiga, lytta m. fl.	1159	2117	+ 958	+82
Summa	58914	69591	+10677	+18,2
Folkmängd män o. kvinnor	187873	207705	+19832	+10,5

[1] Av ungdom 15< ingår endast sådan hos ståndspersoner, handlande och fabrikörer. – [2] För år 1805 "Privatas betjäning". Se texten p. 7.

ofördelaktiga. (Här bör sägas att Sundbärgs genialiska kompletteringar och korrigeringar av folkmängden gäller rikssiffror och att den tiondel som bodde i städerna omöjligt kan justeras med samma metoder på grund av omflyttningen. Om hans enstaka skattningar av stadssiffror se Noter, Karlskrona på sid. 319).

För att lättare överblicka tabellen kan man plocka ut några viktiga grupper:

	1751	1772	Ökning %
Ståndspersoner	13262	14298	7,8
Handel	9738	9982	2,5
Hantverk/industri	47011	49417	5,0
Fria arbetare m. fl.	19107	18646	−2,4
Tjänare	21111	26639	26,2

Man finner nära stillastående eller svagt ökande grupper utom för tjänarna som ökats kraftigt. Det här går väl ihop med bilden av en konjunkturnedgång. Barnen visar likväl en ganska god tillväxt 14–17%.

En grupp som har vuxit bra men är känslig för växlingar är den lilla gruppen fabrikörer, där Stockholm dominerar. I hantverket har gesällhustrurna blivit 60% flera, en förändring som väl inte enbart är beroende av konjunkturerna, och det var alltjämt under slutåret tre gesäller ogifta av fyra.

Eftersom Stockholm svarar för 40% av de 58 städernas folkmängd har det intresse att undersöka huvudstadens betydelse ytterligare. Publika staters betjänte har vuxit med ca 50%. Men det visar sig att så gott som hela ökningen ligger på Stockholm, vars tal nära fördubblats. Troligen är det inte fråga om en reell ökning. Den förut nämnda raden för högre tjänstemän bland ståndspersonerna (kronobetjänter, fogdar, länsmän etc.) har inte använts i Stockholm vare sig 1751 eller 1769. Man kan tänka sig att de högre befattningshavarna dels är adel som har sin rad och dels ingår i "andra ståndspersoner", men båda har en tillväxt som är nära medeltalet. En del av de högre måste väl ingå åtminstone i 1769 års höga tal för de publika staternas folk, men hur stor den delen är kan svårligen skattas.

Frihetstidens lilla industri – övervägande i textilbranschen med mest kvinnor anställda – hade trots statsunderstöd svårigheter under den tiden. Det visar nedgången med 1361 eller 22% i kvinnoarbetet, under det männens tal är stillastående. Större delen av kvinnominskningen (924) kommer på Stockholm och 189 på Norrköping. I Stockholm är nedgången också stark hos kvinnorna av fria arbetare m. fl. Bergshanteringen var en världsindustri, men den var helt förlagd till landsbygden utom i Falun, Sala och Norrköping samt någon annan stad.

Soldaterna har utan tvivel ett av de mest rörliga yrkena. I noterna nämns hurusom militären med familjer hade uteslutits ur tabellverket först och främst i Stockholm och Karlskrona, jämte Göteborg de viktigaste garnisonsorterna, men även i Kalmar, Malmö och andra städer. Klart är att särskilt de värvade förbanden har blivit ganska osäkert redovisade i mina tabeller. Ett krig har utkämpats under perioden. Ett husarregemente som överfördes från Pommern 1772 och förlades med en skvadron i var och en av sex städer i Sydsverige, har bara upptagits i Helsingborg. Prästerna syns ha trott att regementspastorn skulle lämna tabell, och den har ofta uteblivit. Minskningen av soldathustrurna är synlig på de flesta håll och bör väl hänga ihop med nyrekrytering, särskilt efter pommerska kriget.

Gruppen fattiga, lytta m. fl., mest äldre, består som känt är på grund av männens överdödlighet både förr och nu främst av kvinnor. Under den här perioden har dock

ökningen varit stark på den manliga sidan och svag på den kvinnliga. Männens ökning, 36%, faller mest på Stockholm och Göteborg.

Tabellverkets *andra period* (1775–1800) har i mina stadstabeller utvidgats till 1772–1805. Sundbärg talar om tiden i sin konjunkturöversikt. Missväxterna i början av 1770-talet kunde inte på långt när uppvägas av en kraftig uppblomstring 1774–75. Den gustavianska tiden (1776–1815) är i det hela en ganska olycklig period i Sveriges ekonomiska historia, men åren 1774–80 och 1791–95 utgör dock fördelaktiga undantag.

Att studera utvecklingen under andra perioden kompliceras av flera orsaker. Omstöpningen från 1805 av primäruppgiften är mycket grundlig. Äntligen skils landsbygd och städer genom olika formulär. Men yrkesutövarna delas samtidigt på många fler yrken utan att deras familjer uppdelas på samma sätt: de hålls ihop i större grupper ungefär som förut. Av den orsaken har endast de vuxna männen kunnat jämföras med talen för tredje periodens början.

När nu i tab. 1 C på sid. 292 jämförelsen anställs har man således 15 grupper manliga yrkesutövare i stället för första periodens 22×2. Däremot har *antalet* städer ökats med tio, – dvs. de som redovisar för 1772 men inte 1751.

Medan enligt sista raden totala folktalet på de ca 33 åren har tillväxt med 10%, har yrkesutövarnas antal vuxit med hela 18%. För att förklara den differensen får man kalkylera på följande sätt. Man känner visserligen inte åldersfördelningen inom städerna för 1805. Men könen är kända för 1800 års städer, könsproportionen är 1 186/1 000. Den kan knappast ha ändrats nämnvärt till 1805. Är den lika för de 68 städerna blir männens antal runt 95 000. Av dem är enligt tabellen ca 69 600 vuxna och återstoden 25 400 gossar <15. Gör man samma operation för 1772 får man 28 100 gossar. Följaktligen har under andra perioden antalet gossar minskats med 10% liksom utan tvivel flickorna i samma grad. Här finner man förklaringen till de vuxna männens (och kvinnornas) ökade andel av befolkningen i de 68 städerna. Visserligen inkluderade bägge summorna av gossar något tusental ungdom 15<, men det torde inte märkbart inverka på kalkylen.

Gruppen andras tjänstefolk var i de äldre blanketterna delad på män och kvinnor, men från 1805 hade männen förts för sig, och för dem fanns tjänstefolk endast under rubriken "privatas betjäning". Jämförelse bör sålunda göras med summan av ståndens och andras manliga tjänare.

En ovanlig stegring i antalet av hantverkets lärlingar, 78%, förklaras till jämnt hälften av att Stockholm i sin tabell för 1805 har räknat över 1 500 drängar i hantverket till lärlingarna. Ett hundratal sådana har också räknats in i Norrköpings siffra.

För de flesta andra grupper har perioden betytt en jämn ökning av 15–30%. Men handelsfolket utgör ett undantag med bara 6% ökning. Och om Stockholm frändrages visar handeln en *minskning* av lika många procent. Handelsmännen i huvudstaden har i stället höjt sitt antal med 53% till 937. Ändå är häri inte inkluderade krögerskor. Se noter 3, 4 och 7 för Stockholm. Här skall tilläggas att krögaryrket saknades på 1700-talets blanketter och det är troligen också senare ofullständigt medtaget. Det är speciellt i huvudstaden som krogarna sköts av kvinnor. Saken är utförligt behandlad i min artikel för tredje perioden (Statistisk tidskrift 1980:1, sid. 30).

Fria arbetare m.fl. är en restgrupp och innefattar därför konjunkturkänsliga yrken, som nu företer en minskning med 10%.

Att publika staters betjänte har gått ner i antal med nära 20% beror helt på den kategorin i Stockholm, minskningen där är 40%. Den kanske ger stöd åt min förmodan ovan, att år 1769 däri innefattats en del högre tjänstemän som hade bort föras till ståndspersoner. Man kan tänka sig att 1805 års formulär har hjälpt till en bättre uppdelning av tjänstemännen.

Slutligen måste uppmärksammas att gruppen fattiga, lytta m. fl. stigit i antal från 1 159 till 2 117 eller med 82%. Den framträder särskilt i Göteborg som mer än fördubblad (såvida inte de äldre talen där är för låga), i Stockholm som dubblerad, med 60% ökning i Norrköping och rätt förhöjd i flera andra större städer. Här anar man verkningar av de oroliga tiderna. Till sist visar tab. 1 D hela folktalet yrkesvis i de 68 städerna år 1772 och ger tab. 1 E en grov skattning av ståndS- och yrkesgrupperna i hela stadsbefolkningen för 1751 och 1772. Utslagen i 1 E är betydligt starkare än de

Tab. 1 D. Stånd och yrken i 68 städer 1772: alla grupper

	Män	Kvinnor		Män	Kvinnor
Ståndspersoner	4 727	3 975	Fria arbetare m. fl.	8 069	11 520
Ungdom	866	1 699	Publika staters betj.	3 287	3 577
Barn <15	2 524	2 303	Föreg. gr:s ungdom 15<	2 410	5 609
Handelsmän, rådmän	2 814	2 888	Föreg. gr:s barn <15	17 947	18 472
Bodsvenner/ungdom	1 949	1 052	Ståndens betj. 15<	2 653	5 255
Barn <15	1 585	1 642	Andras tj.-folk 15<	4 363	13 825
Fabrikörer	583	570	Tjänstefolk <15	1 200	1 905
Arbetare/ungdom	4 176	4 879	Fattiga, lytta m. fl.	1 159	4 489
Barn <15	1 886	1 897			
Soldater och båtsmän	7 342	4 186	Summa	86 991	100 882
Skeppare och sjömän	2 753	3 246	Folkmängd	187 873	
Hantverkare	6 228	6 775	Därav barn<15	25 142	26 219
Gesäller	4 433	1 110	Summa barn<15	51 361	
Lärlingar 15<	4 037	8			

Tab. 1 E. Skattning av ståndS- och yrkesgrupper i hela stadsbefolkningen omkring 1751 och 1772

	1751	1772	Förändring		%	
			+	−	+	−
Ståndspers. m. b. o. u.	15 500	18 000	2 500		16	
Handel m. b. o. u.	11 500	13 000	1 500		13	
Fabriker m. b. o. u.	16 800	15 400		1 400		8
Soldater	14 000	12 700		1 300		7
Sjöfolk	4 800	6 600	1 800		38	
Hantverkare	20 200	25 100	4 900		24	
Fria arbetare	22 700	21 600		1 100		5
Publ. st. betj.	5 200	7 700	2 500[1]		48	
Barn och ungdom	40 000	49 200	9 200		23	
Tjänstefolk	24 800	32 400	7 600		31	
Fattiga	5 500	6 200	700		13	
Summa	181 000	207 900	30 700	3 800		
			+26 900		+15	

[1] Troligen för högt tal, se sid. 315 ang. Stockholm.

motsvarande i 1 B. Förklaringen är att de tillagda städerna i 1 E har en vida kraftigare folkökning är de kända, nämligen för de 15 hela 33% och för de tio 27% emot 12% för de 58. De beräkningarna är som anförts grundade på mantalets utveckling.

8. Källförteckning

Tryckta källor

Almqvist, D.: Tillståndet i Sveriges städer 1747 (Hist. tidskr. 1949).

Arosenius, E.: Bidrag till det svenska tabellverkets historia. Stockholm 1928.

Betänkande ang. reglering av oregelbundenheter i rikets administrativa, ecklesiastika och kommunala indelningar, avg. 9/10 1882.

Boger, G.: Ulricehamns krönika. Ulricehamn 1951.

Bromé, J.: Karlskrona stads historia. 1–3. Karlskrona 1930–34.

Dahlgren, S. (red.): Nyköpings stads historia 1–2, Nyköping 1973.

Ericsson, Birgitta: Bergsstaden Falun 1720–69. Uppsala 1970.

Grill, C.: Statistiskt sammandrag af svenska indelningsverket. 1855–58. Faks. Göteborg 1978.

Hannerberg, D.: Mantalsskriven folkmängd och total folkmängd. (Gothia 5), Göteborg 1940.

Hillbom, R.: Karlskrona 300 år. 1679–1862. Karlskrona 1979.

Hjelt, A.: Det svenska tabellverkets uppkomst, organisation och tidigare verksamhet. Helsingfors 1900.

Lext, G.: Mantalsskrivningen i Sverige före 1860. Göteborg 1968.

Nicander, H.: Om Tabell-Värkets Tillstånd i Sverige och Finland ifrån 1772 till och med 1795. Angående Folk-Numern i Städerna (KVA, Handl. 1800).

Nilzén, G.: Arkivfynd i Vadstena landsarkiv. Material till belysning av ståndstabellerna (Hist. tidskr. 1974: 4).

Salaminnen, (Nord. Mus. Svenskt liv och arbete. 20). Sala 1954.

Schnell, J.: Nyköpings historia. Nyköping 1963.

Schück, A. m. fl.: Skänninge stads historia. Linköping 1929.

Sjöborg, N. H.: Utkast till Blekings historia och beskrifning. 1–2. Lund 1792–93.

Sjöstrand, J.: Anteckningar rörande den bohuslänska sillfiskeperioden 1747–1808 (Medd. Geogr. för. i Göteborg. 2. 1917).

Sundbärg, G.: Rikets folkmängd fördelad efter ålder och kön för åren 1750–1900 (Statistisk tidskrift 1903).

Otryckta källor

Generalmönsterrullor.

Mantalslängder för de flesta län för olika år 1747–1795.

Tabellverkets folkmängdstabeller för städerna omkring 1751 och 1772, i enstaka fall senare.

Nycklar till huvudtabellen

Första periodens formulär (sid. 285)

Huvudtabell rad	Tabellverkets ståndstabell rader	Huvudtabell rad	Tabellverkets ståndstabell rader
1 Ståndspersoner	1 a, 2 a, 3 a–b, 24	12 Hantverkare	7 a
		13 Deras gesäller	7 b
2 Deras ungdom 15<	1 b, 2 b, 3 c	14 Deras lärlingar 15<	7 c
3 Deras barn <15	1 c, 2 c, 3 d	15 Fria arbetare m. fl.	8, 10 a–b, (15, 16 a–b)[1]
4 Handelsidkare, rådmän	5 a–d		
5 Deras bodsvenner/ungd.	5 e	16 Publika staters betj.	9
6 Deras barn <15	5 f	17 Föregående ungdom	13 a, (23 a)[1]
7 Fabrikörer	6 a, 22	18 gruppers { barn <15	7 d, 13 b, (23 b)[1]
8 Deras arbetare/ungd.	6 b, 21 a–b	19 Ståndens betjän. 15<	4 a–b
9 Deras barn <15	6 c	20 Andras tjänstefolk 15<	14 a
10 Soldater o. båtsmän	19, 25	21 Tjänstefolk <15	4 c, 14 b
11 Skeppare o. sjömän	12	22 Fattiga, lytta m. fl.	11, 26 a–b, 27 a–b, 28–32

[1] I ett fåtal fall fanns torpare med familjer och någon gång bönder. Se t. ex. Västervik.

Andra periodens formulär (sid. 286–288)

Använt av 11 städer, däribland Nyköping, Växjö, Visby, Malmö, Helsingborg, Göteborg och Kristinehamn. – Personer med mer än ett yrke skulle dubbelföras. Det skedde dock praktiskt taget blott i Visby och Trosa. Se noterna! – Enstaka mindre grupper har måst placeras eller delas skönsmässigt.

Huvudtabell rad	Tabellverkets ståndstabell rad		Huvudtabell rad	Tabellverkets ståndstabell rad	
	Första sidan	Andra sidan		Första sidan	Andra sidan
1			15	3:tio 14, 15	
2[1]	1:mo 1, 2:do 1–3	A:a–e, k	16		A:f
3			17[1,2]		
4			18[2]		
5[1]	3:tio 1–3, 5,		19		A:k, n
6	11		20	3:tio 16	
7			21	(3:tio 16)	
8[1]	3:tio 4, (11)		22		B:a–d
9					
10		A:ff			
11	3:tio 13				
12					
13	3:tio 6, 7, 8,				
14	10				

[1] Grupperna ogifta 15< måste delas på vuxna och ungdom. Rikstabellen för 1772 i M gav relationen ungdom/barn cirka 60 % och den för ungdom/äldre ca 40 %. – [2] Ungdom 15<, resp. barn <15, dock ej till ståndspersoner, handel eller fabriker hörande.

* Ståndstabellen omfattade en stad och en socken. Staden har här brutits ut.
** D:o ett helt prosteri. D:o.

Anm. Talet för 1805 (rad 2) inom parentes har på blanketten rubriken "Studerande" oavsett ålder och stånd och motsvarar därför endast delvis talen för de tidigare åren.

Huvudtabell. Stånd och yrken i städerna efter kön omkring 1751 och 1772 samt för män

	STOCKHOLM					SIGTUNA*				
	1751–57[1]		1769		1805	1751		1772		1805
	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd
1 Ståndspersoner	2 369[2]	1 890	2 429	1 678	2 560	5	8	4	10	7
2 Deras ungdom	54[2]	390	155	703	127	4	3	1	3	–
3 Deras barn < 15	867	939	761	912	.	6	5	–	3	.
4 Handelsidkare, rådmän	559[10]	544	612[3,4]	584	937[7]	8	7	7	8	10
5 Deras bodsvenner/ungd.	604	288	502[10]	50	878	–	–	–	4	–
6 Deras barn < 15	328	312	159	96	.	7	5	4	3	.
7 Fabrikörer	198	166	306	293	259	–	–	–	–	–
8 Deras arbetare/ungd.	1 983	3 397	2 121	2 473	2 308	–	–	–	–	–
9 Deras barn < 15	534	454	693	643	.	–	–	–	–	.
10 Soldater och båtsmän	3 588[1]	2 444	3 160[5]	1 902	3 842[8]	–	–	–	–	1
11 Skeppare och sjömän	1 035	1 122	1 107	1 466	678	9	8	9	7	16
12 Hantverkare	1 257	1 362	1 390[6]	1 604	1 817	31	31	54	52	35
13 Deras gesäller	1 953	488	2 152	715	2 557	7	1	9	–	6
14 Deras lärlingar 15 <	1 142	–	1 411	7	3 518[9]	9	–	12	–	10
15 Fria arbetare m. fl.	2 804	4 138	2 623	3 320[4]	2 092	20	45	27[1]	57	31
16 Publika staters betj.	932	980	1 814	2 043	1 077	5	5	4	2	5
17 Föregående gruppers ungdom	175	1 217	397	1 821	.	1	8	10	14	.
18 Föregående gruppers barn < 15	5 480	5 354	6 742	6 675	.	49	57	85	66	.
19 Ståndens betjän. 15 <	1 193	2 097	1 413	3 037	2 188	3	6	5	5	6
20 Andras tjänstefolk 15 <	1 006	4 016	1 334	4 695		–	13	5	23	
21 Tjänstefolk < 15	161	387	239	577	.	1	1	1	1	.
22 Fattiga, lytta m. fl.	297	1 276	358	1 602	714	–	11	2	20	6
23 Summa	28 519	33 261	31 878	36 896	.	165	214	239	278	.
24 Folkmängd	61 780		68 774		72 652	379		517		434
25 Därav barn < 15	7 370	7 446	8 594	8 903	.	63	68	90	73	.

	NORRTÄLJE (1751)		(1772)[3]			VAXHOLM				
1 Ståndspersoner	5	5	12	12	14	7	13	19	24	11
2 Deras ungdom	–	1	1	4	–	–	8	1	6	–
3 Deras barn < 15	6	2	10	11	.	15	8	11	11	.
4 Handelsidkare, rådmän	7	6	9	9	23	8	8	10	8	15
5 Deras bodsvenner/ungd.	2	–	4	3	5	–	–	–	3	3
6 Deras barn < 15	5	6	7	5	.	9	6	7	4	.
7 Fabrikörer	2	3	2	2	2	2	2	2	1	2
8 Deras arbetare/ungd.	62[1]	77	52[4]	33	89	–	–	–	–	–
9 Deras barn < 15	34[2]	33	32[4]	20	.	–	–	–	–	.
10 Soldater och båtsmän	10	7	1	1	4	2	–	24	16	94
11 Skeppare och sjömän	–	–	1	–	4	17	13	20	15	51
12 Hantverkare	19	12	27	23	77	13	10	13	13	73
13 Deras gesäller	–	–	3	2	12	–	–	2	–	10
14 Deras lärlingar 15 <	1	–	4	–	28	1	–	3	–	4
15 Fria arbetare m. fl.	75	73	63[5]	92	13	45	74	80	130	21
16 Publika staters betj.	4	3	3	3	6	3	3	8	7	10
17 Föregående gruppers ungdom	16	18	14	20	.	4	25	30	28	.
18 Föregående gruppers barn < 15	75[2]	87	53	67	.	51	53	90	73	.
19 Ståndens betjän. 15 <	–	–	4	10	21	2	18	3	14	30
20 Andras tjänstefolk 15 <	14	36	8	15		–	22	7	51	
21 Tjänstefolk < 15	6	7	–	–	.	–	–	–	3	.
22 Fattiga, lytta m. fl.	3	5	2	16	6	1	6	16	5	34[1]
23 Summa	346	381	312	348	.	180	269	346	412	.
24 Folkmängd	727		660		876	449		758		1 048
25 Därav barn < 15	126	135	122	116	.	75	68	108	91	.

1805 jämte folkmängd (forts.)

UPPSALA					ENKÖPING					
1751		1773[4]		1805	1751		1772		1805	
Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd	
79	89	118	107	149	6	7	16	20	16	1
13[1]	41	27[1]	54	(514[5])	3	4	4	4	1	2
56	59	53	49	.	1	6	4	4	.	3
29	30	36	29	39	14	15	17	17	30	4
25	22	31	–	38	6	3	10	4	11	5
42	35	11	17	.	12	15	6	11	.	6
3	3	3	1	–	–	–	4	3	1	7
9	7	16	6	35	–	–	1	3	6	8
10	7	1	–	.	–	–	1	1	.	9
76	92	21	33	12	12	12	11	12	8	10
–	–	22	20	33	–	–	–	–	14	11
107	132	114	208	140	32	29	53	60	156[1]	12
77	11	120	33	138	5	–	18	1	23	13
66	–	115	–	175	5	–	17	–	27	14
239	432	206	430	313	138	196	119	167	48	15
100	109	71	76	38	15	14	18	19	13	16
100	162	80	272	.	32	42	41	69	.	17
302	329	373	427	.	135	135	134	117	.	18
60	154	95	190	223	–	2	5	16	33	19
62	189	151[3]	206		33	69	21	87		20
43	34	43	35	.	1	3	3	4	.	21
3	66	3	23	15	6	32	6	24	6	22
1501	2003	1710	2216	.	456	584	509	643	.	23
3504		3926		4897	1040		1152		1174	24
453[2]	464	481[3]	528	.	149	159	148	137	.	25

NYKÖPING* (1749)[1]	(1775–80)[2]			MARIEFRED*					
	53	54	39	2	7	5	6	10	1
166	32	11	(31)	–	–	1	3	–	2
	82[3]	28	.	5	5	3	3	.	3
	45	54	30	8	7	13	12	6	4
122	26	13	17	1	1	4	2	4	5
	46	31	.	10	3	5	5	.	6
	5	4	5	1	1	2	2	–	7
247	168	183	78	6	3	2	3	11	8
	60	75	.	–	–	2	3	.	9
.	5	2	33	7	6	6	6	9	10
46	20	25	29	–	–	–	–	–	11
	107	121	124	38	39	32	26	58	12
526 m. barn	70	8	50	4	–	4	–	30	13
	51	–	96	3	–	9	–	4	14
335 m. barn	66	93	116[4]	33[1]	52	38[1]	59	16	15
191 m. barn	28	34	35	9	16	12	9	3	16
.	15	31	.	11	21	2	12	.	17
.	135	109	.	61	68	34	58	.	18
	18	28	91[5]	2	3	4	10	31	19
426	64	219		9	27	3	23		20
	10	10	.	7	7	2	1	.	21
147	66	151	20	–	2	–	2	1	22
	1172	1284	.	217	268	183	245	.	23
2206	2456		2380	485		428		597	24
	333	253	.	83	83	51	77	.	25

Stånd och yrken i städerna efter kön omkring 1751 och 1772 samt för män 1805

	STRÄNGNÄS*					TORSHÄLLA*				
	1751		1772[2]		1805	1751[1]		1772[2]		1805
	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd
1 Ståndspersoner	31[1]	27	34[1]	53	27	3	6	4	4	11
2 Deras ungdom	4	16	13	18	(15)	–	–	2	3	–
3 Deras barn < 15	18	23	21	23	.	3	5	5	4	.
4 Handelsidkare, rådmän	9	8	11	10	6	20	21	20	21	9
5 Deras bodsvenner/ungd.	2	–	3	2	7	3	6	8	17	1
6 Deras barn < 15	4	3	5	4	.	12	13	9	11	.
7 Fabrikörer	1	1	–	–	–	–	–	–	–	–
8 Deras arbetare/ungd.	2	4	–	6	13	–	1	–	–	–
9 Deras barn < 15	2	1	2	3	.	–	–	–	–	.
10 Soldater och båtsmän	19	18	12	14	9	13	11	10	10	12
11 Skeppare och sjömän	–	–	–	–	6	14	14	7	16	15
12 Hantverkare	40	39	62	64	71	23	22	40	40	34
13 Deras gesäller	6	–	26	–	15	1	–	2	–	10
14 Deras lärlingar 15 <	19	–	27	–	20	5	–	13	–	5
15 Fria arbetare m. fl.	111	174	109	125	92	46	54	36	44	26
16 Publika staters betj.	13	13	8	9	7	11	12	4	16	9
17 Föregående { ungdom	12	26	27	32	.	14	38	22	37	.
18 gruppers { barn < 15	117	110	111	96	.	64	60	72	82	.
19 Ståndens betjän. 15 <	7	11	11	39	41	5	3	2	2	23
20 Andras tjänstefolk 15 <	7	46	13	87		5	24	15	24	
21 Tjänstefolk < 15	2	5	21	16	.	–	–	–	–	.
22 Fattiga, lytta m. fl.	5	21	4	18	3	4	16	1	5	1
23 Summa	431	546	520	619	.	246	306	272	336	.
24 Folkmängd	977		1 139		1 076	552		608		502
25 Därav barn < 15	143	142	160	142	.	79	78	86	97	.

	TROSA (1750)		(1775)			LINKÖPING[1, 3]				
1 Ståndspersoner	3	4	16	13	14	56	50	50	54	105[4]
2 Deras ungdom	–	1	2	–	–	20	16	8	27	(13)
3 Deras barn < 15	3	4	9	8	.	45	36	23	20	.
4 Handelsidkare, rådmän	8	7	9	6	5	24	24	34	26	20
5 Deras bodsvenner/ungd.	1	–	2	1	6	14	13	12	–	18
6 Deras barn < 15	–	–	2	4	.	15	23	25	28	.
7 Fabrikörer	–	–	–	–	–	3	3	7	1	3
8 Deras arbetare/ungd.	–	–	–	–	–	2	2	6	5	15
9 Deras barn < 15	–	–	–	–	.	9	5	1	–	.
10 Soldater och båtsmän	6	4	–	1	–	33	51	33	51	52
11 Skeppare och sjömän	–	–	1	2	–	–	–	1	–	–
12 Hantverkare	6	7	82[2]	82	72[3]	103	110	125	113	159
13 Deras gesäller	1	–	1	–	4	53	1	65	–	72
14 Deras lärlingar 15 <	1	–	2	–	20	79	–	54	–	135
15 Fria arbetare m. fl.	87[1]	115	5	16	4	62	118	145	239	97
16 Publika staters betj.	4	4	5	4	5	48	51	45	56	67
17 Föregående { ungdom	12	24	11	34	.	4	50	18	60	.
18 gruppers { barn < 15	67	72	64	58	.	180[2]	162	236[1]	237	.
19 Ståndens betjän. 15 <	–	–	–	–	–	25	35	34	28	113
20 Andras tjänstefolk 15 <	16	15	12	24	.	35	162	67	212	
21 Tjänstefolk < 15	6	4	3	3	.	5	14	5	14	.
22 Fattiga, lytta m. fl.	–	–	6	11	5	2	102	2	134	16
23 Summa	221	261	232	267	.	817	1 028	996	1 305	.
24 Folkmängd	482		499		490	1 845		2 301		2 915
25 Därav barn < 15	70	76	78	73	.	254	240	290	299	.

jämte folkmängd (forts.)

NORRKÖPING					SÖDERKÖPING					
1751		1772		1805	1751		1772		1805	
Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd	
56	76	86	116	113	16	28	19	30	25	1
14	57	15	47	(28)	1	5	2	15	(17)	2
59	44	47	70	.	12	10	10	11	.	3
87	91	88	79	173	14	12	13	10	12	4
40	24	91	44	111	4	12	7	5	6	5
70	61	65	59	.	3	11	10	9	.	6
18	16	69	74	28	2	2	2	1	1	7
526[1]	814	550[2]	625	725	5	4	16	12	31	8
191	214	255	223	.	2	1	3	4	.	9
65	73	27	27	186	10	23	3	3	15	10
69	58	137	163	144	3	5	8	5	2	11
182	205	205	245	279	46	50	66	60	68	12
129	–	127	7	221	14	–	20	1	11	13
141	–	159	–	357[3]	19	–	18	–	14	14
160	281	341	551	322	26	52	42[1]	92	21	15
91	98	118	127	100	11	21	15	13	16	16
44	112	43	136	.	5	14	18[2]	25	.	17
533	491	741	698	.	86	80	97[2]	81	.	18
64	127	43	176	261	5	18	12	44	62	19
59	350	108	569		6	53	14	55		20
14	43	9	37	.	3	5	1	7	.	21
38	178	54	270	30	2	6	14	20	5	22
2650	3413	3378	4343	.	295	412	410	503	.	23
6063		7721		9428	707		913		870	24
867	853	1117	1087	.	106	107	121	112	.	25

SKÄNNINGE*[1,2,3]					VADSTENA[1,3]					
9	15	7	13	28	12	13	31	36	27	1
5	4	2	2	(20)	1	5	2	10	(22)	2
13	14	–	3	.	6	9	17	18	.	3
20	22	42	48	20	29	28	26	31	20	4
6	4	3	–	12	22	42	11	8	11	5
9	19	4	28	.	40	36	34	34	.	6
–	–	1	1	–	1	1	2	3	–	7
–	–	3	–	1	2	2	2	1	–	8
–	–	1	3	.	1	4	2	5	.	9
2	3	2	3	19	21	26	47[4]	25	13	10
–	–	–	–	–	–	–	–	–	–	11
45	47	42	40	57	44	52	66	73	59	12
7	–	16	–	21	24	–	22	7	15	13
18	–	19	–	20	18	–	18	–	26	14
43	67	25	38	22	38	86	52	80	89	15
15	13	8	8	20	24	23	34	35	46	16
2	19	1	5	.	31	58	25	52	.	17
65	78	72	73	.	102[2]	112	139[5]	155	.	18
27	22	4	12	133[4]	5	19	7	37	42	19
32	64	32	75		32	52	25	95		20
10	4	10	8	.	46	26	1	28	.	21
2	10	11	22	10	1	27	55[6]	57	40	22
330	405	305	382	.	500	621	618	790	.	23
735		687		875	1121		1408		1412	24
97	115	96	115	.	195	187	192	212	.	25

Stånd och yrken i städerna efter kön omkring 1751 och 1772 samt för män 1805

	JÖNKÖPING[1]					EKSJÖ				
	1751		1772		1805			1772[1]		1805
	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd
1 Ståndspersoner	64	55	69	55	104			15	15	19
2 Deras ungdom	6	23	17	21	(73)			3	4	(38)
3 Deras barn < 15	50	50	28	35	.			5	5	.
4 Handelsidkare, rådmän	32	37	35	37	37			16	26	9
5 Deras bodsvenner/ungd.	13	7	6	8	9			10	15	5
6 Deras barn < 15	26	30	22	26	.			20	21	.
7 Fabrikörer	3	3	14	17	8			1	1	3
8 Deras arbetare/ungd.	6	1	9	6	63			17	7	10
9 Deras barn < 15	–	–	13	11	.			13	3	.
10 Soldater och båtsmän	86	93	88	86	76			2	2	16
11 Skeppare och sjömän	–	–	–	–	–			–	–	–
12 Hantverkare	201	221	217	228	148			29	30	77
13 Deras gesäller	110	7	46	1	87			2	3	58
14 Deras lärlingar 15 <	139	–	124	–	144			30	–	72
15 Fria arbetare m. fl.	65[2]	114	72[4]	125	36			22	44	19
16 Publika staters betj.	29	32	31	31	16			9	17	14
17 Föregående ungdom	33[3]	88	96[5]	111	.			80	40	.
18 gruppers barn < 15	243[3]	225	331[5]	361	.			70	90	.
19 Ståndens betjän. 15 <	18	46	37	59	119			10	15	31
20 Andras tjänstefolk 15 <	35	188	45	230				24	116	
21 Tjänstefolk < 15	14	71	7	30	.			7	19	.
22 Fattiga, lytta m. fl.	14	38	20	55	51			5	20	–
23 Summa	1 187	1 329	1 327	1 533	.			390	493	.
24 Folkmängd	2 516		2 860		2 964			883		1 075
25 Därav barn < 15	333	376	401	463	.			145	138	.

	GRÄNNA*					VÄXJÖ*				
								(1790)[1]		
1 Ståndspersoner	5	6	4	6	16			39	46	38
2 Deras ungdom	–	–	1	1	(4)			20	21	(36)
3 Deras barn < 15	2	2	3	3	.			44	42	.
4 Handelsidkare, rådmän	13	11	13	20	12			45[2]	59	19
5 Deras bodsvenner/ungd.	–	3	6	8	3			12	32	6
6 Deras barn < 15	9	6	11	8	.			43	35	.
7 Fabrikörer	1	1	3	3	–			–	–	1
8 Deras arbetare/ungd.	–	–	7	4	–			–	–	3
9 Deras barn < 15	1	1	7	7	.			–	–	.
10 Soldater och båtsmän	–	–	–	–	5			6	9	9
11 Skeppare och sjömän	2	3	–	–	–			–	–	–
12 Hantverkare	53	43	45	49	43			66	67	80
13 Deras gesäller	1	–	5	–	6			46	4	53
14 Deras lärlingar 15 <	5	–	12	–	13			65	–	64
15 Fria arbetare m. fl.	26	47	10	27	20			22	50	47
16 Publika staters betj.	6	7	11	10	9			16	16	18
17 Föregående ungdom	5	10	4	12	.			12	20	.
18 gruppers barn < 15	79	60	43	49	.			120	106	.
19 Ståndens betjän. 15 <	4	5	4	6	21			8	13	33
20 Andras tjänstefolk 15 <	19	28	14	29				36	157	
21 Tjänstefolk < 15	1	7	–	4	.			–	–	.
22 Fattiga, lytta m. fl.	6	20	3	8	5			8	37	10
23 Summa	238	260	206	254	.			608	714	.
24 Folkmängd	498		460		517			1 322		1 188
25 Därav barn < 15	92	76	64	67	.			207	183	.

jämte folkmängd (forts.)

KALMAR					VIMMERBY*					
1751		1772		1805[7]	1751		1772		1805	
Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd	
51	44	54	58	37	4	5	6	7	15	1
10	23	13	30	(150[5])	3	4	3	5	(17)	2
20	25	21	16	.	2	2	6	6	.	3
59	60	66[2]	77	59	13	13	56	59	1	4
47	20	42	–		6	12	23	13	–	5
46	47	33[2]	41	.	14	18	19	20	.	6
5	5	19	17	2	–	–	2	1	1	7
20	11	8	13	4	–	–	–	–	1	8
16	6	–	–	.	–	–	2	1	.	9
185[1]	170	459[3]	107	68	–	–	–	–	1	10
50	58	33	43	299[6]	–	–	1	–	–	11
106	122	120	133	137	21	19	38	39	49	12
49	3	54	13	45	5	–	13	–	9	13
48	–	86	–	88	9	–	16	–	17	14
41	73	40	57	44	75	67	31	49	167[1]	15
54	57	26	29	50	10	8	15	14	7	16
20	65	91	207	.	29	35	24	33	.	17
288	312	328[4]	477	.	65	87	93	90	.	18
23	59	32	62	48	7	7	8	16	35	19
34	170	109	202		33	51	32	89		20
22	39	94	122	.	14	25	20	18	.	21
19	43	20	13	12	3	9	–	10	3	22
1213	1412	1748	1717	.	313	362	408	470	.	23
2625		3465		3656	675		878		863	24
392	429	476	656	.	95	132	120	132	.	25

VÄSTERVIK[1]					VISBY*					
					(1780)[1]					
20	18	25	22	41	54	70	56	1		
5	14	15	50	(1)	60	44	(37)	2		
10	18	11	10	.	54	47	.	3		
40	48	35[3]	32	35[5]	34	76	37	4		
17	28	22	28	40[5]	34	57	27	5		
43	45	23	15	.	30	39	.	6		
2	2	3	3	8	1	1	8	7		
16	13	28	32	47	20	–	15	8		
12	5	14	27	.	1	2	.	9		
9	12	–	–	39	312	218	84	10		
143	115	203	175	269	190	264	236	11		
78	85	85	82	92	167	154	254[4]	12		
39	2	28	2	50	63	2	60	13		
41	–	35	–	53	83	–	123	14		
91[2]	139	151[4]	151	289	105[2]	135	43	15		
32	31	14	12	32	33	54	55	16		
21	83	37	120	.	29	157	.	17		
218	154	174	150	.	429	404	.	18		
17	17	2	3	80	6[3]	26	106	19		
55	171	43	277		95	384		20		
37	67	96	114	.	14	42	.	21		
3	35	5	34	10	20	146	8	22		
949	1102	1049	1339	.	1834	2322	.	23		
2051		2388		2932	4156		3819	24		
320	289	318	316	.	528	534	.	25		

Stånd och yrken i städerna efter kön omkring 1751 och 1772 samt för män 1805

	KARLSKRONA[1]					KARLSHAMN				
	1751[2]		1766[6]		1805	1751		1772		1805
	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd
1 Ståndspersoner	324[3]	205	360	215	370	28	25	30	32	24
2 Deras ungdom	16	100	20	100	(18)	11	20	9	18	(3)
3 Deras barn < 15	170	135	180	140	.	15	24	23	27	.
4 Handelsidkare, rådmän	97	60	100	60	103	39	45	61	73	60
5 Deras bodsvenner/ungd.	97	–	110	–	116	24	17	56	34	38
6 Deras barn < 15	40	50	40	50	.	49	35	38	60	.
7 Fabrikörer	9	9	9	9	1	1	1	3	3	1
8 Deras arbetare/ungd.	41	6	45	5	54	6	1	5	1	82
9 Deras barn < 15	6	12	10	15	.	4	6	15	22	.
10 Soldater och båtsmän	1400[4]	840	1300[7]	800	1453	6	13	–	–	111
11 Skeppare och sjömän	33	30	15	11	37	58	71	89	81	134[1]
12 Hantverkare	210[5]	165	160	110	185	90	93	108	117	143
13 Deras gesäller	132	20	150	25	144	45	7	32	–	50
14 Deras lärlingar 15 <	160	–	181	–	186	58	–	67	–	125
15 Fria arbetare m. fl.	400	785	500	815	439	155	205	117	158	132
16 Publika staters betj.	45	30	50	30	46	26	32	40	39	41
17 Föregående ung-dom	130	200	140	260	.	31	76	55	150	.
18 gruppers barn < 15	1154	1238	1300	1625	.	269	246	260	306	.
19 Ståndens betjän. 15 <	130	300	140	280	145	13	21	2	3	93
20 Andras tjänstefolk 15 <	.	480	.	490		26	215	101	158	
21 Tjänstefolk < 15	.	65	.	.	.	–	21	40	50	.
22 Fattiga, lytta m. fl.	6	370	.	350	93[8]	12	63	9	42	36
23 Summa	4600	5100	4810	5390	.	966	1237	1160	1374	.
24 Folkmängd	9700		10200		10553	2203		2534		3387
25 Därav barn < 15	1500	1700	1530	1830	.	337	332	376	465	.

	SÖLVESBORG[1]		(1773)			SIMRISHAMN				
1 Ståndspersoner	6	10	5	5	4	8	7	7	10	9
2 Deras ungdom	–	4	2	2	(1)	1	4	1	–	(3)
3 Deras barn < 15	1	3	–	2	.	2	2	8	9	.
4 Handelsidkare, rådmän	6	6	8	7	11	17	22	18	23	15
5 Deras bodsvenner/ungd.	1	–	–	–	5	6	30	31	28	–
6 Deras barn < 15	8	4	6	8	.	20	16	13	7	.
7 Fabrikörer	–	–	2	2	–	–	–	1	1	3
8 Deras arbetare/ungd.	–	–	–	–	2	–	–	2	2	1
9 Deras barn < 15	–	–	5	3	.	–	–	–	–	.
10 Soldater och båtsmän	8	5	–	–	8	–	–	4	4	10
11 Skeppare och sjömän	3	3	–	–	3	27	27	24	32	34
12 Hantverkare	25	24	32	30	72	43	42	35	50	60[1]
13 Deras gesäller	6	1	6	–	38	12	–	7	–	24
14 Deras lärlingar 15 <	4	–	2	–	50	18	–	17	–	8
15 Fria arbetare m. fl.	18	25	20	38	18	33	34	14	29	4
16 Publika staters betj.	7	7	8	8	9	5	5	7	6	9
17 Föregående ung-dom	2	9	7	27	.	20	42	3	10	.
18 gruppers barn < 15	56	47	58	46	.	69	90	68	97	.
19 Ståndens betjän. 15 <	–	4	4	8	–	–	–	1	4	30
20 Andras tjänstefolk 15 <	7	25	4	17	.	19	42	16	20	
21 Tjänstefolk < 15	–	2	–	–	.	–	–	7	6	.
22 Fattiga, lytta m. fl.	–	1	–	1	.	–	–	–	1	3
23 Summa	158	180	169	204	.	300	363	284	339	.
24 Folkmängd	338		373		707	663		623		677
25 Därav barn < 15	65	56	69	59	.	91	108	96	120	.

jämte folkmängd (forts.)

ÄNGELHOLM						MALMÖ					
1751		1772[1]		1805				1773–80[1, 2, 3]		1805	
Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd		Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd	
7	4	11	8	4				66	65	53	1
–	3	1	–	(2)				39	10	(12)	2
4	5	13	7	.				135[4]	50	.	3
7	7	4	3	5				79	73	91	4
–	2	4	1	4				58	43	60	5
9	6	5	3	.				86	91	.	6
–	–	1	2	–				18	15	16	7
–	–	6	1	1				–	7	97	8
–	–	3	1	–				12	16	.	9
–	–	–	–	4				319[2]	132	258	10
1	–	–	1	–				22	23	37	11
47	47	61	60	48				203	228	135	12
22	–	31	–	31				152	39	153	13
30	–	35	–	20				107	–	99	14
14	32	12	21	9				48	129	258[5]	15
2	2	2	2	9				10	11	55	16
6	8	7	9	.				37	65	.	17
43	56	53	66	.				333	358	.	18
3	2	9	11	25				27	72	224	19
11	31	61	66					133	349		20
–	–	–	–	.				8	20	.	21
–	–	1	2	4				21	87	25	22
206	205	320	264	.				1913	1883	.	23
411		584		491				3796		4932	24
56	67	74	77	.				574	635	.	25

LUND*[1]						HELSINGBORG					
		(1766)						(1780)[1]			
60	48	37	50	95				56	57	44	1
9	25	19	55	(40)				20	11	(29)	2
48	37	65	58	.				59	40	.	3
21	21	26	23	27				24	22	19	4
9	6	5	–	20				7	1	16	5
11	9	6	8	.				14	34	.	6
3	2	1	1	7				5	4	6	7
10	3	–	–	31				–	–	24	8
4	5	2	2	.				5	4	.	9
8	9	–	–	46				54	11	94	10
–	–	–	–	–				27	26	23	11
101	109	99	106	145				76	77	91	12
79	16	30	18	130				43	6	73	13
43	–	29	–	127				55	–	78	14
156	184	78[2]	100	193				16	28	67	15
25	24	30	30	27				12	14	16	16
20	43	32	49	.				8	15	.	17
219	252	136[3]	138	.				85	98	.	18
31	64	67	54	222				10	5	75	19
67	109	72	76					40	109		20
39	44	137[4]	124	–				1	21	.	21
19	66	–	–	20				23	85	6	22
982	1076	871	892	.				640	668	.	23
2058		1763		3224[5]				1308		1955	24
321	347	346	330	.				164	197	.	25

Stånd och yrken i städerna efter kön omkring 1751 och 1772 samt för män 1805

	YSTADS S:TA MARIA[1]					SKANÖR MED FALSTERBO[1]				
	1751		1772		1805	1751		1772		1805
	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd
1 Ståndspersoner	17	20	18	17	21	4	2	4	5	9
2 Deras ungdom	–	9	5	14	(19)	–	–	3	5	–
3 Deras barn < 15	9	8	21	19	.	–	1	11	8	.
4 Handelsidkare, rådmän	44	60	44	61	36	5	4	6	5	–
5 Deras bodsvenner/ungd.	26	32	14	27	21	–	2	2	2	–
6 Deras barn < 15	39	46	55	68	.	4	6	2	3	.
7 Fabrikörer	8	10	10	10	2	1	1	1	1	–
8 Deras arbetare/ungd.	10	7	1	–	9	–	–	–	–	–
9 Deras barn < 15	8	4	4	3	.	–	–	–	–	.
10 Soldater och båtsmän	24	44	4	6	18	4	3	–	–	–
11 Skeppare och sjömän	35	41	39	44	48	25	9	50	27	116[2]
12 Hantverkare	78	92	79	82	103	11	11	8	8	8
13 Deras gesäller	46	12	48	19	82	–	–	2	–	–
14 Deras lärlingar 15 <	46	1	47	–	105	1	–	–	–	2
15 Fria arbetare m. fl.	88	128	76	134	73	93	102	73	80	65
16 Publika staters betj.	25	30	31	30	17	3	3	7	6	2
17 Föregående gruppers ungdom	30	68	38	56	.	18	20	12	31	.
18 Föregående gruppers barn < 15	216	200	204	185	.	94	83	82	80	.
19 Ståndens betjän. 15 <	7	15	2	4	74	5	5	3	10	50
20 Andras tjänstefolk 15 <	29	156	48	176		2	4	19	12	
21 Tjänstefolk < 15	9	19	4	16	.	7	6	2	–	.
22 Fattiga, lytta m. fl.	–	23	1	33	20	–	2	4	5	1
23 Summa	794	1025	793	1004	.	277	264	291	288	.
24 Folkmängd	1819		1797		2053	541		579		718
25 Därav barn < 15	281	277	288	291	.	105	96	97	91	.

	HALMSTAD*		(1773)			LAHOLM*[1,2]				
1 Ståndspersoner	19	13	23	33	34	11	14	14	14	16
2 Deras ungdom	–	7	5	25	(32)	4	6	4	1	(2)
3 Deras barn < 15	12	9	21	13	.	16	5	5	7	.
4 Handelsidkare, rådmän	18	16	17	16	23	12	11	17	15	3
5 Deras bodsvenner/ungd.	8	3	8	9	15	1	9	3	4	–
6 Deras barn < 15	24	36	16	15	.	7	6	12	12	.
7 Fabrikörer	2	1	1	1	–	2	–	2	1	1
8 Deras arbetare/ungd.	14	32	1	–	–	7	–	5	–	.
9 Deras barn < 15	14	20	–	–	.	2	3	–	–	.
10 Soldater och båtsman	–	–	–	–	22	–	–	–	–	1
11 Skeppare och sjömän	–	–	1	1	25[2]	–	–	–	–	1
12 Hantverkare	54	47	67	70	66	22	22	49	43	39
13 Deras gesäller	22	–	19	1	20	16	1	20	8	14
14 Deras lärlingar 15 <	22	–	28	–	43	8	–	11	–	21
15 Fria arbetare m. fl.	72	84	89	134	32	108	132	92	132	156[2]
16 Publika staters betj.	23	19	18	22	27	5	4	8	7	12
17 Föregående gruppers ungdom	22	46	15	39	.	12	26	27	39	.
18 Föregående gruppers barn < 15	63[1]	55	104	104	.	82	77	101	96	.
19 Ståndens betjän. 15 <	24	29	14	12	72	3	4	17	34	39
20 Andras tjänstefolk 15 <	14	80	36	88		28	25	18	38	
21 Tjänstefolk < 15	10	12	–	4	.	4	6	–	2	.
22 Fattiga, lytta m. fl.	2	12	2	27	15	–	7	–	7	9
23 Summa	439	521	485	614	.	350	358	405	460	.
24 Folkmängd	960		1099		1324	708		865		834
25 Därav barn < 15	123	132	141	136	.	111	97	118	117	.

jämte folkmängd (forts.)

FALKENBERG*					VARBERG				
1751[1]		1772[2]		1805			1763–72[1]		1805
Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd
6	5	4	4	12			24	16	14 1
1	–	–	1	(2)			8	13	(45) 2
3	3	6	4	.			22	26	. 3
11	10	14	14	11			15	30	15 4
3	–	2	–	6			17	–	13 5
8	16	11	10	.			9	3	. 6
–	–	–	–	2			5	6	1 7
–	–	–	–	9			10	–	2 8
–	–	–	–	.			–	–	. 9
–	–	–	–	3			18	14	– 10
–	–	3	3	23[3]			48	48	60 11
16	16	25	30	33			55	76	71 12
1	–	7	–	13			32	5	17 13
2	–	8	–	13			39	–	49 14
68	76	70	129	19[3]			44[2]	70	65 15
3	5	2	3	10			28	20	21 16
29	21	16	12	.			30	62	. 17
56	52	54	68	.			185	180	. 18
–	–	–	–	21			10	22	59 19
15	18	8	17				42	122	20
–	–	–	2	.			2	7	. 21
–	1	–	–	6			26[3]	51	6 22
222	223	230	297	.			669	771	. 23
445		527		642			1 440		1 331 24
67	71	71	84	.			208	216	. 25

KUNGSBACKA					GÖTEBORG 1757 (1750)		(1795)		
8	13	4	4	14	190	158	211	175	164 1
4	2	3	6	–	46	103	141	50	(17) 2
8	11	6	4	.	81	93	215	124	. 3
11	13	11	9	7	132	138	238	245	305 4
5	1	–	–	6	152	78	254	117	308 5
6	6	4	2	.	147	116	191	223	. 6
2	–	–	1	–	14	19	14	32	15 7
2	1	3	2	–	146	113	6	14	121 8
–	–	–	–	.	62	78	15	26	. 9
3	2	–	–	4	1 089	736	1 279	523	1 305 10
38	16	29	10	10[1]	150	150	89	116	152 11
8	7	14	12	15	255	242	255	265	268 12
–	–	4	–	4	322	30	303	90	483 13
1	–	–	–	–	157	–	244	1	245 14
12	28	23	38	37[1]	400	511	413	452	307 15
2	1	8	5	8	111	125	150	152	166 16
3	10	6	17	.	47	286	116	166	. 17
24	25	31	28	.	901	821	824	886	. 18
8	21	–	–	25	102	227	152	93	470 19
10	13	12	27		196	721	513	1 361	20
1	2	2	2	.	31	66	224	300	. 21
–	2	–	2	6	164	731	273	486	635 22
156	174	160	169	.	4 895	5 542	6 120	5 897	. 23
330		329		384	10 437		12 017		11 939 24
39	44	43	36	.	1 222	1 174	1 469	1 559	. 25

Stånd och yrken i städerna efter kön omkring 1751 och 1772 samt för män 1805

	KUNGÄLV*					MARSTRAND				
	1751		1772[1]		1805	1751		1772[1]		1805
	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd
1 Ståndspersoner	4	5	11	20	11	11	13	8	18	16
2 Deras ungdom	–	8	4	4	–	–	6	1	7	(2)
3 Deras barn < 15	6	3	9	8	.	8	10	9	11	.
4 Handelsidkare, rådmän	8	10	17	15	26	9	9	16	23	21
5 Deras bodsvenner/ungd.	–	1	5	4	2	5	12	14	23	9
6 Deras barn < 15	6	4	–	–	.	5	9	18	22	.
7 Fabrikörer	–	–	1	–	2	1	1	–	–	4
8 Deras arbetare/ungd.	–	–	–	1	12	–	–	–	–	5
9 Deras barn < 15	–	–	–	4	.	–	–	–	–	.
10 Soldater och båtsmän	12	10	28	28	8	7	5	6	10	6
11 Skeppare och sjömän	–	–	1	1	9	16	15	8	9	15
12 Hantverkare	16	17	38	42	33	11	12	36	41	158[2]
13 Deras gesäller	2	–	20	3	8	–	–	18	–	23
14 Deras lärlingar 15 <	–	–	3	–	6	–	–	8	–	9
15 Fria arbetare m. fl.	47	56	89	132	72	38	70	75	94	29
16 Publika staters betj.	9	12	6	5	13	5	6	25	22	15
17 Föregående ungdom	3	10	8	23	.	14	29	27	42	.
18 gruppers barn < 15	33	33	138	110	.	38	58	113	105	.
19 Ståndens betjän. 15 <	3	11	–	–	25	–	–	2	10	13
20 Andras tjänstefolk 15 <	4	29	19	56		–	30	16	55	
21 Tjänstefolk < 15	21	20	2	11	.	2	4	4	19	.
22 Fattiga, lytta m. fl.	12	29	2	–	–	2	9	–	17	9
23 Summa	186	258	401	467	.	172	298	404	528	.
24 Folkmängd	444		868		853	470		932		1 189
25 Därav barn < 15	66	60	149	133	.	53	81	144	157	.

	STRÖMSTAD[1, 2]					BORÅS				
1 Ståndspersoner	7	6	16	24	14	6	7	13	13	21
2 Deras ungdom	1	1	1	5	(21)	1	8	3	7	(1)
3 Deras barn < 15	11	5	9	8	.	7	4	7	7	.
4 Handelsidkare, rådmän	6	8	12	8	21	194[1]	244	136[5]	120	92
5 Deras bodsvenner/ungd.	1	5	4	5	2	154[2]	2	99[6]	178	66
6 Deras barn < 15	2	6	8	7	.	104[3]	101	115[7]	109	.
7 Fabrikörer	–	–	2	2	–	6	7	10	9	1
8 Deras arbetare/ungd.	–	–	–	–	–	7	4	8	7	19
9 Deras barn < 15	–	–	–	–	–	5	5	9	14	.
10 Soldater och båtsmän	–	–	1	1	6	–	–	–	–	12
11 Skeppare och sjömän	2	4	9	4	72[3]	–	–	–	–	–
12 Hantverkare	7	7	16	11	75[3]	44	56	63	61	80
13 Deras gesäller	–	–	–	–	4	49	–	35	4	30
14 Deras lärlingar 15 <	–	–	1	–	10	49	–	28	–	45
15 *Fria arbetare m. fl.*	26	37	38	59	35	57	177	62	141	76
16 Publika staters betj.	15	13	19	17	18	27	30	23	19	16
17 Föregående ungdom	9	8	18	12	.	55	138	23	59	.
18 gruppers barn < 15	36	36	61	63	.	67	93	71[8]	70	.
19 Ståndens betjän. 15 <	3	9	1	11	23	2	1	1	–	85
20 Andras tjänstefolk 15 <	2	11	2	16		201	257[4]	167	235	
21 Tjänstefolk < 15	1	2	2	1	.	51	51	13	10	.
22 Fattiga, lytta m. fl.	–	–	3	4	5	1	6	1	45	7
23 Summa	129	158	223	258	.	1 087	1 191	887	1 108	.
24 Folkmängd	287		481		1 085	2 278		1 995		1 792
25 Därav barn < 15	50	49	79	78	.	234	254	215	210	.

jämte folkmängd (forts.)

ALINGSÅS					ÅMÅL*					
1751		1772		1805	1751		1772		1805	
Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd	
7	7	8	4	10	4	3	7	10	15	1
2	4	1	1	–	–	–	2	1	–	2
4	9	4	5	.	2	2	3	3	.	3
8	7	30	31	16	16[1]	17	11[1]	15	8	4
4	1	3	7	2	9	3	–	–	1	5
4	2	7	10	.	17	16	2	6	.	6
22[1]	18	5[1]	2	–[3]	–	–	2	1	–	7
29	93	44	58	32	–	–	1	–	8	8
24	29	58	66	.	–	–	2	1	.	9
4	4	–	–	5	4	3	–	–	4	10
–	–	–	–	–	–[1]	–	30[1]	31	32	11
36	33	62	68	52	29	30	39	44	46	12
51	28	54	30	27	3	–	14	5	22	13
42	–	20	–	22	11	–	28	–	23	14
42	69	58	91	19	58	85	44	53	34	15
7	6	12	10	2	6	7	9	10	10	16
5	17	30	21	.	12	14	13	18	.	17
77	93	81	71	.	70	84	105	115	.	18
10	6	10	15	24	1	–	7	10	76	19
15	65	19	41		21	33	16	59		20
5	4	7	19	.	5	26	3	9	.	21
–	8	8	21	15	–	3	1	2	2	22
398	503	521	571	.	268	326	339	393	.	23
901		1 092		829	594		732		847	24
114	137	157[2]	171	.	94	128	115	134	.	25

ULRICEHAMN*[1, 2, 4]					MARIESTAD					
							(1763)[3]			
10	11	13	10	7	11	11	15	15	30	1
4	6	3	5	(38)	1	1	6	9	(40)	2
14	17	11	6	.	12	7	9	7	.	3
9	10	17	15	16	18	14	16	19	11	4
5	1	2	1	1	4	10	3	11	8	5
3	5	1	1	.	6	7	14	13	.	6
–	–	–	–	3	2	2	3	3	–	7
–	–	1	–	5	4	3	34	26	12	8
–	–	4	–	.	2	3	20	11	.	9
–	–	1	1	2	10	8	–	–	4	10
–	–	–	–	–	12	12	13	16	19	11
18	16	34	33	43	49	56	61	58	51	12
1	–	10	–	6	12	–	17	1	19	13
5	–	6	–	5	14	–	35	–	40	14
17	29	27[3]	29	136[5]	58[1]	74	14[4]	53	61	15
5	7	4	4	11	12	9	18	16	19	16
3	9	7	17	.	8	10	2	28	.	17
17	26	51	53	.	69[2]	78	91[5]	94	.	18
2	5	4	6	21	5	9	10	14	26	19
20	24	27	44		32	55	23	87		20
–	5	3	8	.	–	–	1	2	.	21
–	–	1	2	4	1	11	2	5	3	22
133	171	227	235	.	342	380	407	488	.	23
304		462		789	722		895		1 101	24
34	53	70	68	.	89	95	135	127	.	25

Stånd och yrken i städerna efter kön omkring 1751 och 1772 samt för män 1805

	SKARA*					SKÖVDE*				
	1751		1772[2]		1805	1751		1772		1805
	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd
1 Ståndspersoner	16	14	34	53	32	14[1]	15	8[1]	9	10
2 Deras ungdom	–	4	5	5	(9)	2[1]	7	1[1]	4	(32)
3 Deras barn < 15	7	9	24	9	.	8[1]	12	10[1]	8	.
4 Handelsidkare, rådmän	16	16	14	14	15	15	17	14	13	11
5 Deras bodsvenner/ungd.	1	4	9	1	4	–	–	6	5	1
6 Deras barn < 15	–	–	1	–	.	2	1	4	4	.
7 Fabrikörer	1	1	1	2	–	–	–	–	–	–
8 Deras arbetare/ungd.	2	2	–	–	3	1	2	–	–	2
9 Deras barn < 15	–	–	–	1	.	–	–	–	–	.
10 Soldater och båtsmän	–	–	3[3]	2	7	17[2]	15	14[2]	8	5
11 Skeppare och sjömän	–	–	–	–	–	–	–	–	–	–
12 Hantverkare	48	43	55	67	35	26	26	24	23	26
13 Deras gesäller	15	4	17	3	15	3	–	10	1	17
14 Deras lärlingar 15 <	12	–	23	–	15	3	–	10	–	–
15 Fria arbetare m. fl.	20[1]	32	34[1]	52	86	32	58	38	53	19
16 Publika staters betj.	2	2	9	9	6	14	13	9	8	7
17 Föregående ungdom	4	6	25	41	.	9	15	13	12	.
18 gruppers barn < 15	37	34	90[4]	78	.	38[1]	46	50[1]	42	.
19 Ståndens betjän. 15 <	9	6	17	42	61	7	11	13	15	29
20 Andras tjänstefolk 15 <	12	28	28	61		13	24	23	24	
21 Tjänstefolk < 15	–	–	2	5	.	2	8	9	14	.
22 Fattiga, lytta m. fl.	–	7	–	10	8	7	28	2	20	3
23 Summa	202	212	391	455	.	213	298	258	263	.
24 Folkmängd	414		846		950	511		521		447
25 Därav barn < 15	44	43	117	93	.	50	67	73	68	.

	KARLSTAD*					KRISTINEHAMN (1775)[1]			
1 Ståndspersoner	29[1]	21	20[1]	19	49	33	38	21	
2 Deras ungdom	2[1]	7	6[1]	8	(2)	20[1]	15	(40)	
3 Deras barn < 15	31[1]	19	19[1]	16	.	43	21	.	
4 Handelsidkare, rådmän	17	20	21	21	18	33	30	45	
5 Deras bodsvenner/ungd.	12	9	10	2	13	16[1]	21	16	
6 Deras barn < 15	18	21	2	4	.	31	41	.	
7 Fabrikörer	3	3	–	–	3	1	1	1	
8 Deras arbetare/ungd.	10	4	–	–	13	–	–	8	
9 Deras barn < 15	20	4	–	–	.	–	4	.	
10 Soldater och båtsmän	2[2]	2	–	–	56	–	–	4	
11 Skeppare och sjömän	32	29	20	17	41	38	54	66	
12 Hantverkare	71	73	87	94	123	47	53	57	
13 Deras gesäller	20	–	17	–	44	33	2	22	
14 Deras lärlingar 15 <	24	–	41	–	70	37	–	38	
15 Fria arbetare m. fl.	86	132	111	136	62	32	49	96	
16 Publika staters betj.	33	31	28	23	23	22[2]	26	16	
17 Föregående ungdom	33	59	50	78	.	12	34	.	
18 gruppers barn < 15	144[3]	132	148	168	.	72	112	.	
19 Ståndens betjän. 15 <	21	51	14	29	90	4	10	113	
20 Andras tjänstefolk 15 <	44	136	39	151		61	102		
21 Tjänstefolk < 15	11	29	6	12	.	10	14	.	
22 Fattiga, lytta m. fl.	6	42	16	68	20	–	2	11	
23 Summa	669	824	655	846	.	545	629	.	
24 Folkmängd	1493		1501		2205	1174		1430	
25 Därav barn < 15	224	205	175	200	.	162	192	.	

jämte folkmängd (forts.)

FILIPSTAD*					ÖREBRO*					
1751		1772[2]		1805	1750		1773		1805	
Män	Kvinnor	Män	Kvinnor	Män o. folkmängd	Män	Kvinnor	Män	Kvinnor	Män o. folkmängd	
32[1]	26	16[1]	11	19	48	54	51	39	62	1
9[1]	12	3[1]	6	–	10	33	6	20	(8)	2
24[1]	26	10[1]	10	.	33	29	21	31	.	3
3	3	22[3]	18	17	42	43	30	30	32	4
3	–	8[3]	–	18	13	19	21	16	33	5
2	2	11[3]	12	.	23	22	31	15	.	6
1	1	–	–	–	3	2	3	2	1	7
1	–	–	–	3	–	–	2	2	58	8
1	–	–	–	–	–	–	1	–	.	9
–	–	–	–	7	22	42	14	14	46	10
1	1	–	–	–	39	39	42	42	58	11
31	29	19[3]	20	18	159	167	155	156	146	12
6	–	10	–	8	62	–	53	–	56	13
27	–	15	–	14	58[1]	–	96	–	101	14
20	54	12	18	40	75	116	103	210	147	15
10	9	13	10	13	30	32	38	34	50	16
7	9	1	16	.	29	72	32	89	.	17
49	45	29[4]	30	.	247	207	217	232	.	18
31[1]	34	24[1]	36	44	43	77	53	70	194[2]	19
17	42	21	54		43	168	70	254		20
6	9	7	2	.	7	11	6	6	.	21
–	–	–	12	4	21	75[1]	10	49	19	22
281	302	221	255	.	1007	1208	1055	1311	.	23
583		476		655	2215		2366		3242	24
82	82	57	54	.	310	269	276	284	.	25

NORA*[1, 2, 4]					VÄSTERÅS* (1751)[1]		(1769)[2]			
10[3]	10	17[3]	14	25	33	37	46	45	59	1
–[3]	1	3[3]	4	–	10	11	6	19	(52)	2
6[3]	5	10[3]	9	.	22	21	17	22	.	3
9	7	24	19	12	34	43	32	29	20	4
–	–	1	–	8	47	16	42	21	19	5
3	3	8	8	.	39	41	25	27	.	6
–	–	–	–	–	2	–	5	4	–	7
–	–	–	–	2	5	2	9	7	12	8
–	–	–	–	.	–	–	7	12	.	9
7	7	–	–	5	47	64	20	39	55	10
–	–	–	–	–	35	39	62	60	41	11
28	28	45	42	40	91	131	110	126	154	12
8	–	5	–	6	65	3	65	5	84	13
11	–	12	–	9	58	–	87	–	81	14
26	38	37	61	19	149	252	142	200	55	15
7	7	6	5	24	51	70	58	65	71	16
46	83	17	20	.	34	88	29	66	.	17
60	68	30	27	.	256	239	273	238	.	18
6	5	8	6	48	30	53	23	94	132	19
39	62	51	93		37	176	55	232		20
4	7	29	21	.	3	10	7	10	.	21
10[3]	41	2[3]	10	1	11	71	18	84	60[3]	22
280	372	305	339	.	1059	1367	1138	1405	.	23
652		644		720	2426		2543		2953	24
73	83	77	65	.	320	311	329	309	.	25

Stånd och yrken i städerna efter kön omkring 1751 och 1772 samt för män 1805

	KÖPING[*1]					SALA[1]				
	1757		1769		1805	1752		1772		1805
	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd
1 Ståndspersoner	25	23	12	15	22	36	36	48[5]	44	39
2 Deras ungdom	2	7	2	4	(29)	5	12	4	14	–
3 Deras barn < 15	13	11	4	8	.	14	9	13	15	.
4 Handelsidkare, rådmän	32	39	29	28	13	18	12	22	18	19
5 Deras bodsvenner/ungd.	12	13	14	6	13	6	4	12	12	8
6 Deras barn < 15	17	24	5	6	.	7	8	7	12	.
7 Fabrikörer	3	2	7	7	1	3	3	2	3	104[7]
8 Deras arbetare/ungd.	16	4	1	1	6	133[2]	222[3]	161[6]	137	201
9 Deras barn < 15	7	3	2	–	.	–[4]	5	–[4]	–	.
10 Soldater och båtsmän	–[1]	–	7	18	27	4	2	–	–	23
11 Skeppare och sjömän	33	40	35	41	37	–	–	–	–	–
12 Hantverkare	65	70	80	83	94	54	57	75	87	90
13 Deras gesäller	27	2	32	6	20	24	1	25	5	20
14 Deras lärlingar 15 <	28	–	33	–	21	27	–	32	–	43
15 Fria arbetare m. fl.	86	132	82	115	27	88	106	87	166	39
16 Publika staters betj.	14	19	25	27	16	5	5	19	27	5
17 Föregående ungdom	24	35	32	49	.	73	72	34	56	.
18 gruppers barn < 15	128	147	183	169	.	223	208	219	200	.
19 Ståndens betjän. 15 <	9	26	5	16	57	1	–	31[5]	28	139
20 Andras tjänstefolk 15 <	12	80	16	86		64	114	33	126	
21 Tjänstefolk < 15	1	5	–	4	.	–	6	–	–	.
22 Fattiga, lytta m. fl.	1	17	–	24	8	4	56	4	27	3
23 Summa	555	699	606	713	.	789	938	828	977	.
24 Folkmängd	1254		1319		1232	1727		1805		2116
25 Därav barn < 15	166	190	194	187	.	244	236	239	227	.

	FALUN*					HEDEMORA*[1, 3]				
	(1751)		(1772)			(1751)				
1 Ståndspersoner	122	122	98	133	57	18	13	20	16	19
2 Deras ungdom	37	56	35	63	(7)	2	3	4	10	–
3 Deras barn < 15	91	87	99	93	.	8	20	4	9	.
4 Handelsidkare, rådmän	44	43	37	43	33	21	18	21	23	10
5 Deras bodsvenner/ungd.	17	16	9	9	15	5	2	5	–	12
6 Deras barn < 15	21	32	23	18	.	8	9	–	–	.
7 Fabrikörer	1	1	3	3	18	–	–	–	–	–
8 Deras arbetare/ungd.	799	1264	767	1148	565	–	–	–	–	3
9 Deras barn < 15	600	610	565	617	.	–	–	–	–	.
10 Soldater och båtsmän	6	–	10	14	44	4	4	7	5	23[4]
11 Skeppare och sjömän	–	–	–	–	–	–	–	–	–	–
12 Hantverkare	101	104	97	109	125	45	41	49	47	67
13 Deras gesäller	54	–	45	11	81	4	–	13	–	18
14 Deras lärlingar 15 <	60	–	53	–	81	10	–	18	–	32
15 Fria arbetare m. fl.	130	236	117	160	68	54	76	73	93	22
16 Publika staters betj.	38	37	26	25	32	12	11	25	21	5
17 Föregående ungdom	22	50	23	73	.	9	27	49	67	.
18 gruppers barn < 15	112	97	117	118	.	71	46	92	132	.
19 Ståndens betjän. 15 <	148	180	95	174	201	28	34	25	40	44
20 Andras tjänstefolk 15 <	49	120	13	74		16	60	36	45	
21 Tjänstefolk < 15	2	2	–	–	.	77[2]	62	11	11	.
22 Fattiga, lytta m. fl.	22	103	15	89	8	6	14	5	20	4
23 Summa	2476	3160	2247	2974	.	398	440	457	539	.
24 Folkmängd	5636		5221		4790	838		996		767
25 Därav barn < 15	824	826	804	846	.	164	137	107	152	.

jämte folkmängd (forts.)

SÄTER*					GÄVLE*					
1752		1772		1805	1751		1772		1805	
Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd	
9	8	8	10	11	43	51	42	48	71	1
1	1	5	5	(1)	40	32	6	30	(263[2])	2
2	4	7	4	.	87	45	35	27	.	3
15	15	13	17	9	53	56	63	74	93	4
8	3	9	4	2	32	39	48	40	65	5
14	5	8	6	.	33	49	33	33	.	6
–	–	1	–	1	14	9	3	2	3	7
–	–	2	–	23	24	16	18	29	128	8
–	–	–	–	.	27	7	34	36	.	9
8[2]	6	8[2]	6	5	16	12	10	13	84	10
–	–	–	–	–	85	58	138	161	441[3]	11
51	54	45	50	51	78	87	103	122	375[3]	12
11	2	28	7	21	41	–	43	–	116	13
15	1	15	–	5	42	–	71	–	160	14
14	19	7	16	4	324	453	338	425	106	15
8	9	5	4	5	32	33	52	76	66	16
9	26	8	15	.	70	152	137	145	.	17
61	39	55	61	.	309	340	361[1]	369	.	18
5	6	10	16	15	52	84	30	52	104	19
20	34	25	50		54	235	35	324		20
–	1	–	–	.	2	18	52	66	.	21
1[1]	12	2	7	1	8	34	4	29	26	22
252	245	261	278	.	1466	1810	1656	2101	.	23
497		539		430	3276		3757		5930	24
77	49	70	71	.	458	459	572	543	.	25

SÖDERHAMN (1751)[1]					HUDIKSVALL					
16	15	15	25	10	20	26	23	22	18	1
2	4	1	9	(2)	3	7	6	8	(49)	2
5	11	11	12	.	10	25	9	12	.	3
36	32	32	41	21	22	22	39	34	22	4
13	24	16	7	11	8	8	13	11	19	5
27	14	14	22	.	6	7	11	15	.	6
3	2	6	4	–	2	2	1	1	–	7
10	6	11	12	54	–	–	2	2	6	8
2	–	6	4	–	–	–	–	–	.	9
7	7	4	8	10	13	17	–	–	9	10
16	15	24	31	35	8	9	15	21	8	11
102	131	128	157	127	26	28	45	45	71	12
18	–	13	2	30	8	1	13	2	55	13
32	–	24	–	27	6	–	18	–	–	14
74	112	69	112	21	86	134	108	133	161	15
7	9	11	15	10	14	16	17	20	15	16
38	116	43	103	.	45	59	29	64	.	17
99	130	176	158	.	93[1]	87	122[1]	124	.	18
11	22	11	24	48	6	24	6	14	82	19
37	80	23	66		37	66	26	53		20
–	1	4	1	.	1	1	–	2	.	21
3	*4*	–	9	*4*	*1*	*10*	2	4	*1*	22
558	735	642	822	.	415	549	505	587	.	23
1293		1464		1435	964		1092		1429	24
133	156	216	202	.	110	120	142	153	.	25

Stånd och yrken i städerna efter kön omkring 1751 och 1772 samt för män 1805 jämte folkmängd

	HÄRNÖSAND*					UMEÅ**				
	1751		1772[2]		1805[3]			1795[1]		1805
	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd	Män	Kvinnor	Män	Kvinnor	Män o. folk-mängd
1 Ståndspersoner	29	27	35	36	41			37	30	24
2 Deras ungdom	8	8	13	10	(36)			16	10	.
3 Deras barn < 15	15	20	24	26	.			44	14	.
4 Handelsidkare, rådmän	41	50	33	43	53			23	24	27
5 Deras bodsvenner/ungd.	36	15	31	13	37			36	12	–
6 Deras barn < 15	19	23	17	22	.			16	14	.
7 Fabrikörer	–	–	3	3	1			3	3	–
8 Deras arbetare/ungd.	12	16	5	4	11			–	1	2
9 Deras barn < 15	2	–	3	1	.			1	2	.
10 Soldater och båtsmän	–[1]	–	–	–	8			2	1	6
11 Skeppare och sjömän	10	11	21	22	56			33	28	38
12 Hantverkare	26	27	44	62	160[4]			58	59	72
13 Deras gesäller	4	–	21	7	41			14	4	18
14 Deras lärlingar 15 <	22	–	29	–	65			31	–	23
15 Fria arbetare m. fl.	86	127	87	132	58			24	28	22
16 Publika staters betj.	5	5	25	30	23			13	10	18
17 Föregående ungdom	21	60	25	53	.			12	5	.
18 gruppers barn < 15	78	86	188	166	.			67	82	.
19 Ståndens betjän. 15 <	1	29	14	46	41			8	5	74
20 Andras tjänstefolk 15 <	34	109	25	120				33	111	
21 Tjänstefolk < 15	–	5	2	5	.			7	4	.
22 Fattiga, lytta m. fl.	1	13	3	21	7			4	7	5
23 Summa	450	631	648	822	.			482	454	.
24 Folkmängd	1081		1470		1842			936		1010
25 Därav barn < 15	114	129	234	220	.			135[2]	116	.

	LULEÅ			PITEÅ**		
	(1780)[1]			(1790)[1, 2, 3]		
1 Ståndspersoner	17	16	17	13	19	14
2 Deras ungdom	7	9	–	16	10	–
3 Deras barn < 15	17	13	.	24	9	.
4 Handelsidkare, rådmän	69	82	20[2]	21	19	20
5 Deras bodsvenner/ungd.	44	47	14[2]	22	8	12
6 Deras barn < 15	55	58	.	10	6	.
7 Fabrikörer	–	–	–	–	–	–
8 Deras arbetare/ungd.	–	–	–	–	–	–
9 Deras barn < 15	–	–	–	–	–	–
10 Soldater och båtsmän	–	–	8	–	–	4
11 Skeppare och sjömän	24	24	50[2]	44	60	31
12 Hantverkare	16	16	67[2]	26	29	117[3]
13 Deras gesäller	5	2	21[2]	3	–	6
14 Deras lärlingar 15 <	4	–	18	6	–	16
15 Fria arbetare m. fl.	9	11	34	37	40	10
16 Publika staters betj.	–	–	7	9	10	12
17 Föregående ungdom	4	5	.	31	43	.
18 gruppers barn < 15	16	19	.	87[3]	106	.
19 Ståndens betjän. 15 <	–	7	38	2	4	20
20 Andras tjänstefolk 15 <	37	84		24	52	
21 Tjänstefolk < 15	–	11	.	–	–	.
22 Fattiga, lytta m. fl.	1	1	–	12	18	–
23 Summa	325	405	.	387	433	.
24 Folkmängd	730		947	820		899
25 Därav barn < 15	88	101	.	121	121	.

Noter till huvudtabellen

STOCKHOLM

1751. [1] Redan Wargentin, den förste mannen vid tabellverket, meddelar att i tabellen sannolikt felas Stockholms garnison. Enligt en uppgift för 1757 utgjorde den 5 340 personer, varav 2 885 män och 2 455 kvinnor och barn. Männens tal går fint ihop med generalmönsterrullorna och gäller tydligen nummerstyrkan, ej officerare eller underofficerare. Här har följaktligen antagits att befälet måste skattas och ytterligare läggas till stadens folktal (340+200 med barn 100+100, lagda till ståndspersoner). Av soldatfamiljerna har 1 240 beräknats vara hustrur, 15+100 ungdom och 550+550 barn <15 år. Av manskapet syns sålunda drygt 40% vara gifta. Man har ett jämförelsetal för 1762 för Svea livgarde – ungefär 2/3 av garnisonen – nämligen 59% gifta. Men då året är slutåret för pommerska kriget får talet anses onormalt högt på grund av nyrekrytering och giften. I kalkylen här har soldatfamiljerna delats så att hustrurna blir 44% av männen och barnen <15 år ej fullt ett per familj. En högre giftesprocent skulle ha krävt ett ännu lägre barnantal.

[2] Vuxna ståndspersoner anges till 1515+1 483 (utom adel och präster), varför 98% av männen skulle tyckas vara gifta. Men bland de vuxna kvinnorna måste ingå åtskilliga andra familjemedlemmar än hustrurna. Motsvarande tal för 1769 är likväl 1786+1 268 och 71%. Å andra sidan har ståndspersonernas kvinnliga ungdom mångdubblats på 18 år. Den låt vara obestämda gränsen mellan vuxna och ungdom har således dragits helt olika de båda åren. En överföring från äldre till ungdom av 200 kvinnor har synts rimlig.

1769. [3] Därav på tillskriven rad krögare 1+1. Verkliga antalet var likväl enligt Stockholms handelskollegium under 1770-talet 700. Men krogarna ägdes och sköttes av kvinnor (se min föregående artikel i Stat. tidskr. 1980: 1, sid. 30). I 1700-talets tabellformulär finns ingen plats för dessa kvinnor. Den enda manliga krögaren har fått representera en stor yrkeskår. – De 700 krogarna minskades under 1800-talet till 400.

[4] Det fanns mer handelsfolk som inte fångades av tabellverket: månglerskorna. De mönstrades en gång om året och fördes i rulla hos handelskollegiet med namn och varuslag samt om de var kringgående eller hade torgplats. Antalet var ca 24(?) år 1751 och 400 år 1769. De gifta är troligen förda tillsammans med sina män. Änkor och ogifta ingår måhända i Fria arbetare m. fl. som har stort övertal för kvinnor. Vid pass 1/3 hade fast plats.

[5] Nu har garnisonen införts. De vuxna männen har en obetydlig minskning till 2 709. Den grupp som då omfattade familjerna i nummerstyrkan svarar väl emot vad som nu anges för kvinnor, 2 582. Även nu bör för den skull siffran innefatta soldatbarn av båda könen. De har skattats till 500+582 och lagts till den allmänna barngruppen, rad 18. Efter den förskjutningen finner man nära lika könsproportion i huvudstaden 1757 och 1769, 115–116/100. Barnprocenten blir 24,0 och 25,5 respektive. – Är nummerstyrkan redovisad bör rimligtvis också befälet ingå men var de räknats in är inte klart. En stor del bör dock tillhöra adeln som ingår i ståndspersonerna.

[6] Ett litet överskott av hantverksmästare, 57, har MTL 1770 jämfört med 1769 års ståndstabell. Enär gruppens kvinnor visar ett nästan exakt lika underskott, 61, förefaller det som om några änkor som fortsatte mannens rörelse hade i MTL räknats in bland männen. – För gesäller och lärpojkar tillsammans stämmer jämförelsen med MTL så när som på något tiotal.

1805. [7] Utom 745 kvinnor, flertalet krögerskor. – [8] Militära hantverkare samt stads- och brandvakter 795, förut utan tvivel mestadels uteslutna ur militären. – [9] Därav drängar 1572. – [10] Rådmännen har i Stockholm förts till ståndspersoner.

SIGTUNA

* Staden, ansedd som Sveriges äldsta, hade i sin tabell S:t Olofs och S:t Pers landsförsamlingar som här har brutits ut. Numerären av rena stadsyrken och av allmogen på landet ger vid pass 60% åt landsbygden. Av ståndspersonerna har adeln förts till landet och prästerna till staden, övriga har de-

lats enligt ovan, så också de fattiga och lytta.

1772. [1] Av folkmängden enligt ålderstabellen, 1 030, hade 66 uteslutits ur ståndstabellen. Bristen har till 2/3 ansetts komma på staden och summan tillagts Fria arbetare m. fl. och deras barn.

NORRTÄLJE

1751. [1] Därav i "stadens kamulls spinnskola" 2+47. – [2] En brist i ståndstabellen på 80+80 har rättats sålunda: På prickade rader hade redovisats gevärsfaktoriets personal med hustrur (58+30) men utan barn. De felande 160 antages vara barn och delas ungefär lika på faktoriets barn och den allmänna barngruppen. – *1772.* [3] Uppgiven folkmängd har ökats med 21, i huvudsak enligt ståndstabellen. – [4] Gevärsfaktoriet. –[5] Därav torpare 3+3.

VAXHOLM

1805. [1] Därav fångar 30.

UPPSALA

[1] Studenterna vid akademien är, såsom skrivna i hemorterna, inte här inräknade. – *1751.* [2] Barn under 15 år är enligt ålderstabellen 87 färre än de i ståndstabellen. Det antalet har nu överförts till ungdom. Barnprocenten blir då 26 eller nära lika med 1773 års. – *1773.* [3] Gruppen tjänstefolk 15< (utom ståndens) har osannolikt höga tal – män 281 eller 5 ggr 1751 års tal, kvinnor 290 eller 60% ökning. Genom överföring till barngrupperna har talen blivit något rimligare. – [4] Folkökningen från 1751 faller helt på akademistaten som har stigit från 599 till 989. – *1805.* [5] Studenterna har felaktigt medräknats.

ENKÖPING

1805. [1] Antalet hantverksmästare har femfaldigats under halvseklet.

NYKÖPING

* Staden består av två församlingar, vardera förenad och redovisad med en landskommun.

1749. [1] Tabell ur magistratens handlingar, återgiven i stadens historia (1973), uppenbarligen grundad på tabellverket. – *1775–80.* [2] Med hjälp av MTL och genom kartstudium har stadsförsamlingarna kunnat brytas ut (Östra 1775, Västra 1780). Men MTL ger inte yrkesfördelningen utom beträffande gruvorna och den för tiden rätt livliga industrin. Totalfolktalet stämmer rätt väl. (För socknarna, S:t Nikolai och Helgona, är det ett hundratal fler i kyrkböckerna än i skattelängderna.) Staden har däremot efter min kalkyl ca 70 personer färre än vad mantalskommissarierna redovisar. Yrkesgrupper med låga tal är särskilt osäkra, i synnerhet som de är överförda från andra till första periodens formulär. Ett överskott i förhållande till MTL i Västra församlingen motsvaras av ett ungefär lika stort underskott i den Östra. Det gäller mellan 150 och 200 personer och kan inte här förklaras. – Inga biyrken har redovisats i Västra, troligen inte heller i Östra församlingen. – [3] Här kan ingå skolbarn från landsbygden. – *1805* [4] Därav arbetskarlar 77. – [5] Därav arbetsdrängar 53, tjänstgossar 19.

MARIEFRED

* Prästens tabeller innefattar Kärnbo landsförsamling med staden år 1751 som den större men 1772 den mindre delen. (Kärnbo hade under tiden ökats med 60%, bl. a. genom hovets täta besök på Gripsholm.) Endast tabellens större grupper kan anses tillförlitliga.

1751 och 1772. [1] Här ingår några få bönder som tillhör staden.

STRÄNGNÄS

* Var förenad med Strängnäs landsförsamling men drygt 60% kom på staden.

1751 och 1772. [1] Gruppen ståndspersoner, gemensam för stad och land, är här ovanligt stor. Biskopen bor i landsförsamlingen. De flesta övriga i gruppen har med ledning av MTL förts till staden. – *1772.* [2] Kvinnorna är 57 flera än i ålderstabellen. Eftersom ingen grupp syns onormal, har med tvekan den gemensamma folkmängden höjts med sagda antal. Könsproportionen i staden är även efter höjningen lägre, 119/100, än den var 1751, 127/100.

TORSHÄLLA

* Tabellen inkluderade Torshälla landsförsamling. Hushållens fördelning på land och stad skiftar mycket i Torshälla under 1700-talet. Lanthushållens medelstorlek syns vara 10–11 och stadshushållens 5 under båda tabellåren. Se vidare texten bl.a. p. 6.

1751. [1] Ståndstabellen hade ett underskott på 134. Det har fyllts genom tillägg på barn- och ungdomsgrupper men nästan helt på landsdelen. – *1772.* [2] Här hade ståndstabellen i stället överskott, 39. Vid jämkning blev 22 frånräknade stadens folk, några de fria arbetarna, andra ungdomsgruppen, rad 17.

TROSA

1750. [1] Här måste ingå många fiskare. Antalet är tyvärr okänt. – *1775.* [2] Enär biyrken skulle anges från och med året, har 80 personer dubbelförts. Med stöd av 1751 års tal samt MTL har som dubbelförda avdragits 15 gifta par av hantverkare, vilkas huvudyrke antagits vara fiske, ävensom 10+40 ogifta fiskare, som nu räknats till hantverket. De 40 flickorna förmodas ha berett fisk såsom säsongarbete. Fiske som huvudyrke beräknas ha utövats av 44+46 personer med 80 barn över och under 15 år. – *1805.* [3] Därav 34 fiskare.

LINKÖPING

1751. [1] Ståndstabellen tar upp fler personer än ålderstabellen: 18 manliga och 56 kvinnliga. Dessa har nu dragits från ett par större yrkesgrupper under jämförelse med 1772. – [2] Därav 22 lärgossar. – *1772.* [3] Ståndstabellens senare del ej ifylld; ca 200 människor har nu satts in till samma antal som 1751. Dessutom hade ca 100 gossar < 15 år och lika många flickor utelämnats, men de har nu lagts till skönsmässigt. – *1805.* [4] Ståndspersonernas antal är givetvis särskilt osäkert, när blanketten har ändrats, så att den rubriken inte längre förekommer. En fördubbling från 1772 är knappast trolig.

NORRKÖPING

För rikets fjärde stad gör ståndstabellen ett vårdat intryck. Slutsummorna överensstämmer i huvudsak med dem för ålder och civilstånd.

1751. [1] Därav i bergsbruk 161+133. Övriga 681 kvinnor i gruppen torde till största delen arbeta i textilindustrin. – *1772.* [2] Därav i bergsbruk 204+173. I övrig industri har kvinnorna nu minskats med en tredjedel till 452. – *1805.* [3] Därav drängar 118.

SÖDERKÖPING

1772. [1] Därav torpare 2+2. – [2] Dessas ungdom 1+8 och barn 29+33 inkluderar tjänstefolk.

SKÄNNINGE

* Omfattade staden och Allhelgona landsförsamling. Utbrytningen av den senare ger ovisst resultat för ståndspersonerna. Stadens stora marknader har nog gett befolkningen en särskild rörlighet.

1772. [1] Hälften av kvinnorna hade uteslutits ur ståndstabellen och måste spridas ungefärligt över de större yrkesgrupperna. – [2] Stadshistoriken ger folktalet 687. Folkminskningen 6% är troligen en kvarstående följd av branden år 1760, då halva staden ödelades. – *1805.* [3] Folktalet har nu stigit till 875 eller med 27%. MTL för 1806 slutar likväl på 675. Möjligen är det fråga om ett område som har felförts till Allhelgona men som helt tillhörde staden enligt 1882 års regleringskommitté. – [4] Därav 110 arbetsdrängar. Ökningen är ovanligt stor.

VADSTENA

1751. [1] Krigsmanshuset och hospitalet syns vara uteslutna. MTL ger inte underlag för rättelse. – [2] Därav soldatbarn 12+24. – *1772.* [3] Det är uppenbart att varken krigsmanshus eller hospital har räknats med. De utgjorde särskilda församlingar men har införts efter MTL. – [4] Därav i krigsmanshuset 43+3 gratialister. – [5] D:o barn till gratialister 18. – [6] Därav i hospitalet 102, här delade lika på könen.

JÖNKÖPING

[1] Slottsförsamlingen är inkluderad.

1751. [2] Därav torpare 15+18. – [3] Deras ungdom 2+22 och barn 31+47 kan innefatta tjänstefolk. – *1772.* [4] Därav torpare 11+14.

– [5] Deras ungdom 14+25 och barn 23+32 innehåller möjligen tjänstefolk.

EKSJÖ

1772. [1] Tabellen är en konstruktion. MTL för året – som bara tar upp de i mantal förda, fastän efter 1765 de befriade skulle komma till – ger nästan lika många hushåll som 1805 års folkmängdstabell. Husbönderna och änkorna har nödtorftigt kunnat hämtas ur MTL, likaså de vuxna familjemedlemmarna. Barnen har därefter delats någorlunda jämnt. Jämförelse har också gjorts med mellanliggande 1795 års ståndstabell.

GRÄNNA

* Gränna landsförsamling var sammanförd med staden. Staden hade enligt hushållsredovisningen en knapp fjärdedel. Ståndspersonerna har fördelats skönsmässigt på land och stad. De fattiga och lytta har delats i angiven proportion.

VÄXJÖ

* Växjö landsförsamling hade inräknats.

1790. [1] Den första folkmängdstabell som finns bevarad är ifylld då gällande andra periodens blankett och måste transformeras till den för första perioden, som är väsentligt olika. På övligt sätt har den så prövats mot MTL och landsyrkena avskilts. Av ståndspersonerna har prästerskapet helt hänförts till staden. Publika staters betjänte har alla lagts dit. De vuxna, som därefter har fördelats på stad och land, utgör knappt 150. Personer med mer än ett yrke skulle under andra perioden dubbelföras, men sådana fall var här fåtaliga och kunde elimineras. – [2] Därav "avsigkomna" borgare 30+41.

KALMAR

Slottsförsamlingen ingår.

1751. [1] Artilleritrupp 114 man var icke införd men upptages efter rullor för 1745, hustrur enligt antalet gifta och barn skattade 50+50, rad 18. – *1772.* [2] Ca 10 köpmän och 30 köpdrängar i "stadens köpingar" (7 st.) kan antas ingå här liksom i MTL. – [3] Garnisonen var ej heller nu inräknad men tillförs efter rullor för 1777 (1772 års är inte tillgängliga) med familjer omkring 750 personer. Folkökningen blir hela 32% 1751–72, varav dock nära hälften är garnisonens förstärkning. – [4] Därav soldatbarn minst 105+116. – *1805.* [5] Antalet 150 verkar väl högt. Kan det vara skolans hela elevskara med barn också från landsbygden? – [6] Det nära tiofaldigt ökade sjöfolket avspeglas inte i totalsiffran för städerna, som dock har en topp nära år 1800. – [7] TK har för året lagt slottsförsamlingen till landsbygden, 598 invånare här tillförda.

VIMMERBY

* Vimmerby socken ingår i ståndstabellen. Staden har att döma av hushållens antal 42% av folktalet. "Om alla utfattge skulle utelämnas, så bleve icke många hushåll. De uslaste äro lämnade". Hur detta inverkar på delningen stad – landsbygd är okänt. Min beräkning ger hälften av ståndspersonerna med betjäning åt vardera. Inalles får staden 33–37%. – *1805.* [1] "Borgare" 116.

VÄSTERVIK

1751. [1] Därav 14 män med familjer i Gamleby – enligt Västerviks MTL, se nedan. – [2] Därav 20 torpare med hustrur. Deras barn och tjänare på motsvarande rader. – *1772.* [3] Därav enligt MTL 11 män i Gamleby, flertalet troligen provisionshandlande med betjäning och familjer, ca 100 personer. – [4] Därav 31 torpare m. h. Deras barn som ovan. – *1805.* [5] Därav 13 grossörer med 20 betjäning, sannolikt motsvarande sagda handlande.

VISBY

* Den lilla Visby socken med ca 150 invånare har dragits av.

1780. [1] Här har man flitigt tillämpat andra periodens föreskrift om redovisning av biyrken. Det är 376 fall, de flesta gifta par – båda dubbelförda – men därtill 75 ogifta män och några andra. MTL ger anledning tro att några har kombinationen borgerligt yrke + magistraten resp. publika staters betjänte eller fiskare, måhända också yrke + fattiga, lytta etc. De nu jämkade talen är givetvis särdeles osäkra. – [2] Därav fiskare

13+16. – [3] En del torde ingå i allmänna tjänargruppen. – *1805.* [4] Fiskare räknades under tredje perioden i städerna till hantverket, 13 män.

KARLSKRONA

[1] Den totala folkmängden saknas – utom för 1766, se nedan – ända till 1795. Födda och döda finns däremot ganska fullständigt från början. Den summariska "folknummern" 13000 enligt Sjöborgs Blekinge-historik av år 1792 är uppenbart för hög, så även TKs och Sundbärgs beräkningar. Jag har utgått från de föddas tal som är ganska jämna, och folkmängden har kalkylerats från en födelsefrekvens av 4,2%, den kända i staden omkring 1800.

De särskilda yrkesgrupperna har antagits proportionella mot 1805 års, så också könen och de särredovisade barnen. Krigsfolkets så erhållna tal har jämförts med rullornas. Den mest stabila delen syns ha varit garnisonen ("marinierna"), medan "täta omväxlingar månatligen skedde av volontärer och timmermän". Också indelta båtsmän togs hit från olika delar av landet. Sjukbarackerna fick i krigstid utökas med hela sjukläger. "Karlskrona hyser en framför andra städer överlägsen mängd tiggare och fattiga", säger borgerskapets äldste omkring år 1800.

Det är klart att dessa förhållanden gjorde kyrkobokföringen i den stora Amiralitetsförsamlingen svår och rentav hindrade redovisning till tabellverket. Tabellens poster är därför i högsta grad osäkra men kanske de bästa som står att få.

1751. [2] Rikssiffrorna har ett högre födelsetal vid tabellverkets början med en långsam nedgång under halvseklet. Om den utvecklingen skulle ha gällt här, borde 1751 års folktal ha varit ännu lägre än 9700. – Hillbom beräknar 9000 ungefär denna tid. Sundbärg har 12000. – [3] Därav officerare 150 m., 100 kv. – [4] Därav skeppstimmermän o. a. hantverkare ca 500 m., 300 kv. – [5] Enligt J. Bromés stadshistorik.

1766. [6] Året har valts av följande skäl. Ehuru staden var befriad från att utgöra mantalspenningen från grundandet ända till 1806, kom mantalslängder in ändå för det året. Hela befolkningen var rubricerad som "avkortade mantal". Den var delad i sex grupper: män, hustrur, söner, döttrar, drängar och pigor. Sönerna och döttrarna omfattade tillsammans 37% av totalsumman. Det är väl högt och får motivera att 400 av dem har överförts till ungdom över 15 år. – [7] En säkerligen ofullständig siffra ur rullorna ger 2400 soldater, båtsmän och skeppshantverkare som har tjänstgjort *någon gång* under året. Kanske en sådan siffra har ingått i den förhöjda "folknummern".

1805. [8] Därav 88 fångar.

KARLSHAMN

1805. [1] Skeppare och sjömän. Se not 6 till Kalmar.

SÖLVESBORG

1805. [1] Folkökning från 1772 är 90%.

SIMRISHAMN

1805. [1] Därav fiskare 15.

ÄNGELHOLM

1772. [1] Folkökningen från 1751 är 42%.

MALMÖ

1773–80. [1] S:t Petri 1775, Caroli 1780, garnisonen 1773–74. – [2] Ståndstabellen har ett underskott av 1066 i förhållande till ålderstabell m. fl., uppenbarligen garnisonsförsamlingen. Militären har nu erhållits ur GMR, hustrur och barn samt de militära yrkesmännen har skattats med stöd av 1810 års tabell. – [3] S:t Petri har klumpsummor för "övr. borgerskapet på ö. och s. förstäderne" 500 personer, hälften gifta. De har fördelats proportionellt över stadsyrkena. – [4] Här kan ingå skolgossar från landsbygden. – *1805.* [5] Fria arbetare m. fl. är en restgrupp, och det är svårt att säga varför den är 5 ggr så stor som 1773–80.

LUND

*Lunds landskommun med 10% av folktalet har kunnat brytas ut. Studenterna vid akademien är skrivna i hemorterna.

1751. [1] Totalfolkmängden är lätt justerad. Det var väl högt barnantal, särskilt hos Fria arbetare m. fl. som dock var ovanligt talrik.

– *1766.* [2] Fria arbetare m. fl. minskad med 60% på 15 år. – [3] Härav 40 lärgossar. – [4] Ovanligt många tjänare under 15 år, därav ca 50 + 50 hos ståndspersoner. Hela barnantalet ca 30% av folket. – *1805.* [5] Folkökning från 1766 hela 85%.

HELSINGBORG

1780. [1] Slutsummorna av gifta, förut gifta etc. i ståndstabellen stämmer dåligt med civilståndstabellens grupper. Ej heller totalfolkmängden stämmer. En jämförelse med Malmö S:t Petri pekar på ståndstabellen som den bättre. Den senare har jämförts också med MTL för tre närliggande år ifråga om hela folktalet och för grupper som hustrur, barn <15 år och hantverkare. Jämförelsen stöder antagandet att ingen nämnvärd dubbelföring av yrkesutövare med biyrke har ägt rum. Slutligen befinns att de i MTL ingående ca 15 torpare och gatuhusmän samt mjölnare inte har införts på sina ställen i ståndstabellen. Det kan betyda att folktalet bör höjas med ca 60.

YSTADS S:ta MARIA

[1] S:ta Maria församling utgör ca 3/4 av staden. Hela staden hade följande antal invånare: ca 2400 år 1751 (MTL), 2338 år 1771 (d:o) och 2648 personer år 1805. Antalet barn <15 var enligt MTL så högt som 729 eller 33% men 1805 nere i 555 (24%) enligt längden.

SKANÖR MED FALSTERBO

[1] De båda städerna har senare sammanslagits. Falsterbo hade 118 invånare år 1751 och 131 år 1772. – *1805.* [2] Skeppare och sjömän har nära femfaldigats sedan 1751.

HALMSTAD

*Landskommunen Övraby med 20–27% av befolkningen har kunnat brytas ut.

1751. [1] Därav 6 lärgossar. – *1805.* [2] Märk den starka ökningen av sjöfolket från 1773.

LAHOLM

*Laholms landskommun har avskilts med ledning av 1772, se nedan. Ungefär halva folktalet kommer på den.

1772. [1] Stadens och socknens invånare har i ruta där bägge förekommer angivits skilda. Även för övriga poster har markerats vart de hör. – *1805.* [2] "Åkare och diverse borgare". MTL har 42 "arbetsborgare" och 4 fiskare, vilka torde ingå.

FALKENBERG

*Skrea socken utgör ungefär halva folket i 1751 års tabell men har kunnat avskiljas.

1751.[1] Ålderstabellen har ett överskott av kvinnor över män av 67, ingenting ovanligt. Men ståndstabellen har jämvikt mellan könen och fördelningen bär sannolikhetens prägel (utom att bondhustrurna i Skrea hade uteslutits, men det är en annan historia). Ståndstabellen antages som den rätta. – *1772.* [2] Ytterst ofullständig för kvinnorna med luckor för över hälften i förhållande till ålderstabellen. Kompletteringar har gjorts på fri hand. – *1805.* [3] Stark ökning av sjöfolket men än större minskning av Fria arbetare m. fl.

VARBERG

1763–72. Stadsförsamlingen 1763, slottsförsamlingen 1772. – [1] De flesta hustrur hade uteslutits ur ståndstabellen (1/4 av hela antalet kvinnor enligt ålderstabellen). De felande har kunnat fördelas hjälpligt. – [2] 12 bönder och torpare, måhända i Träslövs enklav inom staden (se Folkräkningen 1880: 2) har här dragits av. – [3] Därav "livsfångar" 25.

KUNGSBACKA

1805.[1] Här är det stor minskning från 1772 av sjöfolket och ungefär lika stor ökning av fria arbetare m. fl., en förändring som står i motsats till den i Falkenberg under samma tid.

GÖTEBORG

1757 (1750). Tal föreligger klara bara för två av de tre större församlingarna, Domkyrkan och Kristine. För den tredje, Garnisonen, har istället här använts talen för 1750, som inte syns avvika mer än något hundratal neråt. Två små församlingar, Hospitals och Fattighus, har med talet 432 från Lext för 1757 och skönsmässigt delade på kön lagts till sista tabellraden. – Speci-

ellt för Göteborg är att församlingarna inte var territoriella och därför visar oregelbundna förändringar från 1757 till 1795 (Domkyrkan – 500, Kristine + 1 000 eller mer än fördubbling). Ståndspersonerna har ökats med 36%, motsvarande inemot 1% om året men gruppen är ju obestämd. Den allmänna tjänargruppen, har mer än fördubblats, bl. a. därför att den i Kristine 1757 angavs till bara 3 personer.

1795. Den dubbelföring av personer med mer än ett yrke, som förutsattes under andra perioden, har inte tillämpats nämnvärt i staden. Däremot syns det vara ett fel i könsproportionen, dock bara i ståndstabellen: den har så gott som jämvikt – det är annars klart kvinnoöverskott i städerna med i den här undersökningen tre undantag. Ålderstabellens könsfördelning får därför anses vara den riktiga, 115/100. Det medför att ca 230 överförs från män till kvinnor. Jämförelser med Kristine och med Stockholm ger stöd för antagandet att felet ligger i gruppen tjänstefolk, där ju kvinnorna brukar vara mångdubbelt fler (Stockholm 1769, 3,5 ggr). Korrigering av könsfördelningen gjorde också Sundbärg på riksnivå när han reviderade tabellverket.

1750–95. Den stora sillfiskeperioden 1747–1808 måste här nämnas, då sillen gick in i Bohusläns fjordar och skärgård. Under den perioden fördubblades folktalet på många håll genom fiske, salterier, trankokerier m. m. (jfr Marstrand, Strömstad och Kungälv). Huvuddelen av sillexporten gick över Göteborg, där Sjöstrand har procenttalet 66 för 1790-talet. Själva fisket bedrevs väsentligen av bohuslänningar. För transporter och förädling kom årligen massor av främlingar under säsongen, som var hösten och del av vintern. Att det har blivit ett allmänt uppsving kan utläsas av folkökningen.

KUNGÄLV

*Annexförsamlingen Rödbo ingår med omkring 36% men har kunnat avskiljas med hjälp av MTL. "Bohus fästnings slottsstat" ingår i staden, men det visade sig för sent att den saknades i ståndstabellen. Tillägg: rad 10: 120 + 22, rad 18: 14 + 8, summa 164.

1772. [1] Den fördubblade folkmängden sammanhänger troligen med den stora sillfiskeperioden. Se närmare Göteborg.

MARSTRAND

1772. [1] Folktalet blev fördubblat på de 21 åren, säkerligen en följd av sillfiskeperioden 1747–1808. Jfr Göteborg. – *1805.* [2] Därav fiskare 122. Visserligen har det yrket inte funnits på tidigare blanketter, så att man kunde ha jämfört. Men det är ändå tydligt att stadens största utveckling har kommit under det intensiva sillfisket.

STRÖMSTAD

1772. [1] Folkökning med två tredjedelar. Utan tvivel beror ökningen mest på det periodiska sillfisket. Se Göteborg och Marstrand. Att MTL upptar 620 personer eller ytterligare 129 (tab. 1 A på sid. 283–284) kan möjligen bero på att MTL-skrivningen sker på senhösten, just under sillfisket. – *1805.* [2] Men ökningstakten är än kraftigare efter 1772. De största sillfångsterna togs i det här området efter 1770. – [3] Nu framträder också en väldig frammarsch av sjöfolket, som på ett tredjedels sekel har sjufaldigats, liksom av hantverkarna, vilka nu är fem gånger så många.

BORÅS

1751. [1] Resehandlare ingår med ca 140 män och kanske lika många kvinnor. – [2] Drängar hos resehandlare är inkluderade med beräknat 150 och [3] resehandlares barn med skattat 100 + 100. De är flyttade från rad 20 resp. 18. – [4] Möjligen är några av de 261 pigorna biträden åt resehandlare. – *1772.* [5] Resehandlare 117 +103. – [6] I huvudsak resehandlares drängar och [7] resehandlares barn, några måhända deras kvinnliga biträden. – [8] Därav 12 lärgossar.

ALINGSÅS

1751 och 1772. [1] Före 1750 var stadens industri betydande. År 1751 redovisas 22 fabrikörer med 122 arbetare/ungdom men senare bara 5 fabrikörer som dock hade över 100 i den andra gruppen. Huvudsakligen gäller det textilbranschen. Är det ett prov

på mellanformen mellan hantverk och industri – "manufakturen"? – [2] För 1772 har MTL (1774) 244 personer färre. En stor del av underskottet ligger på barn <15 år. De är i MTL 220 men här 328. Möjligen har något företag med manufakturprivilegier uteslutits ur MTL. – *1805*. [3] Nu är ingen fabrikör kvar. Här som i många andra städer får man ett intryck av företagens kortvarighet.

ÅMÅL

*Åmåls landskommun ingår men har dragits av med 716 personer.

1751–1772. [1]Yrkesgrupperna har förändrats åtskilligt på de 21 åren: bl. a. har handelns folk minskats till hälften, men 30 sjöfolk angivits 1772, som inte alls togs upp 1751.

ULRICEHAMN

*Annexet Vist ingår med 75 personer som har avskilts.

1751. [1] Staden förlorade sin rätt till boskapshandel 1739, men "ännu vid seklets mitt var den handeln Ulricehamns huvudnäring" (Boger). Det är svårt att av stadshistoriken få en bild av den handeln. Än svårare är det att finna hur den redovisats i ståndstabellen. – *1772*. [2] Folkökning från 1751 hela 52%. – [3] Därav torpare 9 + 9. – *1805*. [4] På 33 år en ökning med 70%. – [5] Därav "borgare" 34 och arbetskarlar 100. Enligt Boger var tiden 1800–60 stadens "knalletid": "En stor del av de handlande lantborna i Kind, Redväg och Ås står skrivna i staden som handelsbetjänter eller handlande borgare". En uppgift från Kommerskollegium från 1787 nämner emellertid 43 resehandlare "som ej bo i staden".

MARIESTAD

1751. [1] Därav 1 "seminant" (bonde) med hustru. – [2] Därav 5 lärgossar. – *1763*. [3] På de 12 åren emellan har folket ökats med 25%. – [4] Därav 3 torpare med hustrur. Denna restgrupp har fluktuerat starkt. – [5] Därav 10 lärgossar.

SKARA

*Det är särskilt osäker delning på Skara stad och Skara landskommun, med omkring 64% på staden.

1751. [1] Enligt MTL 1764 ingår några bönder och torpare i staden. De är skattade till ca 70 med familjer och förda till restgruppen fria arbetare m. fl. – *1772*. [2] Dubblerad befolkning på 21 år med den starkaste tillväxten på ståndspersonerna och därnäst på nämnda restgrupp. – [3] Enligt indelningsverket skulle 3 soldater finnas i landskommunen och t. o. m. 1748 lika många båtsmän i staden. Av de 6 som togs upp i 1772 års tabell har alltså tre räknats till staden såsom avskedat krigsfolk. – [4] Därav 8 lärgossar.

SKÖVDE

*Skövde landskommun har brutits ut med ca 55%. MTL har 103 personer färre. Något av det kan bero på fel vid landskommunens utbrytning.

1751 och 1772. [1] Ståndspersonerna är jämförda med MTL för staden och landskommunen och approximativt delade därefter. – [2] Soldaters och båtsmäns tal är osäkra. Jämförelse med Grill visar avvikelser. Det kan bl. a. bero på att krigsmannen inte bor på sitt torp, kanske inte ens har något ("Jordägaren – – – kan slippa medelst erläggande av några lumpna riksdaler såsom hyresersättning"). Grills kritik gäller dock möjligen mest 1800-talet (Grill III: 39, not).

KARLSTAD

*Den lilla Karlstads landskommun (ca 350) ingick i ståndstabellen.

1751 och 1772. [1] Ståndspersonerna är jämförda med MTL för de båda kommunerna och ungefär överensstämmande med dem. Med stöd härav har gruppen med barn och tjänare nästan helt förts till staden. – *1751*. [2] Soldaters och båtsmäns tal är osäkra. År 1748 blev enligt Grill 11 stadsbåtsmän indragna. De båda för 1751 antecknade är således avskedat folk och hämtade från MTL. – [3] Därav 16 lärgossar.

KRISTINEHAMN

1775. [1] Året tillhör andra perioden. Transformationen till första periodens formulär innebär att den ungdom över 15, som skall

särskiljas för ståndspersoner m. fl. grupper, måste skattas. Se texten p. 2. – [2] Inkl. "stadens hederligare betjäning" ca 15 + 15.

FILIPSTAD

*I ståndstabellen ingick Färnebo landskommun med omkring 5/6.

1751 och 1772. [1] Stånden med betjäning prövade mot MTL och ungefärligt delade på land och stad. – *1772.* [2] På de 21 åren minskas folktalet i staden med 18%, medan landskommunens ökas med 20%. [3] Samtidigt märks i staden en övergång från hantverk till handel. – [4] Därav lärgossar 12.

ÖREBRO

*Ståndstabellen för 1773 omfattade även Ånsta och Längbro socknar. Stöd för avskiljandet av dem har varit tabellerna för 1750 och 1751, eftersom den förra var utan, den senare med socknarna.

1750. [1] Ståndstabellens slutsummor var 14 resp. 29 högre än ålderstabellens män och kvinnor. Differenserna har dragits från hantverkets lärgossar resp. fattiga och lytta kvinnor. – *1805.* [2] Därav 109 arbetsdrängar och 37 tjänstgossar. Se vidare texten p. 7 angående bristande jämförbarhet.

NORA

*Nora landskommun ingår i ståndstabellen och har först 5/6, sedan 7/8 av folkmängden.

[1] Bergverkens personal, ett hundratal, har med stöd av 1805 års tabell helt lagts till landsbygden. – [2] Som ofta i tabellverkets början har ståndstabellen för 1751 inte tagit upp hela folktalet, underskottet är 77 män och 182 kvinnor. Men hela felet har ansetts komma på socknen, som är utesluten. – [3] Ståndspersonerna har delats på land och stad – efter jämförelse med MTL och 1805 – ungefär lika, så även fattiga och lytta. – [4] 1751 års könsproportion är ovanligt hög, 133/100, emot 111/100 för 1772.

VÄSTERÅS

*S:t Ilians landsförsamling är under 1700-talet vanligen inräknad i stadens folkmängdstabeller. När den redovisas särskild har den på 1760-talet ca 330 invånare men 1775 och senare ca 470. Den lägre siffran kan vara ofullständig. År 1751, då S:t Ilian var med, har den vid avdraget skattats till 420.

1751. [1] Könsproportionen är ganska hög, 129/100, inte på grund av talrika tjänare men möjligen av många änkor. – *1769.* [2] Året har valts för att S:t Ilian inte ingår. – *1805.* [3] Därav 20 "gäldstuguboer".

KÖPING

(*)I regel är Köpings landsförsamling inräknad under första perioden. 1757 års tabell är ett undantag och har därför valts. År 1769 omfattar båda delarna, men alla rubriker utom ett par små har särskilda summor för vardera delen, så att stadstabell lätt har kunnat göras.

1757. [1] Några f. d. stadsbåtsmän torde dock ha funnits, eftersom 13 ordinarie sådana drogs in år 1748.

SALA

[1] Från 1752 är ståndstabellen utan Sala landsförsamling. Sala silvergruva drevs tidvis av kronan före 1682 men sedan enbart av Sala bergslag. Flertalet fastighetsägare i staden var delägare i gruvan och hade därtill jord i ytterområdena. Sala var till ytan länge Sveriges största stad. Bergsmännen fanns inom olika yrken och stånd. Men bergsmän, kallades enligt Salaminnen (1954) i dagligt tal endast de som levde huvudsakligen på jordbruk och körslor. I ståndstabellen nämns de inte under 1700-talet. Däremot upptar MTL titeln (1751 för 37, 1773 för 60 och 1796 för 45 män), och i ståndstabellen för 1830 nämns 45 "arbetande bergsmän".

1752. [2] Vid silververket 133 manliga arbetare, varav ca 100 i gruvan och resten i hyttan. – [3] Därav ca 60 ullspinnerskor. – [4] Silververkets barn <15 ingår i den allmänna barngruppen rad 18, så även 1772.

1772. – [5] En brist i ståndstabellen av 92 personer har föranlett, med något stöd av MTL, att ståndspersonerna har ökats med 23 + 10 och ståndens betjäning har införts med 31 + 28. – [6] Enligt MTL 1773 var anställda i gruvan något över 100 och i hyttan 90 manliga arbetare. – *1805.* [7] "Bergsmän och borgare".

FALUN

*Falun var rikets femte stad. Den består av Kristine församling, som helt ingår, och en del av Stora Kopparbergs församling med drygt hälften av dess befolkning. Mantalslängderna för staden saknar uppgifter om fördelningen på områdena. Den främsta källan till en delning har varit prästens i Kopparberg data rörande stadshushåll och lanthushåll. De har givit ramen, totalbefolkningen delad på stad och land. Kopparbergs-delen har ca 30% av stadens folk.

Jämförelsen med MTL har inte haft mycket att ge för 1751, därför att de från mantal befriade saknas där – fullt lagligt för den tiden. Men för 1772 har den på hushållen baserade delningen visat sig nära stämma med MTL. Vidare har påträffats ståndstabeller avgivna av borgmästare och råd dels för åren 1749–52 och dels för år 1769 (Birgitta Ericsson, Bergsstaden Falun 1720–69). Dock utesluter de barn <15 och uppenbarligen också de för "andra orsaker" (ålder, fattigdom och sjukdom) befriade. För vuxna grupper har dessa källor emellertid visat god överensstämmelse med mina kalkyler utom för några få grupper. Det har för den skull inte ansetts behövligt att byta ut de här redovisade åren mot Ericssons närliggande år.

Talen har prövats på flera sätt, bl.a. genom jämförelse av barnprocenterna för olika områden och studium av folktalets förändringar i tiden. Barnprocenten har i Kopparbergs-delen stigit från 29 till 34 men eljest varit jämn.

Den nedgång i folktal som har ägt rum förklaras av Ericsson. Det var kris under 1760-talet, först på grund av stora bränder 1760, följda av nödår, tilltagande fattigdom och folkminskning, och i slutet genom en ekonomisk kris för kopparhanteringen. Nedgången 1751–72 ca 400 svarar mot en jämnstor ökning i Stora Kopparbergs socken som där betyder 70%. Ericsson talar om de många ensamstående kvinnorna, särskilt änkor, och de fader- och moderlösa barnen såsom ett svårt problem i frihetstidens Falun.

HEDEMORA

*Förenad med Hedemora landsförsamling, som var omkring 5 ggr så stor.

1751. [1] Liksom i länet mest var jordbrukarhustrurna uteslutna ur ståndstabellen. Änkorna med jordbruk syns tillagda männen. Också ålderstabellen tycks ha ett underskott (ca 600). För att kunna skilja land och stad måste man först justera för felen. Därefter har gemensamma yrkesgrupper delats i förhållandet 1:5 och något jämkats för ståndspersoner. – [2] Troligen för höga tal; flera bör vara över 15 år.

1772. [3] Här saknas tabell för de båda församlingarna men finns för Hedemora prosteri. Stadens andel – både enligt hushållssiffrorna och efter vad de rena stads- och landsyrkena visar – är knappt 6%. Den andelen har tagits för de gemensamma yrkena. Möjligen har ett hundratal för många förts till staden.

1805. [4] Indelta soldater i landskommunen 93, men liksom i hela länet saknar de torp (Grill), och en del bor i staden.

SÄTER

*Säters landsförsamling ingick i ståndstabellen med tre fjärdedelar av folktalet.

1752. [1] Länets felaktiga redovisning av jordbruksbefolkningen (se Hedemora) gjorde det nödvändigt med ganska stora skönsmässiga justeringar av dessa grupper samt av de fattiga i både land och stad.

1752 och 1772. [2] 8 + 6 soldater och båtsmän. Där var intill 1748 sex ordinarie och lika många fördubblingsbåtsmän. Siffrorna torde gälla de kvarlevande av dessa.

GÄVLE

*Två byar i Valbo socken hör i kyrkligt avseende till staden. Dessa hade något under 200 invånare som ingick i 1700-talets tabellverk. Byborna är här frånräknade med stöd av Landsarkivets arkivförteckning och bilaga till 1805 års ståndstabell.

1772. [1] Därav 27 lärgossar. – *1805.* [2] I ålderstabellen är mankönet övertaligt med 125 i 10–15 och 24 i 15–20 års ålder. Orsaken kan vara att inemot hälften av skolgossarna rad 2 är från landsbygden men har inräknats. – [3] Ovanligt stark folkökning,

58% på 33 år. Sjöfolket är mer än tredubblat liksom hantverkarna men de fria arbetarna är nu bara tredjedelen.

SÖDERHAMN

1751. [1] Totalsumman av mankön var i ståndstabellen 27 mindre än i ålderstabellen. För kvinnkönet var i stället ståndssumman 34 högre än ålderssumman. Till rättelse har några större grupper jämkats.

HUDIKSVALL

1751 och 1772. [1] Ståndstabellen för 1751 har ett överskott av 37 personer, 1772 års tabell ett mindre underskott i förhållande till ålderstabellen. Med hänsyn till barnprocent och folkökning och med ledning av Söderhamns motsvarande data har barnantalet jämkats på rad 18.

HÄRNÖSAND

*Härnö socken ingår med ca 10% av folktalet. Den har kunnat skiljas ut utan större svårighet med hjälp av MTL.

1751. [1] Det är dock inte klart om 9 soldater och båtsmän och 6 avskedade d:o med hustrur hör till Härnö – som här antagits – eller till staden. – *1772.* [2] Folkökning 36%. Slutsummorna i ståndstabellen avviker från motsvarande summor i ålderstabellen: manliga är 17 färre och kvinnliga 4 fler i den förra. Justering har skett på de största grupperna. – *1805.* [3] Folkökning 25%. – [4] Därav fiskare: mästare 78, gossar 29. Yrket tidigare ej specificerat. Jfr Piteå, not 3.

UMEÅ

** Ur tabell för Umeå prosteri har stadsyrkena brutits ut. De gemensamma ståndspersonerna har delats på stad och land, delvis enligt MTL, betjäningen hos stånden likaledes. Gruppen fattiga, lytta m. fl. har med en tiondel lagts till staden – lite mer än som motsvarar folkmängden – och gruppen publika staters betjänte med en fjärdedel. Dessa som rätt godtyckligt har gjorts till stadsbor är tillsammans ca 70.

1795. [1] Andra periodens dubbelföring av personer med biyrke omfattar här ett stort antal bönder med familjer, vid pass 3 500 eller inemot hälften. Antagligen är biyrket här fiske. De därutöver dubbelförda 500 har ansetts helt höra till landsbygden. Troligen gäller det mest vargeringsmän, dvs. militära reservkarlar. (Hela gruppen krigsmän på landet utgör med familjer 1 400, medan Grill, som inte tar med vargeringskarlarna, har blott 106 soldater motsvarande kanske 300 personer.) – [2] Barnen <15 är 27% av befolkningen emot 37% i prosteriets landsdel. Vid seklets mitt (1747) uppgives staden ha 32,5% barn.

LULEÅ

1780. [1] Inga dubbelföringar förekommer, dvs. biyrken saknas på denna andra periodens tabell. – *1805.* [2] Sedan 1780 märks en stark minskning av handelsfolket och en kraftig ökning av både sjöfolket och hantverkarna.

PITEÅ

**Tabell har erhållits genom att stadsyrkena har dragits ur tabellen för Piteå prosteri.

1790. [1]MTL har givit någon ledning att tilllägga 91 ståndspersoner med familjer. Härtill kommer 30 fattiga, lytta m. fl. efter stadens andel av folktalet, något över 5%. – [2] De som dubbelförts för mer än ett yrke är bara 500 och kan anses höra till prosteriets landsbefolkning. – [3] Man finner i hela prosteriet blott 3 "fiskare och skärkarlar", säkerligen på landet. För fiskare i stad fanns då ingen rubrik. Inte heller har sådana nämnts i MTL för 1794, den närmaste tillgängliga. Men yrket har troligen funnits, dolt i någon annan grupp, ty så kort efter som 1805 redovisas 70, som då räknas till hantverkarna.

Systematic Bias in Hours Worked?

by Iiris Niemi[1]

1. Introduction

The International Labour Organization has recommended the collection of data on hours actually worked with surveys. Employed people are asked to estimate how many hours they have worked during a reference week. But is this data unbiased and valid? Can people accurately recall how many hours they have worked during a certain week when asked afterwards? Are there perhaps some systematic differences in the estimation of hours worked among population groups? To study these questions labour force survey results are evaluated in the light of data obtained with detailed time diaries, which has proved to be a more accurate method to measure the use of time (see e.g. Robinson 1977, and 1982).

2. Methods compared

Data on hours worked have been compiled in the Central Statistical Office of Finland with labour force surveys by postal inquiries until 1982 (from 1983 on by interview). The sample has pertained to 18000 persons 15–74-year old persons per month. These monthly inquiries have been completed with yearly interview surveys conducted in autumn. This survey has consisted of a separate sample of 10000 persons 15–64-years of age. A subsample of 7000 persons was selected from this interview study sample in 1979 in order to perform a time use study. Thus data are available both on employment status and the use of time for these persons.

With the time budget method all of a person's activities are systematically recorded during a day. In the 1979 time use study persons themselves entered in time diaries what they did during two consecutive days. The fieldwork was done in September-November. The diaries were centrally coded in the Central Statistical Office and so it was possible to obtain a uniform coding of the hours worked.

3. Hours worked by industry

In this comparison hours of work refer to the time used both for primary and secondary employment, including overtime work but excluding lunch hours and trips between job and home. In both cases, the results were calculated for each employed person by excluding absences, including holidays, from the hours of work.

Measured with the aid of the labour force survey, the total for working hours is somewhat greater than that obtained in the time use study. A similar difference has been observed in the USA in comparisons of the results of a time budget study to labour statistics (Robinson 1982, 39).

Table 1 indicates, however, that measurement discrepancies are not parallel in all branches of industry. A labour force survey yields higher working hours in branches in which the labour force partici-

[1] Central Statistical Office of Finland, Helsinki. Prepared for the Workshop on Time budgets, Institute for Social Research, The University of Michigan, Ann Arbor, May 20-21 1982.

pation rate of women is of major importance (finance and insurance, community, social and personal services). Hours of work measured in a time use study are higher in branches dominated by men (construction and transport). Agriculture and forestry is the exception in that the discrepancy in working hours is considerable, even though this branch is not dominated by women. Thus the sex structure of a branch of industry seems to affect measurement. The following investigate hours of work performed within different branches of industry by sex.

Table 1. Hours worked (main work, extra work, overtime) by branch of industry in autumn 1979. Hours per week and per employed

Industry	Labour Force Survey	(N)	Time Use Study	(N)	Share of Women of the Employed in the Branch of Industry
Total	38.0	(41821)	36.8	(7395)	47%
Agriculture and forestry	50.2	(5389)	41.0	(978)	41%
Manufacturing	37.9	(10717)	37.6	(2064)	35%
Construction	37.8	(2802)	38.7	(557)	9%
Trade	37.4	(6097)	37.7	(1036)	60%
Transport	36.1	(3250)	37.8	(529)	29%
Finance, insurance	34.3	(2331)	33.1	(342)	62%
Community, social and personal services	34.9	(11105)	33.1	(1889)	68%

Table 2 shows that according to the labour force survey the hours of work for women are higher on the average than those obtained on the basis of the time use study. The trend is the reverse for men when measured with time-budget technique, the average hours of work for men being slightly longer than those recorded in the labour force survey. Viewed by branch of industry, exceptions to this general trend emerge. In agriculture and forestry hours of work are longer in the labour force survey both for men and women, whereas among women in construction the working hours obtained with a time-budget technique are longer. This may be due to random error, because weekly hours of work in construction have been aggregated from a limited sample in the time use study (N=67).

Differences in measurement results are significant in agriculture and forestry. According to the labour force survey, the hours of work have turned out to be as

Table 2. Hours worked by branch of industry and sex. Hours per week and per employed

	MALES		FEMALES	
	Labour Force Survey	Time Use Study	Labour Force Survey	Time Use Study
All industries	39.7	40.4	36.0	32.9
Agriculture and forestry	47.8	46.9	53.6	32.1
Manufacturing	39.0	39.2	35.9	35.1
Construction	38.5	38.2	31.6	43.2
Trade	40.5	40.4	35.3	35.9
Transport	37.6	40.1	32.2	31.2
Finance, insurance	37.7	42.2	32.3	26.7
Community, social and personal services	37.7	38.2	33.7	31.0

Table 3. Hours worked by industrial status and sex. Hours per week and per employed

	TOTAL		MALES		FEMALES	
	Labour Force Survey	Time Use Study[1]	Labour Force Survey	Time Use Study[1]	Labour Force Survey	Time Use Study[1]
Total	38.0	37.0	39.7	40.5	36.0	33.3
Self-employed and family workers	50.4	41.9	49.6	48.9	51.6	32.7
Workers	36.3	36.1	38.4	38.9	34.1	33.3
– white collar	35.7	36.0	38.4	39.4	34.0	33.7
– blue collar	37.0	36.3	38.4	38.5	34.5	32.7

[1] In table 3 the definition for employment status in the time use study diverges from that in the labour force study. The time use study classifies students as those who have been employed and who weekly use more time for studying than for gainful employment. Their share came to 1.9 per cent of those defined as employed in the labour force survey. The concept used in the labour force survey is called actual activity while that in the time use study is usual activity. According to actual activitiy, the definition of the employed results in slightly fewer hours of work, since students engaged in part-time employment are included. This difference is 0.2 hours weekly , or 0.1 hours for men and 0.4 hours for women.

much as 67 per cent longer for women engaged in agriculture and forestry than figures produced by the time-budget method.

4. Hours worked by industrial status

The following investigates the relationship of industrial status and measurement disparities.

Table 3 reveals that the method employed exerts considerable influence for the self-employed and family workers. Wage and salary earners arrive at nearly the same hours worked regardless of the method used.

The methods differ in respect to the accuracy of measurement. The time-budget method examines time spent on work by means of diaries recording the use of time in 10-minute intervals during the study days in question. The results therefore fairly accurately correspond to actual hours of work. Respondents replying to the postal questionnaire in the labour force survey estimate their weekly work input ex post. The labour force survey can be assumed to measure a person's image of his weekly working hours, or experienced hours of work.

The observation that 84 per cent of the respondents in the labour force survey have reported the same number of hours worked for successive study weeks supports the concept of experienced hours of work ($r = 0.78$). The study was carried out by random sampling from among the 1981 postal inquiry forms in January-July (N=3872).

Wage and salary earners have a fairly clear conception of their actual hours of work. Women, however, estimate their weekly hours of work to be somewhat longer than they are in reality. Conversely, men do not seem to count as working hours all of the time categorized as hours of work in the time use study (meetings, seminares and courses and related trips, etc.).

The self-employed experience longer days of work than they do when measured in the time use study. This difference is most apparent for women engaged in agriculture. The following figure, describing the working rhythm of own-account male and female farmers and family workers, illustrates reasons for hours of work thought to

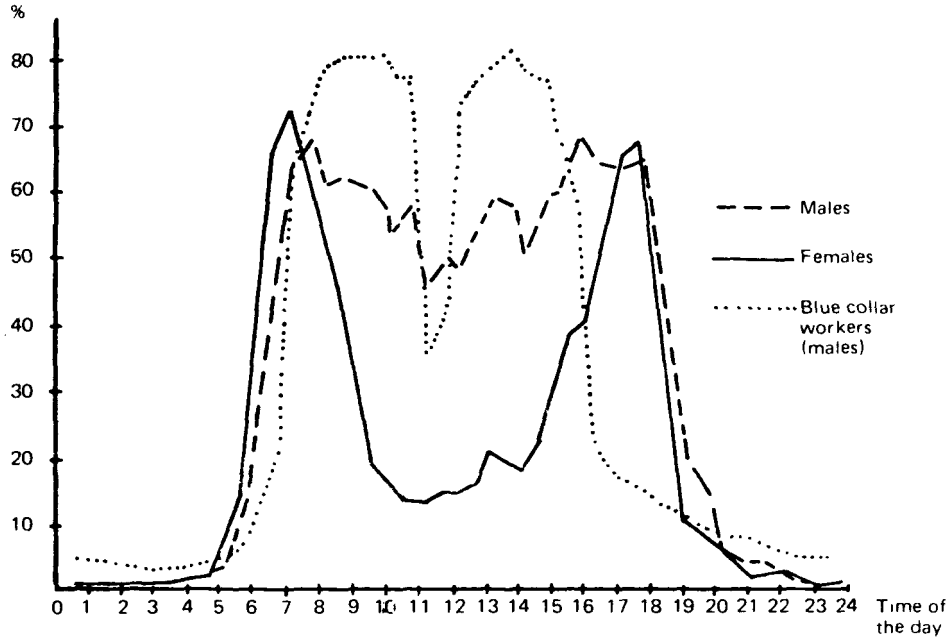


Fig. 1. Daily working rhythm of farmers and family workers on workdays according to sex

be longer than they actually are.

The working rhythm of farmers on own account and family workers differs from that of the wage and salary earners, for which blue collar workers are given as an example in the figure. The hours worked by farmers on own account and family workers are not as clearly demarcated from other use of time in the same manner as they are for wage and salary earners. Work on the other hand typically centers on mornings and evenings due to the nature of work performed on livestock farms. The two peaks in the graphs of working rhythm are particularly distinct for women. The time between these work peaks is mainly used for domestic work. On the basis of labour input data obtained from the labour force survey, the majority of domestic work seems to be regarded as connected with farm work, even though work done for the household is to be excluded according to the instructions for filling in the form. Another reason why these two forms of work are not readily discerned is that both are experienced as an obligation performed as overlapping work.

Occupation groups performing part of their gainful employment at home have faced the same problem. In the autumn of 1979 a survey of the work load of teachers was conducted with diaries and the results were compared to data obtained in the time use study (Askelo 1981). It was observed that men's working hours did not significantly differ between the teacher survey (39.4 hrs) and the time use study (38.7 hrs). But the difference was apparent among women, being 38.2 hours in the teacher survey and 31.7 hours in the time use study. This discrepancy can perhaps mainly be explained by how simultaneous activities at

home have been recorded; in the teacher survey attention has primarily been focused on gainful employment, while different activities were assigned the same value in the time use study. Gainful employment at home seems to complicate the estimation of hours worked especially among women.

5. Concluding remarks

Measuring hours actually worked with weekly estimates seems to be problematic. Due to the coarseness of measurement, minor deviations from normal hours of work are disregarded, such as beginning and ending work at slightly different times as well as performing work elsewhere than at the actual place of employment. Measurement functions differently for men and women as well as for wage and salary earners and the self-employed. Weekly estimates seem to measure experienced hours worked. Men in Finland seem to work slightly longer actual hours than their experienced working hours reveal. This emerges when comparing weekly estimates to a more accurate measurement method, the time budget. Women, on the contrary, perform less actual work than they experience.

Differences stemming from the method employed appear greater in the hours of work for the self-employed than for wage and salary earners. This is naturally a consequence of the fact that the working hours of wage and salary earners are more regulated and controlled than those of the self-employed. The twin-peaked working day of women engaged in agriculture is caused by the fact that working hours are experienced as being longer than the active hours used for gainful employment are in reality. Household obligations are sensed as a part of work. Measured with the weekly estimate, the working hours for female own-account farmers are 67 per cent longer than those measured with the time-budget method, which is able to segregate with accuracy time spent on gainful employment from time used for domestic work.

The boundary between gainful employment and domestic work is also blurred for other women performing gainful employment at home.

International statistics on labour input are mainly based on weekly estimates, which vary in measurement accuracy according to the character of work. Measurement is more precise for a regular salaried job than for the irregular hours of work done by the self-employed. The accuracy of measurements performed by weekly estimates should be improved by means of time-diaries in order to achieve greater validity for statistics on working hours.

6. References

Askelo, S.: Lärarnas arbetsmängd 1979, Studies No. 66, Central Statistical Office of Finland, Helsinki 1982.

Labour Force Survey 1979, Central Statistical Office of Finland, unpublished tables.

Niemi, I., Kiiski, S. & Liikkanen, M.: Use of time in Finland 1979, Central Statistical Office of Finland, Studies No. 65, Helsinki 1981.

Robinson, John P.: How Americans Use Time, A Social Psychological Analysis of Everyday Behaviour, New York 1977.

Robinson, John P.: Alternative Methods of Measuring Time-Use. Chapter 3. Unpublished Monograph from 1975 Time Use Study. Survey Research Center, Institute for Social Research, the University of Michigan. Ann Arbor 1982.

Time-Use Study 1979, Central Statistical Office of Finland, unpublished tables.

Meddelanden

Europeiska statistiska möten

Metodfrågor i miljöstatistiken (ECE)

Inom ECE har startats en serie informella möten om metodproblem i miljöstatistiken. Efter förberedelser i Genève 1981[1] hölls det första mötet i Helsingfors 2–5 maj 1983 med deltagande från åtta länder. De svenska deltagarna var avdelningschef Knut Medin och avdelningsdirektör Thomas Polfeldt, SCB.

En kanadensisk rapport om metoder för att identifiera miljöindikatorer gav upphov till en livlig diskussion om hur miljöstatistiken kan läggas upp. I en rapport om miljötillståndet, som nu förbereds i Kanada, vill man bygga på en grundindelning i ett femtontal ekologiska zoner. Man tar upp olika typer av miljöpåverkan i dessa (där t. ex. industriutsläpp, lantbruk, väg- och vattenbygge ingår) och avläser effekterna på olika sätt. Ett system med sju generella typer av biologiska reaktioner till följd av miljöpåverkan (bl. a. förändringar i arternas mångfald och storleksfördelning, näringskoncentration, ändrad primärproduktivitet) är av stort allmänt intresse. Försök göres att på detta sätt tydligare belysa samverkan mellan olika faktorer.

Från Finland och Sverige förelåg rapporter om olika aspekter av användning av slumpmetodik i miljöstatistiska undersökningar. En sammanställning i tabellform av existerande sådana undersökningar och deras karakteristika kommer efter svenskt mönster att genomföras av andra deltagarländer. En stor del av diskussionen kretsade kring användning av övervakningsdata, varvid vi bl. a. tog upp stratifieringsteknik och möjligheten att tolka mätpunkterna som slumpmässigt utlagda.

En tredje programpunkt behandlade mätstationsdata som sådana – problem med analysfel och kontroll av mätningarnas tillförlitlighet, föroreningshalter under mätbarhetsgränsen, samt redovisning av sådana data diskuterades, med utgångspunkt i bidrag från Frankrike och USA.

Nästa möte i serien planeras till våren 1984. Man avser då att speciellt behandla (i) generaliseringsproblem (bl. a. ytterligare aspekter på hur man kan gå från övervakningsdata till relevant statistik), och (ii) metoder för att länka ekonomisk statistik och miljöstatistik.

Thomas Polfeldt

[1] Se Statistisk tidskrift 1982: 1, sid 56.

Nordiska statistiska utskott

Utskottet för befolkningsstatistik (NUB)

Det nordiska utskottet för befolkningsstatistik (NUB) höll ett möte i Stockholm den 18–20 januari 1983. I mötet deltog – helt eller delvis – för Sverige Curt Nilsson, Arne Arvidsson, Anne-Marie Bolander, Rolf Dahlström, Lars-Gunnar Hörte och Åke Nordin.

Utskottet deltog också i det möte som SCB arrangerade med anledning av att det 1983 är 200 år sedan Pehr Wargentin – den svenska statistikens fader – dog.

Migrationsstatistiken

Utskottet diskuterade statistiken över migrationen mellan de nordiska länderna. Utgångspunkt för diskussionerna var ett finskt dokument rörande de nordiska ländernas migrationssiffror för åren 1975–1981.

De förbättringar av kunskapen om migrationsstatistiken som tidigare föreslagits ske genom utbyte av individlistor över in- och utflyttade i resp. land hade via korrespondens diskuterats inom utskottet. En väg att gå vidare med projektet ansåg utskottet vara att NUB tar kontakt med den samar-

betsgrupp som bildats mellan cheferna för befolkningsregistren i de nordiska länderna och med dem diskuterar hur man praktiskt bör gå tillväga för att kunna genomföra projektet.

Statistik över samlevnad utan äktenskap

Utskottet diskuterade statistiken över samlevnad utan äktenskap.

I Danmark används omnibusundersökningar för att belysa antalet samboende utan äktenskap. Hittills har antalet sådana samboendeförhållanden varit stadigt stigande. Genom den nya familjestatistiken kommer det att bli möjligt att via register få information om ogifta samboende med gemensamma barn. Det är också möjligt att registervägen få information om personer av olika kön som bor på samma adress och som inte är gifta. Däremot kan man inte registervägen särskilja t.ex. samboende syskon. De uppgifter Finland publicerar över samboende utan äktenskap grundar sig på SCs arbetskraftsundersökningar. Uppgifterna från dessa undersökningar tyder på att ca 10% av alla sammanboende ej är gifta. I Island registrerade befolkningsregistret tidigare om man var sammanboende eller ej. Numera registreras endast samboende utan äktenskap med gemensamma barn. I FoB 80 ställdes en fråga om samboende utan äktenskap. Norges siffror över samboende utan äktenskap är låga i jämförelse med de övriga nordiska länderna enligt de resultat som framkommit vid den senaste folkräkningen.

I Sverige överensstämmer definitionen av samboende utan äktenskap i FoB 80 med den som använts i Norge, dvs. de skall vara registrerade på samma adress. Vid de undersökningar om befolkningens levnadsförhållanden (ULF) som utförs regelbundet ställs en fråga om samboendeförhållanden. En jämförelse mellan resultat från ULF-undersökningarna och FoB 80 visar att FoB 80 redovisar färre samboende utan äktenskap än ULF-undersökningarna (18% i ULF mot 15% i FoB 80).

Ett gemensamt nordiskt problem tycks vara att många samboende ej anmäler flyttning utan förblir skrivna på var sin adress vilket leder till att registerbearbetningar ger färre antal samboende än de urvalsundersökningar som mäter faktiskt samboende.

Fruktsamhetsutvecklingen

Varje land lämnade en kortfattad redogörelse över fruktsamhetsutvecklingen.

I Danmark har fertiliteten minskat med hela 7–8% mellan åren 1980–1981. I Finland ökade antalet födslar under 1982 med 2–2,5%. Island har fortfarande den högsta nativiteten i Norden. Det har dock skett en minskning av antalet födda barn per kvinna från 2,9 år 1979 till 2,3 år 1981. Norge har haft en relativt stabil fruktsamhet efter 1977, dock med en tendens till en svag nedgång.

Sverige lämnade en kort redogörelse om fruktsamhetsutvecklingen i västvärldens industriländer under 1900-talet. Fruktsamheten sjönk under första delen av 1970-talet från nivåer kring 2,3–2,6 barn per kvinna till 1,6–1,8 år 1975. Därefter har fruktsamheten legat i stort konstant eller visat en svag ökning. Undantag är Danmark och Västtyskland.

Specialdatabaser

I Finland finns det i dag tre databaser som kan användas inom befolkningsstatistiken:
- Prognosdatabaser
- Databasen över den svenskspråkiga befolkningen i Finland
- Regionaldatabasen.

Norge har i dag en databas för folk- och bostadsräkningen 1980. Härutöver planeras ett antal nya databaser:
- SSBs centrala personregister
- En regionalstatistisk databas
- En kommunalekonomisk databas
- Databeredskap för persondata
- SSB har dessutom levererat data till två tabelldatabaser hos Norsk samfunnsvitenskapelig datatjeneste – kommundatabasen och kretsdatabasen.

Danmark, Island och Sverige skall efter mötet skriva ett kort inlägg där databaserna och deras användningsområden beskrivs.

Yrkes- och dödlighetsundersökningar

I samband med NUB-mötet sammanträffade experter inom området yrke/dödlighet. Gruppen utväxlade information och erfarenheter av kommande och utförda undersökningar.

Utskottet diskuterade möjligheterna att åstadkomma en gemensam nordisk dödlighetstabell och att göra en gemensam nordisk publikation avseende yrke/dödlighet under perioden 1971–1980. Sekretariatet fick i uppdrag att närmare undersöka de praktiska detaljerna kring det senare förslaget.

Uppföljning av FoB 80

Folk- och bostadsräkningen i Danmark utfördes den 1 januari 1981. Insamlingen, som var helt registerbaserad, gjordes från åtta olika register. Folk- och bostadsräkningen i Finland ägde rum den 1 november 1980. Frågeformulären skickades till varje hushåll som sedan i sin tur skickade svaret till kommunens insamlingsställe. Materialet kodades därefter i kommunerna. Folk- och bostadsräkningen i Island ägde rum den 31 januari 1981. I Norge skedde optisk registrering av alla uppgifter från FoBen med undantag för näring, yrke och arbetsställe. Genom att indela landet i ett antal grupper och på det förtryckta svarskuvertet ge var och en av dessa grupper en egen postbox fick man genom postverkets försorg en första grovsortering av materialet.

I samband med arbetet på FoB 80 i Sverige infördes en hel del nyheter jämfört med tidigare folk- och bostadsräkningar:

- Löpnummerkodning av materialet. Metoden, som närmare beskrivs i ett Metodmeddelande utgivet av SCB, innebär att för varje förvärvsarbetande person anges ett löpnummer till personens arbetsställe. Kodningen har därvid gjorts i kommunerna. Vid en samkörning med företagsregistret har sedan uppgift hämtats om näringsgren, sektortillhörighet och arbetsplatsens belägenhet.
- Samordning med fastighetstaxeringen. Bostadsdata hämtades från fastighetstaxeringen. Denna del av FoB 80 ledde till att extra insatser behövde göras för att kvaliteten på resultaten inte skulle bli alltför dålig. Trots detta kvarstår vissa kvalitetsproblem i bostadsstatistiken från FoB 80.
- Automatisk yrkeskodning. Metoden innebär att yrkesuppgiften i klartext registreras varefter materialet matchas mot ett yrkeslexikon. Utifrån detta sätts därefter koden. I ca 70% av alla fall har metoden fungerat utan manuellt uppföljningsarbete.

Vid FoB 80 har SCB vidare satsat en hel del på grafisk presentation av materialet. Ett speciellt system har konstruerats för detta ändamål, Graf-80. SCB har också bildat en speciell arbetsgrupp för marknadsföring av FoB-data.

Specialpublikationer från FoB 80

Förslaget att göra en gemensam nordisk publikation med data från FoB 80 hade godkänts av chefsstatistikermötet. Publikationen kommer endast att innehålla uppgifter på riksnivå. Däremot kommer årsrapporten från basprojektet inom den nordiska ämbetsmannakommittén för regionalpolitik att innehålla tabeller med regionala data. Tabellinnehållet i dessa båda publikationer diskuterades vid utskottsmötet.

Longitudinella studier på basen av folkräkningarna 1960–1980

På grund av såväl jämförbarhets- som resursproblem har man i Danmark för närvarande ansett sig tvungen att lågprioritera arbetet med longitudinella undersökningar. I Finland hade Statistikcentralen ur 1960 års blankettmaterial tagit ett urval som sedan samkörts med FoB 70. Uppföljning med uppgifter från 1975 och 1980 pågår. I Island har hittills icke utförts några longitudinella undersökningar. Norge införde personnummersystemet i mitten på 1960-talet, som i efterhand påfördes FoB 60-materialet. Arbete pågår för att under 1983 få fram en publikation med jämförelser mellan folkräkningarna 1960, 1970 och 1980.

Efterfrågan på longitudinella undersökningar ökar starkt i Sverige, men för närvarande pågår inget särskilt projekt. Det utförs dock en hel del undersökningar på upp-

dragsbasis där folk- och bostadsräkningarna 1960, 1965, 1970, 1975 och 1980 används.

Nordisk statistisk årsbok

Utskottet diskuterade såväl definitionsmässiga skillnader i de nordiska ländernas befolkningsstatistik och dessas konsekvenser för årsboken som de enskilda tabellerna i årsboken.

Utskottets kommande arbete

Utskottets nästa möte planeras äga rum i Oslo under verksamhetsåret 1984/85. På dagordningen till detta möte planeras följande ämnen ingå:

- löpande befolkningsstatistik (fruktsamhet, dödlighet, migration och civilståndsförändringar)
- yrke/dödlighet
- longitudinella studier över befolkningen
- folk- och bostadsräkningar
- nordisk statistisk årsbok
- presentation och bearbetning av hushållsstatistik (eventuellt)
- statistik över boendeförhållanden.

Därutöver beslöt utskottet föreslå att tre speciella möten skall få hållas, nämligen

- expertgruppsmöte om yrke/dödlighet (våren 1984)
- kontaktmöte om registerbaserade folk- och bostadsräkningar (hösten 1983 eller våren 1984)
- kontaktmöte om grafisk presentation av statistik.

Curt Nilsson

Utskottet för miljöstatistik (NUM)

Utskottet för miljöstatistik (NUM) sammanträdde den 8–9 juni 1983 i Sundvollen, Norge. I mötet, som hölls i anslutning till ett nordiskt seminarium om vattendata och vattenstatistik[1], deltog från Sverige avdelningschef Knut Medin och statistikchef Bengt Malmås, SCB.

Vid mötet lämnades kortfattade redovisningar av utvecklingsläget i stort inom miljöstatistiken i resp. länder. Det framgick av dessa bl. a. att man i Danmark för närvarande inte har möjlighet att bedriva något mera omfattande arbete på området. Vissa arbeten pågår dock inom markanvändningsstatistiken, varifrån lämnades en del intressanta uppgifter om de problem man stött på vid användningen av administrativa register som datakällor. Från finsk sida upplystes om att Statistikcentralens arbete med miljöstatistiken förlagts till en nyinrättad byrå för levnadsförhållandestatistik. I planen för arbetet under de närmaste två åren ingår bl. a. att publicera en ny årgång av "miljöstatistisk årsbok" och en rapport med de uppgifter om olika miljöförhållanden som Finland redovisat till OECD-projektet the State of the Environment. I Norge är nya utgåvor av "miljöstatistisk årsbok" och en publikation med referenser till naturresurs- och föroreningsdata under färdigställande för publicering andra halvåret 1983. I övrigt är det norska arbetet i stor utsträckning koncentrerat till utvecklingen av "resursräkenskaper" för bl. a. mark, vatten, skog och energi. Från svensk sida slutligen informerades om arbetet på en ny utgåva av Miljöstatistisk årsbok hösten 1983 och om det övriga utvecklingsarbetet som tar sikte på att förbättra den officiella statistiken över emissioner, luft- och vattenkvalitet m. m.

Formerna för uppföljning av arbetet med en "nordisk arealklassificering" diskuterades vid mötet mot bakgrund av att ett preliminärt klassificeringsförslag nu används försöksvis i de fyra länderna. Det uppdrogs åt en tidigare arbetande "kontaktgrupp" för arealklassificeringsarbetet att till hösten 1984 genomföra en utvärdering och åt Sverige att svara för sekretariatsfunktionen i detta arbete.

Till mötet hade länderna utarbetat skriftliga rapporter om vilka geografiska områdesindelningar som används (eller övervägs) inom miljöstatistiken i det egna landet. Vid diskussionen kring dessa redovisningar konstaterades bl. a. att de s. k. avrinningsområdena (flodområdena) kunde komma att bli ett centralt redovisningsbegrepp och att dessa därför kunde behöva uppmärksammas särskilt i det nordiska statistiksamarbetet.

Olika frågor som rör klassificeringen av

[1] Se särskild redogörelse sid. 335 f.

avfall, i första hand det s.k. fasta avfallet, diskuterades med utgångspunkt i en rapport från Finland. Diskussionen föranledde inget särskilt ställningstagande från utskottet, men det konstaterades rent allmänt att det, bl.a. som förberedelse för ett senare ECE-arbete på området, var värdefullt att frågan om den statistiska redovisningen av avfall hade aktualiserats i utskottet och att man i Finland var beredd att arbeta vidare med klassificeringsproblematiken.

Andra frågor som diskuterades vid mötet gällde det övriga internationella samarbetet inom miljöstatistiken och den miljöstatistiska redovisningen i Nordisk statistisk årsbok. I anknytning till den förra frågan konstaterades att det internationella miljöstatistiska samarbetet på förhållandevis kort tid hade fått en betydande omfattning och att det numera krävdes ganska stora resurser för att aktivt kunna delta i denna samverkan.

I anslutning till att det internationella statistiksamarbetet berördes vid mötet informerades också om utvecklingen av det särskilda samarbetet i metodfrågor som förekommer mellan Finland, Ungern och Sverige.

Nästa möte skall enligt utskottets förslag hållas i Stockholm hösten 1984. Preliminärt har följande frågor uppförts på programmet för detta möte:

- Uppföljning av det nordiska arealklassificeringsprojektet
- Naturresursräkenskaper
- Erfarenheter av tidsserieredovisning
- Utnyttjande av industrins näringsgrensindelning i miljöstatistiken
- Avfallsklassificering
- Uppföljning av seminariet om vattenstatistik i juni 1983 (särskilt med avseende på möjligheten att redovisa statistik för avrinningsområden)
- Förberedelser för ett metodseminarium som NUM skall arrangera tillsammans med det särskilda utskottet för metodfrågor (KOMET)
- Uppföljning av OECDs miljörapporteringsprojekt (the State of the Environment).

I utskottets arbetsplan ingår vidare att undersöka behov och övriga förutsättningar för utgivning av en särskild nordisk miljöstatistisk rapport.

Bengt Malmås

Seminarium om vattendata

I regi av det nordiska utskottet för miljöstatistik och nordiska kontaktgruppen för miljödata[1] hölls ett seminarium om vattendata och vattenstatistik på Sundvollen i Norge den 6–8 juni 1983. Ett sextiotal nordiska deltagare lyssnade till och diskuterade drygt trettiotalet inlägg i det späckade programmet. Från SCB deltog avdelningschef Knut Medin, avdelningsdirektörerna Ingegerd Fängström och Thomas Polfeldt och byrådirektör Lars-Gösta Wahlberg.

Efter några inledande föredrag, där det bl.a. ingick ett referat från Nordforsks nittonde vattendatasymposium (maj 1983), var inläggen grupperade under fyra rubriker.

Register och databanker för vattendata finns i alla de fyra nordiska länderna. Med undantag för Finland, där vattenstyrelsen har ett samlat ansvar, är splittringen stor, t.ex. med hydrologiska data på ett ställe, vattenkvalitetsdata på ett annat, brunnsdata på ett tredje osv. Arbete på att samordna de olika registren pågår emellertid och några föredrag tog upp problem i detta sammanhang. I Norge har man åstadkommit en god överensstämmelse mellan gränserna för avrinningsområden och s.k. grundkretsar, som är de minsta areella enheterna för administrativt och statistiskt bruk. I det svenska bidraget underströks bl.a. att statistisk användning av data ofta ställer särskilda krav på registren, och att en medvetenhet om detta vid registeruppbyggnaden behövs.

Ett genomgående intryck vid seminariet var att det på länsnivå finns stor datatillgång och möjligheter att göra avancerade bearbetningar. I Telemark fylke i Norge har man t.ex. i samarbete med distriktshögskolan i länet startat ett omfattande projekt för en vattenatlas över länet. Åtta detaljerade tematiska kartor är hittills färdiga.

Under rubriken *databehandlingsmetoder*

[1] Se särskild redogörelse, sid. 334 f.

och statistiska metoder presenterades föredrag om trendanalys av vattendata, om fördelarna med stratifiering vid beräkning av vissa genomsnittsvärden för större sjöar och om begreppet "standardiserad flodkilometer" (där vattenföring och flödeshastighet beaktas) och dess användning. Detta begrepp, som tagits fram i Norge, har bl. a. använts försöksvis för att skaffa ett slumpmässigt stickprov från allt rinnande vatten i ett avrinningsområde. Vidare redogjordes för en norsk intervjuundersökning om skador på fisk vid vattenreglering och för det finska vattenövervakningsnätet. Som avslutning på denna avdelning kom ett inlägg som täckte flera av seminariets rubriker. Det handlade om vattendata i vattenanvändningsplanering, med översikter över de norska fylkenas uppgifter och databehov på området. Principer som bör gälla för användningen av vattendata i detta sammanhang formulerades.

Andra dagens eftermiddag ägnades åt föredrag med temat *databehandlingsmetoder – modeller*. I en dansk studie redovisades möjligheterna att förutsäga syrekoncentrationen i ett vattendrag med utgångspunkt från mätningar av vattentemperatur, ljusinstrålning och vissa hydrologiska variabler. Problem med att testa vattenkvalitetsmodeller med hjälp av existerande data togs upp i ett finskt föredrag. Bristen på data om föroreningsbelastningar ansågs särskilt besvärlig, men det behövs också en mer genomtänkt insamling av data om vattenkvalitet.

En resursräkenskap för vatten som skall kunna kopplas till nationalräkenskapssystemet och till resursräkenskaper för areal och energi beskrevs i ett norskt föredrag. En liknande modell har utnyttjats i den svenska undersökning som presenterades i föredraget "Vatten som begränsande faktor för ekologisk och ekonomisk produktion på Gotland". Med hjälp av input-output-analys belystes de olika ekonomiska sektorernas både direkta och indirekta beroende av vattentillgången.

För att kunna förutsäga effekter av ökade vattenuttag från grundvattnet har ett par olika danska hydrologiska modeller utvecklats. Med hjälp av geologiska, hydrologiska och meteorologiska data kan dagliga beräkningar av avdunstning, perkolation, grundvattennivåer och avrinning beräknas. Sista föredraget för dagen behandlade ett ADB-system för beräkning av föroreningsbelastningar. Systemet ansågs egentligen vara alltför bra i förhållande till existerande data. Även osäkra indata skall emellertid kunna användas om osäkerheten för dessa kan anges.

Vattendata/vattenstatistik i förvaltningssammanhang var ämnet för sista dagens föredrag. Ett projekt för samordning av den decentraliserade datainsamlingen i Danmark redovisades. För vattenstyrelsen i Finland utarbetas för närvarande ett nytt program för perioden 1985–1995. Från norsk sida presenterades ett program för övervakning av vattenkvalitet i fjordar och vattendrag. Statens forurensningstilsyn arbetar här med att förbättra metoderna för val av mätstationer, parametrar och mätfrekvens.

Hur ytvattnets beskaffenhet bedöms enligt ett antal olika kriterier beskrevs i en intressant redovisning från Kronobergs län. Två danska föredrag om system för vattenplanering inom olika amtskommuner avslutade det innehållsrika programmet.

Seminariet gav en mångsidig och stimulerande genomgång av olika vattendataproblem och var också ett utmärkt tillfälle för nya kontakter och informella diskussioner.

Ingegerd Fängström *Thomas Polfeldt*

Utskottet för utbildnings- och kulturstatistik

Nordiska utskottet för utbildnings- och kulturstatistik (NUUK) sammanträdde i Köpenhamn 4–6 maj 1983. Mötets två första dagar ägnades åt utbildningsstatistik och den tredje dagen åt kulturstatistik. Mötet hölls i Danmarks Statistiks lokaler (dag 1 och 2) och vid sekretariatet för nordiskt kulturellt samarbete (dag 3). Från Sverige deltog avdelningsdirektör Leif Gouiedo och statistikchef Olof Jos.

Mötets dagordning omfattade under dag 1 och 2 följande huvudpunkter.

1. Befolkningens utbildningsnivå. Jämförbarhet mellan registreringsrutiner och resultat.
2. Utbildningsindikatorer och sociala indikatorer på utbildningsområdet.
3. Orientering om nya statistikprodukter och statistikprojekt inom utbildningsområdet i de nordiska länderna.
4. Avsnittet om utbildning i Nordisk statistisk årsbok.

Under dag 3 togs upp bl. a. följande.

5. Avsnittet om kultur i Nordisk statistisk årsbok.
6. Nationella kulturstatistiska översiktspublikationer.
7. Nordiskt deltagande i UNESCOs arbete med kulturstatistik.

Under punkt ett redovisades hur arbetet med utbildningsuppgifterna gått vid folk- och bostadsräkningarna 1980 i Danmark, Finland och Norge. Sådana uppgifter hade inhämtats från befintliga register, ungefär på det sätt som föreslagits av SCBs s. k. FOBALT-utredning för framtida folk- och bostadsräkningar i Sverige. Ett gemensamt problem konstaterades vara att man ur befintliga register ej kunnat hämta uppgifter om invandrare från nordiskt land. En lösning på detta problem som diskuterats är att försöka få en bestämmelse om att uppgift om högsta utbildning skall tas in på det för de nordiska länderna gemensamma flyttningsbetyget. NUUK ansåg att denna möjlighet borde penetreras och att de nordiska chefsstatistikerna således borde låta utreda detta.

Man enades om att försöka ta fram en tabell för Nordisk statistisk årsbok med uppgifter om befolkningens högsta utbildning 1980 grundade på resp. lands folk- och bostadsräkningsmaterial. Fördelning skulle ske efter ålder och kön. För Sveriges del skulle motsvarande uppgifter hämtas från arbetskraftsundersökningen i februari 1980.

Utbildningsindikatorer och sociala indikatorer på utbildningsområdet var den punkt på dagordningen som upptog mest tid. Som förberedelse hade bl. a. gjorts en genomgång av utbildningsavsnitten i resp. lands publikationer med resultat och analyser från levnadsnivåundersökningarna. En slutsats av denna genomgång och diskussion vid mötet i anslutning därtill var att det enligt NUUK vore önskvärt med bättre samordning av t. ex. uppdelningar efter ålder. Här borde man följa de internationella rekommendationer som finns på utbildningsstatistikens område, och att det vore bra med en nordisk jämförelse som inledning till utbildningsavsnitten i resp. lands publikation. Som förberedelse för mötets behandling av frågorna om utbildningsindikatorer hade också utarbetats ett notat som refererade existerande teori inom detta område. Den allmänna meningen var att det för närvarande brister i teoretisk förankring av de indikatorer som används. Åtminstone borde förhållandet mellan teori och praktik belysas bättre. Man enades om att NUUK bör arbeta vidare med detta och att det inför utskottets nästa möte skall tas fram ett underlag innehållande förslag till indikatorer och hur dessa anknyter till en modell över utbildningssystemet och till ett antal teoretiska kriterier. Syftet är att NUUK skall kunna enas om att rekommendera ett antal indikatorer, nordiska nyckeltal, som är teoretiskt förankrade eller vars teoretiska egenskaper åtminstone är belysta.

Bland de nyheter inom utbildningsstatistiken som togs upp under punkt tre kan nämnas planerna i Finland på att upprätta en utbildningsstatistisk databas, för distribution av statistik, som skulle innehålla uppgifter från utbildnings-, befolknings- och arbetsmarknadsstatistiken m. m. Det gäller alltså en databas för utbildningssektorns planeringsbehov m. m. Från norsk sida orienterades om den omläggning av insamlingen av uppgifter för den löpande utbildningsstatistiken som skall ske på grund av att det nya systemet i Norge för intagning av elever i "vidaregående" skolor innehåller ajourhållna elevregister i varje enskilt fylke. I dessa regionala register ingår även individuppgifter om elever i grundskolans årskurs 9. Vidare orienterades om en undersökning av deltagare i vuxenutbildning som genomförs av SSB (Statistisk Sentralbyrå) och Norsk Voksenpedagogisk Institutt som en tilläggsundersökning till arbeidskraftsundersökelsen 2a kvartalet 1983.

Från svensk sida orienterades bl. a. om den nya statistiken över grundutbildningen för vuxna och satsningen på förbättring av publiceringen av utbildningsstatistiken.

Förslag till ett antal nya tabeller för utökning av innehållet om utbildning i Nordisk statistisk årsbok förelåg. Det gällde uppgifter om specialundervisning, lärare och utbildningskostnader. Det konstaterades på nytt, frågan har diskuterats vid tidigare NUUK-möten, att någon jämförbar statistik om specialundervisning och lärare för närvarande ej finns. Vad gäller specialundervisning synes det ej heller inom överskådlig tid komma att finnas eller vara möjligt att utveckla någon sådan statistik. Vad gäller lärarstatistik skulle vidare penetrering och överväganden ske. Om det är möjligt att i Nordisk statistisk årsbok få med några uppgifter om utbildningskostnader skulle också undersökas vidare genom att tabeller upprättas och genom att det redovisas hur resp. lands uppgifter därvid tagits fram. Detta skulle undersökas genom att nordiska statistiska sekretariatet inhämtar uppgifter från resp. land. Uppgifter om utbildningskostnader lämnas för närvarande till UNESCO, men NUUK ansåg inte detta innebära att uppgifterna härigenom automatiskt skulle anses vara "kvalificerade" för Nordisk statistisk årsbok.

Avsnittet om kultur i Nordisk statistisk årsbok togs således upp under tredje sammanträdesdagen. Det beslöts att ett antal kompletterande upplysningar skulle inhämtas och i tillämpliga fall i notform tas med i vissa av de föreslagna tabellerna. Bland föreslagna nya tabeller gällde en tabell antal föreningar och medlemsantal för olika idrotter. Här ansåg NUUK att ingen meningsfull nordisk statistik kunde uppnås med förslaget. Vad som däremot skulle vara av stort intresse vore om statistik kunde tas fram om utövande av olika idrotter och om storlek och fördelning av ekonomiska satsningar på olika idrotter. Andra förslag till nya tabeller som NUUK ej ansåg meningsfulla ur kulturpolitisk synvinkel gällde uppgifter om antal försålda exemplar av videobandspelare och fonogram.

NUUK har ansett det viktigt att nationella översiktspublikationer med kulturstatistik tas fram, vilket också skett, och i fortsättningen utarbetas löpande med viss periodicitet. Utskottet vill därutöver verka för att dessa publikationer samordnas i viss utsträckning. Man har därför enats om att eftersträva gemensamt publiceringsår, gemensam uppdelning på olika kulturområden och inkludering av en nordisk översikt i resp. utgåvor. Läget vad gäller dessa samordningssträvanden behandlades under punkt sex. Det konstaterades med tillfredsställelse att konkreta resultat synes möjliga att uppnå även härvidlag.

De nordiska länderna tar aktiv del i UNESCOs arbete med utveckling av kulturstatistik. Arbetet avser för närvarande utveckling inom olika delområden av ett Framework for Cultural Statistics. Arbetsläget diskuterades under punkt sju på dagordningen.

Mötet avslutades med en kort information från svensk sida om försöket att få till stånd ett samarbete om utbildningsstatistik på institutionell bas med folkrepubliken Kina och med en genomgång av ett preliminärt arbetsprogram inför nästa utskottsmöte, som avses hållas i Norge våren 1984.

Olof Jos

OECD-möten

Möte om miljöstatistik

Ett möte i OECDs arbetsgrupp om the State of the Environment hölls i Paris den 9–11 maj 1983. 18 nationer deltog. Från Sverige deltog Ulla Swarén, socialdepartementet, som också var ordförande för mötet, Jan-Erling Larsson, statens naturvårdsverk samt Inger Öhman, SCB.

Huvuddelen av mötet ägnades åt att diskutera dokument som utgör underlag för nästa rapport om the State of the Environment. År 1982 insamlades från medlemsländerna siffermaterial på omfattande frågeformulär belysande olika aspekter av naturmiljön. Detta material har nu delvis ställts samman av OECD-sekretariatet. Man har också engagerat ett antal konsulter som skrivit textavsnitt avsedda att vara underlag för den slutliga rapporten. Följande dokument

togs upp till diskussion:

- Land and the Environment
- Agriculture and the Environment
- Wild Life
- Marine Environment
- Solid Wastes.

Dessutom behandlades outlines till avsnittet om "Transport and the Environment" samt avsnitt om "Energy and the Environment".

Avsikten är att rapporten om the State of the Environment skall ges ut i slutet av 1984. Man bestämde vid mötet att det statistiska materialet skall ställas samman i preliminär form i ett särskilt statistiskt kompendium, vilket skall tillställas OECDs miljökommitté i slutet av 1983. Detta kräver dock en hel del arbete av såväl OECD-sekretariatet som medlemsländerna. Miljöstatistiken är outvecklad och detta faktum speglas naturligtvis i rapporteringen till OECD genom att uppgifter ofta saknas, att uppgifterna är grova uppskattningar eller att definitionerna varierar mellan länderna.

Vid mötet kom man vidare överens om att påbörja arbetet med en publikation om naturresurser. Där skall redovisas de franska och norska erfarenheterna av naturresursräkenskaper samt erfarenheter av att använda fjärranalys för miljöinformation.

Inger Öhman

Contact 83

CONTACT 83 var en mässa för marknadsförare som ägde rum i Svenska Mässans lokaler i Göteborg 31 maj–3 juni 1983. SCB deltog och disponerade en monter på 20 m^2, som bemannades av sammanlagt nio personer. Utställarnas antal var ca 75 och antalet besökande ca 4 500, varav åtskilliga besåg SCBs monter. För ca 700 besökare fick SCB en närmare kontakt genom att dessa visade sig intresserade av våra erbjudanden.

I första hand gjorde SCB reklam för marknadsprofiler och annan marknadsstatistik, för centrala företagsregistret samt för SCB som allmän konsult. Antalet intresserade blev långt över förväntan och de på förhand uppställda målen för deltagandet i mässan uppnåddes mer än väl.

De ca 700 adresserna vi fick kommer nu att tas om hand för återkontakt. Det gäller att vidarebearbeta dem till uppdragskunder. I några fall har man "sålt" direkt vid montern, och dessa kommer omgående att få en offert på det efterfrågade.

Arne Orrgård

SCBs yttrande över underhållsfondsutredningens slutbetänkande "Återinvestera i bostäder" (SOU 1982: 65)

Utredningen föreslår i ett principbetänkande ett nytt finansieringssystem för det periodiska underhållet av hyres- och bostadsrättslägenheter. Systemet innebär en samordnad belåning av nybyggnad, ombyggnad och periodiskt underhåll. Vidare redovisas alternativa modeller för hyresgästinflytande.

SCB är central myndighet för den statliga statistikproduktionen med uppgift att utarbeta statistik, föra register och göra prognoser. Mot denna bakgrund begränsar sig verket i sitt yttrande till frågor som anknyter till denna verksamhet.

Utredningen har i sitt arbete använt statistik från bl. a. SCB för att belysa kostnadsutvecklingen för periodiskt underhåll. Man konstaterar därvid att intäkts- och kostnadsundersökningen (IKU) inte ger ett tillräckligt bra underlag för detta syfte. Möjligheterna att åstadkomma en bättre redovisning är emellertid begränsade, speciellt för enskilt ägda fastigheter. Ägarna till dessa behöver inte göra någon särskild redovisning för periodiskt resp. löpande underhåll i samband med hyressättning. Skattemyndigheterna ställer inte heller krav på sådan uppdelning.

Utredningen säger vidare att det för närvarande inte finns några tillfredsställande pris- eller kostnadsindex för periodiskt underhåll. För index som är aktuella i detta sammahang vill SCB meddela följande.

På byggnadsindexområdet pågår för närvarande en översyn av entreprenadindex för husbyggnadsentreprenader (H 63). Av-

sikten är att H 63 under 1983 och 1984 skall ersättas av en reviderad entreprenadindex, benämnd H 83. H 83 kommer förutom nybyggnader av bostadshus även att omfatta ombyggnader.

Så snart som översynen av entreprenadindex är genomförd avser SCB att företa en genomgripande översyn av faktorprisindex för bostäder, som i huvudsak baseras på entreprenadindex.

I samband med den kommande översynen av faktorprisindex för bostäder avser SCB att närmare undersöka förutsättningarna för en revision av indexen för underhållskostnader. En index av typ faktorprisindex bedöms även fortsättningsvis ligga närmast till hands. Det innebär således en sammanvägd index för ingående produktionsfaktorer och utan hänsyn till produktivitets- och vinstutvecklingen. Vidare förutsätts oförändrad byggnadsteknik- och organisation.

Framtagandet av en ny och bättre faktorprisindex för underhållskostnader hänger nära samman med möjligheterna att få fram underlag för en någorlunda realistisk kostnadsbudget. Fastighetsunderhållets olika delar utförs relativt utspritt över tiden och insatsen av produktionsfaktorerna kan skilja sig kraftigt från projekt till projekt. Det medför problem när det gäller att entydigt fastställa omfattningen av olika komponenter i en tänkt indexbudget för periodiskt underhåll. Prisuppgifterna till en eventuell index förutsätts kunna bygga på befintlig prisinsamling.

Nuvarande faktorprisindex för nybyggnation av bostäder torde tills vidare kunna användas som mått på underhållskostnadernas utveckling. Detta gäller i synnerhet vid jämförelser över längre tidsperioder. Faktorprisindex har nämligen visat sig relativt okänslig för förändringar i budgeten.

Skördestatistiska nämnden

Fr. o. m. 1 juli 1983 består den skördestatistiska nämnden av följande personer: generaldirektör Sten Johansson (ordf.), agronom Eskil Arvidsson, lantbruksdirektör Gunnar Jeppsson, statsagronom Bertil Johnsson, byråchef Hans Lebert, avdelningschef Knut Medin (v. ordf.) samt direktör Erik Swedborg.

Personliga suppleanter är statistikchef Rune Tryggveson (för Johansson), agronom Bengt-Erik Jansson (för Arvidsson), byråchef Lennart Lindqvist (för Jeppsson), departementssekreterare Rein Karm (för Johnsson), avdelningsdirektör Ulf Tillgren (för Lebert), statistikchef Per Olof Olofsson (för Medin) samt civilekonom Ulf Brangenfeldt (för Swedborg).

Personalnotis

Tidigare redaktören för Statistisk tidskrift Thore Grönqvist har fr. o. m. 1 september 1983 gått i pension. Som vikarie på Thore Grönqvists tjänst har utsetts Nils Welander.

Summary in English

Two cartographic methods based on the delimitation of urban areas at Statistics Sweden

by H. Ylander

The relation between maps and statistics is twofold. Statistical information can be presented in maps, and maps can be used as a source of statistical information. In the first case we talk about thematic maps. The article deals with the second case where land use information in maps is transformed into statistical tables. The types of land use studied are:

1. The expansion of urban settlement on agricultural land.

2. The areas accessible to the urban population for recreational purposes in private property forests near the cities.

As the two methods described are based on the delimitation of all urban areas (localities) in Sweden performed in connection with the population and housing censuses, an introduction is given of the statistical coverage of rural-urban settlement during the last 100 years. Since the 1960 census, the technique has been improved by the use of largescale maps for the delimitation of all urban areas.

The material used in both cases is the delimitation lines of the urban areas drawn on transparent sheets in the scale 1:10000, in combination with the economic map, printed in colour and in the same scale.

In the first study of urban settlement on agricultural land the transparent sheets contained both the borderlines of the 1975 census and the borderlines of the 1980 census. By superimposing these sheets on the corresponding economic map, it was possible to determine the land use in the expansion zone between the 1975 and the 1980 delimitation line of the urban settlements.

The use of these maps makes it possible to observe whether the urban expansion has taken place on arable land or on some other type of land and to measure the areas in question by means of a planimeter. This kind of statistical information makes it possible to observe if the restrictions given by the government in 1977 concerning further urban settlement on agricultural land have had the intendet effect in different parts of the country.

In the second study the aim has been to measure the areas accessible to the urban population for recreational purposes in the neigbourhood of the cities. These areas have been defined – with some simplification of reality – as the green (nonagricultural) areas on the economic maps lying in one-kilometre-zones outside and parallel to the delimitation lines of the urban areas. With the use of a transparent grid sheet furnished with the parallel zones and superimposed on the economic map, the points in the grid that fall on green areas are recorded as accessible for recreational puposes. In addition, some water areas (lakes) and beaches within the zones are also recorded as areas of special interest.

The Swedish Town Population in the 18th Century; 68 Towns by Occupation and Social Status

by Y. Fritzell

The Statistical Committee ("Tabellverket"), believed to be the oldest statistical government agency in the world, commenced its work in 1749. Sixty years earlier, the clergy had been instructed to keep registers on marriages, births and deaths in their parishes. They were now charged with

the task of tabulating these data, together with data on social status (estate) and occupation, and regularly submitting the tables to higher authorities for further tabulations up to the national level. For the first few years the population statistics were kept secret, but since the 1760's they were published, albeit irregularly, in the records of the Royal Academy of Sciences. Publication became regular only in the first years of the 19th century.

At the beginning, the tables on social status and occupation were submitted every three years, but since 1775 this was done with five-year intervals. The tables, which also included a breakdown of the population by age and sex, passed through many intermediaries before national tables could be compiled. Thus the Governor of each province made summaries for his area and then passed these on to Stockholm, where national tables were prepared by the Statistical Commission. Urban and rural areas were combined in the provincial summaries, and not until 1805 were specific data for each town published. As a consequence of this and although the history of especially many big towns, including their demographic development, has in fact been written, there exists no total description of the occupational structure of the towns during the 18th century.

It has been the author's aim to fill in this gap. From the regional archives I have obtained copies of the primary tables for the towns and compared these with taxation registers for the same periods and areas. As some tables have unfortunately disappeared, the study only covers 68 out of the 83 towns which existed at the time. To avoid being swamped by the material I have selected two periods, one from the first years of the work of the Statistical Committee, around 1751, and one around 1772. The findings have also been compared with the statistics of 1805.

One of the difficulties I have encountered has been that for almost half the number of the towns the civil registration records also covered a rural district. In these cases I have had to make an estimation of the town's share of some of the occupations, the others being either urban or rural.

The classification by estates was important during the period studied. In addition to occupation proper, "gentry" was also distinguished as a group, although there were no clear-cut definitions even in the 18th century. Paupers and invalids constituted another category outside the occupations.

In the article some ten major occupational groups have been distinguished. Both the gentry and the occupations have been broken down by adults, youths and children, all of them subdivided by sex.

During the second half of the 18th century, economic conditions were generally declining. The number of servants increased more than other population groups, and manufacturing, which had flourished before 1750, receded (mining was a world business but located almost entirely in the rural areas).

An estimate is made of the total urban population in 1751 and 1772. The towns had 10 percent of the people.

Translation of words in the tables

Yrken	Occupations
Städer	Towns
Män	Men
Kvinnor	Women
Folkmängd	Population
Mantalslängd	Taxation list
Mantalsfri	Free of taxation
1 Ståndspersoner (stånd)	Gentry
2 Deras ungdom	Their youths
3 Deras barn < 15	Their children

4	Handelsidkare, rådmän	Commerce
5	Deras bodsvenner/ungd.	Their clerks and youths
6	Deras barn < 15	Their children
7	Fabrikörer	Manufacturers
8	Deras arbetare/ungd.	Their workers and youths
9	Deras barn < 15	Their children
10	Soldater och båtsmän	Soldiers
11	Skeppare och sjömän	Skippers and sailors
12	Hantverkare	Master artisans
13	Deras gesäller	Journeymen
14	Deras lärlingar 15 <	Apprentices
15	Fria arbetare m. fl.	Free workers, etc.
16	Publika staters betj.	Lower officials
17	Föregående ungdom	Youths among soldiers, artisans etc.
18	gruppers barn < 15	Children among soldiers, artisans etc.
19	Ståndens betjän. 15 <	Servants with the gentry
20	Andras tjänstefolk 15 <	Other servants
21	Tjänstefolk < 15	Servants < 15
22	Fattiga, lytta m. fl.	Paupers, invalids etc.

Litteratur

Nyare svensk och utländsk litteratur

Nyförvärv (urval vid SCBs bibliotek oktober 1982–mars 1983)

Enskilda arbeten

Applied time series analysis. Proceedings of the international conference held at Houston, Texas, August 1981/ed. by O. D. Anderson and M. R. Perryman. – Amsterdam; New York: North – Holland Pub. Co.; New York: Sole distributors for the U. S. A. and Canada, Elsevier Science Pub. Co., 1982. – viii, 528 p.: ill. Papers from the second American Time Series Meeting (5th ITSM), Houston, Tex., Aug. 6–7, 1981. – Includes bibliographies ISBN 0-444-86424-5 (2D/698)

Bergman, Lars R., [1945–], Om datainsamling i surveyundersökningar: vilken metod är bäst och vad får vi egentligen veta?/Lars R. Bergman, Bo Wärneryd. – Stockholm: Statistiska centralbyrån: Liber/Allmänna Förlag (distr.), 1982. – 241 s.: diagr., ill., tab. ISBN 91-38-07233-5 (2D/708)

Bergström, Reinhold, Prognoser i praktiken/Reinhold Bergström. – 1. uppl. – Stockholm: LiberFörlag, 1982. – 162 s.: diagr., tab. Omslagstitel: Statistik: prognoser i praktiken ISBN 91-38-60002-1 (2D/697)

Boen, James R., The human side of statistical consulting/James R. Boen, Douglas A. Zahn. – Belmont, Calif.: Lifetime Learning publ., 1982. – 196 s. ISBN 0-534-97949-1 (2D/710)

Büringer, Helmut, Nonparametric sequential selection procedures/H. Büringer, H. Martin, K.-H. Schriever. – Boston: Birkhäuser, cop. 1980. – 488 s. ISBN 3-7643-3021-X (2D/727)

Classifying social data: [new applications of analytic methods for social science research]/Herschel C. Hudson and associates [Mark S. Aldenderfer ...]. – 1. ed. – San Francisco, Calif.: Jossey-Bass, 1982. – 270 s.: ill., tab. – (Jossey-Bass social and behavioral science series, NBN 99-0150100-8) S. 231-259: References ISBN 0-87589-505-0 (2D/687)

Devijver, Pierre A., [1938–], Pattern recognition: a statistical approach/Pierre A. Devijver and Josef Kittler. – Englewood Cliffs; London: Prentice-Hall, cop. 1982. – xiv, 448 s.: ill. Includes bibliographies and index ISBN 0-13-654236-0 (2D/716)

Edwards, Chris, [1947–], Developing microcomputer-based business systems/ Chris Edwards. – Englewood: Prentice-Hall in assoc. with the Institute of cost and management accountants (ICMA), cop. 1982. – 202 s. Bibliography: p. 196–198. – Includes index ISBN 0-13204560-5 (6/781)

Encyclopedia of statistical science ... Vol. 2: Classification to Eye estimate. – cop. 1982. – 613 s. ISBN 0-471-05547-6 (Oi)

Forsman, Gösta, Effektiva strategier för estimation av förändringar och nivåer vid föränderlig population/av Gösta Forsman och Tomas Garås. – Stockholm, 1982. – 51 s. – (Promemorior från P/STM, NBN 99-0223458-5; 7) NBN 99-0354913-X (PIA/87-7)

Govindarajulu, Z., The sequential statistical analysis of hypothesis testing, point and interval estimation, and decision theory/Z. Govindarajulu. – Columbus, Ohio: American Sciences P., cop. 1981. – 680 s. – (American series in mathematical and management sciences; 5) Includes indexes. – Bibliography: p. 629–667 ISBN 0-935950-02-8 (2D/718)

Hamburger, Henry, Games as models of social phenomena/Henry Hamburger. – San Francisco: Freeman, cop. 1979. – xi, 264 s.: ill. S. 255–258: Bibliography ISBN 0-7167-1010-2 (2D/717)

Harvey, Joan Margaret, [1918–], Statistics

Europe: sources for social, economic and market research/Joan M. Harvey. – Ed. 4, rev. and enl. – Beckenham: CBD Research, 1981. – xiv, 508 s. Previous ed.: 1976. – Includes index ISBN 0-900246-36-7 (Ref Aaoi)

Heyer, Herbert, Theory of statistical experiments/H. Heyer. – New York; Berlin: Springer, cop. 1982. – 289 s. – (Springer series in statistics, NBN 99-0204827-7) New ed. of a book, Mathematische Theorie statistischer Experimente, cop. 1973 ISBN 0-387-90785-8 (2D/706)

Holst, Ulla, [1946–], Recursive M-estimators with dependent observations/Ulla Holst. – Lund, 1982. – [6], 10 s. Diss. (sammanfattning) Lund: Univ. – Härtill 3 uppsatser. – Rapportkod: LUTFD2/(TFMS-1001)/1-90/(1982) NBN 99-0315613-8 (2D/696)

Jacobsen, Martin, Statistical analysis of counting processes/Martin Jacobsen. – New York; Berlin: Springer, cop. 1982. – 226 s. – (Lecture notes in statistics, NBN 99-0198235-9; 12) ISBN 0-387-90769-6 (2D/711)

Klaassen, C. A. J., Statistical performance of location estimators/C. A. J. Klaassen. – Amsterdam: Mathematisch centrum, 1981. – 100 s. – (Mathematical centre tracts, ISSN 0569-9657; 133) ISBN 90-6196-209-9 (2D/719)

Lantbruksekonomiska samarbetsnämnden. Sektorsgruppen, Utredning om de objektiva skördeuppskattningarnas omfattning/Lantbruksekonomiska samarbetsnämndens sektorsgrupp. – Stockholm: Nämnden, 1982. – [200] s. (var.pag.). NBN 99-0346135-6 (2D/693)

Lindell, Jan, [1938–], Mätning av upplevelser i arbetsmiljön: om kartläggning av psykosocial arbetsmiljö/Jan Lindell. – Stockholm: Management media, 1982. – 88 s.: diagr., tab. ISBN 91-7722-000-5 (2D/704)

Lohse, Heinz, Prüfungsstatistik: ein programmierter Lehrgang/Heinz Lohse und Rolf Ludwig. – 1. Aufl. – Leipzig: VEB Fachbuchvlg, cop. 1982. – 236 s. + Beiheft. NBN 99-0345714-6 (2D/689)

The making of statisticians/ed. by J. Gani. – New York; Berlin: Springer-Vlg, 1982. – 263 s.: ill. ISBN 0-387-90684-3 (2D/700)

Martz, Harry F., Bayesian reliability analysis/Harry F. Martz and Ray A. Waller. – New York: Wiley, cop. 1982. – 745 s. – (Wiley series in probability and mathematical statistics, ISSN 0271-6232) ISBN 0-471-86425-0 (2D/720)

Measuring social judgments: the factorial survey approach/Peter H. Rossi, Steven L. Nock ed. – Beverly Hills: Sage, cop. 1982. – 255 s. Includes bibliographies ISBN 0-8039-1816-X (2D/715)

Mitchell, Brian R., International historical statistics: Africa and Asia/B. R. Mitchell. – London: Macmillan, 1982. – 761 s. – (Macmillan reference books) ISBN 0-333-31630-4 (ALLM)

Muirhead, Robb J., [1946–], Aspects of multivariate statistical theory/Robb J. Muirhead. – New York: Wiley, cop. 1982. – 673 s. – (Wiley series in probability and mathematical statistics, ISSN 0271-6232) ISBN 0-471-09442-0 (2D/721)

Nelson, Wayne, Applied life data analysis/Wayne Nelson. – New York, 1982. – 634 s. – (Wiley series in probability and mathematical statistics, ISSN 0271-6232) ISBN 0-471-09458-7 (2D/709)

Newell, Gordon Frank, Applications of queueing theory/G. F. Newell. – 2. ed. – London: Chapman and Hall, 1982. – 303 s. – (Monographs on statistics and applied probability) Previous ed.: 1971. – Bibliography: p. 293–298. – Includes index, ISBN 0-412-24500-0 (2D/725)

Nisbet, R. M., Modelling fluctuating populations/ R. M. Nisbet and W. S. C. Gurney. – Chichester: Wiley, cop. 1982. – xiii, 379 p.: ill. A Wiley-interscience publ. – Bibliography: p. 361–370. – Includes index ISBN 0-471-28058-5 (2D/731)

Nordisk symposium i anvendt statistik og EDB, 1982 (Lyngby) Nordisk symposium i anvendt statistik og EDB Lyngby [1982] Red.: Agnar Högskuldsson ... – Lyngby: NEUCC, 1982. – 411 s.: tab. Omslagstitel: Nordic symposium in applied statistics and data processing ISBN 87-88257-00-2 (2D/703)

Pannonian symposium on mathematical, statistics, 2 (Bad Tatzmannsdorf 1981) Probability and statistical inference/ proceedings of the 2nd Pannonian symposium on mathematical statistics, Bad Tatzmannsdorf, Austria, June 14–20, 1981; ed. by Wilfried Grossmann ... – Dordrecht: Reidel, cop. 1982. – 389 s.: ill. Includes bibliographies and index ISBN 90-277-1427-4 (2D/714)

Pfanzagl, Johann, Contributions to a general asymptotic statistical theory/J. Pfanzagl; with the assistance of W. Wefelmeyer. – New York; Berlin: Springer, cop. 1982. – 315 s. – (Lecture notes in statistics, NBN 99-0198235-9; 13) Includes indexes. – Bibliography: p. 300-310. ISBN 0-387-90776-9 (2D/723)

Quantitative population dynamics/ed. by Douglas G. Chapman, Vincent F. Gallucci. – Burtonsville, Md.: International co-operative publ. house, cop. 1981. – 290 s. – (Statistical ecology, NBN 99-0203011-4; 13) ISBN 0-89974-010-3 (2D/712)

Queues and point processes/Peter Franken ... Chichester: Wiley, 1981. – 208 s. – (Wiley series in probability and mathematical statistics, ISSN 0271-6232) ISBN 0-471-10074-9 (2D/726)

Reid, Constance, Neyman – from life/by Constance Reid. – New York; Berlin: Springer, cop. 1982. – 298 s.: ill. ISBN 0-387-90747-5 (2D/713)

Runyon, Richard P., How numbers lie: a consumer's guide to the fine art of numerical deception/by Richard P. Runyon. – 1 ed. – Lexington, Mass.: Lewis, cop. 1981. – 182 s. – (A Rand corporation research study) Portions of this work have been excerpted from Winning with statistics by Richard P. Runyon, cop. 1977 ISBN 0-86616-000-0 (2D/685)

Statistics and probability/essays in honor of C. R. Rao/ed. by G. Kallianpur, ... – Amsterdam; New York: North-Holland Pub. Co.; New York: Sole distributors for the U. S. A. and Canada, Elsevier North-Holland, 1982. – xi, 722 s.: ill. "Publications of C. R. Rao": p. 713–722. – Includes bibliographies ISBN 0-444-86130-0 (2D/701)

Statistiska centralbyrån, Perspektiv på välfärden 1982/Statistiska centralbyrån = [Perspectives on Swedish welfare in 1982]/[National central bureau of statistics]. – Stockholm: LiberFörlag/Allmänna förl., 1982. – 348 s.: diagr., tab. – (Levnadsförhållanden, ISSN 0347-7193; 33) (Sveriges officiella statistik, NBN 99-0111480-2) ISBN 91-38-06956-3 (P3A/68-33) (Oi-c)

Statistiska centralbyrån, "Utveckling och nuläge" (SOS), signum P3/984 Sverige ALLM och år 1977 under beteckning "Levnadsförhållanden, rapport nr 6, utveckling och nuläge enl. tillgänglig statistik 1976" (SOS), (P3A/68 -6)

Statistiska centralbyrån, Våra inkomster: en sammanfattning byggd på 1979 års inkomstförhållanden/Statistiska centralbyrån = [Our income]: [an outline based on the income conditions in 1979]/[National central bureau of statistics]. – Stockholm: LiberFörlag/Allmänna förl., 1982. – 216 s.: diagr., tab. – (Sveriges officiella statistik, NBN 99-0111480-2), With English summaries ISBN 91-38-07259-9 (4E/973) (Oi-c)

Tigelaar, H. H., Identification and informative sample size/H.H. Tigelaar. – Amsterdam, 1982. – 144 s. – (Mathematical centre tracts, ISSN 0569-9657; 147) ISBN 90-6196-235-8(2D/722)

United States. Federal committee on statistical methodology Report on exact and statistical matching techniques/prep. by Subcommittee on matching techniques, Federal committee on statistical methodology. – Washington, D.C.: GPO, 1980. – 57 s. – (Statistical policy working paper, NBN 99-0214681-3; 5) NBN 99-0291297-4 (2D/695)

Unwin, David, Introductory spatial analysis/David Unwin. – London: Methuen, 1981. – 212 s. Includes bibliographies and index ISBN 0-416-72190-7 (2D/736)

Bokanmälningar

Hofsten, E., Befolkningslära – demografi för lekmän. Studentlitteratur. Lund 1982. 118 s. Kr. 88:–.

Erland Hofsten har etablerat sig som författare av läroböcker med demografisk inriktning. Hans senaste tillskott till denna litteratur är ett försök att popularisera och göra lättare tillgänglig den demografiska teorin för personer med en begränsad matematisk bakgrund. Boken tar upp de centrala begreppen inom demografi och beskriver hur man mäter dödlighet, fruktsamhet, reproduktion osv. Hofsten gör detta utan att egentligen använda några matematiska formler. Det är förvånansvärt hur mycket man kan få med på ett så begränsat utrymme som det här är fråga om; boken omfattar ca 100 sidor. Förutom presentationen av demografiska grundbegrepp finns en diskussion av befolkningsmodeller och folkomflyttning. Ett stort kapitel behandlar förklaringar, följder och befolkningspolitiska åtgärder. Det finns med en historisk översikt av befolkningsutvecklingen där befolkningspolitik, barnafödandet i industriländer, befolkningsutvecklingens följder samt befolkningspolitik och annan politik diskuteras. För dem som ändå vill ha lite formler finns i en bilaga några demografiska exempel på beräkning av dödsrisker, fruktsamhetstal och livslängdstabeller samt en kort beskrivning av hur befolkningsprojektioner går till.

Ingvar Holmberg

Sachs, L., Applied Statistics. A handbook of techniques. Springer-Verlag. Berlin 1982. 706 s. US $ 47:20.

Sachs inleder sin bok med följande definition:
"Statistics is the art and science of data: of generating or gathering, describing, analyzing, summarizing and interpreting data to discover new knowledge."
Detta förefaller vara en god formulering av statistikens plats i den vetenskapliga gemenskapen.

Boken är enligt författaren avsedd för "icke-matematiker, särskilt för tekniker, ingenjörer, beslutsfattare, studenter, läkare såväl som forskare i andra ämnen. Den ger en matematiker som är intresserad av statistikens praktiska användningar en allmän orientering om ämnet. Praktisk tillämpning är dess huvudtema."

Vad innehåller då boken? I en intressant inledning utvecklas och fördjupas den ovan nämnda definitionen av statistikens uppgift. I kap. 0 ges en översikt över elementära matematiska begrepp samt några beräkningstekniska råd. I det relativt omfattande kap. 1 förklaras sannolikhetslärans och statistikens teoretiska grunder. I kap. 2 beskrivs översiktligt några (som det förefaller rätt godtyckligt valda) tillämpningsområden: medicinsk statistik, sekventiella testplaner, uppskattning av biologiskt aktiva ämnen, statistik inom ingenjörsvetenskapen och operationsanalys. I kap. 3 och 4 presenteras ett stort antal testförfaranden, dels sådana som bygger på ett normalfördelningsantagande dels icke-parametriska dito. Kap. 5 behandlar olika mått på association: korrelation och regression. I kap. 6 beskrivs analys av (och tester som bygger på) tvåvägstabeller. Kap. 7, slutligen, presenterar tekniker för variansanalys och principer för försöksplanering.

En första fråga som inställer sig är de motiv som styrt urvalet av metodområden. I förhållande till den breda kategori av grupper som boken är avsedd för är det trots allt ett ganska snävt urval. Områden som (nästan) inte alls behandlas är t.ex. survey sampling, stokastiska processer och tidsserieanalys samt multivariat analys. I fråga om survey sampling ges några spridda kom-

mentarer här och var i boken men någon samlad behandling kommer aldrig till stånd. Vad gäller tidsserier så är det avklarat på en sida med ett test för förekomsten av trend utan att dessförinnan tidsserieanalys som eget metodområde över huvud taget uppmärksammats. Beträffande multivariat analys ges ett antal referenser och konstateras att man måste börja med att lära sig matrisalgebra.

Det är rimligt att anta att bokens begränsningar styrts av dels författarens eget erfarenhetsområde, som uppenbarligen är medicinska och tekniska tillämpningar, dels av ambitionen att hålla framställningen på en matematiskt elementär nivå. Men även med den senare ambitionen borde det t.ex. ha varit möjligt att inkludera ett avsnitt om deskriptiv tidsserieanalys och säsongrensning.

För en stor del av den uppsatta målgruppen bör dock boken kunna vara av värde. Det gäller läkare, tekniker och naturvetare som i forskning eller annat arbete behöver använda sig av statistiska metoder. För samhällsvetare och beslutsfattare torde dock denna bok vara mindre lämplig, eftersom en betoning av andra metodområden (som t.ex. survey sampling) vore av större betydelse för dessa än vissa av de i boken presenterade.

Jörgen Dalén

Utkom från trycket den 31 oktober 1983.

STATISTICAL REVIEW

PUBLISHED BY STATISTICS SWEDEN, STOCKHOLM, SWEDEN

Legally responsible publisher: Sten Johansson. Editor: Nils Welander
Guest Editor: Lars Lyberg
Address: Karlavägen 100. Mailing address: S-115 81 Stockholm, Sweden
Telephone: 08/14 05 60

Subscription: Liber Distribution, Subscription order
S-162 89 Stockholm, Sweden
Five issues a year

Third series. Vol. 21 No. 5 1983

Contents

Preface 5

Works by Tore E. Dalenius

A. Scholarly and other publications 7

B. Research reports from the project FEL I UNDERSÖKNINGAR ("Total Survey Design") 15

C. Research reports from the project PERSONINTEGRITET VID STATISTISKA UNDERSÖKNINGAR ("Confidentiality in Surveys") 16

D. Research reports from the project ACCESS TO INFORMATION THROUGH CENSUSES AND SURVEYS 17

Articles

Knut Medin:
The Introduction of Probability Sampling in Swedish Acreage and Livestock Statistics in 1950. A Note on the Background and Later Developments 19

Bertil Matérn:
On the Efficiency of Sampling with Unequal Probabilities 27

Carl-Erik Särndal and Göran Råbäck:
Variance Reduction and Unbiasedness for Small Domain Estimators 33

Sven Berg:
Factorial Series Distributions and Estimation Based on the Distinct Units of a Sample 41

Myron J. Katzoff:
Two Sequential Procedures for Obtaining Subdomain Estimates 49

Claes M. Cassel:
Statistical Matching – Statistical Prediction What is the Difference? 55

N. Anders Klevmarken:
Pooling Incomplete Data Sets 67

Jan M. Hoem:
Balancing Bias in Vital Rates Due to an Informative Sampling Plan 81

Ingrid Lyberg:
On the Practical Use of Subsampling Among Nonrespondents 89

Jan Lanke:
Hot Deck Imputation Techniques that Permit Standard Methods for Assessing Precision of Estimates 105

Bengt Swensson:
On Double Sampling for Stratification with Non-Response in the Second Phase 111

Jan Håkan Wretman:
On Variance Estimation in Survey Sampling with Measurement Errors 117

Patrick D. Bourke:
Randomized Response Designs With Attribute-Based Randomization 125

Anders Christianson:
Balancing Recall Bias Against Sampling Error in the Swedish TV Audience Surveys 133

Lars Lyberg:
The Development of Procedures for Industry and Occupation Coding at Statistics Sweden 139

Thomas B. Jabine:
Teaching Questionnaire Design 157

Maria Elena Gonzales:
Organization of the U.S. Federal Committee on Statistical Methodology 167

Ove Frank:
Statistical Disclosure Control 173

Lawrence H. Cox:
Some Mathematical Problems Arising from Confidentiality Concerns 179

Leo K. Lichtig:
Confidentiality Issues in Health Statistics:
An American Perspective 191

Joseph W. Duncan:
Confidentiality in a Decentralized Statistical System 203

Leif Björk and Jan Hagberg:
Probability – Knowledge – Action: Comments on Some Recent Controversies
on Applications of Probabilistic Methods 207

Thomas Polfeldt and Bo Justusson:
On Applications of Random Techniques in Environmental Statistics 217

Preface

It is a privilege for Statistics Sweden to be able to present this Special Issue of the Statistical Review in honour of Professor Tore E. Dalenius.

Tore Dalenius was born on November 21, 1917 in Jukkasjärvi, Sweden. After studies at Uppsala University he obtained his fil. kand. in 1942 and fil. lic. in 1948.

During the 40's he held positions in various Swedish government agencies and private firms, and he also spent a couple of years studying statistics in the United States. In 1950 he joined the Swedish National Board of Health and Welfare and in 1952 he moved to the National Central Bureau of Statistics (now Statistics Sweden). In 1956 he left in order to finish his Ph. D thesis, "Sampling in Sweden", which became one of the most, if not the most, frequently cited theses in the field of survey sampling. 26 years after its publication it is still relevant for survey researchers and it still offers an abundance of ideas for future research, often hinted at in a single line only.

Since 1956 he has been a faculty member of the Department of Statistics, Stockholm University, during the last 20 years as Professor of Statistics with a special concern for official statistics production. As a consequence a community of interest grew up between him and Statistics Sweden. In the beginning of the 70's he moved to Brown University, Providence, Rhode Island, USA, but his interest in Statistics Sweden has never ceased.

Over the years this interest has manifested itself in different ways. In the early 50's he organized a course in sampling for the professional staff with Morris Hansen and P. V. Sukhatme as two of the teachers. This and other initiatives were fundamental elements in the work to introduce sampling methodology at Statistics Sweden. He founded the Research Institute that today performs many regular and ad-hoc surveys of individuals and households as well as other surveys on a commission basis. He has contributed 41 articles to the Statistical Review and his contacts with Statistics Sweden have been intense and extensive. The saying is that everybody at Statistics Sweden has a Tore Dalenius file.

However, the communications have not always been accompanied by marks of courtesy. Tore Dalenius has not only been interested in what Statistics Sweden actually has been working on, but also in things he thinks Statistics Sweden *should* have done. Thus, he has functioned as a constant goad, sometimes irritating but always rewarding.

This special issue of the Statistical Review attempts to reflect his main methodological interests over the years, viz. sampling, non-sampling errors and problems associated with privacy and confidentiality. His own contributions within these fields are extensive, as is apparent from the subsequent list of works, with theory and practice closely integrated.

Now Tore Dalenius has reached the age of retirement and I join with the contributors to this special issue in congratulating him on, and thanking him for, his great contributions to our science.

For me personally it has been very rewarding to be the guest editor of this special issue in honour of my dear friend and teacher Tore Dalenius. I also want to thank all the contributors and the editorial staff who have helped in this secret mission.

Statistics Sweden, however, considers his retirement as merely formal. We know that he will not sit under a tree feeding the birds. We count on him in the future and wish him a very busy retirement.

LARS LYBERG

Works by Tore E. Dalenius

A. Scholarly and other publications

1 Hålkortsmetodens användning vid opinionsmätningar och marknadsundersökningar. IBM-Nyheter, 1944.

2 Barnkläder. En undersökning bland 1 273 svenska familjer med barn i åldern 1–15 år. Aktiv Hushållning, 1946.

3 Om stickprovsundersökningens teknik och metodik. Affärsekonomi skriftserie nr 25, Stockholm, 1949.

4 Användningen av stickprovsundersökningar vid studiet av "Consumer Finances" i Förenta Staterna. Ekonomisk Revy, 1949.

5 Den nyare utvecklingen inom teorin och metodiken för stickprovsundersökningar. Förhandlingar vid Nordiska statistikermötet, Helsingfors 1949.

6 (With O. Zetterberg) Nya metoder för skördeuppskattning. Jordbruksekonomiska Uppgifter, 1949.

7 Staten och statistikproduktionen. Tiden, 1949.

8 Glücksrad statt Urnenschema. Mitteilungsblatt für Mathematische Statistik, 1950.

9 (With G. Jacobsson) En metod för kortsiktiga prognoser beträffande äktenskapsbildningen. Statistisk Månadstidskrift för Stockholms stad, 1950.

10 Modern Sampling. Den svenska marknaden, 1950.

11 Kostnadsproblemet vid sampling. Den svenska marknaden, 1950.

12 The Problem of Optimum Stratification. Skandinavisk Aktuarietidskrift, 1950.

13 Försvar för statistiken. Tiden, 1950.

14 Opinionsmätningarnas seger och kris. Fackföreningsrörelsen, 1950.

15 Bibliography on Sampling. S A F, Stockholm 1950.

16 Användningen av stickprovsmetodik vid 1950 års arealinventering och kreatursräkning. Jordbruksekonomiska Meddelanden, 1951.

17 Sampling och geografi. Ymer, 1951.

18 (With M. Gurney) The Problem of Optimum Stratification II. Skandinavisk Aktuarietidskrift, 1951.

19 Om planering av stickprovsundersökningar. Sociala Meddelanden, 1951.

20 Bortfallsproblemet vid statistiska undersökningar. Den svenska marknaden, 1951.

21 Användbarheten av en organisation för stickprovsundersökning. Stencil, 1951.

22 Statistikproduktions mål och medel. Ekonomisk Revy, 1952.

23 (With O. Zetterberg) Ny metod för skördeuppskattning. Statistisk tidskrift, 1952.

24 Statistikermötet i Indien. Sociala Meddelanden, 1952.

25 The Problem of Optimum Stratification in a Special Type of Design. Skandinavisk Aktuarietidskrift, 1952.

26 Operationsanalys. Statistisk tidskrift, 1952.

27 En granskning av socialstyrelsens lönestatistik. Sociala Meddelanden, 1952.

28 Eine einfache geometrische Veranschaulichung der Theorie des geschichteten Stichprobenverfahrens. Mitteilungsblatt für Mathematische Statistik, 1952.

29 (With O. Zetterberg et al.) Provundersökningarna rörande ny skördeberäkningsmetod. Statistisk tidskrift, 1953.

30 Representativitet. Statistisk tidskrift, 1953.

31 Något om metoder för objektiva skördeberäkningar. Kungl. Lantbruksakademiens Tidskrift, 1953.

32 The Multi-Variate Sampling Problem. Skandinavisk Aktuarietidskrift, 1953.

33 The Economics of One-Stage Stratified Sampling. Sankhyā, 1953.

34 Review of Kungl. Socialstyrelsen: Levnadskostnaderna i tätortshushåll år 1948. S O S, 1953. Statistisk tidskrift, 1953.

35 Review of M. H. Hansen, W. N. Hurwitz and W. G. Madow: Sample Survey Methods and Theory. Journal of the American Statistical Association, 1954.

36 Statistiska centralbyråns utredningsinstitut. Statistisk tidskrift, 1954.

37 The Problem of Not-at-Homes. Statistisk tidskrift, 1955.

38 Ett svenskt statistiskt system. Tiden, 1955.

39 Statistikermötet i Brasilien. Statistisk tidskrift, 1955.

40 Review of P. V. Sukhatme: Sampling Theory of Surveys with Applications. Statistisk tidskrift, 1955.

41 Approximativa variansberäkningar. Statistisk tidskrift, 1956.

42 Review of W. A. Wallis and H. V. Roberts: Statistics: A New Approach. Sociala Meddelanden, 1956.

43 "Stadiga läsare av sista numret". Kritik av en populärpressanalys. Info, 1956.

44 The Survey Research Centre of the Central Bureau of Statistics, Sweden. Sankhyā, 1956.

45 Possibilities and Limits of Sampling in Regional Inquiries. Bulletin of the International Statistical Institute, 1957.

46 Sampling in Sweden. Ph. D. Thesis, Almqvist & Wiksell, Uppsala 1957.

47 Basurvalet vid detaljhandelsundersökningar. Pris- och kartellfrågor, 1957.

48 (With J. L. Hodges, Jr) The Choice of Stratification Points. Skandinavisk Aktuarietidskrift, 1957.

49 Internationella statistiska institutets stockholmsmöte: Det vetenskapliga programmet. Statistisk tidskrift, 1957.

50 Posibilidades y Limitaciones del Muestreo en Encuestas Regionales. Estadística, Journal of the Inter-American Statistical Institute, 1957.

51 En geometrisk representation av medelavvikelsen för en tvåpunktsfördelad variabel. Statistisk tidskrift, 1958.

52 Internationella statistiska institutets möte i Bryssel: Det vetenskapliga programmet. Statistisk tidskrift, 1958.

53 Review of Ph. J. McCarthy: Introduction to Statistical Reasoning. Sociala Meddelanden, 1958.

54 Review of Estadística: Journal of the Inter-American Statistical Institute, Vol. 15, 1957. A Subject Issue on Sampling. Statistisk tidskrift, 1958.

55 Review of H. Strecker: Moderne Methoden in der Agrarstatistik. Einzelschriften der Deutschen Statistischen Gesellschaft. Review of the International Statistical Institute, 1958.

56 Review of B. Varenius: Intervjuundersökningsteknik vid marknadsstudier och opinionsmätningar – metodik och felrisker. Info, 1959.

57 Marknadsundersökningar. Ekonomisk Revy, 1959.

58 Operationsanalys. Ekonomisk Revy, 1959.

59 Statistisk teori och officiell statistikproduktion. Ekonomisk Tidskrift, 1959.

60 Förslag till organisation av ett svenskt statistiskt system. Ekonomisk Tidskrift, 1959.

61 (With J. L. Hodges, Jr) Minimum Variance Stratification. Journal of the American Statistical Association, 1959.

62 Något om automatisk databehandling. Statistisk tidskrift, 1959.

63 El Problema de la Estratificatión Óptima. Estadística, Journal of the Inter-American Statistical Institute, 1959.

64 El Muestreo Estratificaco Multiparamétrico. Estadística, Journal of the Inter-American Statistical Institute, 1959.

65 Ett mätfelsproblem. Nordisk Tidskrift for Industriel Statistik, 1959.

66 Review of F. F. Stephan and Ph. J. McCarthy: Sampling Opinions. An Analysis of Survey Procedures. Teknisk Tidskrift, 1959.

67 Potentiella användningar av sampling i ett statistiskt system. Statistisk tidskrift, 1960.

68 Review of G. Dannerstedt: Seriestorlek och beställningsnivå. Elementarfallstabeller och diagram. 1960.

69 Om användningen av ett speciellt slag av information vid planeringen av en stickprovsundersökning. Nordisk Tidskrift for Industriel Statistik, 1960.

70 Training in Sampling for a Government Statistical System. Bulletin of the International Statistical Institute, 1960.

71 (With B. Ajne) Några tillämpningar av statistiska idéer på numerisk integration. Nordisk Matematisk Tidskrift, 1960.

72 Monte Carlo-metoden. Svensk Naturvetenskap, 1960.

73 Treatment of the Non-Response Problem. Journal of Advertising Research, 1961.

74 (With J. Hájek and S. Zubrzycki) On Plane Sampling and Related Geometrical Problems. Proceedings of the Fourth Berkeley Symposium. Berkeley and Los Angeles, 1961.

75 Stickprovsundersökningarnas teoretiska och metodologiska grundvalar. Kap. 4 i Georg Karlsson: Lärobok i sociologisk metod I, 1961.

76 Utlåtande om Gunnar Fredrikssons avhandling "Kriminalstatistiken och kriminologien". 1962.

77 Statistisk kvalitetskontroll av en kontinuerlig produktionsprocess. Statistisk tidskrift, 1962.

78 Review of Statistisches Bundesamt: Stichproben in der amtlichen Statistik. Statistisk tidskrift, 1962.

79 Automatic Estimation of Missing Values in Censuses and Sample Surveys. Statistisk tidskrift, 1962.

80 Något om köteorin och dess tillämpning på sjukvårdens område. Nordisk Medicin, 1962.

81 Recent Advances in Sample Survey Theory and Methods. Annals of Mathematical Statistics, 1962.

82 Ramproblemet vid urvalsundersökningar. Statistisk tidskrift, 1963.

83 Ett exempel på bestämning av optimal urvalsstorlek. Nordisk Tidskrift for Industriel Statistik, 1963.

84 Det svenska postverkets urvalsundersökning av posttrafiken före och efter senaste portoändringen. Nordisk Post Tidskrift, 1963.

85 The Panel Approach to Continuous Sample Survey Activity. Metrika, 1963.

86 Synpunkter på artikeln om "Datamaskinell optimering m.m.". Svenska Läkartidningen, 1964.

87 (With B. Matérn) Is There a Need for a Unified Theory of Random Experiments? Metrika, 1964.

88 Review of W. G. Cochran: Sampling Techniques. Annals of Mathematical Statistics, 1964.

89 Progresos Recientes en la Teoria y Metodos de las Encuestas por Muestra. Estadística, Journal of the Inter-American Statistical Institute, 1964.

90 Lattice Sampling by Means of Lahiri's Sampling Scheme. In Rao, R. C. (Ed): Contributions to Statistics Presented to Professor P. C. Mahalanobis on the Occasion of his 70th Birthday. Calcutta, 1964.

91 Internationella statistiska institutets 34:e session. Statistisk tidskrift, 1964.

92 Potential Research Objects in Sample Survey Theory and Methods. Colloquium on the Application of Mathematics to Economics, Budapest 1963. Budapest 1965.

93 Current Trends in the Development of Sample Survey Theory and Methods. In Proceedings of the *IBM* Scientific Computing Symposium on Statistics, October 21–23, 1963. 1965.

94 The Mode – A Neglected Statistical Parameter. Journal of the Royal Statistical Society, Series A, 1965.

95 Principer och metoder för planering av samplingundersökningar. Intern handbok nr 4, Statistiska centralbyrån, 1965.

96 Use of Sampling Methods for the Estimation of Areas. In Zarkovic, S. S.: Estimation of Areas in Agricultural Statistics. Rome 1965.

97 Time and Survey Design. Journal of Advertising Research, 1965.

98 Review of L. Kish: Survey Sampling. New York 1965. Statistisk tidskrift, 1966.

99 Vetenskapliga metoder och praktiska problem. Svensk naturvetenskap, 1966.

100 (With O. Persson) En redogörelse för en kurs i statistik inriktad på virkesmätning. Nordisk Tidskrift for Industriel Statistik, 1966.

101 Intuition versus Analysis in Problem Solving. Report DAC 56044. Douglas Aircraft Division, 1966.

102 Operationsanalys inom sjukvården. SJURA. Rådet för sjukhusdriftens rationalisering. Stockholm 1966.

103 Uso de los metodos de mestro para la estimacion de superficies. Estadística, 1966.

104 (With S. P. Ghosh) Some Theory of Sampling when the Stratification is Subject to Error. Technometrics, 1967.

105 Official Statistics and Their Uses. Review of the International Statistical Institute, 1968.

106 A Feasible Approach to General Purpose Sampling Design. Management Science, 1968.

107 Några kommentarer till fil.lic. Hans L. Zetterbergs artikel "System Veckobuss". Sociologisk Forskning, 1968.

108 (With O. Frank) Control of Classification. Review of the International Statistical Institute, 1968.

109 Konsumentupplysning och statistik. (Appendix nr 3, Konsumentupplysning. Principer och riktlinjer. SOU 1968:58), 1968.

110 Las estadisticas officiales y sus usos. Estadística, 1968.

111 Review of M. N. Murthy: Sampling Theory and Methods. Statistical Publishing Society, 1967. Annals of Mathematical Statistics, 1969.

112 Designing Descriptive Sample Surveys. In N. L. Johnson and H. Smith, Jr. (Ed): New Developments in Survey Sampling. Wiley-Interscience, 1969.

113 Modern surveymetodik. Intern handbok nr 8, Statistiska centralbyrån, 1969.

114 En matrismultiplikationstabell. Elementa, 1969.

115 Towards a Survey Measurement System. Zastosowania matematyki, 1969.

116 (With B. A. Bailar) Estimating the Response Variance Components of the U. S. Bureau of the Census' Survey Model. Sankhyā, Series B, 1969.

117 Theory and Methods of Descriptive Surveys. Some Recent Advances. In Paul Lindblom (Ed): Theory and Methods in Behavioural Sciences. Scandinavian University Books, 1970.

118 Some Observations on my Research Process. In T. Dalenius, G. Karlsson and S. Malmquist: Scientists at Work. Festschrift in Honour of Herman Wold. Almqvist & Wicksell, 1970.

119 Kommunurval. In J. Westerståhl: Den kommunala självstyrelsen 1. Ett forskningsprogram. Almqvist & Wicksell, 1970.

120 Survey Sampling in a Computerized Environment. Review of the International Statistical Institute, 1971.

121 Principer och metoder för planering av samplingundersökningar. Intern handbok nr 4, Statistiska centralbyrån, 1971 (rev.).

122 En granskning av två undersökningar av partisympatier. I Valbarometern – Ett Radionämndsärende. Sveriges Radios Förlag 1971.

123 a Kritik av min kritiker. Statistisk tidskrift, 1973.

b Kritik av min kritiker II. Statistisk tidskrift, 1973.

c Slutreplik till Rune Sandström. Statistisk tidskrift, 1973.

124 (With O. Frank) Sampling Populations with Partial Variate Patterns. Scandinavian Journal of Statistics, 1974.

125 Statistics from Professor Stockfelt's Standpoint. Statistisk tidskrift, 1974.

126 The Invasion of Privacy Problem and Statistics Production – An Overview. Statistisk tidskrift, 1974.

127 Confidentiality in Surveys. A Research Program. Journal of Statistical Research, 1976.

128 Sample-Dependent Estimation in Survey Sampling. In Walter Joh. Ziegler, (Ed). Contributions to Applied Statistics. Birkhäuser Verlag, Basel and Stuttgart, 1976.

129 (With P. D. Bourke) Some New Ideas in the Realm of Randomized Inquiries. International Statistical Review, 1976.

130 Aktuell metodutveckling på surveyområdet. Statistisk tidskrift, 1976.

131 Bortfallsproblemet vid statistiska undersökningar. Marknadsvetande, nr 4, 1976.

132 Privacy Transformations for Statistical Information Systems. Journal of Statistical Planning and Inference, 1977.

133 Towards a Methodology for Statistical Disclosure Control. Statistisk tidskrift, 1977.

134 Strain at a Gnat and Swallow a Camel: Or the Problem of Measuring Sampling and Non-Sampling Errors. American Statistical Association, Proceedings of the Social Statistics Section, 1977, Part I.

135 Discussion of: Goldfield, E. D., Turner, A. G., Cowan, C. D. and Scott, J. C.: Privacy and Confidentiality as Factors in Survey Response. American Statistical Association, Proceedings of the Social Statistics Section, 1977, Part I.

136 The Swedish Data Act and Statistical Data. Statistisk tidskrift, 1978.

137 Utbildningen i statistik. Statistisk tidskrift, 1978.

138 Med mitt mått mätt eller Om slumpmässig längdmätning. Elementa, nr 3, 1978.

139 (With S. P. Reiss) Data-Swapping – A Technique for Disclosure Control. American Statistical Association, Proceedings of the Section on Survey Research Methods, 1978.

140 Computers and Individual Privacy: Some International Implications. In: Bulletin of the International Statistical Institute, XLVII (1), 1977.

141 Some Personal Recollections of Jaroslav Hájek. In: Contributions to Statistics. (Jaroslav Hájek memorial volume). Academia, Prague, 1979.

142 (With R. A. Vitale): A New Randomized Response Design for Estimating the Mean of a Distribution. In: Contributions to Statistics. (Jaroslav Hájek memorial volume). Academia, Prague, 1979.

143 Heuristics for Survey Sampling Design. In: Contributions to Statistics. (Jaroslav Hájek memorial volume). Academia, Prague, 1979.

144 Data-Swapping: A Technique for Disclosure Control when Releasing Micro-Statistics. Statistisk tidskrift, 1979.

145 Data Protection Legislation in Sweden: A Statistician's Perspective. Journal of the Royal Statistical Society, Series A, Part 3, 1979.

146 En elementär framställning av Monte Carlo metoden. Elementa, nr 3, 1979.

147 Informed Consent or R.S.V.P. In: Madow, W. G. and Rizvi, M. H.: Preliminary Proceedings of the Symposium on Incomplete Data, August 10–11, 1979, printed by the U. S. Department of Health, Education and Welfare, Washington, D. C., 1979.

148 Some Reflections on the Problem of "Missing Data". In: Madow, W. G. and Rizvi, M. H.: Preliminary Proceedings of the Symposium on Incomplete Data, August 10–11, 1979, printed by the U. S. Department of Health, Education and Welfare, Washington, D. C., 1979.

149 Randomized Response. In: Boruch, R. F., Ross, J. and Cecil, J. S. (Eds) Proceedings and Background Papers: Conference on Ethical and Legal Problems in Applied Social Research. Department of Psychology, Northwestern University, Evanston, Illinois, 1979.

150 Large-Sample vs Small-Sample Survey Practice – A Heretical Note. Statistisk tidskrift, 1980.

151 U. S. Bureau of the Census' Conference on Census Undercount. Statistisk tidskrift, 1980.

152 Privacy and Telephone Surveys. American Statistical Association, Proceedings of the Business and Economic Statistics Section, 1980.

153 Review of Holmberg, S. and Petersson, O.: Inom felmarginalen. En bok om politiska opinionsundersökningar. Liber förlag, Stockholm, 1980. Statistisk tidskrift, 1981.

154 A Simple Procedure for Controlled Rounding. Statistisk tidskrift, 1981.

155 (With V. N. Nair) Sampling from Structured Populations: Some Issues and Answers. The Bell System Technical Journal, No. 7, 1981.

156 The Survey Statistician's Responsibility for Both Sampling and Measurement Errors. In Kreski, D., Platek, R. and Rao, J. N. K. (Eds): Current Topics in Survey Sampling. Academic Press, New York, 1981.

157 Release of Microdata – The International Picture. Interner Bericht IPES 81.204. Gesellschaft für Mathematik und Datenverarbeitung MBH, Bonn, 1981.

158 The Dissemination of Official Statistics – Technology. Invited paper. 43rd Session of the International Statistical Institute, Buenos Aires, 1981.

159 (With S. P. Reiss) Data-Swapping: A Technique for Disclosure Control. Journal of Statistical Planning and Inference, 1982.

160 (With D. E. Denning) A Hybrid Scheme for Release of Statistics. Statistisk tidskrift, 1982.

161 Privacy and Statistics: Some Potential Research Topics. In Raab, C. D. (Ed.): Data Protection and Privacy. Proceedings of a conference. Social Research Association, London, 1982.

162 Disclosure Control of Magnitude Data. Statistisk tidskrift, 1982.

163 A Sample of Ideas for Research and Development in the Theory and Methods of Sample Surveys. Utilitas Mathematica, Vol. 21 A, 1982.

164 Official Statistics in the People's Republic of China. Statistisk tidskrift, 1982.

B. Research reports from the project
FEL I UNDERSÖKNINGAR
("Total Survey Design")
supported by a grant from the Bank of Sweden Tercentenary Foundation, Stockholm.
(The numbers assigned are the report numbers.)

1 Towards a Survey Measurement System.

2 Designing Enumerative Sample Surveys.

3 (With O. Frank) Control of Classification.

5 Non-Sampling Errors in Censuses and Sample Surveys.

6 Some Comments on the Paper: "System Veckobuss" by fil. lic. Hans L. Zetterberg.

7 (With M. Gurney) The Multivariate Approach to Survey Design and Estimation.

9 (With B. A. Bailar) Estimating the Response Variance Components of the U. S. Bureau of the Census' Survey Model.

10 Theory and Methods of Descriptive Surveys. Some Recent Advances.

25 (With L. Peiram) Sample – Dependent Estimation.

29 (With U. Grenander) Training in Survey Sampling Using Interactive Computing.

30 Survey Sampling in a Computerized Environment.

31 (With R. Vitale) Interactive Computation as a Means of Training in Survey Sampling.

32 A List of Ideas for Research and Development in the Realm of Total Survey Design and Practice.

33 A General Notion of a Survey.

34 Information for Survey Design.

35 Nonresponse and Inadequate Coverage as Sources of Errors in Surveys.

36 Thinning Statistical Archives.

41 An Evaluation of Two Surveys of Political Opinions.

53 A New Approach to Sampling on Two Occasions.

54 Computer-Supported Instruction in Survey Sampling.

58 (With O. Frank) Sampling Populations with Partial Variate Patterns.

59 (With A. Fisz) A Note on Sample-Dependent Estimation.

60 (With A. Klevmarken) Model-Based Survey Procedures.

61 Heuristics for Survey Sampling Design: A Case Study.

63 Criticism of my Critic.

64 Survey Sampling in Developing Countries.

65 Criticism of my Critic – II.

68 (With P. D. Bourke) Multi-Proportions Randomized Response Using a Single Sample.

69 The Invasion of Privacy Problem and Statistics Production – An Overview.

71 (With P. D. Bourke) Randomized Response Models with Lying.

72 (With P. D. Bourke) A Note on Inadmissible Estimates in Randomized Inquiries.

77 Protecting Confidentiality by Means of Weighing Designs.

78 (With R. Vitale) A New Randomized Response Design for Estimating the Mean of a Distribution.

81 Two Measures of Unemployment. A Study of Swedish Labor Market Statistics.

C. Research reports from the project
PERSONINTEGRITET VID STATISTISKA UNDERSÖKNINGAR
("Confidentiality in Surveys")
supported by a grant from the Bank of Sweden Tercentenary Foundation, Stockholm.
(The numbers assigned are the report numbers.)

1 Confidentiality in Surveys. A Research Program.

5 (With P. D. Bourke) Some New Ideas in the Realm of Randomized Inquiries.

9 Privacy Transformations for Statistical Information Systems.

19 Towards A Methodology for Statistical Disclosure Control.

21 Matching Personal Files – A Threat to Individual Privacy?

22 The Swedish Data Act and Statistical Data.

23 Computers and Individual Privacy – Some International Implications.

24 Data Base Insecurity and the Statistical Disclosure Problem.

27 (With S. Geman) Application of the Combined Question Technique to Non-Reversible Privacy Transformations.

28 The Swedish Data Act and Statistical Data – II.

29 (With J. Silverstein) Public Key Cryptosystems – An Overview.

31 (With S. Reiss) Data-Swapping – A Technique for Disclosure Control.

32 Informed Consent or. R.S.V.P.

D. Research reports from the project
ACCESS TO INFORMATION THROUGH CENSUSES AND SURVEYS
supported by STIFTUNG VOLKSWAGENWERK
(The numbers assigned are the report numbers.)

1/1 The Notions of Cross-Sectional and Longitudinal Studies.

2/1 The Notion of Privacy Transformations.

2/2 Non-Reversible Privacy Transformations – Scheme A.

2/3 Non-Reversible Privacy Transformations – Scheme B.

2/4 (With M. Post and S. Reiss) Non-Reversible Privacy Transformations.

2/5 (With M. Post) A Regression Technique for Protection of Micro-Statistics.

3/1 The Notion of Statistical Database Security.

3/2 Two Classes of Schemes for Counteracting Attempts at Compromising a Statistical Database.

3/3 Database Security: Some Small-Scale Computational Illustrations.

3/4 Statistical Database Security Through Matrix Sampling.

3/5 Matrix Sampling: A Small-Scale Computational Illustration.

3/6 A Procedure for Partial Permutation.

4/1 The Notion of Portable Statistical Disclosure Control.

4/2 Elementary Disclosure Problem – 1.

4/3 Elementary Disclosure Problem – 2.

4/4 Measuring the Impact of Rounding Counts in Contingency Tables.

4/5 (With M. Post) Software for Control of Elementary Disclosure Problem – 1.

4/6 (With M. Post) Software for Control of Elementary Disclosure Problem – 2.

4/7 Elementary Disclosure Problem – 3.

5/1 The Notion of Quasi-Identifiers.

5/2 Coping with the Problem of Quasi-Identifiers: Preliminary Considerations of a Special Case.

5/3 An Empirical Study of the Problem of Quasi-Identifiers.

5/4 Extreme Values and Disclosure Control.

5/5 Addressing the Quasi – Identifier Problem.

6/1 The Notion of an Intruder.

6/2 The Motivation of the Type-2 Intruder.

7/1 (With E. Denning) A Hybrid Scheme for Release of Statistics.

The Introduction of Probability Sampling in Swedish Acreage and Livestock Statistics in 1950. A Note on the Background and Later Developments

by Knut Medin[1]

1. Introduction

The February issue of the Journal of Agricultural Economics for 1951, which is published by the National Agricultural Market Board in Sweden, contained three articles. One of them dealt with the balance between agricultural production and food exports, one with fodder import losses and one with *the Use of Sampling Methods in the 1950 Acreage and Livestock Survey*. The latter was written by Tore Dalenius (1951).

I have mentioned all the articles in that issue because they together show the environment in which the methodological work in agricultural statistics was carried out around 1950.

In this note I shall take as my starting-point the article on the 1950 Acreage and Livestock Survey. My intention is to give a brief description of the methods used in that survey and to show in a somewhat longer perspective its place in Sweden's acreage and livestock statistics as part of this country's official statistics. I shall make special reference to developments after 1950.

2. Swedish agricultural statistics 1920–1950

Swedish agricultural statistics date back to the eighteenth century. Progress was, however, slow until about 1920 when intensive development work started at the Central Bureau of Statistics under the immediate direction of Ernst Höijer, who was later to become head of the Central Bureau (now Statistics Sweden). In a noteworthy paper Höijer (1921) suggested that Sweden should be divided into what he called "natural farming districts". This division, with some modifications, is still in use in Sweden's agricultural statistics.

The first census of agriculture was taken in 1927. From 1930 onwards annual acreage and livestock surveys were carried out with data from a non-random sample of about 10 per cent of all the farms in Sweden as their basis. In principle these data were to be supplied by the farmers themselves on special forms which were distributed and collected in July each year by the local branches of the County Agricultural Societies, which have a semi-official status.

Decisions on agricultural policy made by the Swedish Government during the economic crisis in the early 1930s highlighted the need to improve agricultural statistics. Improvements were made in the statistics of the means of agricultural production and of its results, and also of the economic situation of the farmers.

Consequently, in contrast with the situation on the eve of the First World War, at the outbreak of the Second World War Sweden had at its disposal much improved statistical information on agriculture and food production.

For neutral Sweden, too, the war led to a growing demand for statistics. New types of data were needed. There was an increasing call for reliability and timeliness. Further-

[1] Head of the Department for Area Statistics, Statistics Sweden.

more, a detailed regional breakdown of the statistics became necessary. Considerable resources, both finance and staff, were made available for this statistical work.

One significant change, among many others, was the replacement of the annual sample survey of acreages and livestock in the summer of 1940 by annual complete enumerations. The data were collected by the local Food Offices which had by then been set up.

Beyond any doubt agricultural statistics played an important part not only in the national planning of food production and in ensuring a fair distribution of the available food supplies, but also as a basis for the price regulation of agricultural produce. The problems caused by the war in no way ceased when it ended; they in fact characterized the whole of the 1940s.

The experience gained during these years had for a long time a pronounced effect on development work in agricultural statistics. However, it may be of some general historical interest to note that all the efforts expended on agricultural statistics in the 1940s had practically no lasting influence on statistical methodology[1] in this field. What merely happened was that procedures that were already well known were applied on a much larger scale.

3. The 1950 Acreage and Livestock Survey

When the food shortage eased, it was possible to reduce the scale of the agricultural statistics and thus the costs involved. In 1950 it was felt that it would be possible to return to sampling techniques for the annual acreage and livestock statistics. However, the "representative" method used during the 1930s was not applied again. Instead, use was made of the probability sampling methods developed during the 1930s and 1940s. Tore Dalenius was the foremost advocate in Sweden of these methods. In Dalenius' article (1951), which I mentioned at the beginning of this paper, he summarized the choice of technique and described how the work was actually carried out. In his doctor's thesis *Sampling in Sweden,* Dalenius (1957) also gave a detailed description in English of the planning and carrying out of the survey. I shall only take up a few of the main features here.

In brief, what had to be done was to compile statistics on the areas under different crops and on the numbers of certain kinds of animals for each of the some 60 "natural farming districts" in which Sweden was divided in the agricultural statistics. The results obtained could easily be combined into statistics for the 24 counties.

The basic data were collected from the individual farmers by the local municipal authorities. The statistics referred to the situation on a certain date at the beginning of June and were to be available at the beginning of August.

The sampling frame for the 1950 survey consisted of the forms from the 1949 survey which, at least in principle, comprised all the farms in Sweden (with a very low cut-off limit). The sample selected was used not only in 1950 but also for several surveys during the years that followed, and for certain other surveys, too.

It was decided that the sample should contain the same proportion of farms as the surveys in the 1930s, that is about 10 per cent, or 27 000, of all farms in Sweden.

For each "natural farming district" the farms were stratified according to the area of arable land. Within each stratum the units in the frame (i.e. the 1949 forms) were arranged roughly in order according to their total area of arable land. Systematic sampling, with the first unit selected at random, was then applied.

[1] If one takes together not only basic statistics but also their application, the *total calculation of agricultural income and costs* (totalkalkylen) and certain other *balances* may be regarded as lasting methodological contributions from the 1940s.

4. What did probability sampling mean to Swedish agricultural statistics?

The views on sample surveys put forward during the 1930s and 1940s brought about radical changes in official statistics all over the world. The principal significance of these changes is well documented. But what were the consequences for Swedish agricultural statistics? I shall just make some brief comments on the question here from my own personal experience of the field, which covers the period from 1952.

When statistical investigations were being planned and their findings used, it was a very great advantage that *more specific objectives* for the work could be set by using well defined concepts such as population, parameters, etc. From the point of view of an outsider, the changes may perhaps have not appeared to be very important. Nevertheless, they had a profound effect on the way in which agricultural statisticians argued and worked. The latter were now provided with an *efficient educational instrument.* As a result the users of agricultural statistics soon got used to expressing themselves with a greater degree of exactitude than they had before.

One important consequence of these developments was an *improvement in the internal co-ordination* between the various sections of our agricultural statistics. On this point, however, the work is not complete and developments are still in progress on the basis of the methodological renewal about 1950.

On the other hand, it cannot be denied that the new developments have sometimes caused a certain *lack of sensibility* among statisticians, for instance in their formulation of the objectives for new investigations. The need for co-ordination often conflicts with the need for flexibility in the approach to new problems.

Agricultural statistics deal traditionally with the number of farms in different size-groups, the total area under different crops, the average yield per hectare, the total number of certain animals, the size of the labour force and the average income per farm, etc. However, for various reasons a more detailed statistical picture of the agricultural situation is being increasingly asked for. The picture required may, for instance, be the distribution between different regions, between different income brackets, between different types of farming, etc. Sometimes what is wanted is the identification of problem areas of various kinds. For these requirements the modern survey technique is generally suitable. Nevertheless, the consequence is a heavy burden of responsibility for the statisticians who have to be able to re-formulate the problems in statistical terms.

I pointed out above that the acreage and livestock surveys of the 1930s were based on data drawn from a 10 per cent sample of farms. The latter were selected locally in accordance with certain directives but without a probability theory approach. In the estimation procedure the farms were divided into size-groups and they were then weighted with the aid of the statistics from the most recent agricultural census. This procedure can be regarded as an intuitive approach. The new sample survey theory provided a *scientific foundation* for the selection and estimation procedure.

Of course the statisticians who were working at the time, as well as the users of the statistics, had some doubts about the new technique when it was launched. However, once introduced, it was rapidly accepted. Some people even had such *blind faith* in the new technique that random sampling was almost regarded as the ultimate solution to all statistical problems. It may also be mentioned that many people thought that random sampling meant that *all* the farms in Sweden must inevitably be regarded as the survey population.

One important consequence of the new sampling technique was that the *sampling error could be calculated.* On this point developments later on – at least as far as Swedish

agricultural statistics are concerned – were not what might have been expected. In the official statistical publications the standard error is normally given in the hope that users will find the information useful. It is, however, obvious that even statistically qualified users do not use the information to any great extent. It has also been difficult to obtain a constructive dialogue on the sampling error between the producers and users of the statistics. When users are questioned about this, the immediate comment is frequently "You statisticians know best what precision is needed". Reliable point estimates are what are required. Above them some uncertainty is vaguely accepted.

5. The later development of the annual acreage and livestock surveys

The technique introduced in the 1950 Acreage and Livestock Survey was used with certain modifications for almost twenty years. I shall now take up some of the modifications as well as some of the methodological experiences gained.

Since the sample used was based on the complete enumeration in 1949, it eventually became out of date. From the middle of the 1950s it became possible to use as a sampling frame a register of farms kept by a nation-wide farmers' organization, the National Farmers' Union. This register contained particulars such as a farmer's name and the total area of his arable land. The register was continuously kept up to date, at least in principle. As a rule, a new sample was drawn every three years for the annual acreage and livestock survey.

A rotating sample scheme was introduced in 1963. In the new scheme a sample representing one third of the municipalities was renewed each year. With the old procedure large "leaps" in the statistical series appeared, especially at county level, when a new sample was selected. With the new sampling scheme the effects of the changes were more evenly distributed over time.

EDP was introduced in the surveys in the early 1960s. For instance a rather large system was developed for automatic editing and correction (Medin and Larson (1964)). The coordination between automatic and manual routines was studied by Wilson (1967).

Another innovation at this time was the attempt to construct a system for the analysis and control of non-sampling errors. The question of non-sampling errors in the surveys had previously been dealt with by Dalenius (1951, 1957). A regular system of area measurements on a subsample of farms in the surveys was introduced in the sixties (Medin, Wilson and Karlström (1966), see also Polfeldt (1977)). Certain tests were made to check the statistics of the number of animals by the use of interpenetrating samples. While data from certain farms were obtained in the usual way, i.e. through the municipal authorities which at least during the 1960s normally used the regular mail service for the distribution and collection of forms, data from other farms were obtained through interviews. A similar technique was used to estimate the effect of changes in the crop classification.

I must stress that the changes and measures I have mentioned here, as well as others I have had to omit, can all be seen as being complementary to the 1950 plan whose principal structure remained unchanged until 1967. Before describing the new farm register system, which was introduced in 1968, I must however mention a special problem.

During the 1950s and the 1960s Swedish agriculture went through quite considerable structural changes. In a twenty-year period there was about a fifty per cent decrease in the number of farms because of amalgamations and the withdrawal of land from cultivation. As a result, when the survey was taken, the farms sampled had often disappeared entirely, or were completely changed and quite different from when the sampling frame was created. We had to develop special rules in

order to be able to determine which farm or farms, if any, were to be regarded as corresponding to each of the farms in the original sample.

We intended that these rules should also make it possible for us to produce statistics on the structural changes in agriculture, i.e. gross changes in the number of farms in different size-groups. We spent much time and effort on this work for a number of years. However, when we compared our findings with the results of the agricultural censuses every fifth year, we found that all our efforts had been in vain. It became obvious that structural changes could not be measured by analysing the available information on changes in sampled farms. It was clear that this was not due to logical shortcomings in our survey technique but that it was connected with the difficulty, or impossibility, of obtaining correct and comprehensive information on changes in the selected units in a sample survey for which the basic particulars were collected by mail. In this field, however, quite new potentialities emerged when a Register of Farms was set up in 1968.

6. The Register of Farms

A general crop insurance scheme was introduced in Sweden in 1961. By this scheme the damage to an individual farm's yield is calculated on the basis of (1) the area under different crops on the farm and (2) the deviation from normal of the actual yield value per hectare for each crop as estimated statistically for the crop-yield district to which the farm belongs. In the crop insurance scheme the whole of Sweden is divided into about 400 crop-yield districts. The actual yield is obtained by means of the objective crop-yield survey, which is one of the largest statistical undertakings carried out by Statistics Sweden. The normal yield is calculated on the basis of historical data. These are the main rules which have of course to be supplemented in practice. An outline of this scheme is available in English (see Medin (1965)); it is however not entirely up to date. Right from the outset the technical administration of the crop insurance scheme has been the responsibility of Statistics Sweden.

When the crop insurance scheme was first introduced, acreage data for the individual farms were collected on a special application form. Because of the construction of the insurance scheme – the damage per hectare is estimated per district and not for the individual farm – many farmers were not certain whether they should submit an application for compensation or not. It was found that many farmers did not get the sums they were entitled to. After a thorough review of the matter, the submission of an application form was made compulsory for all farms from 1968 and co-ordinated with the collection of data for agricultural statistics.

In the new Register of Farms system, which has been described by Sandqvist (1971), all farms having at least two hectares of arable land, on a certain date in June each year, must supply information by mail to Statistics Sweden on partially preprinted forms.

This information is used for

- the crop insurance scheme,
- certain other administrative purposes,
- the annual statistics of farm structure, acreages, crops, animals, etc.,
- extended statistics every five years (called the agricultural census years).

The statistics are extended in census years also to cover, first, farms with at most two hectares of arable land or with forest land only and, second, a few additional variables. In addition, the findings are published in greater detail.

Since the annual statistics from the Register of Farms system have now replaced the annual acreage and livestock survey, some comparisons can be made between the two ways of producing agricultural statistics.

First of all, I must stress that not much has changed with regard to types of statistics, definitions and classifications. There has, however, been one important change; a far-reaching regional breakdown can now be made annually, not only for census years. This is of importance primarily because of the greater need for statistical data for local government purposes.

The new system has one serious drawback in that it leads to severe delays in publication. The issuing of preliminary estimates has to some extent remedied this problem.

Studies have been made of the quality of the statistics from the old surveys and they have been found to be reasonably good. Nevertheless, the introduction of the Register of Farms system has undoubtedly made certain improvements possible. In the Register system the data collection for a certain year is based on addresses, etc. collected and checked in the previous year. Since all farms with at least two hectares of arable land are included, a thorough follow-up of almost all land transfers between the farms can be regularly made. We now have much less difficulty in identifying the units of investigation. As a consequence the statistics are of good quality not only for the individual years but also for the changes from year to year.

The latter improvement is an important one. A systematic longitudinal analysis results in statistics of gross changes, that is the number of farms in all combinations of size-groups for two consecutive years. The technique applied has been described by Medin and Wilson (1974). The future size distribution of Swedish farms can be forecasted on the basis of these statistics; this has been described by Thorburn (1981). My remarks here about the size distribution of farms can also relate to the size of herds, types of farming and the age distribution of farmers. In those fields development work is now in progress.

The non-response rate in the annual acreage and livestock surveys was normally between 1 and 5 per cent. Since one of the aims of the Register system is to make it possible to follow individual farms over time, the non-response rate must be kept down. Through the joint efforts of Statistics Sweden and the 24 County Agricultural Boards, the annual non-response rate has been kept down at 0.1–0.3 per cent.

The Register of Farms can be used as a sampling frame for sample surveys in agriculture. In the Register a consistent identification system is applied. Efficient sampling and estimation procedures can be used since the Register contains comprehensive information on each individual farm. Owing to the existence of the Register, it has been possible to coordinate to a high degree the parts of agricultural statistics that are based on data from or for the farms.

To sum up: the introduction of probability sampling in the 1950 Acreage and Livestock Survey was an outstanding event in the long history of agricultural statistics in Sweden. For numerous and varying reasons there have often been changes in the statistical technique applied. This can be illustrated by the following list, which only covers rather more than 50 years.

1930–1939	Non-random sample survey
1940–1949	Complete enumeration
1950–1967	Probability sample survey
1968–	Complete enumeration = Register of Farms.

7. References

Dalenius, T. (1951): The Use of Sampling Methods in the 1950 Area Survey and Cattle Census. (Swedish with summary in English.) Jordbruksekonomiska Meddelanden 1951(2), pp. 59–75, 44.

Dalenius, T. (1957): Sampling in Sweden. Contributions to the Methods and Theories of Sample Survey Practice. (Diss. Uppsala.) Stockholm/Uppsala 1957.

Höijer, E. (1921): Sveriges uppdelning på naturliga jordbruksområden. (In Swedish.) Statsvetenskaplig tidskrift 1921, pp. 209–231, pp. 321–348.

Medin, K. (1965): Crop Yield Estimation and Crop Insurance in Sweden. Review of the International Statistical Institute, Vol. 33(3), pp. 414–442.

Medin, K. and Larson, B. (1964): The New System for Agricultural Statistics. II. Automatic Control and Correction of Primary Data. (Swedish with summary in English.) Statistisk tidskrift 1964:6, pp. 393–403, pp. 466–467.

Medin, K. and Wilson, B. (1974): Farm Structure in Figures. A Study in Statistical Methodology. European Review of Agricultural Economics 1(4), pp. 461–481.

Medin, K., Wilson, B. and Karlström, F.O. (1966): The New System for Agricultural Statistics. IV. Field Surveys for Control of the Area Data Obtained. (Swedish with summary in English.) Statistisk tidskrift 1966:3, pp. 181–207, pp. 251–255.

Polfeldt, T. (1977): Reexamination of the Area Checking Surveys. (Swedish with summary in English.) Statistisk tidskrift, 1977:3, pp. 202–218, pp. 269–271.

Sandqvist, E. (1971): The Agricultural Enterprise Register. (Swedish with summary in English.) Statistisk tidskrift 1971:5, pp. 397–411, pp. 434–437.

Wilson, B. (1967): The New System for Agricultural Statistics. V. Manual Routines. (Swedish with summary in English.) Statistisk tidskrift 1967:1, pp. 36–46, pp. 64–66.

Thorburn, D. (1981): Projections by Means of Longitudinal Studies and Transition Matrices. (Swedish with summary in English.) Statistisk tidskrift 1981:3, pp. 189–201, pp. 238–239.

On the Efficiency of Sampling with Unequal Probabilities

by Bertil Matérn[1]

1. Introduction

Sampling with unequal probabilities appeared in the 1940's, cf. the survey by Hanif and Brewer (1980). An early example of the use of this device in Sweden was presented by Dalenius (1953).

We consider a population of N units. The ith unit, U_i, has two quantitative characters, X_i, and Y_i, where X_i is > 0 and can be thought of as a measure of size. The population totals are X and Y. The sizes X_i are known, whereas the Y_i are unknown. To estimate the total Y we select a sample of n units. The selection is made with probabilities proportional to the X_i. The method is usually called "pps-sampling", where pps stands for "probability proportional to size". We estimate Y by multiplying the known total X by the sample mean of the ratios Y_i/X_i.

An actual survey can have a complicated structure involving stratification, multi-stage and multi-phase procedures, etc. The pps-sampling can then be related to a particular step in the survey such as the choice of one type of units from one stratum. The background of the present article are some not too complicated applications in forestry. An example is the following: N is the number of stems in a forest (or possibly in a first phase sample), X_i is the basal area (i.e. the area of the cross-section at breast height), and Y_i is the volume of the ith stem. The selection of sample trees with probabilities proportional to basal area is usually achieved automatically by means of Bitterlich's "relascope". (See for example Kendall and Moran (1963), p. 47.) In the forestry examples n and N/n are typically rather large.

Sampling with pps is *one* way of using the information existing on the X_i. Applying instead equal probability sampling, the X_i can be used for stratification, for forming ratio or regression estimates, etc. Comparisons with such methods are found in the literature. (See e.g. Cochran (1977), section 9 A.5, Foreman and Brewer (1971), and papers quoted by these authors.) The present article aims at giving some additional illustrations.

2. A general expression for the variance in the large sample case

We assume, as indicated above, that *both n and N/n are large.* We can then confine our attention to *asymptotic expressions.* We can hence also assume that the sampling is *with replacement.*

In each one of n draws U_i has the probability p_i (> 0) of selection. Let t_i denote the number of times that U_i appears in the sample ("Cornfield's indicator"). The expressions given in the sequel can all be derived from the two simple formulas

$$E(t_i) = np_i \quad \text{and}$$

$$\mathrm{Cov}(t_i,t_j) = np_i(\delta_{ij}-p_j),$$

where δ_{ij} is the Kronecker delta (1 if $i = j$, 0 otherwise). We first introduce X' and Y' by the relations

[1] Professor Emeritus of Forest Biometry, Swedish University of Agricultural Sciences.

$$nX' = \Sigma(t_i X_i/p_i) \text{ and}$$

$$nY' = \Sigma(t_i Y_i/p_i).$$

Here, as in the following formulas, the summation is over the whole population. We then consider the estimator

$$\hat{Y} = Y' + k(X-X') \qquad (1)$$

of Y. In (1) k is either given a priori or based on the sample. However, in the large sample case, we need not distinguish between the two cases. We assume that in the latter case k is an efficient estimator of the corresponding population value. In the asymptotic variance expression, we can interpret k as signifying the population parameter. We then have the asymptotic expression

$$n \operatorname{Var}(\hat{Y}) = \Sigma(Y_i - p_i Y - k(X_i - p_i X))^2/p_i.$$

If we let (y,x,p) denote a random variable representing the triple of values (Y_i,X_i,p_i) attached to a randomly (equal probabilities) selected unit in the population, we can replace the expression above by

$$(n/N) \operatorname{Var}(\hat{Y}) = E((y-pY-k(x-pX))^2/p). \qquad (2)$$

3. Comparison with the ratio estimator

The usual pps-estimator in sampling with replacement (the Hansen-Hurwitz-estimator) is obtained by choosing

$$p_i = X_i/X$$

in (1). In this case the value of k is immaterial and $\hat{Y} = Y'$.

We denote the estimator $\hat{Y}_{\text{pps}}$. It is given by

$$n\,\hat{Y}_{\text{pps}} = X\Sigma(t_i Y_i/X_i).$$

The variance is found from

$$(n/N^2) \operatorname{Var}(\hat{Y}_{\text{pps}}) = Ex\, E((y-Rx)^2/x), \qquad (3)$$

where $R = Y/X$.

The ratio estimator, $\hat{Y}_{\text{ra}}$, is obtained from (1) by choosing $p_i = 1/N$ for all i and $k = Y'/X'$. The asymptotic variance is obtained from

$$(n/N^2) \operatorname{Var}(\hat{Y}_{\text{ra}}) = E((y-Rx)^2) = \\ = Ex\, E((y-Rx)^2/x) + \operatorname{Cov}(x,(y-Rx)^2/x). \qquad (4)$$

Thus the relative efficiency of $\hat{Y}_{\text{pps}}$ with respect to $\hat{Y}_{\text{ra}}$ is

$$\text{eff(pps/ra)} = 1 + \frac{\operatorname{Cov}(x,(y-Rx)^2/x)}{Ex\, E((y-Rx)^2/x)} \qquad (5)$$

for fixed n. In practical cases of pps the covariance on the right hand side is usually positive.

As an illustration, we consider a simple parametric case in which we also introduce an asumption about costs. (The assumptions may be thought of as valid in a superpopulation or as approximately valid in a large actual population.) In the conditional distribution for given x, y is assumed to have the expectation Rx and the residual variance proportional to x^d. The marginal distribution of x is supposed to be of the Γ-type with density proportional to $t^{\lambda-1}\exp(-t)$. The cost of including U_i in the sample is supposed to be proportional to X_i^c. To compare cases with the same total cost we assume that the number of units in the two samples has the relation

$$n_{\text{pps}}/n_{\text{ra}} = Ex\, Ex^c/Ex^{c+1} = \lambda/(\lambda+c).$$

The relative efficiency, under constant total costs, is then found to be

$$\text{eff(pps/ra)} = \frac{\lambda+d-1}{\lambda+c}$$

or, noting that the coefficient of variation of x is $V_x = 1/\sqrt{\lambda}$,

$$\text{eff(pps/ra)} = \frac{1+(d-1)V_x^2}{1+cV_x^2}. \qquad (6)$$

Although the assumptions are chosen for analytical convenience, (6) seems to give a good illustration of how the comparison is affected by costs (parameter c), residual variance (parameter d) and the relative variation in size (V_x). It seems likely – when the assumptions are at all realistic – that $1 < d < 2$ (Foreman and Brewer (1971)) and that $0 < c < 1$. Thus, for pps to gain we must have

$$d - 1 > c$$

so no general conclusion can be reached. Formula (5) also illustrates the obvious fact that when $V_x \to 0$ the two schemes tend to be equally efficient, and of course not more efficient than the simple estimator Y'.

4. Comparison with the regression estimator

In the previous example it was assumed that the proportionality – basic to the two schemes compared – was valid, i.e. that the regression of y on x was linear and passed through the origin. However, the formulas (3)–(5) are general. Roughly speaking, they indicate that $\hat{Y}_{pps}$ and $\hat{Y}_{ra}$ are affected in much the same way if the regression behaves otherwise, cf. Foreman and Brewer (1971).

In a practical survey many variables of the type y may be under observation simultaneously, and not even an approximate proportionality can be valid in every case. Returning to the forestry example, we note that a much used model is the so called volume-basal area line (also named "Kopetzky's Massenlinie")

$$y = A + Bx \tag{7}$$

expressing the average relation between volume (y) and basal area (x) of the stems in a forest stand. Here A is usually $\neq 0$.

In other cases the relation may be more complicated. We take now, however, the simple equation (7) as representative of the cases in which the regression is *not* $y = Rx$.

A commonly used estimator of Y under equal probability sampling is then the regression estimator which one gets from (1) by putting $k = b$, the sample regression coefficient. We denote this estimator $\hat{Y}_{reg}$. Its asymptotic variance is found from

$$(n/N^2)\ \mathrm{Var}\,(\hat{Y}_{reg}) = E((y-A-Bx)^2). \tag{8}$$

This is a generally valid expression if we interpret A and B as the parameters of the *mean square regression line* in the population. (On this concept, see Cramér (1945), p. 302.)

We now suppose, however, that $y = A + Bx$ is the conditional expectation of y given x. We can then rewrite (3) in the following way:

$$(n/N^2)\ \mathrm{Var}(\hat{Y}_{pps}) = Ex\ E((y-A-Bx)^2/x) + \\ + (R-B)^2\ Ex\ E((x-Ex)^2/x). \tag{9}$$

The first term on the right hand side represents the influence from the residual variation around the regression line. The second term represents the bias of the line $y = Rx$ when the true regression is $y = A + Bx$. It can also be written $Ex\ E((A+Bx-Rx)^2/x)$. In comparison, $\mathrm{Var}\,(\hat{Y}_{reg})$ contains only an expression for the residual variation, which means that $\hat{Y}_{pps}$ can have a much lower precision than $\hat{Y}_{reg}$. Cf. also the illustration below.

In a situation in which the pps-sampling is a fact we can hence often have reasons to replace $\hat{Y}_{pps}$ by another estimator. As already indicated, (1) offers no possibilities since $X' = X$ under pps-sampling. However, an obvious choice is to make a ratio estimation with X_i replaced by $Z_i = a + bX_i$. We thus introduce the modified pps-estimator

$$\hat{Y}_{mod} = \hat{Y}_{pps}\, Z/Z', \tag{10}$$

where

$$Z' = (aX/n)\Sigma(t_i/X_i) + bX,$$

$$Z = aN + bX$$

and $y = a+bx$ is a sample estimate of the population regression $y = A+Bx$. If a and b are efficient estimates, we can use the following formula to find the large sample variance of $\hat{Y}_{mod}$:

$$(n/N^2)\,\mathrm{Var}(\hat{Y}_{mod}) = Ex\,E((y-A-Bx)^2/x). \tag{11}$$

It is seen that (11) bears the same relation to (8) as does (3) to (4).

We now make similar parametric assumptions as in Section 3. It is assumed that the conditional expectation and variance of y for given x are $A+Bx$ and $d_0\,x^d$, respectively, that the unit cost is proportional to x^c, and that x has the density function

$$t^{\lambda-1}\exp(-t)/\Gamma(\lambda).$$

We find comparing $\hat{Y}_{mod}$ and $\hat{Y}_{reg}$:

$$\mathrm{eff(mod/reg)} = \frac{1+(d-1)V_x^2}{1+cV_x^2} \tag{12}$$

where the right hand side is the same as in (6). If $d-1>c$, pps is a better allocation than equal probability sampling, and (12) shows that $\hat{Y}_{mod}$ utilizes the advantage of this allocation. To compare $\hat{Y}_{mod}$ and $\hat{Y}_{pps}$ we first introduce

$$\varphi = E((Rx-A-Bx)^2)/E((y-A-Bx)^2)$$

as an expression of the relation between bias and residuals. We note that we must assume $\lambda>1$, in other words $V_x<1$ for $\mathrm{Var}(\hat{Y}_{pps})$ to be finite. We then find

$$\mathrm{eff(mod/pps)} = 1+\varphi+\varphi d\,V_x^2/(1-V_x^2) \tag{13}$$

showing that the gain of $\hat{Y}_{mod}$ over $\hat{Y}_{pps}$ can be substantial. (The relative efficiency eff(pps/reg) is found by dividing (12) by (13).)

In conclusion of this section we note that pps-sampling can be advantageous also in cases when there is no overall proportionality between y and c. However, the usual pps-estimator should then be modified.

5. Comments on the small sample case

In many applications n is quite small, for example in stratified sampling with pps used in each one of many strata. Studying the literature on sampling with unequal probabilities one also finds that a substantial part of the theoretical investigations are devoted to this case. For example, a rather large number of different methods of selecting pps-samples with $n = 2$ units without replacement have been proposed, see again the review by Hanif and Brewer (1980). (These proposals are relevant when also N/n is small.)

The interest in the small sample case can be justified by the fact that the common ratio and regression estimators (with equal probability sampling) are not robust: Although these estimators are consistent they can be seriously biased when n is small. The difference estimator – formula (2) with k chosen a priori – can be used for $n \geq 1$, and stratification is a possible alternative if $n \geq 2$.

Of the previously listed variance formulas, (3) and (9), referring to $\hat{Y}_{pps}$, and (4) interpreted as related to a difference estimator, are valid for all n. However, as is well known, the asymptotic expressions of the ratio and regression estimates are useless for very small n. To illustrate this, we can consider the mean square error of the ratio estimate in the case $n = 1$, which is

$$x^2\,E((\frac{y-Rx}{x})^2).$$

If we make the same parametric assumptions as in Section 3, we find (with $n = 1$ in both cases):

$$\mathrm{eff(pps,ra)} = \frac{1}{1+V_x^2(d-2)}$$

which should be compared with the numerator

$$1+(d-1)V_x^2$$

in (6). Thus low values of d favour pps when $n = 1$ but favour $\hat{Y}_{ra}$ when $n \to \infty$.

6. Summary

In estimating the mean or total of a character y when the values of a supplementary variable x (>0) are known in the whole population, use is often made of pps-sampling, i.e. sampling with probabilities proportional to x. In this article comparisons are made – in the large sample case – with other methods of using the supplementary information. When there is not an overall proportionality between y and x, pps can still be used, but the common pps-estimator should be modified. It is further argued that the case for pps is stronger in small samples.

7. References

Cochran, W.G. (1977): Sampling Techniques. 3rd edition. Wiley.

Cramér, H. (1945): Mathematical Methods of Statistics. Geber.

Dalenius, T. (1953): Some Aspects of the Use of Objective Methods for Crop Estimation. (In Swedish, with summary in English.) Kungl. Lantbruksakademiens Tidskrift 92, pp. 99–118.

Foreman, E.K. and Brewer, K.R.W. (1971): The Efficient Use of Supplementary Information in Standard Sampling Procedures. Journal of the Royal Statistical Society, Series B 33, pp. 391–400.

Hanif, M. and Brewer, K.R.W. (1980): Sampling with Unequal Probabilities without Replacement: A Review. International Statistical Review 48, pp. 317–335.

Kendall, M.G. and Moran, P.A.P. (1963): Geometrical Probability. Griffin.

Variance Reduction and Unbiasedness for Small Domain Estimators

by Carl-Erik Särndal[1] and Göran Råbäck[2]

1. Introduction

Dalenius (1977), with his flair for poignant expression, has entitled one of his articles "Strain at a gnat and swallow a camel, or, the problem of measuring sampling and non-sampling errors", an item out of many in a distinguished and influential stream of publications. There he argues rightly that the sampling error in survey estimates is often small and may receive too much attention in comparison to the more elusive non-sampling errors, including those due to imprecise measurement and nonresponse. However, one area where a discussion of the sampling error's magnitude is highly relevant is the estimation of parameters for small domains of a sampled finite population.

In a survey with a total sample size of the order of $n = 10\,000$, estimates formed at the level of the whole population can indeed be expected to be very precise from the sampling error point of view, but it goes without saying that domains which account for .001 or less of the population will receive so few of the total sample's observations that all hope for a negligible sampling error must be abandoned when estimates are made at the domain level. In addition, the non-sampling errors may be as large as ever.

Not surprisingly, therefore, model oriented procedures attempting to exploit strong structural relationships and any available sources of auxiliary information are currently very much in vogue for small domains. Such procedures include the well known synthetic estimation and SPREE techniques. Thus an important area of research has opened up, the interest in such approaches having been spurred by growing demands for estimates useful in decisions at local levels. It appears that in most large scale surveys, estimates for small local areas have the status of a somewhat neglected by-product, perhaps necessarily so. Budgetary decisions are often made with a view to producing solid estimates for the population at large, or for large parts thereof. For small domains one ends up doing "the best one can". There is so much greater reason to adopt a critical attitude towards errors of estimates, in particular their sampling error.

Purcell and Kish (1979) point out that the choice of estimation technique for small domains is intimately connected with domain size. They classify domains into four types: Major domains (comprising 10 % of the population or more), Minor domains (between 10 % and 1 % of the population), Mini domains (between 1 % and .01 % of the population), and Rare types of individuals (less than .01 % of the population).

Our study is concerned essentially with Minor domains; to see how well the techniques hold up for Mini domains requires further work. It is our purpose to confront, by empirical testing through Monte Carlo methods, two kinds of procedures both of which attempt to alleviate the scanty data problem by use of auxiliary information. If we wish, this amounts to confronting two current philosophies of estimation for surveys. They

[1] Professor of Statistics, Université de Montréal, Canada.
[2] Senior statistician, Statistics Sweden.

are (a) approximately design unbiased (ADU) methods; (b) model dependent, design biased methods, whose redeaming feature would be a smaller variance.

The ADU procedures are founded theoretically on the concept of asymptotic design unbiasedness; as any sample is necessarily finite, there is always a "technical bias" which turns out to be unimportant even in moderately large samples, provided the domain size is not exceptionally small. The methods we call model dependent include the popular synthetic estimator, originally proposed by the U.S.National Center for Health Statistics. Model dependent estimators may succeed well in some domains, but may have a large design bias in other domains, the consequence being that a confidence interval centered on the point estimate will almost certainly fail to contain the true point under estimation.

The quality of an estimator is measured here in the "traditional" way, namely, by design variance, that is, variance over repeated samples drawn by the same fixed (but arbitrary) sampling design. Consequently, a confidence coefficient is interpreted as the long-run rate of coverage of the true unknown point over repeated draws of samples. We emphasize this at the outset, since some may prefer "the other way", not used here, which proceeds by means of a model-based variance, that is, a variability arising under the stochastic structure contained in a hypothetical model assumption.

2. Estimators under study

Let $U=\{1,\ldots, k,\ldots, N\}$ be a finite population from which a sample s of fixed size n is selected by a given sampling design admitting the inclusion probabilities of first and second orders, π_k and π_{kl}, respectively. Let $U_{\cdot q}$ $(q=1,\ldots, Q)$ be domains with known sizes $N_{\cdot q}$ and let $n_{\cdot q}$ be the (random) number of observations contained in $s_{\cdot q}$, the part of s that happens to fall in $U_{\cdot q}$. We seek estimates of the unknown domain means

$$\bar{y}_{\cdot q} = \Sigma_{U_{\cdot q}} y_k / N_{\cdot q} (q=1,\ldots, Q).$$

First, the ADU estimators of $\bar{y}_{\cdot q}$ considered below are of the kind also known as generalized regression estimators. They capitalize on a variance reducing explanation of y in terms of the auxiliary variable vector $\underset{\sim}{z}$, whose values $\underset{\sim}{z}_1,\ldots, \underset{\sim}{z}_N$ are known. The estimator of $\bar{y}_{\cdot q}$ is of the general form (Särndal, 1981, 1982):

$$\hat{T}_q = \Sigma_1^N c_{kq}\hat{y}_k + \Sigma_s c_{kq} e_k/\pi_k, \qquad (2.1)$$

where $e_k = y_k - \hat{y}_k = y_k - \underset{\sim}{z}_k' \hat{\underset{\sim}{\beta}}$ with

$$\hat{\underset{\sim}{\beta}} = (\Sigma_s w_k \underset{\sim}{z}_k \underset{\sim}{z}_k')^{-1} \Sigma_s w_k \underset{\sim}{z}_k y_k \qquad (2.2)$$

is the residual arising in fitting the linear model (ξ, say) such that the y_k are independent (throughout) and $E_\xi(y_k) = \underset{\sim}{z}_k' \underset{\sim}{\beta}$ to the n data points of the sample. Here the weights in $\hat{\underset{\sim}{\beta}}$ are chosen as $w_k = 1/\pi_k\sigma_k^2$, where $\sigma_k^2 = V_\xi(y_k)$. Moreover, $c_{kq} = 1/N_{\cdot q}$ if $k \in U_{\cdot q}$ and $c_{kq} = 0$ otherwise. For a given sampling design, each model fit gives rise to a different ADU estimator.

An approximate variance estimate of T_q is given by the Yates-Grundy formula in $c_{kq}e_k$, that is

$$\hat{V}(\hat{T}_q) = \sum_{\substack{\sum\sum \\ k<l \\ s}} \Delta_{kl} \left(\frac{c_{kq}e_k}{\pi_k} - \frac{c_{lq}e_l}{\pi_l} \right)^2, \qquad (2.3)$$

where $\Delta_{kl} = (\pi_k\pi_l - \pi_{kl})/\pi_{kl}$. An approximate $100(1-\alpha)\%$ confidence interval for $\bar{y}_{\cdot q}$ is therefore

$$\hat{T}_q \pm u_{1-\alpha/2} \{\hat{V}(\hat{T}_q)\}^{1/2} \qquad (2.4)$$

where $u_{1-\alpha/2}$ is the unit normal deviate. Under repeated draws of samples s by the given design, roughly $100(1-\alpha)\%$ of all confidence intervals calculated by (2.4) should cover the true value $\bar{y}_{\cdot q}$. As revealed by the particular

cases covered by our Monte Carlo study, this contention is confirmed, except for very small domains.

The procedure is applicable for an arbitrary sampling design, but in the study reported below, the simple random sampling (srs) design is used; thus $\pi_k = f = n/N$ for all k and $\pi_{kl} = n(n-1)/N(N-1)$ for all $k \neq l$. The fit of a model usually requires a certain amount of auxiliary information. Here we assume that $x_1,\ldots, x_N$ are known positive values of an auxiliary variable x strongly correlated with y. The N values of x are size ordered and the population units are grouped, by intervals of x, into H disjunct groups $U_{h\cdot}$ of known size $N_{h\cdot}$ $(h=1,\ldots, H)$. The population is thereby cross-classified into HQ cells U_{hq} of known size N_{hq}. At the sample level, the corresponding notation is $s_{h\cdot}$, $n_{h\cdot}$, s_{hq} and n_{hq}. Thus,

$$n = \sum_{h=1}^{H} n_{h\cdot} = \sum_{q=1}^{Q} n_{\cdot q} = \sum_{h=1}^{H}\sum_{q=1}^{Q} n_{hq},$$

relations which also hold for the big N's.

It is assumed, for the practical applications, that the $N_{h\cdot}$ be relatively large and that $n_{h\cdot} > 0$ for all h.

We consider the following ADU estimators obtained from the general formula (2.1) when the sampling design is srs,

$$\hat{T}_{1q} = N_{\cdot q}^{-1} \Sigma_1^H (N_{hq}\bar{y}_{s_{h\cdot}} + f^{-1}\Sigma_{s_{hq}} e_{1k}) \qquad (2.5)$$

and

$$\hat{T}_{2q} = N_{\cdot q}^{-1} \Sigma_1^H (\hat{\beta}_h \Sigma_{U_{hq}} x_k + f^{-1}\Sigma_{s_{hq}} e_{2k}) \qquad (2.6)$$

where, for $k \in s_{h\cdot}$, $e_{1k} = y_k - \bar{y}_{s_{h\cdot}} = y_k - \Sigma_{s_{h\cdot}} y_k/n_{h\cdot}$ and $e_{2k} = y_k - \hat{\beta}_h x_k$ with $\hat{\beta}_h = \Sigma_{s_{h\cdot}} y_k/\Sigma_{s_{h\cdot}} x_k$. Moreover, when the variance is estimated by means of (2.3), we use $e_k = e_{1k}$ and $e_k = e_{2k}$, respectively.

The estimators $\hat{T}_{1q}$ and $\hat{T}_{2q}$ arise from (2.1) for the models stating that the y_k are such that, respectively ,

$$E_\xi(y_k) = \beta_h \,;\; V_\xi(y_k) = \sigma_h^2 \qquad \text{all } k \in U_{h\cdot} \quad (2.7)$$

(that is, $\beta_h = \underset{\sim}{z}_k'\underset{\sim}{\beta}$ where $\underset{\sim}{z}_k$ is a vector of 0–1-variables indicating group adherence), and

$$E_\xi(y_k) = \beta_h x_k \,;\; V_\xi(y_k) = \sigma_h^2 x_k \qquad \text{all } k \in U_{h\cdot}\,, \qquad (2.8)$$

where $x_1,\ldots, x_N$ are known and positive. An interesting point is that (2.5) and (2.6) produce an estimate even with zero observations in a domain.

Other ADU estimators under srs are the "simple direct estimator"

$$\hat{T}_{3q} = \Sigma_{s_{\cdot q}} y_k/n_{\cdot q} = \bar{y}_{s_{\cdot q}} \qquad (2.9)$$

and the "combined cell mean estimator"

$$\hat{T}_{4q} = \Sigma_1^H N_{hq}\bar{y}_{s_{hq}}/N_{\cdot q}. \qquad (2.10)$$

The former requires for its calculation that there be at least one observation in the domain; it is the basic procedure, but inefficient, for it uses no auxiliary information. The latter is frequently undefined for a given domain q, namely, as soon as a cell count n_{hq} appearing in the denominator of $\bar{y}_{s_{hq}} = \Sigma_{s_{hq}} y_k/n_{hq}$ is zero for some h. In other words, to make $\hat{T}_{4q}$ operational, a simple and practical rule for collapsing cells would first have to be defined, a complication not dealt with here.

Now $\hat{T}_{3q}$ and $\hat{T}_{4q}$ are also obtainable from the general formula (2.1) under obvious model formulations; their respective variance estimates are given by (2.3) with $e_k = e_{3k} = y_k - \bar{y}_{s_{\cdot q}}$ and $e_k = e_{4k} = y_k - \bar{y}_{s_{hq}}$.

Turning now to the model dependent estimators, we shall consider the much debated synthetic estimator

$$\hat{T}_{5q} = \Sigma_1^H N_{hq}\bar{y}_{s_{h\cdot}}/N_{\cdot q}. \qquad (2.11)$$

It requires the same auxiliary information as (2.5) above and (2.12) below, that is, know-

ledge of the population cell counts N_{hq}. Comparing (2.5) with (2.11), we may consider the additional term in (2.5) as a bias removal term; that this procedure for eliminating the design bias of the synthetic estimator works well is shown by the results of our Monte Carlo study.

Another group of model dependent alternatives that deserve attention for comparative purposes are those arising from the predictive approach, (Holt, Smith and Tomberlin (1979), Laake (1979)), that is, from the general predictor formula,

$$\hat{T}_q = \Sigma_s c_{kq} y_k + \Sigma_{U-s} c_{kq} \hat{y}_k,$$

where $\hat{y}_k$ is the predicted y_k-value under the model. For the models (2.7) and (2.8) we get, respectively (with $w_k = 1/\sigma_k^2$ in (2.2))

$$\hat{T}_{6q} = N_{\cdot q}^{-1} \Sigma_1^H (N_{hq} \bar{y}_{s_h} + \Sigma_{s_{hq}} e_{1k}) \qquad (2.12)$$

and

$$\hat{T}_{7q} = N_{\cdot q}^{-1} \Sigma_1^H (\hat{\beta}_h \Sigma_{U_{hq}} x_k + \Sigma_{s_{hq}} e_{2k}) \quad (2.13)$$

with e_{1k} and e_{2k} as in (2.5) and (2.6), respectively.

It is unclear what variance estimate to use for (2.11)–(2.13); suggestions made for the synthetic estimator (2.11) (Gonzalez and Waksberg (1973)) do not seem to carry the usual repeated sampling interpretation.

For (2.12) we try the variance estimator pooled over all groups h and suggested by Holt, Smith and Tomberlin (1979),

$$\hat{V}(\hat{T}_{6q}) =$$

$$N_{\cdot q}^{-2} d_q \sum_{h=1}^{H} \sum_{q=1}^{Q} \Sigma_{s_{hq}} (y_k - \bar{y}_{s_h})^2 / (n-H), \qquad (2.14)$$

where

$$d_q = \sum_{h=1}^{H} (N_{hq} - n_{hq})(N_{hq} - n_{hq} + n_{h\cdot})/n_{h\cdot}.$$

This is a model based variance estimate, and although we use it below in our Monte Carlo study for the construction of confidence intervals, we can expect a priori that the coverage rate achieved in repeated samples by these intervals will not generally come near the theoretical confidence level aimed at.

3. Description of the Monte Carlo simulation

We identified a population consisting of $N=1287$ Swedish households taken from the Household Income Survey known as HINK. The households were classified into $Q=12$ domains defined by household type (based on characteristics such as number of household members, age and occupation of head of household). Let y_k and x_k denote, respectively, disposable income and net income (as per the income tax return) of the k:th household. For every unit $k=1,\ldots,1287$, we assume that the following information is available: (1) the value of x_k obtained from the income tax return and (2) the identity of the domain to which k belongs. The disposable income y_k differs from x_k because of child allowances, rent subsidies, etc., and y_k can only be observed for households selected for the survey.

The domain sizes $N_{\cdot q}$ are thus known. The relative domain sizes $N_{\cdot q}/N$ varied between 1.3 % and 29.5 %. In the discussion below we distinguish the "six larger domains" varying in relative size from 7.1 % to 29.5 %, and the "six smaller domains" varying in relative size from 1.3 % to 4.7 %.

The target of estimation is the mean disposable income in each domain, $\bar{y}_{\cdot q} = \Sigma_{U_q} y_k / N_{\cdot q}$ $(q=1,\ldots,Q)$.

The x-variable was used to create $H=5$ net income classes. For our population, the correlation between x and y is .79. It is expected, therefore, that the x-dimension will have a considerable variance reducing effect. That is, the estimators $\hat{T}_{1q}$ and $\hat{T}_{2q}$, which capitalize on the x-variable, will be better than estimators

such as the simple direct formula $\hat{T}_{3q}$ which do not. Now, $\hat{T}_{5q}$, $\hat{T}_{6q}$ and $\hat{T}_{7q}$ also benefit from the x-information to achieve small variance, but their MSE is often high due to design bias.

Our Monte Carlo simulation consisted in the selection of 1000 samples, each of size $n=200$, from the population of 1287. Thus, the expected number of observations in a domain ranged from 2.6 for the smallest to 59 for the largest domain.

For each sample from 1 to 1000, and for $i=1,..,7$; $q=1,..,12$, we calculated the estimator value $\hat{T}_{iq}$, its estimated variance $\hat{V}_{iq} = \hat{V}_{srs}(\hat{T}_{iq})$, and the confidence interval

$$\hat{T}_{iq} \pm u_{1-\alpha/2}\, \hat{V}_{iq}^{1/2}$$

for $1-\alpha = 80\ \%, 90\ \%$ and $95\ \%$. In summary of the whole experiment we calculated the mean, ME_{iq}, and the variance, VA_{iq}, of the 1000 estimator values, as well as the coverage rate CR_{iq} defined as the fraction of the 1000 intervals that covered the true point $\bar{y}_{.q}$ (at nominal confidence levels of 80 %, 90 % and 95 %). With 1000 repetitions, ME_{iq}, VA_{iq} and CR_{iq} can be taken as fairly good measures of the true mean, variance and coverage rate of the respective procedures. The CR rate is conditional on there being at least so many observations in a domain that a variance estimate can be obtained; for example, it suffices with one observation for $\hat{T}_{1q}$ and $\hat{T}_{2q}$. The ME and VA calculation is conditional on the estimator being defined.

We discuss the main findings with respect to the concepts measured, (a) design bias; (b) design variance; (c) coverage rate. In Tables 1 to 3, results are given separately for "six larger domains" and "six smaller domains". For domains in each of these two categories, "max" and "min" gives, respectively, the largest and smallest value of the concept in question. This gives some indication of the deterioration of the quality of an estimator as the domain size decreases.

As expected, $\hat{T}_{4q}$ turned out to be undefined for many of the 1000 samples (because of at least one zero cell count n_{hq}) and $\hat{T}_{4q}$ is therefore excluded from the comparisons below.

(a) *The design bias.* For all domains, including the very small ones, ME_{iq} should according to theory be close to the true point $\bar{y}_{.q}$ for the ADU estimators ($i=1,2,3$), but not likely so for the design biased ones ($i=5,6,7$). This was confirmed by the Monte Carlo results reported in Table 1, which gives values of the relative bias $100\,(ME_{iq} - \bar{y}_{.q})/\bar{y}_{.q}$.

It should be noted that the sample size here is $n=200$, which is large enough to obtain the desired approximate design unbiasedness for the ADU group of estimators, even though the sample count in the smallest domains varies around an expected value as low as about 3 observations.

Table 1 reveals a striking contrast between the three ADU estimators on the one hand and the three model dependent estimators on the other hand.

The relative design bias in the ADU group is never important; the max and min values tend, however, to be slightly larger in the smaller domains, a not unexpected tendency, in view of the instability caused by a lack of observations.

Earlier claims that the design bias of the synthetic estimator $\hat{T}_{5q}$ can not be estimated from the sample are unfounded, considering the results of this study; the term $f^{-1}\Sigma_{s_{hq}} e_{1k}$ of $\hat{T}_{1q}$ given by (2.5) is shown to effectively remove the design bias of the synthetic estimator (2.11). For the model dependent group, the direction and magnitude of the relative design bias is highly unpredictable; one may guess that particularly large biases arise if the underlying model fits the population badly.

(b) *Design variance.* Among the ADU methods, $\hat{T}_{1q}$ and $\hat{T}_{2q}$ capitalize on the information contained in the x-variable and are expected to give smaller variances than the simple direct method $\hat{T}_{3q}$. On the other hand, we expect

Table 1. Features of the relative design bias 100 $(ME_{iq} - \bar{y}_{\cdot q})/\bar{y}_{\cdot q} = BI_{iq}$

Estimator	Six larger domains		Six smaller domains	
	Max BI_{iq} q	Min BI_{iq} q	Max BI_{iq} q	Min BI_{iq} q
ADU:				
$\hat{T}_{1q}$	.73	–.41	1.63	–1.62
$\hat{T}_{2q}$	.59	–.39	2.00	–1.59
$\hat{T}_{3q}$	.04	–.48	2.70	–1.16
Model dependent:				
$\hat{T}_{5q}$	26.0	–15.0	14.6	–34.1
$\hat{T}_{6q}$	22.1	–12.7	12.1	–30.0
$\hat{T}_{7q}$	20.9	–13.1	2.0	–24.1

Table 2. Features of the efficiency $EF_{iq} = \{VA_{1q}/VA_{iq}\}^{1/2}$ of $\hat{T}_{iq}$ relative to $\hat{T}_{1q}$

Estimator	Six larger domains			Six smaller domains		
	Max EF_{iq} q	Min EF_{iq} q	$\overline{EF}_{iq}$[1]	Max EF_{iq} q	Min EF_{iq} q	$\overline{EF}_{iq}$[1]
ADU:						
$\hat{T}_{2q}$	1.16	.77	1.00	1.09	.55	.93
$\hat{T}_{3q}$	.99	.69	.84	.91	.65	.75
Model dependent:						
$\hat{T}_{5q}$	2.53	.98	1.64	17.26	3.03	7.00
$\hat{T}_{6q}$	2.34	1.08	1.66	5.43	2.90	4.17
$\hat{T}_{7q}$	2.41	1.09	1.64	5.68	2.72	3.95

[1] The overbar indicates an avarage of over the domains concerned.

the model dependent methods $\hat{T}_{5q}$, $\hat{T}_{6q}$ and $\hat{T}_{7q}$ to profit even more strongly from the variance reducing effect of the x-dimension, unhampered as they are by the ADU-ness constraint.

These expectations are confirmed by the findings shown in Table 2, which uses the standard error of $\hat{T}_{1q}$ as the point of reference and shows max, min and average values of the efficiency measure $\{VA_{1q}/VA_{iq}\}^{1/2} = EF_{iq}$, separately for the six larger domains and for the six smaller domains. Within the ADU group, $\hat{T}_{1q}$ and $\hat{T}_{2q}$ perform about equally well. One notes that $\hat{T}_{1q}$ is better than $\hat{T}_{3q}$ in all domains, and the tendency is more pronounced in the smaller domains. If the gains of $\hat{T}_{1q}$ over $\hat{T}_{3q}$ seem relatively modest, we must consider that the .79 correlation between x and y is good but not extremely strong.

The variance reduction achieved by the model dependent methods is of more dramatic proportions and is also especially apparent in the smaller domains, where the synthetic estimator $\hat{T}_{5q}$ performs particularly well with an average efficiency measure of 7.0 relative to $\hat{T}_{1q}$.

(c) *Coverage rate.* Table 3 shows coverage rates, at nominal values $1-\alpha$ of 90 % and 95 %, achieved by the confidence intervals centered on $\hat{T}_{1q}$, $\hat{T}_{2q}$, $\hat{T}_{3q}$ and $\hat{T}_{6q}$. (Results for the nominal level 80 % are not shown, as they confirm more or less the tendencies found in the cases of 90 % and 95 %.)

The intervals were constructed according to

Table 3. Features of the coverage rate CR_{iq}. The upper (lower) half of the table refers to the nominal rate of $1-\alpha = 90$ % *(95 %)*

Estimator	Six larger domains		Six smaller domains	
	Max CR_{iq} q	Min CR_{iq} q	Max CR_{iq} q	Min CR_{iq} q
ADU:				
$\hat{T}_{1q}$	89.7	84.4	89.5	59.0
$\hat{T}_{2q}$	92.5	83.3	90.9	67.7
$\hat{T}_{3q}$	88.3	77.1	82.0	61.8
Model dependent:				
$\hat{T}_{6q}$	99.7	0.0	99.7	0.0
ADU:				
$\hat{T}_{1q}$	94.2	89.9	94.3	66.6
$\hat{T}_{2q}$	96.9	88.3	96.5	72.1
$\hat{T}_{3q}$	94.6	83.1	87.0	64.4
Model dependent:				
$\hat{T}_{6q}$	100.0	0	100.0	.3

(2.4), where $\hat{V}(\hat{T}_{iq})$ is given by the srs special case of the Yates-Grundy formula (2.3) for the estimators $\hat{T}_{1q}$, $\hat{T}_{2q}$ and $\hat{T}_{3q}$, and by (2.14) for $\hat{T}_{6q}$.

The coverage rates achieved by the ADU methods have a tendency to fall short of the desired nominal rate. For domains of relative size 5 % or more, the drop is of no great importance, but for our smallest domains the tendency becomes highly noticeable. The achieved CR drops, in the worst case, to about 60 % (for an intended 90 %) and 65 % (for an intended 95 %). This is unsatisfactory but not so surprising, considering that the smallest domains receive extremely few non-zero values $c_{kq}e_k/\pi_k$ on which to base the calculation of the estimated variance.

The coverage rates of the design biased method $\hat{T}_{6q}$ are highly unpredictable and usually bear no resemblance to the nominal rates aimed at. We conclude that a model based variance estimate such as (2.14) is of little use in attempting to construct confidence intervals with the customary in randomization theory interpretation with repeated draws of samples s under the given sampling design. The model dependent estimators $\hat{T}_{5q}$ and $\hat{T}_{7q}$ were omitted from the coverage rate study because of the apparent lack of a confidence interval procedure that carries the randomization theory interpretation.

4. Conclusions

Our study shows that proper use of auxiliary information can considerably alleviate the dilemma of lack of precision in estimates for small domains. The strength of the auxiliary information is in fact all-important for the success of the ADU estimation techniques.

Our study is limited in that our smallest domains are not really that small, considering that domain sizes .001 or less occur often in practice. On the other hand, the correlation in the population between the variable of interest, y, and our auxiliary variable, x, was "only" .79, a decent but not exceptionally powerful relationship. Even at this, our ADU techniques improved distinctly over the

simple direct procedure. But at the same time, the confidence intervals for the ADU methods $\hat{T}_{1q}$ and $\hat{T}_{2q}$ were often, for our smallest domains, so wide that, seen in relation to the value of the interval's central point, they convey little information for the user. On the other hand, the model dependent intervals are often equally uninformative because they are completely off target, due to design bias.

A lesson to be learned for ADU estimation seems to be that auxiliary variables explaining .95 or more of the variable of interest are strongly needed. And even under such condition, can it carry us to useful, informative interval estimates for domains of relative size, say, in the .005 area or less? Put differently and more generally, does any technique exist that will allow us to reduce sampling errors that are almost limitless?

5. References

Dalenius, T. (1977): Strain at a Gnat and Swallow a Camel: Or the Problem of Measuring Sampling and Non-Sampling Errors. American Statistical Association, Proceedings of the Social Statistics Section, pp. 21–25.

Gonzalez, M.E. and Waksberg, J. (1973): Estimation of the Error of Synthetic Estimates. Paper presented at the 1st meeting of the International Association of Survey Statisticians, Vienna.

Holt, D., Smith, T.M.F. and Tomberlin, T.J. (1979): A Model-Based Approach to Estimation for Small Subgroups of a Population. Journal of the American Statistical Association, 74, pp. 405–410.

Laake, P. (1979): A Prediction Approach to Sub-Domain Estimation in Finite Populations. Journal of the American Statistical Association, 74, pp. 355–358.

Purcell, N.J. and Kish, L. (1979): Estimation for Small Domains. Biometrics 35, pp. 365–384.

Särndal, C.E. (1981): Design Consistent versus Model Dependent Estimation for Small Domains. To appear, Journal of the American Statistical Association.

Särndal, C.E. (1982): Implications of Survey Design for Generalized Regression Estimation of Linear Functions. Journal of Statistical Planning and Inference, 7, pp. 155–170.

Factorial Series Distributions and Estimation Based on the Distinct Units of a Sample

by Sven Berg[1]

1. Introduction

A survey population consists of a finite number of identifiable units. Certain sampling designs involve possible repetitions of the same unit from the finite population, either because the sampling frame contains multiple listings or because sampling is done with replacement. The repetitions can be ascertained through the use of identifying labels. In the absence of measurement errors, replicates clearly provide no additional information about the population parameters. On the contrary, they tend to increase the error variance of estimates. Therefore, if possible, estimates of population parameters should be based on the distinct units obtained through the sampling design. In a seminal paper, Basu (1958) showed that in simple random sampling with replacement, the mean based on the distinct units is a better estimate of the population mean than the one based on all the units of the sample. Since then quite a few papers have been devoted to the application of the concepts of sufficiency and admissibility to survey sampling problems. Some of these papers deal with estimation based on the distinct units for specific sampling designs.

In this essay we show that some of the problems that arise in connection with estimates based on the distinct units of the sample can be fruitfully discussed within the framework of a class of discrete distributions, called factorial series distributions. First, in Section 2, the class is presented and a useful moment property is derived, giving a simple formula for computing reciprocal moments. Then, in Section 3, a class of designs is introduced, in which the number of distinct units follows a factorial series distribution. A number of well known designs are shown to be members of this class. In particular, interpenetrating subsampling, post-cluster sampling (a design we owe to Dalenius (1957)), and sampling from overlapping frames are interesting special cases.

[1] Department of Statistics, University of Lund, Sweden.

2. Factorial series distributions

The discrete random variable (r. v.) X is said to have a factorial series distribution (FSD) if

$$Pr(X=x) = p(x) = N^{(x)}a_x/g(N),\ x=0,1,\ldots,N, \tag{1}$$

where $N^{(x)}$ is a descending factorial $N^{(x)} = N(N-1)\ldots(N-x+1)$, and N is the integer-valued parameter. In analogy with power series distributions, FSDs are generated by expanding a parametric function $g(N)$ in a factorial series. The function $g(N)$ is called the series function (s. f.) and, finally, a_x is independent of the parameter N. Expressed in terms of "differences of zeros", a_x in (1) can be written:

$$a_x = \frac{1}{x!}\Delta^x g(N)\,/N=0 = \frac{1}{x!}\Delta^x g(\mathbf{0}) \tag{2}$$

where Δ is the forward difference operator: $\Delta f(x) = f(x+1)-f(x)$.

Berg (1974) introduced the family of FSDs and applications of this concept in urn sampling and capture-recapture sampling are described by Johnson and Kotz (1977) and

Seber (1981), respectively.

In statistical applications of FSDs, the parameter N appearing in (1) typically stands for the size of a finite population, while the r. v. X describes the number of distinct units from the population observed during a sampling experiment. For instance, X may describe the total number of different animals observed during a mark-capture experiment, N being the size of the population.

Moment properties of FSDs are simple. Thus if the r. v. X has an FSD, then the binomial moments of the complementary variable: $Z=N-X$ turn out to be very neat, namely

$$\beta_i = E\binom{Z}{i} = E\binom{N-X}{i} = \binom{N}{i} g(N-i)/g(N), \quad i=1,2,... \tag{3}$$

In particular, the expected value of an FSD-variable is given by

$$E(X) = \mu_N = N(1-g(N-1)/g(N)). \tag{4}$$

For use in what follows we will need another moment property. Let us denote by $\mu_{k,N}$ the common k:th moment, $k=\pm1,\pm2,...$ When reciprocal moments are involved it is tacitly understood that zero is excluded from the range of the r.v. However, this is no serious limitation, since we note that a zero-truncated FSD is an FSD in its own right. The following recursion relation is useful.

Property. Let the discrete r.v. X have an FSD with s. f. $g(N)$. Then the following holds for moments of X

$$\nabla(g(N)\mu_{k,N}) = \frac{g(N)}{N}\mu_{k+1,N}, \quad k=\pm1,\pm2,... \tag{5}$$

where ∇ is the backward difference operator: $\nabla f(x) = f(x)-f(x-1)$. In particular, if $k=-1$, then (5) reduces to an equation which can be solved to give

$$\mu_{-1,N} = E(1/X) = \sum_{i=1}^{N-1} g(N-i)/g(N)(N-i). \tag{6}$$

Proof: Formula (4) is readily obtained if we observe that $\nabla\binom{N}{x} = \binom{N}{x}\frac{x}{N}$. Whence

$$\nabla(g(N)E(X^k)) = \Sigma\binom{N}{x}\frac{x}{N}x^k a_x = \frac{g(N)}{N}E(X^{k+1}).$$

To prove the special case (6), we expand formally:

$$\nabla^{-1} = (1-E^{-1})^{-1} = 1+E^{-1}+E^{-2}+ ...$$

where E^{-1} is the unit shift operator: $E^{-1}f(x) = f(x-1)$. If we let the sum of operators operate on $g(N)/N$, we get the expression (6).

Remark: In survey sampling literature one can find reciprocal moments for specific discrete distributions, such as the binomial or hypergeometric (Cf., e.g., Govindarajulu (1964)). Using a more general approach, Pathak (1961) derived an expression for the reciprocal moment in terms of exclusion probabilities. Now in sampling applications, the ratios $g(N-i)/g(N)$, appearing in (6), are just that: the probability of having i specific units excluded from the sample. The more general result (5) was first shown in an unpublished technical report by the present author (Berg (1975)).

3. Applications to survey sampling

Let $U = (u_1,...,u_N)$ be a finite population with a variate Y defined on U such that $Y(u_i) = y_i$, $i=1, ...,N$. We are interested in inference about the population mean: $\bar{Y} = \frac{1}{N}\Sigma y_i$. For simple random sampling without replacement we will use the conventional abbreviation SRSWOR.

We are concerned here with a class of sampling designs with the following properties:

1) The design is such that the number of distinct units, or the effective sample size, is an r. v. having an FSD (1).

2) Conditional on a given number of distinct units, ν say, the sample of distinct units is equivalent to an SRSWOR of size ν from U.

We will use the symbol D_g to denote such a design and D_Γ for the class, where $g(N)$ is the FSD which determines the effective sample size. Typically, but not necessarily, the design involves replacement sampling and repetitions of the same unit. However, the actual element-by-element sampling is of no concern here. To be absolutely correct, the s. f. characterizing a design should be written $g(N,\Theta)$, where Θ is a known vector of design parameters, e. g., a nominal sample size, an inclusion probability, etc.

The estimate of the population mean $\bar{Y}$, based on the distinct units will be written: $\bar{y}_\nu = \frac{1}{\nu}\Sigma\, y_i$, where the summation involves the distinct units of the sample only. The expected number of different units, or the expected effective sample size, is by formula (4): $E(\nu) = N(1-g(N-1)/g(N))$. The estimate $\bar{y}_\nu$ is clearly an unbiased estimate of the population mean for all designs in the class D_Γ.[1] We may now use formula (6) in the previous section to obtain an expression for the variance of this estimate. First, the conditional variance is

$$V_g(\bar{y}_\nu/\nu) = \frac{N-\nu}{N}\frac{S^2}{\nu}$$

where $S^2 = \frac{1}{N-1}\Sigma\,(y_i - \bar{Y})^2$. By formula (6), the overall variance is

$$V_g(\bar{y}_\nu) = \frac{S^2}{g(N)}\sum_{i=1}^{N-1} g(N-i)/(N-i). \qquad (7)$$

We note that the summation in (7) starts with $i=1$.

[1] If we divide the sample sum by $E(\nu)$, we get another unbiased estimate of Y, an estimate of Horvitz-Thompson-type. This estimate, however, is usually associated with unequal probability sampling, and will not be discussed here.

Efficiency criterion. Given two members of the class D_Γ, with series functions $g(N)$ and $h(N)$, respectively, and the same effective expected sample size, which is equivalent to $g(N-1)/g(N) = h(N-1)/h(N)$, then the difference in variance is given by the sum

$$V_g(\bar{y}_\nu) - V_h(\bar{y}_\nu) =$$

$$S^2\sum_{i=2}^{N-1}\left(\frac{g(N-i)}{g(N)} - \frac{h(N-i)}{h(N)}\right)\frac{1}{N-i}. \qquad (8)$$

Note that here the summation starts at $i=2$.

In actual practice, however, it often turns out to be extremely difficult to even determine the sign of the difference (8), let alone its magnitude. We will now look in more detail at a few special cases of designs belonging to the class D_Γ.

Example 1. Simple random sampling without replacement (SRSWOR).

In our first example, we will show that SRSWOR is indeed a member of our class, albeit a special one. In an SRSWOR of size n from U, the total number of samples is $\binom{N}{n}$. Formally, the effective sample size ν may now be taken as the degenerate r. v. having as s. f. $g(N) = \binom{N}{n}$. Then, of course, there is only one non-zero term in the factorial series expansion of $g(N)$, and

$$p(\nu) = \begin{cases} 1 \text{ if } \nu = n \\ 0 \text{ otherwise.} \end{cases}$$

Turning now to the variance formula (7) we first have

$$V_D(\bar{y}_n) = S^2\sum_{i=1}^{N-1}\left(\frac{N-i}{n}\right)/(N-i)\binom{N}{n}.$$

Then using a summation formula in Feller (p. 62, (1961)), we get

$$\Sigma \binom{N-i-1}{n-1}/n\binom{N}{n} = \sum_{0}^{N-1} \binom{j+n-1}{n-1}/n\binom{N}{n} =$$

$$\binom{N-1}{n}/n\binom{N}{n} = \frac{N-n}{nN}.$$

Hence for SRSWOR the general variance formula (7) does reduce to the well known expression:

$$V(\bar{y}_n) = \frac{N-n}{N}\,\frac{S^2}{n}.$$

The SRSWOR design is in a certain sense optimal in the class we consider: For a (non-degenerate) r. v. positive X it holds that $E(1/X) > 1/E(X)$, so if for an arbitrary member $D_g \in D_\Gamma$, $E(\nu)$ is an integer $= n$, say, then we know that $V_g > (\frac{1}{n}-\frac{1}{N})\,S^2$, equality only if $g(N) = \binom{N}{n}$.

Example 2. Bernoulli sampling.

For each of the N units of U, a Bernoulli-trial is used to decide if the unit is included or not in the sample. The N trials are independent and the probability for exclusion, say q $(0<q<1)$, is the same throughout. In this case, the discrete distribution governing the effective sample size is binomial, $b(N, 1-q)$, or, if the outcome $\nu=0$ is excluded, the corresponding zero-truncated binomial. The design belongs to our class with the s. f. taken to be $g(N) = = 1/q^N - 1$.

Let us now compare SRSWOR with Bernoulli sampling, supposing for simplicity that we can set $g(N) = 1/q^N$ (i. e. we disregard the zero-truncation) and further that $E(\nu) = = N(1-q)$ is an integer. We then have the two variance expressions

$$V_1 = S^2 \Sigma\, q^i/(N-i) \text{ and}$$

$$V_2 = S^2\, q/N(1-q) = S^2 \Sigma\, q^i/N,$$

respectively. Thus the difference in variance is seen to be of the order $0(1/N^2)$.

Example 3. Simple random sampling with replacement (SRSWR).

If n units are selected with replacement and with equal probabilities from U, then the total number of different samples is N^n. This design, commonly abbreviated SRSWR, is the member of our class with s. f. $g(N)=N^n$. The number of distinct units – the effective sample size – has a well known p. f.

$$p(\nu) = \binom{N}{\nu}\Delta^\nu 0^n/N^n,\ \nu=1,\ldots,\min(n,N). \tag{9}$$

The discrete probability distribution (9) is called the classical occupancy distribution by Johnson and Kotz (1977) in their book on urn models. Sometimes one also sees the name *Arfwedson's* distribution. Here formula (7) takes the simple form

$$V_g(\bar{y}_\nu) = S^2 \sum_{i=1}^{N-1} (N-i)^{n-1}/N^n. \tag{10}$$

For the design SRSWR another unbiased estimate of $\bar{Y}$ is the average of *all* units in the sample, whether or not repetitions occur. If $\bar{y}_n$ denotes this estimate, then $V_g(\bar{y}_n) = S^2/n$. It is well known that the latter estimate of the population mean is not admissible. This may be verified, for example, by a direct proof that $V_g(\bar{y}_\nu) < V_g(\bar{y}_n)$ for all $n\geq 2$ and all parameter points. However, appealing to the Rao-Blackwell theorem, as did Basu (1958), is a much easier way to derive this result.

Example 4. Interpenetrating subsamples.

In this design k (≥ 2) independent SRSWOR's, say of equal size n, are taken from the population U. This design was discussed in some detail by Pathak (1961) and recently by, e. g. Sengupta (1980). Pathak showed that $\bar{y}_\nu$ – the estimate based on the distinct units – is an admissible estimate of the population mean $\bar{Y}$. Pathak derived the estimate $\bar{y}_\nu$ by means of

"Rao-Blackwellization" of the mean of all nk units, $\bar{y}_{nk}$. The design is seen to be a member of our class with the s. f. $g(N) = \binom{N}{n}^k$.

The overall variances of the two estimates mentioned are

$$V_g(\bar{y}_v) = S^2 \sum_{i=1} \binom{N-i}{n}^k / (N-i)\binom{N}{n}^k$$

and

$$V_g(\bar{y}_{nk}) = \frac{N-n}{N} \frac{S^2}{nk}$$

respectively. Sengupta (1980) shows by direct comparison that the interpenetrating subsamples design is more efficient than the SRSWR design, subject to the same expected sample size. The latter is given by formula (4) as

$$E(v) = N(1-\binom{N-1}{n}^k / \binom{N}{n}^k) =$$

$$= N(1-(1-n/N)^k).$$

Consider now a subclass of designs $D_g \subset D$ such that

$$D_G = \{ D_g : g(N) = \prod_{i=1}^{k} \binom{N}{n_i}, \Sigma n_i = nk \},$$

where n and k are assumed to be given integers, and $n_i \geq 1$, $k \geq 2$. Thus our class of designs consists of all interpenetrating subsampling designs with k subsamples and such that the total sample size is a constant $= nk$. We first note that

$$\max_{\Sigma n_i = nk} \prod_{i=1}^{k} \binom{N-j}{n_i} / (N-j)\binom{N}{n_i} =$$

$$= \binom{N-j}{n}^k / (N-j)\binom{N}{n}^k \text{ for } j=1,\ldots,N-1.$$

Next, if D_g is an arbitrary member of D_G, then by our variance formula (7) it must hold that

$$V_g(\bar{y}_v) \leq S^2 \sum_{j=1}^{N-1} \binom{N-j}{n}^k / (N-j)\binom{N}{n}^k \qquad (11)$$

with equality in (11) if and only if $g(N) = \binom{N}{n}^k$. A result to this effect is mentioned by Sengupta (1980). The inequality (11) is intuitively reasonable, since a design with subsamples of equal size is apt to lead to more repetitions and thus less precise estimation than a design with an unequal allocation of the total sample size.

Example 5. Post cluster sampling.

This type of design was first studied by Dalenius (1957) in the form of two-phase sampling for two-stage sampling. Ghosh (1963), Berg (1972) and, more recently, Sadasivan and Srinath (1975) have discussed estimation and efficiency problems.

The situation may be briefly described as follows. In a finite population $U = (u_1,\ldots, u_i,\ldots,u_N)$, each u_i is a primary unit or cluster, comprising M_i secondary units, or elements. There is no list of primary units, so a direct cluster-, or two-stage sample is not possible. However, there is available a list of all the elements, and an SRSWOR of elements is possible with subsequent identification of clusters or primary units. We may then proceed to cluster, or two-stage sampling, using the first sample of elements as a provisional sampling frame.

Let us for simplicity assume that the clusters are of equal size $= M$, and that for each cluster identified in the sample it is possible to ascertain the cluster mean. We can then estimate the population mean (the mean per element) using the v distinct clusters observed in the sample. Now this latter quantity will be an r. v. having an FSD with series function $g(N) = \binom{MN}{n}$, so this design is also a member of our class. Let us denote by $\bar{\bar{y}}_v$ the mean of the clusters included in the sample of elements. By formula (7), the variance of this unbiased estimate of the population mean per element is given by

$$V_g(\bar{\bar{y}}_v) = \bar{S}_b^2 \Sigma \binom{(N-i)M}{n} / \binom{(N-i)}{n} \binom{NM}{n} \quad (12)$$

where $\bar{S}_b^2 = \frac{1}{N-1} \Sigma (\bar{Y}_i - \bar{\bar{Y}})^2$, the between cluster variance.

Example 6. Overlapping sampling frames.

As our final example we will take the problem of sampling from overlapping sampling frames. A standard reference here is Hartley (1974). We consider first the situation in which some units have multiple listings. Suppose, for simplicity, that no unit is listed more than twice; for example, our list may be the result of merging two partly overlapping sampling frames. We will use the notation:

N_s = the number of units listed once,
N_d = the number of units listed twice,
$N = N_s + N_d$ = the population size,
$N' = N_s + 2N_d$ = the total number of units on our list.

Suppose further that it is required to estimate the total of the population

$$T = \sum^N y_i = \sum^s y_i + \sum^d y_j .$$

The design is an SRSWOR of size n from the list containing N' units. It is convenient to introduce the following auxiliary variate, defined for all units listed as

$$\acute{y}_i = \begin{cases} y_i \text{ if } u_i \text{ is listed once} \\ \frac{1}{2} y_i \text{ if } u_i \text{ is listed twice.} \end{cases}$$

Clearly, the total of the $\acute{y}$-values, summed over all N' listed units, is the same as that of the original y-values:

$$T' = \sum^{N'} \acute{y}_i = \sum^s y_i + \frac{1}{2} \sum^d 2y_j = T.$$

This leads us to the following simple estimate of T, based on all n units of the sample:

$$\hat{T}_n = \frac{N'}{n} \sum_1^n \acute{y}_i . \quad (13)$$

It is important to note that the estimate (13) presupposes that it is possible to ascertain whether a particular sample unit is listed twice or not.

The variance of the estimate $\hat{T}_n$, defined by (13), can be written:

$$V(\hat{T}_n) = N'(N'-n)\, \acute{S}^2/n \quad (14)$$

where $\acute{S}^2$ is the population variance of the auxiliary variate:

$$\acute{S}^2 = \Sigma (\acute{y}_i - \bar{Y}')^2/(N'-1), \; Y' = \sum^{N'} \acute{y}_i / N' .$$

Now the estimate $\hat{T}_n$ is based on all the n units of the SRSWOR sample, including repetitions of the same unit, and therefore the estimate is not admissible. To remedy the situation, we will need some further notation. Let the number of different units in the sample be: $n_1 + \nu$, where n_1 is the number of units obtained from those listed only once, while ν is the number of different units from the set of units listed twice. We can now give an improved estimate of the population total as

$$\hat{T}_\nu = \frac{N'}{n} \left(n_1 \bar{y}_{n_1} + \frac{1}{2}(n-n_1) \bar{y}_\nu \right) \quad (15)$$

where $\bar{y}_\nu$ is the mean computed on the distinct units, $=0$ if $\nu=0$.

We note the conditional expected value: $E(\hat{T}_n / n_1, \nu) = \hat{T}_\nu$. Hence

$$V(\hat{T}_n) = V(\hat{T}_\nu) + EV(\hat{T}_n / n_1, \nu)$$

and also

$$V(\hat{T}_\nu) \leq N'(N'-n)\, \acute{S}^2 / n. \quad (16)$$

An exact expression for the variance of $\hat{T}_\nu$ may be computed using the results developed here. However, in order not to burden the text

with complicated formulas, we confine ourselves to the following few remarks. For the SRSWOR design we are discussing, the introduced r. v. ν has an FSD with s. f. $g(N_d) = \binom{N'}{n} = \binom{N+2N_d}{n}$. The expected total effective sample size is

$$E(n_1+\nu) = nN_s/N' + N_d(1-\binom{N'-2}{n}/\binom{N'}{n})$$

$$= nN_s/N' + N_d(1-(N'-n)(N'-n-1)/N'(N'-1)). \quad (17)$$

Finally, we address the problem of two independent samples from two overlapping frames, a case considered in some detail by Fuller and Burmeister (1972).

Let us denote by N_1 and N_2 the number of elements in lists 1 and 2, respectively. Let us further write $N_1 = N_{11} + N_{12}$ and $N_2 = N_{22} + N_{12}$, where N_{12} is the size of the overlap domain. Finally, the sample sizes are n_1 and n_2. As before, we can introduce an auxiliary variate and use a "domain estimation" technique to estimate the population total. However, let us instead assume that it is required to estimate the mean $\bar{Y}_{11}$ of the overlap domain. Let $n_{1(12)}$ and $n_{2(12)}$ denote the number of sampled units falling in the overlap domain, with $n_{(12)} = n_{1(12)} + n_{2(12)}$, while the number of distinct units is $\nu \leq n_{(12)}$.

Now we have again two sample averages as possible estimates of the population mean:
- the average of all $n_{(12)}$ units, $\bar{y}_{(12)}$, and
- the average of the distinct units, $\bar{y}_\nu$.

The former estimate, involving duplicated elements, is not admissible. In fact, by symmetry, we will have conditionally:
$E(\bar{y}_{(12)}/\nu \text{ distinct units}) = \bar{y}$.

So the mean of the distinct units is an improved estimate of the overlap domain mean. Moreover, the effective sample size ν is an r.v. having an FSD with series function

$$g(N_{12}) =$$

$$\binom{N_1}{n_1}\binom{N_2}{n_2} = \binom{N_{11}+N_{12}}{n_1}\binom{N_{22}+N_{12}}{n_2}. \quad (18)$$

Consequently, what we are discussing here may be regarded as a special case of the class of designs D_Γ.

4. References

Basu, D. (1958): On Sampling with and without Replacement. Sankhyā 20, pp. 3–16.

Berg, S. (1972): Double Sampling for Cluster- or Two-Stage Sampling. International Statistical Review, Vol. 40, pp. 13–22.

Berg, S. (1974): Factorial Series Distributions with Applications to Capture-Recapture Problems. Scandinavian Journal of Statistics, Vol. 1, pp. 145–152.

Berg, S. (1975): A Note on Reciprocal Moments of Discrete Random Variables. Technical Report 1975:8, Department of Statistics, University of Lund, Sweden.

Dalenius, T. (1957): Sampling in Sweden. Almqvist & Wiksell, Stockholm.

Feller, W.G. (1961): An Introduction to Probability Theory and Its Applications, (Second edition). John Wiley & Sons, New York.

Fuller, W.A. and Burmeister, L.F. (1972): Estimators for Samples Selected for Two Overlapping Frames. American Statistical Association, Proceedings of the Social Statistics Section, pp. 203–206.

Ghosh, S.P. (1963): Post-Cluster Sampling. Annals of Mathematical Statistics, 34, pp. 587–597.

Govindarajulu, Z. (1964): The First Two Moments of the Reciprocal of the Positive Hypergeometric Variable. Sankhyā, Ser. B, 26, pp. 217–236.

Hartley, H.O. (1974): Multiple Frame Methodology and Selected Applications. Sankhyā, Ser. C, 36, pp. 99–118.

Johnson, N.L. and Kotz, S. (1977): Urn Models and Their Application. John Wiley & Sons, New York.

Pathak, P.K. (1961): On the Evaluation of Moments of Distinct Units in a Sample. Sankhyā, Ser. A, 23, pp. 415–420.

Pathak, P.K. (1964): On the Efficiency of the Interpenetrating Subsampling Procedure. Annals of Mathematical Statistics, 35, pp. 795–808.

Sadasivan, S. and Srinath, M. (1975): Some Contributions to Post-Cluster Sampling. Sankhyā, Ser. C, pp. 171–180.

Seber, G.A.F. (1981): The Estimation of Animal Abundance and Related Parameters, (Second edition). Griffin, London.

Sengupta, S. (1980): On the Efficiency of the Interpenetrating Subsampling Procedure. Bulletin of the Calcutta Statistical Association, Vol. 29, pp. 173–177.

Two Sequential Procedures for Obtaining Subdomain Estimates

by Myron J. Katzoff [1]

1. Introduction

This paper presents sample survey theory for two sequential sample-selection procedures or designs which can be employed in the estimation of a finite population subdomain total. Both designs involve two stages of selection and they have the common feature that, at each selection of a primary sampling unit (PSU), the second stage units (SSU) are randomly selected without replacement and one at a time until a predetermined number (or quota) of SSUs belonging to the subdomain of interest have been chosen. In the order indicated, the following are developed for each design: (1) an unbiased estimator of the subdomain total; (2) an expression for the variance of the estimator; and (3) an unbiased estimator of that variance. The results recorded here will eventually be included among those from a larger effort to investigate methods for the sequential estimation of finite population parameters.

In the first procedure, treated in the next section, PSUs are selected without replacement and with known equal or unequal probabilities. It is the author's conjecture that this procedure could be applied, for example, in a special purpose household survey designed to concentrate on the characteristics of a small demographic group. It might be reasonable to plan the design of such a survey to utilize a well-defined PSU structure developed for some routine area or address frame survey for which first-stage inclusion probabilities are known.

At the first stage of selection for the second procedure, examined in the third section of this paper, PSUs are selected with replacement with unknown probabilities proportional to the number of subdomain elements among the SSUs contained in each PSU. (When a PSU is chosen more than once, it is assumed that SSUs are selected independently in each instance.) To be precise, a PSU is chosen with probability $1/M$ (where M is the total number of PSUs in the population) and, then, an SSU contained in that PSU is randomly selected. If the SSU chosen belongs to the subdomain of interest, the characteristic under study is quantified for that SSU and sequential selection of SSUs from the PSU continues until the quota is satisfied, after which the PSU is "replaced" for possible further sampling. If the first randomly chosen SSU of a selected PSU does not belong to the subdomain of interest, the PSU is replaced and another draw of a PSU with probability $1/M$ is undertaken, etc. This process continues until a fixed sample size requirement for PSUs is met.

The reader should note that if the second procedure is applied to a situation where samples are drawn from a frame of commercial and noncommercial telephone numbers and the subdomain of interest is residential telephone numbers, then the design appears to be very much like the one examined by Waksberg (1978). The results of the third section on an unbiased estimator and its variance differ slightly from those of Waksberg because

[1] Assistant Chief for Mathematical Statistics and Sampling Techniques, Statistical Research Division, U.S. Bureau of the Census, Washington, D.C.

SSUs are selected without replacement. The derivation of an unbiased estimator of variance seems to be new.

We now assume that the quota for each sample PSU can be satisfied and introduce the following notation:

- y_{ij} the value of the characteristic under study for the j-th subdomain element chosen from the i-th sample PSU
- k_i the quota to be satisfied in the i-th PSU; this is a constant in the second procedure
- n_i a random variable which is the number of SSUs drawn from the i-th PSU to satisfy the quota, k_i, for that PSU
- m the sample size for the first stage of selection
- N_i the number of SSUs in the i-th PSU; this is a constant in the second procedure
- D_i the number of subdomain elements for the i-th PSU.

2. PSUs selected with known probabilities without replacement

2.1. An estimator of the subdomain total and its variance

If the sampling is carried out in the manner described above, the size of the sample drawn from the i-th PSU, n_i, will be a random variable which follows the negative hypergeometric distribution with probability mass function

$$p(n_i\,;k_i\,,N_i\,,D_i) =$$

$$= \begin{cases} \dbinom{n_i-1}{k_i-1}\dbinom{N_i-n_i}{D_i-k_i}\Big/\dbinom{N_i}{D_i} & \text{for} \\ \quad n_i = k_i\,,\ k_i+1,\ldots,N_i-D_i+k_i \\ 0, \text{ otherwise.} \end{cases}$$

In Katzoff (1979), it was readily established that

$$E((k_i-1)/(n_i-1) \mid \text{PSU}_i) = D_i/N_i$$

so that an unbiased estimator of the subdomain total, Y_D, is

$$\hat{Y}_D = \sum_{i=1}^{m} \pi_i^{-1}\,\hat{Y}_i$$

where π_i is the inclusion probability for the i-th PSU and

$$\hat{Y}_i = [N_i(k_i-1)k_i^{-1}(n_i-1)^{-1}] \sum_{j=1}^{k_i} y_{ij}.$$

Putting $Y_i = E[\hat{Y}_i \mid \text{PSU}_i]$ and $\sigma_{2i}^2 = \text{Var}\,(\hat{Y}_i \mid \text{PSU}_i)$, it follows from theorem 11.1 in Cochran (1977) that

$$\text{Var}\,(\hat{Y}_D) =$$

$$\sum_{i=1}^{M}\left\{[(1-\pi_i)/\pi_i]\,Y_i^2 + 2\sum_{j>i}^{M} Y_i\,Y_j\,[(\pi_{ij}-\pi_i\pi_j)/\pi_i\pi_j]\right\} + \sum_{i=1}^{M} \pi_i^{-1}\sigma_{2i}^2$$

where π_{ij} is the joint inclusion probability for the i-th and j-th PSUs and

$$\sigma_{2i}^2 = x_i^2\,\beta_i^2 + (D_i/N_i)^2\,\beta_i^2 + (N_i\bar{Y}_i)^2\,x_i^2$$

$$x_i^2 = E[[(k_i-1)/(n_i-1)]^2|\text{PSU}_i]-(D_i/N_i)^2$$

$$\hat{Y}_i = D_i^{-1}\sum_{j=1}^{D_i} y_{ij}$$

and

$$\beta_i^2 = N_i^2\,(k_i^{-1}-D_i^{-1})\left(\sum_{j=1}^{D_i} y_{ij}^2 - D_i\bar{Y}_i^2\right)/(D_i-1).$$

To obtain the expression for σ_{2i}^2, notice that, for each PSU in sample,

$$\sum_{j=1}^{k_i} y_{ij} = \sum_{j=1}^{D_i} y_{ij}\,t_{ij},$$

where

$$t_{ij} = \begin{cases} 1, \text{ if the } j\text{-th subdomain element of the PSU is in sample} \\ 0, \text{ otherwise.} \end{cases}$$

Now, the probability that the j-th element of the i-th sample PSU is in sample and that the number of second stage units drawn from that PSU is n_i, is

$$k_i\binom{n_i-1}{k_i-1}(D_i-1)_{k_i-1}(N_i-D_i)_{n_i-k_i}/(N_i)_{n_i}$$
$$= (k_i/D_i)\,p(n_i;k_i,N_i,D_i),$$

where $(H)_t = H!/(H-t)!$ for integers $H>t\geq 0$. Thus, given i, the t_{ij} $(j=1,2,\ldots,D_i)$ are independent of n_i. It is from this independence that the form of σ_{2i}^2 follows.

2.2. An estimator of the variance of $\hat{Y}_D$

According to theorem 11.2 of Cochran (1977), an unbiased estimator of Var $(\hat{Y}_D)$ is

$$\hat{\text{Var}}(\hat{Y}_D) = \sum_{i=1}^{m}\sum_{j>i}^{m}\frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}}\left(\frac{\hat{Y}_i}{\pi_i}-\frac{\hat{Y}_j}{\pi_j}\right)^2$$
$$+\sum_{i=1}^{m}\pi_i^{-1}\hat{\sigma}_{2i}^2$$

where $\hat{\sigma}_{2i}^2$ is some unbiased estimator for σ_{2i}^2. In this subsection, we obtain the form of $\hat{\sigma}_{2i}^2$. In what follows, take the conditioning on the selection of PSU i to be understood.

Clearly, an unbiased estimator for σ_{2i}^2 is

$$\hat{\sigma}_{2i}^2 = \hat{Y}_i^2 - [\text{unbiased estimator for } Y_i^2].$$

Thus, we can achieve our objective by getting unbiased estimators for each of the terms of

$$Y_i^2 = \left(\sum_{j=1}^{D_i} y_{ij}\right)^2 = \sum_{j=1}^{D_i} y_{ij}^2 + \sum_{\substack{j,k=1\\ j\neq k}}^{D_i} y_{ij}\,y_{ik}.$$

An unbiased estimator of $\sum_{j=1}^{D_i} y_{ij}^2$ is

$$\frac{N_i(k_i-1)}{k_i(n_i-1)}\sum_{j=1}^{k_i} y_{ij}^2.$$

To get an unbiased estimator for the cross products, note that the joint (conditional) probability that the j-th and k-th subdomain elements of the i-th PSU are in the sample and that sampling ceases when n_i elements have been drawn is

$$k_i(k_i-1)\binom{n_i-1}{k_i-1}(D_i-2)_{k_i-2}$$
$$\times(N_i-D_i)_{n_i-k_i}/(N_i)_{n_i}$$
$$= \frac{k_i(k_i-1)}{D_i(D_i-1)}\,p(n_i;k_i,N_i,D_i),$$

which implies that

$$E_{j\neq k}[t_{ij}\,t_{ik} \mid n_i \text{ and PSU}_i] = \frac{k_i(k_i-1)}{D_i(D_i-1)}.$$

Since

$$E\left[\frac{(k_i-1)(k_i-2)}{(n_i-1)(n_i-2)} \mid \text{PSU}_i \text{ chosen for subsampling}\right] = \frac{D_i(D_i-1)}{N_i(N_i-1)},$$

an unbiased estimator of the second term in Y_i^2 is

$$\frac{N_i(N_i-1)(k_i-2)}{k_i(n_i-1)(n_i-2)}\sum_{j\neq k}^{k_i} y_{ij}\,y_{ik}.$$

But

$$\sum_{j\neq k}^{k_i} y_{ij}\,y_{ik} = \left(\sum_{j=1}^{k_i} y_{ij}\right)^2 - \sum_{j=1}^{k_i} y_{ij}^2$$

so that if we define

$$A(N_i,k_i,n_i) = \frac{N_i(k_i-1)}{k_i(n_i-1)},$$

then

$$\hat{\sigma}_{2i}^2 = A(N_i,k_i,n_i)\,[A(N_i,k_i,n_i)$$
$$-A(N_i-1,k_i-1,n_i-1)]\left[\sum_{j=1}^{k_i} y_{ij}\right]^2$$
$$-A(N_i,k_i,n_i)\,[1-A(N_i-1,k_i-1,n_i-1)]\sum_{j=1}^{k_i} y_{ij}^2.$$

3. PSUs selected with unknown probabilities with replacement

3.1. Estimator of the subdomain total and its variance

Following Waksberg, it is assumed that the proportion of SSUs in the population that belong to the subdomain of interest (or, equivalently, $D_0 = \sum_{i=1}^{M} D_i$) is known; however, the *individual* D_i are assumed to be *unknown*. The sampling of the PSUs with replacement is carried out with the probabilities

$$z_i = D_i/D_0 \,(i=1,2,\ldots,M'),$$

where $0<M'\leqslant M$ is the number of PSUs with at least one subdomain element. The second stage sample size for the i-th sample PSU now follows a negative hypergeometric distribution with the mass function

$$p^*(n_i;k,N,D_i) = \begin{cases} \binom{n_i-2}{k-2}\binom{N-n_i}{D_i-k}\Big/\binom{N-1}{D_i-1} & \text{for} \\ n_i=k,k+1,\ldots,N-D_i+k \\ 0, \text{ otherwise.} \end{cases}$$

The unbiased estimator whose properties will be examined is

$$\hat{Y}_D{}^* = \sum_{i=1}^{M} \frac{D_0}{km} \sum_{j=1}^{k} y_{ij}\,.$$

(To get the desired results, the material in Chapter 11 of Cochran (1977) is used.) By an argument similar to that used earlier, it can be seen that $E(t_{ij} \mid \text{PSU selection})$ $= (k/D_i)\, p^*(n_i;k,N,D_i)$; and, if x_i denotes the number of times the i-th PSU enters the sample, then $E(\hat{Y}_D{}^*) = E[E(\hat{Y}_D{}^* \mid \text{PSU selection})]$

$$= E\left[\sum_{i=1}^{M} \frac{x_i D_0}{m}\frac{Y_i}{D_i}\right] = E\left[\sum_{i=1}^{M'} \frac{x_i}{mz_i} Y_i\right]$$

$$= \sum_{i=1}^{M'} Y_i = \sum_{i=1}^{M} Y_i = Y_D\,.$$

Thus, $\hat{Y}_D{}^*$ is an unbiased estimator of Y_D.

The variance of $\hat{Y}_D{}^*$ is

$$\text{Var}\,(\hat{Y}_D{}^*) = m^{-1} \sum_{i=1}^{M'} z_i[(Y_i/z_i) - Y_D]^2 + (D_0^2/m) \sum_{i=1}^{M'} z_i(k^{-1} - D_i^{-1})S_{2i}^2$$

where

$$S_{2i}^2 = (D_i-1)^{-1} \sum_{j=1}^{D_i} [y_{ij} - (Y_i/D_i)]^2\,.$$

This is obtained with the help of the results in Cochran (1977) by putting

$$\text{(i) } \hat{Y}_D{}^* = \sum_{i=1}^{m} y_i', \text{ where } y_i' = \frac{D_0}{km} \sum_{j=1}^{k} y_{ij}$$

and

$$\text{(ii) } \hat{Y}' = \sum_{i=1}^{m} Y_i', \text{ where}$$

$$Y_i' = E\,(y_i' \mid \text{PSU selection})$$

$$= Y_i/mz_i$$

so that we get for the variance of $\hat{Y}'$

$$V(\hat{Y}') = 1/m \sum_{i=1}^{M'} z_i\,[(Y_i/z_i) - Y_D]^2$$

and

$$\sigma_{2i}^2 = E[(y_i' - Y_i')^2 \mid \text{PSU selection}]$$

$$= (D_0/m)^2\,(k^{-1} - D_i{}^{-1})\,S_{2i}^2\,.$$

3.2. An estimator of the variance of $\hat{Y}_D{}^*$

By theorem 11.2 in Cochran (1977) it follows that an unbiased estimator for the variance of $\hat{Y}_D{}^*$ is

$$v\,(\hat{Y}_D{}^*) = [m(m-1)]^{-1} \sum_{i=1}^{m} (my_i' - \hat{Y}_D{}^*)^2 + v_2$$

where

$$v_2 = \frac{D_0}{m} \sum_{i=1}^{m} \left\{ \frac{(N-1)(k-2)}{k(N_i-2)} - (1-k^{-1}) \right\} s_{2i}^2,$$

$$s_{2i}^2 = (k-1)^{-1} \sum_{j=1}^{k} (y_{ij}-\bar{y}_i)^2,$$

and

$$\bar{y}_i = m y_i' / D_0 .$$

This follows from:

(1) the observation that the second term in Var $(\hat{Y}_D{}^*)$ can be written as

$$(D_0/m) \sum_{i=1}^{M'} [(D_i/k) - 1] S_{2i}^2;$$

(2) the independence of s_{2i}^2 and n_i; and

(3) the relationship

$$E((k_i-2)/(n_i-2) \mid i\text{-th PSU selected}) = (D_i-1)/(N-1).$$

4. References

Cochran, W.G. (1977): Sampling Techniques, Third Edition. John Wiley.

Katzoff, M.J. (1979): Probability Quota Sampling of a Dichotomous Variable. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 407–410.

Waksberg, J. (1978): Sampling Methods for Random Digit Dialing. Journal of the American Statistical Association, pp. 40–46, Section 4.

Statistical Matching – Statistical Prediction What is the Difference?

An evaluation of statistical matching and a special type of prediction using data from a survey on living conditions

by Claes M. Cassel[1]

In ancient Rome a story was told about professor Gigo (incidentally, he discovered the famous Gigo-principle rediscovered in modern times by professor Dalenius (1964)). One day he wanted to smoke a cigar but found that his lighter did not work. Since he was a heavy smoker he felt the need of asking for fire so he approached the nearest person and said: "Excuse me sir, my lighter does not work. Do you have a statistical match?"

This real life story had the effect of inspiring me to do some research in order to find out more about the technique known as statistical matching. The following is an interim report of my findings. I want to thank Mr. G. Råbäck, Statistics Sweden, and Mr. C. E. Särndal, University of Montreal, Canada, for many helpful suggestions. G. Råbäck has also done the programming work.

0. Introduction

Statistical matching (SM) procedures aim to merge information from different sources into one single file containing the combined information of these sources. The problem of combining different bodies of data y and z from different samples s_1 and s_2, where y is measured exclusively in s_1 and z exclusively in s_2 is claimed to be solved, provided we use proper statistical matching techniques on a set of matching variables x measured in both s_1 and s_2. This magic method would, if it works, certainly open new fields of data for low cost analyses by researchers and statistical authorities.

One could search for some set of basic variables to be measured in both surveys in order to make it possible to connect it to a lot of other surveys. Such variables could include age, sex, income, marital status, and household size. Variables concerning certain living conditions could be included as well. The crucial question is: Does SM work? By this we mean that the SM technique produces data files where the combined information from the two separate files is of high quality. The quality is said to be high if the relations between y and z in the produced file bear a close resemblance to the true relations.

The purpose of this paper is twofold:

i) to try to answer the question: Does statistical matching work?

ii) to compare statistical matching with statistical prediction.

By statistical prediction we mean a special kind of prediction. We will use estimated probability distributions to predict the number of units falling in a cell. The predictor is simply the expected number of units. We will discuss this in some detail in Section 2.3.

1. Statistical matching

1.1. Description

To answer the question we simulate statistical

[1] Head of the Statistical Methods Section, Department for Individual and Household Statistics, Statistics Sweden.

matching on a file containing data from the Survey on Living Conditions carried out regularly by Statistics Sweden. The file consists of about 10 000 individuals. They are randomly divided in two files A and B, each consisting of about 5 000 individuals. The variables in the files are split into three categories y, x, and z variables, and we pretend that the z-information for file A and the y-information for file B have been lost. A file C is created containing y- and x-information from file A and z-information from file B matched on x. The new file C is then compared to file A now including the original z-information. The procedure is shown in Table 1 below.

In the sequel we will occasionally meet expressions such as "a match", "a statistical match", "a file matched on x", "a synthetic unit", and "a synthetic value". We will now briefly discuss these terms. For a more detailed discussion we refer to Radner et al. (1980). By a match we mean a linkage of records from at least two separate files. By a statistical match we mean a match where the records do not necessarily have the same identification. Instead records which are similar to each other are linked together. The similarity is measured by a distance function involving the x-variables. The file A is sometimes referred to as the basefile. The file C which is matched on x is created by the following procedure: A unit is selected from the basefile A. A statistical match between the record for that unit and a record from file B is performed. A new record is constructed containing the y- and x-information from the unit in the A-file and the z-information from the record in the B-file. The new record is said to belong to a synthetic unit. The z-values are called synthetic values. The file C is said to be matched on x if the x-variables are used in the matching procedure to obtain similarity between the records.

1.2. Relation to other techniques

Although they do not serve the same purpose, SM and certain other statistical techniques are closely related. These techniques are

i) divided schedule
ii) imputation for nonresponse.

It seems plausible that all three techniques could be brought together under the structural umbrella of matrix sampling (c. f. Gurney and Dalenius (1969), Dalenius (1982)).

The resemblance between SM and divided schedules is illustrated by Tables 2 and 3 below.

Table 2 shows one typical feature of SM, namely that the two samples usually have no units in common. Also there is no common part of s_1 and s_2 in which we have measured both y and z. Thus, our inference about the relations between y and z must come through the variables in common, that is through x.

Table 1. The general procedure

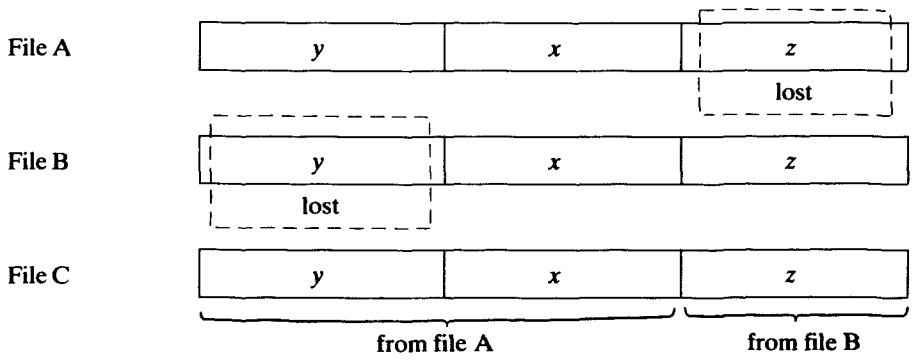


Table 2. SM

	Variables / Unit	$y_1 \ldots y_p$	$x_1 \ldots x_q$	$z_1 \ldots z_r$
s_1	k_1	√ √ √	√ √ √	
	.	√ √ √	√ √ √	
	.	√ √ √	√ √ √	
	.	√ √ √	√ √ √	
	k_{s_1}	√ √ √	√ √ √	
s_2	v_1		√ √ √	√ √ √
	.		√ √ √	√ √ √
	.		√ √ √	√ √ √
	v_{s_2}		√ √ √	√ √ √

Table 3 shows that, contrary to SM, divided schedule techniques allow some common set of units where the variables y and z (and possibly x) are measured jointly. The main difference, however, is that divided schedules are used on purpose to reduce the respondent burden, whereas SM is used to merge information from existing files. Thus, divided schedules aim at minimizing the respondent burden by deliberately choosing not to measure all variables on all units. In doing this we try to maximize the information exchange from the samples, Hocking (1979). If we consider it necessary we will put emphasis on the relation between y and z. By contrast, when using SM we try to make the best out of what we have in order to create artificial but hopefully realistic relations between y and z.

That SM is closely related to imputation for nonresponse is clear when we recognize that imputation means that a missing record or a missing value on a variable is replaced by

Table 3. Divided schedule technique

	Variables / Unit	$y_1 \ldots y_p$	$x_1 \ldots x_q$	$z_1 \ldots z_r$
s_1	k_1	√ √ √	√ √ √	
	.	√ √ √	√ √ √	
	.	√ √ √	√ √ √	
	.	√ √ √	√ √ √	
	.	√ √ √	√ √ √	
	v_1	√ √ √	√ √ √	√ √ √
	.	√ √ √	√ √ √	√ √ √
	.	√ √ √	√ √ √	√ √ √
	k_{s_1}	√ √ √	√ √ √	√ √ √
s_2	.		√ √ √	√ √ √
	.		√ √ √	√ √ √
	.		√ √ √	√ √ √
	v_{s_2}		√ √ √	√ √ √

some record or value created by the imputation technique. But when we in SM create synthetic values of the z:s for all the units on file A it will be sufficient to create synthetic values for the nonresponse in the case of imputation. There is also a major difference between SM and imputation. In SM we have no information on the joint distribution of y and z other than that through x. When we use imputation we do indeed have information on the joint distribution of variables through the respondents. This information is used to create synthetic values for the missing records or the missing values. The synthetic value is created in accordance to the available information for the respondents. For example, we may impute the mean value of z among the respondents in some class specified by x, or we may impute a randomly selected value of z among the respondents in the same class, or we may use some explanatory model to predict a value of z given the value of x.

2. Theoretical properties of SM

2.1. The purpose of SM

Let us now discuss the properties of a good statistical matching procedure. Some relevant references are Rogers and DeVol (1982), Hollenbeck and Doyle (1980), and Sims (1972). Rogers and DeVol (1982) give an extensive list of references concerning SM.

The primary purpose of SM is to create realistic information on the joint distribution of y and z. That is, we want to know one or more of the following:

i) the joint distribution $f(y,z)$, i.e. the frequency-function of (y,z),
ii) some covariance matrix C_{yz},
iii) some crosstabulation of y and z.

In practice we are especially interested in cases ii) and iii) since C_{yz} forms the basis of many statistical methods for multivariate analyses of linear models and since crosstabulation is the usual way of presenting statistical data.

2.2. Three possible matching procedures

Rogers and DeVol (1982) describe two kinds of matching procedures, unconstrained and constrained. Both types of matching procedures use a distance function, denoted D_{ij}, which measures the closeness of an individual i in file A and an individual j in file B. Let D_{ij} be defined by

$$D_{ij} = \sum_{r=1}^{q} w_r \cdot h_r(x_{ir}, x_{jr}),$$

where w_r is a weight reflecting the importance of the matching variable x_r, $r=1...q$. The function $h_r(x_{ir}, x_{jr})$ can be, for example

$$h_r(x_{ir}, x_{jr}) = (x_{ir} - x_{jr})^2.$$

Unconstrained matching chooses for each given i the individual j that produces the smallest D_{ij} irrespective of whether j has been used before or not. This would usually result in a close match in the sense that the distance is small but at the price of not being able to preserve the distribution of z, that is, it could happen that the distribution of z in the file B may be different from the distribution of z in the file C.

Constrained matching is free of this disadvantage. It preserves the z distributions. Let w_i be the weight (sample-weight or other weight) of individual i in file A and let w_j be the corresponding weight in file B. Let w_{ij} be weights assigned to a synthetic unit resulting from a potential match between i and j. A constrained matching procedure is defined as one that minimizes $\sum_i \sum_j w_{ij} D_{ij}$ subject to the constraints

$$\sum_i w_{ij} = w_j \text{ and}$$

$$\sum_j w_{ij} = w_i.$$

This type of matching procedure is likely to be computationally expensive.

Let us also mention the type of matching

used in imputation. The hot-deck methods cluster records in homogenous subgroups defined by some variables (matching variables) and impute a missing record from the records in the same subgroup for example by substituting a record drawn at random within that subgroup (Bailar and Bailar (1979)). Here, all the records in y and z are independent conditional on x.

2.3. The implicit assumption in SM. Prediction approach

Sims (1972), Hollenbeck and Doyle (1980), and Rogers and DeVol (1982), state that the implicit condition for SM to work is that y and z are independent conditional on x.

Let us quote Sims (1972) (in a discussion of a paper by Okner(1972)): "To get the joint distribution of x, y and z, we need an a priori assumption about the joint distribution of y, z conditional on x. The one Okner implicitly makes is that y, z are independent for given x." Sims then touches upon the possibility of estimating the joint distribution of x, y and z directly.

In fact, if y and z were independent conditional on x then

$$f((y,z)|x) = f(y|x)\, f(z|x).$$

But

$$f(x,y,z) = f((y,z)|x) f(x),$$

so

$$f(y,z) = \int_x f(x,y,z)dx =$$

$$\int_x f(y|x) f(z|x)\, f(x)dx.$$

(For simplicity of notation we use continuous variables terminology although the data sets are finite.)

By "prediction approach" we mean the following procedure: Let

$$a = \{(y,z) : ((y \leqslant y_0) \cap (z \leqslant z_0))\}.$$

Then

$$P(a) = \int_a f(y,z)dy\, dz =$$

$$\int_a \int_x f(y|x)\, f(z|x)\, f(x)\, dy\, dz\, dx.$$

Since the distributions $f(x)$ and $f(y|x)$ are known in the file A and since the distribution $f(z|x)$ is known in the file B it is possible to approximate the probability $P(a)$ by

$$\hat{P}(a) = \int_a \int_x f_A(y|x)\, f_B(z|x)\, f_A(x)dy\, dz\, dx,$$

where the subscripts A and B refer to the files A and B, respectively.

This approximation is then used to predict (or estimate) the number of units in the file A which have values (y,z) belonging to a by the predictor $n_A \cdot \hat{P}(a)$, where n_A is the number of units in A. The predictor is simply the expected value of a binomial variable with parameters n_A and $\hat{P}(a)$.

In Klevmarken (1982) another type of prediction is used. The procedure there can be demonstrated by the following example: Suppose that the relation

$$x = B' z + \varepsilon_1$$

holds, where B is a matrix of parameters and ε_1 is a vector of errors. The problem is to estimate the matrix of parameters C in the relation

$$y = C' z + \varepsilon_2$$

where ε_2 is another vector of errors. Klevmarken solves this by a two step method. First the matrix B is estimated from file B. Then the missing z-values in file A are predicted using the estimated matrix B and the known x-values in file A.

The matrix C is estimated in the second step using the known y-values and the predicted z-values from file A.

This method has the advantage that it does not need the assumption of conditional independence between y and z. In my opinion the assumption of conditional independence raises an important question:

Exactly what is meant by the statement that SM works if y and z are independent conditional on x?

Does it mean that the artificial distribution $f(y,z)$ created in file C is a good approximation of

i) a distribution $f(y,z)$ in file A provided it were possible to measure z on the individuals in file A?

ii) a distribution $f(y,z)$ in a given finite population from which the files A and B are probability samples?

iii) a distribution $f(y,z)$ in a hypothetical or real process generating values x, y and z, that is, a superpopulation?

Suppose that the matching procedure is equivalent to the following:

1) For each value of x (or within a small interval of x) select one individual in file A with equal probability $n_A(x)^{-1}$, where $n_A(x)$ is the number of units having the same value x (or having values falling in the interval around x).

2) For each value of x, select one individual with probability $n_B(x)^{-1}$, where $n_B(x)$ is the number of individuals in B having the same value x.

3) For the individuals chosen in 1) and 2), combine the y-value from A and the z-value from B into a value (y,z) given x in a file C.

This procedure would generate combinations (y,z) with probabilities $(n_A(x,y)/n_A(x)) \cdot (n_B(x,z)/n_B(x))$. Thus we would have (at least approximately)

$$f_C((y,z)|x) = f_A(y|x)\, f_B(z|x).$$

Now, the question is: Does $f_C((y,z)|x)$ qualify as an approximation of $f_A((y,z)|x)$ if we assume conditional independence between y and z given x?

Obviously, this is the case only if $f_B(z|x)$ is a good approximation of $f_A(z|x)$. When does this occur?

In case i) it would happen by chance if the files A and B are unrelated.

In case ii) it would happen if the samples A and B were drawn with equal probability sampling so that each unit in file A has the same sample weight as in file B. (If the inclusion probabilities in the sample A are unequal for the units having the same x-value, then the unweighted distribution $f_A((y,z)|x)$ can not in general be expected to serve as an approximation of $f((y,z)|x)$ in the population, since the different inclusion probabilities would distort the distribution. However, the distribution $f((y,z)|x)$ can be estimated from a weighted equivalent of $f_A((y,z)|x)$.

A sampling design such as simple random sampling would do and so would also stratified sampling with simple random sampling within strata provided the intervals around x are used to form the strata.

(If the sample weights in the sets $A(\cdot|x)$ and $B(\cdot|x)$ where

$$A(\cdot|x) = \{i:(i \in A) \cap (x_i = x)\} \quad \text{and}$$

$$B(\cdot|x) = \{j:(j \in B) \cap (x_j = x)\}$$

are not equal, one must use a matching technique other than that described above in order to be able to estimate the distribution $f((y,z)|x)$ in the population.)

In case iii) it would happen if the process which generates the values of the variables acts uniformly and independently on each unit in the population. Then any sample would in expectation have the same distribution $f(z|x)$ irrespective of the sampling designs used for A and for B. Thus, $f_B(z|x)$ could serve as an estimator of $f_A(z|x)$. This is the analogue to what is usually called observations on independent

and identically distributed random variables, which is a standard assumption in traditional inference theory. This assumption seems to be underlying much of the discussion on SM. Economists and sociologists who are used to think in terms of models would normally have no objection against this point of view. Contrary to this, survey sampling practitioners usually prefer to think of A and B as samples from a given, fixed population.

3. Direct estimation of covariances

Rogers and DeVol (1982) state that if the assumption of the conditional independence of y and z given x holds, it is possible to estimate the covariance matrix C_{yz} directly from the knowledge of C_{xz} in file B and C_{yx} in file A through the relation

$$C_{yz} = C_{yx} C_{xx}^{-1} C_{xz}.$$

The statement can of course be adjusted to the discussion regarding the cases i–iii above. For instance, in case ii) where the covariance matrix C_{yz} in the fixed population is of interest, C_{yx}, C_{xx}, C_{xz} should be the estimates of the matrices in the population, which means that C_{yx} and C_{xx} are estimated using the sample weights from sample A and C_{yz} is estimated using the weights from sample B. This seems to be easier than first constructing a file C with appropriate sample weights, and then estimating the matrix C_{yz} in the population.

4. The SM simulation

4.1. Description

In the simulation, the files A and B both contain about 5 000 individuals. The x, y and z variables are

x_1 = degree of education (10 codes, 6 categories)
x_2 = occupation (about 280 codes, 116 categories)
x_3 = hours worked during one week (number of hours)
y_1 = sex (2 codes)
y_2 = age (number of years, 7 categories)
y_3 = sociological-economical group (index, 6 categories)
z_1 = individual income (amount of money earned)
z_2 = income at disposal for the household (amount of money)
z_3 = longlasting illness (2 codes).

The matching variables x_1, x_2, and x_3 consisted of two coded and one continuous but classified variable. We first tried to match A and B using a program for unconstrained matching given in Rogers and DeVol (1982). We chose the distance function $\Sigma_1^3(x_{ir} - x_{jr})^2$, because we wanted to see what happened if the variables were considered equally important. This matching did not work out well. Some consideration makes it clear that it is important that the variables are standardized so that the range of possible values is about the same for all variables in order to prevent that one variable will dominate heavily. Some further consideration makes it clear that the use of coded variables in the distance function is not feasible unless closeness of the codes really means that the units are close to each other in a sense which is important for the distribution $f(y,z|x)$. (This raises the question of how to select, when possible, the x-variables. This question is not studied in the present paper.) Also, the program was not as fast as one could desire. We decided to try another matching procedure, using an imputation technique. We classified the records in the files A and B according to the (discrete) values of the matching variables. When the x-value of a unit i in file A coincided with that of a unit j in file B we considered these units to be a match. When unit j differed from unit i the procedure took the record that was closest to i to be a match. The closeness was measured in the last classification variable given that the other variable values were identical. Thus, it was a hierarchical classification. The hierarchy was

set to be x_1, x_2, and x_3, where x_1 was considered to be the most important variable and x_3 the least. Because of the sorting, this program was very fast. The time needed was less than .1 per cent of the time needed to run the unconstrained matching program. Ruggles and Ruggles (1975) state that "One of the major advantages of this technique is that it easily handles very large data sets." Our findings confirm their statement.

4.2. The empirical evaluation

The aim of the empirical evaluation was primarily to study how well the SM technique could reconstruct the missing variables z in the file A. We also wanted to see how well the prediction approach technique could approximate the SM technique.

The evaluation of the SM was done by comparing crosstabulations of y and z on the matched file C with the true tables from file A. The degree of correspondence between the frequency tables from C and the true tables from A was measured by a chi-square test. Of course, the appropriateness of this test can be questioned, but at least it can serve as a measure not influenced by subjective assessments.

In statistical prediction we used the relation that y and z are independent, conditional on x, that is

$$f(y,z) = \int_x f_A(y|x)\, f_B(z|x)\, f_A(x)\, dx,$$

to create tables where the cells contain predicted numbers of units falling in the cell in question. The distributions $f(y|x)$ and $f(x)$ were estimated from A and $f(z|x)$ was estimated from B. The matching variable x_2 was ignored because the x cells would otherwise be almost empty.

4.3. Result

The matching procedure did well in transferring the marginal distributions for the z:s from file B to file C. Since the files A and B are the result of a random split of one single file, one would expect the marginal distributions in A and B to be similar and accordingly the marginal distributions of C to be close to those of A. This was also the case. None out of three χ^2-tests were significant at the 5 per cent level.

The picture was completely different when we examined the crosstabulations. Only one out of seven tests failed to achieve significance at the 5 per cent level. From this point of view the matching was unsatisfactory.

When we compared the tables from the file C with the tables obtained using statistical prediction we found that the marginal distributions of the z:s were almost the same and this was also the case for the crosstabulations. Totally there was one significant difference out of eight possible. There is virtually no difference between statistical matching and statistical prediction in this case. Thus, in honor of professor Dalenius, if someone asks if you have a statistical match, you are now able to reply: No, but I can predict one.

5. References

Bailar, B.A. and Bailar, J.C. (1979): Comparison of the Biases of the "Hot-Deck" Imputation Procedure with an "Equal-Weights" Imputation Procedure. Symposium on Incomplete Data, August 1979, Washington D.C.

Dalenius, T. (1964): Synpunkter på artikeln om "Datamaskinell optimering m.m.". Svenska Läkartidningen, 61. (Unfortunately in Swedish only.)

Dalenius, T. (1982): A Sample of Ideas for Research and Development in the Theory and Methods of Sample Surveys. Utilitas Mathematica, 21 A, pp. 59–74.

Gurney, M. and Dalenius, T. (1969): The Matrix Method for Survey Design and Estimation. Memo, US Bureau of the Census.

Hocking, R.R. (1979): The Design and Analysis of Sample Surveys with Incomplete Data: Reduction of Respondent Burden.

Symposium on Incomplete Data, August 1979, Washington D.C.

Hollenbeck, K. and Doyle, P. (1980): Distributional Characteristics of a Merged Microdata File. American Statistical Association, Proceedings of the Survey Research Methods Section, pp. 418–420.

Klevmarken, A. (1982): Missing Variables and Two-Stage Least-Squares Estimation from More Than One Dataset. American Statistical Association, Proceedings of the Business and Economic Statistics Section, pp. 156–161.

Okner, B.A. (1972): Constructing a New Data Base From Existing Microdata Sets: The 1966 Merge File. Annals of Economic and Social Measurement, 1, pp. 325–342.

Radner, D.B., Allen, R., Gonzalez, M.E., Jabine, T.B. and Muller, H.J. (1980): Report on Exact and Statistical Matching. Statistical Policy Working Paper 5, US Department of Commerce, US Government Printing Office, Washington D.C.

Rogers, W.L. and DeVol, E.B. (1982): An Evaluation of Statistical Matching. Manuscript, Survey Research Center, The University of Michigan, Ann Arbor, Michigan.

Ruggles, R. and Ruggles, N.D. (1975): The Role of Microdata in the National Economic and Social Accounts. Review of Income and Wealth, 21, pp. 203–216.

Sims, C.A. (1972): Comments (on a discussion of Okner). Annals of Economic and Social Measurement, 1, pp. 342–345.

Example:
File A. True tables

z_1 = individual income, y_2 = age

z_1 \ y_2	19	29	39	49	59	69	99	S
99	3.0	2.4	2.5	1.4	2.9	8.7	4.8	25.7
199	.7	1.5	1.2	.6	.8	.9	.2	5.9
299	.7	1.9	2.3	1.1	1.0	.6	.1	7.7
399	.6	2.3	2.3	1.6	1.6	.7	.0	9.2
499	.3	3.0	2.1	1.7	1.3	.6	.0	9.0
599	.2	4.1	3.5	2.0	1.8	.8	.0	12.4
699	.0	2.9	4.1	2.4	2.0	.7	.0	12.1
799	.0	1.1	2.9	1.4	1.5	.3	.0	7.1
899	.0	.4	1.1	.8	.9	.3	.0	3.6
9999	.0	.1	1.3	1.6	1.1	.3	.0	4.8
S	5.5	20.0	24.1	15.5	15.7	14.2	5.1	100.0

z_3 = longlasting illness

z_3 \ y_2	19	29	39	49	59	69	99	S
0	4.5	16.0	18.3	10.5	7.4	5.2	1.3	63.2
1	.9	4.0	5.7	5.0	8.3	9.0	3.8	36.8
S	5.5	20.0	24.1	15.5	15.7	14.2	5.1	100.0

y_1 = sex

z_1 \ y_1	1	2	S
99	9.3	16.4	25.7
199	1.8	4.1	5.9
299	1.6	6.1	7.7
399	2.4	6.8	9.2
499	2.7	6.3	9.0
599	6.7	5.7	12.4
699	9.2	2.9	12.1
799	5.4	1.7	7.1
899	3.0	.6	3.6
999	2.2	.2	2.4
9999	4.6	.2	4.8
S	45.9	51.1	100.0

z_3 \ y_1	1	2	S
0	30.9	32.3	63.2
1	18.0	18.8	36.8
S	48.9	51.1	100.0

File B. Tables resulting from the prediction approach

z_1 \ y_2	19	29	39	49	59	69	99	S
99	2.5	3.0	3.3	2.4	3.3	7.2	3.4	25.1
199	.4	1.0	1.1	.7	.8	1.1	.5	5.6
299	.4	1.6	1.9	1.3	1.3	1.3	.4	8.2
399	.4	2.0	2.6	1.7	1.8	1.1	.3	9.9
499	.4	2.0	2.4	1.6	1.6	.8	.2	9.0
599	.5	3.3	3.8	2.5	2.3	.9	.1	13.4
699	.4	2.9	3.4	2.2	2.0	.7	.2	11.8
799	.2	1.8	2.2	1.3	1.2	.4	.1	7.2
899	.1	.8	1.1	.6	.5	.2	.0	3.3
999	.1	.5	.8	.4	.3	.1	.0	2.2
9999	.1	.9	1.5	.9	.6	.3	.0	4.3
S	5.5	19.8	24.1	15.6	15.7	14.1	5.2	100.0

z_3 \ y_2	19	29	39	49	59	69	99	S
0	3.0	13.0	16.0	10.0	10.0	8.0	2.0	62.0
1	3.0	7.0	8.0	5.0	6.0	7.0	2.0	38.0
S	6.0	20.0	24.0	15.0	16.0	15.0	4.0	100.0

z_1 \ y_1	1	2	S
99	11.0	14.0	25.0
199	2.0	3.0	5.0
299	3.0	6.0	9.0
399	3.0	7.0	10.0
499	4.0	5.0	9.0
599	8.0	6.0	14.0
699	7.0	4.0	11.0
799	4.0	3.0	7.0
899	2.0	1.0	3.0
999	2.0	1.0	3.0
9999	3.0	1.0	4.0
S	49.0	51.0	100.0

z_3 \ y_1	1	2	S
0	31.0	31.0	62.0
1	18.0	20.0	38.0
S	49.0	51.0	100.0

File C. Tables resulting from the SM approach

z_1 \ y_2	19	29	39	49	59	69	99	S
99	2.2	3.4	4.0	2.6	3.5	6.8	2.8	25.3
199	.3	1.3	1.4	.8	.8	.8	.2	5.6
299	.5	1.7	1.8	1.4	1.4	1.0	.4	8.1
399	.6	2.2	2.4	1.6	1.5	1.1	.3	9.8
499	.5	1.8	2.2	1.5	1.6	1.0	.4	9.0
599	.5	3.1	3.7	2.2	2.4	1.0	.3	13.2
699	.4	2.8	3.2	1.8	2.3	.8	.3	11.6
799	.2	1.5	2.3	1.4	1.1	.6	.1	7.3
899	.0	.7	1.0	.8	.5	.4	.1	3.5
999	.0	.7	.6	.4	.2	.3	.2	2.4
9999	.0	.9	1.7	.8	.5	.4	.0	4.4
S	5.5	20.0	24.1	15.5	15.7	14.2	5.1	100.0

z_3 \ y_2	19	29	39	49	59	69	99	S
0	3.5	13.6	16.0	9.9	9.4	7.5	2.5	62.5
1	2.0	6.4	8.1	5.6	6.3	6.7	2.6	37.5
S	5.5	20.0	24.1	15.5	15.7	14.2	5.1	100.0

z_1 \ y_1	1	2	S
99	11.1	14.2	25.3
199	2.7	2.9	5.6
299	3.7	4.3	8.1
399	4.8	5.0	9.8
499	4.6	4.4	9.0
599	6.8	6.4	13.2
699	5.8	5.8	11.6
799	4.0	3.3	7.3
899	1.8	1.7	3.5
999	1.1	1.3	2.4
9999	2.6	1.7	4.3
S	49.0	51.0	100.0

z_3 \ y_1	1	2	S
0	30.7	31.7	62.5
1	18.2	19.3	37.5
S	48.9	51.1	100.0

Pooling Incomplete Data Sets

by N. Anders Klevmarken[1]

0. Abstract

Data needed in micro studies are not always available from just a single source. In such cases it might be possible to combine two or more independent samples. The problem studied in this paper is to estimate a linear function between y and x from one sample of y-observations and another of x-observations. This is feasible if there are common variables z which can be used to predict x. A two-stage least squares estimator is propounded, which, for the model considered, is also an ML estimator. Simulation experiments show that it has a good relative efficiency and virtually no small sample bias.

1. Introduction

Microdata, i.e. data on individuals, households or firms, have long been used in social science research, and their importance is increasing.[2] Since human behavior sometimes calls for very complex explanations and since social experimentation is rarely a feasible approach, social scientists frequently work with complex models involving many variables in efforts to control for confounding effects. To estimate these models we do not only need large samples of microdata, but ideally we would also need to observe many aspects of behavior for *each* individual, household or firm.

Surveys are the primary source of microdata, but survey research is very expensive and few researchers can afford new surveys. Contributing to the high costs are the increasing difficulties in many countries to gain the cooperation of the respondents. In particular, when many questions are asked and there is a heavy respondent burden, the respondents tend to economize their time. They have also become aware of the privacy issues. The public debate about the use of computers and the risks for invasion of privacy, which in some countries have resulted in data legislation, have made it even more difficult to make the respondents cooperate.

As pointed out in Dalenius (1982), matrix sampling is a class of sampling schemes which may cope with these problems. Since with such a design only a sample of all variables is observed for each selected unit, the respondent burden and the risk for invasion of privacy are reduced. It is, however, obvious that multivariate analysis from such a sample might meet with difficulties.

High costs and nonresponse problems also make us look for alternative data sources, i.e. already existing data files and administrative records. Frequently, however, we are unable to find all the variables needed in those files and records. If feasible, matching of two or more data sets might give us what we need. Matching is also much less expensive than a new survey. However, in most cases exact matching is not possible, either because there is no overlap, i.e. no individual, household or firm can be found in more than one data set, or there is no unique identification term, or the use of this term is prohibited in order to protect personal privacy. In Sweden, for instance, matching data on individuals falls under the

[1] Professor of Statistics, University of Gothenburg, Sweden.

[2] For a discussion of the need for microdata in economics, see Klevmarken (1983a).

Data Act and requires a permit from the National Data Inspection Board.

If neither a new complete data collection nor an exact matching of existing files are feasible solutions, to what extent is it then possible to use non-overlapping data sets? The answer to this question will in the general case depend on the intended analysis and in what sense the data are incomplete. This paper treats the problem of estimating a linear relation from two independent data sets, none of which includes all the relevant variables.

Suppose, for instance, that we want to estimate a model according to which y depends on x. y and x are, however, not observed for the same units, but y- and x-observations are found in two independent samples. This is a hopeless situation unless something more is known about y and x, and there is some common information in both samples. Suppose a third variable, z, which could be used as a predictor of x, is observed in both samples. It is then possible to estimate the predictive relationship between x and z from one sample and use the estimates in the other sample to predict x. The relation between y and x is finally estimated on the basis of these predictions and observed y-values. This procedure requires model assumptions which make it possible to pool two (or more) samples. The properties of the estimates, of course, depend on the assumptions made, and the whole approach is useful only if these are not too constraining. As we shall see, it is possible to work within quite a general class of linear models.

When invasion of privacy is a real issue or nonresponse a severe problem, a complete data collection might not be permissible or feasible. However, if good predictors can be found, each individual, household or firm would only need to contribute partial information, which would be less sensitive, and an analysis would still be feasible with the methods suggested in this paper.

Missing variables is not a new problem. In most textbooks of regression analysis and econometrics it is discussed as a specification error. In the applied literature various *ad hoc* approaches can be found. For instance, in the economic literature on compensating wage differentials, aggregate data on occupational characteristics such as accident rates and aggregate work environment data are matched with individuals on the basis of their occupation and industry. One example is given in Brown (1980). As we shall find, this is a special application of the prediction approach discussed below.

An alternative approach is statistical matching (U.S. Department of Commerce (1980)). The theoretical basis for statistical matching is still undeveloped, and it is also a very expensive approach (Barr et al. (1982), Paass (1982), Rogers and DeVol (1982)). The approach discussed in this paper has the advantages of being based on conventional statistical theory and of requiring no expensive matching of microdata.

In Section 2 of the paper we review some results for a model with stochastically dependent equations. Since interdependent systems have mostly been used in connection with aggregate time-series, while microdata applications are less frequent, it might erroneously be believed that the approach suggested in this paper is only of minor interest in microstudies. However, the importance of this type of model is likely to increase in microstudies as well, when more longitudinal data become available and when ample data make it feasible to analyze joint decisions of microunits, e.g. joint household decisions about work, leisure, consumption and savings. Also, and more important, the estimation approach suggested is not only applicable to interdependent systems but to a much wider class of models. In Section 3 this approach will be used to estimate a linear regression model. One reason to discuss its application to an interdependent system first is that the "predictors" of the unobserved

variables are in a natural way given by theory, which is not obviously the case with a regression model.

Section 4 presents some findings from a simulation study of the small sample properties of the estimator, and Section 5 gives some concluding remarks.

2. Estimation of an interdependent linear model

The estimation of one equation in an interdependent system of equations from two independent samples with missing variables was discussed in Klevmarken (1982). This equation was specified as,

$$y = Y_1 \beta + X_1 \gamma + u; \tag{1}$$

and it was part of the interdependent system,

$$YB' + X\Gamma' = U; \tag{2a}$$

$$E(U) = 0;\ E(U'U) = n\Sigma, \tag{2b, c}$$

where $Y_{n \cdot G}$ is a matrix of n observations on G endogenous variables,

$y_{n \cdot 1}$ is a vector of the n observations on the endogenous variable explained by (1),

$Y_{1, n \cdot g}$ is a matrix of the n observations on the g explanatory endogenous variables in (1),

$X_{n \cdot K}$ is the observational matrix of all K exogenous variables,

$X_{1, n \cdot k}$ is a submatrix of X which includes the k exogenous variables in (1),

$U_{n \cdot G}$ is a matrix of stochastic disturbances,

$u_{n \cdot 1}$ is the vector of stochastic disturbances of (1), one of the columns of U,

$B_{G \cdot G}$ and $\Gamma_{K \cdot K}$ are parameter matrices,

$\beta_{g \cdot 1}$ and $\gamma_{k \cdot 1}$ are vectors of the non-zero parameters in (1),

$\Sigma_{G \cdot G}$ is an unknown positive definite moment matrix.

It is assumed that (1) is identified.

The reduced form of the complete system is,

$$Y = X\pi' + V; \tag{3a}$$

where $\pi = -B^{-1}\Gamma$ (3b)

and $V = U(B')^{-1}$. (3c)

The part of the reduced form corresponding to the endogenous variables to the right in equation (1) is,

$$Y_1 = X\pi_1' + V_1; \tag{4}$$

where π_1 and V_1 are the corresponding $g \cdot K$ and $n \cdot g$ submatrices of π and V, respectively. For later use it is also convenient to introduce an $n \cdot (K-k)$ matrix X_2 defined by,

$$X = \{X_1 \mid X_2\}. \tag{5}$$

Suppose now that data are not available in the form of one complete sample, but that there are two samples, A and B, none of which contains all the variables. Assume that the

data come in the following form,

Sample A: $y^A_{(n_A \cdot 1)}$; $X^A_{(n_A \cdot K)}$

Sample B: $Y^B_{1,(n_B \cdot g)}$; $X^B_{(n_B \cdot K)}$.

n_A and n_B are the two sample sizes. They are not necessarily equal. Since (2c) implies that there is no residual correlation between observational units, the two samples can be treated as independent random samples.

An example to which this problem specification might be applicable is the joint estimation of demand functions for consumer goods and household time-use functions, both derived from a household production type of model. Consumer expenditure data could be obtained from a household expenditure study, while time-use data would have to be taken from a separate time-use survey. There are at present practically no surveys which include both kinds of data. Both kinds of surveys would, however, give income data and other characteristics of the household.

Eq. (1) cannot be estimated from sample A alone, since the Y_1-variables are missing, but, if $g \leq K-k$, the two samples can be combined in the following two-stage procedure:

I. Estimate the reduced form equations (4) from sample B by OLS, which gives the estimates $\hat{\pi}^B_1$. Use these estimates to predict Y_1 in sample A, i.e.

$$\hat{Y}^A_1 = X^A(\hat{\pi}^B_1)'; \tag{6}$$

II. Estimate by OLS from sample A

$$y^A = \hat{Y}^A_1\beta + X^A_1\gamma + (u^A + \tilde{V}^A_1\beta); \tag{7}$$

where $\tilde{V}^A_1 = Y^A_1 - \hat{Y}^A_1$.

Note that $\tilde{V}^A_1$ is not the vector of least squares prediction errors from sample A and thus not necessarily orthogonal to X^A.

With the following notation,

$$\delta' = \{\beta' \mid \gamma'\}_{1 \cdot (g+k)};$$

$$Z = \{\hat{Y}^A_1 \mid X^A_1\}_{n_A \cdot (g+k)};$$

(7) becomes

$$y^A = Z\delta + (u^A + \tilde{V}^A_1\beta); \tag{8}$$

and the estimator of δ is,

$$\hat{\delta} = (Z'Z)^{-1}Z'y^A . \tag{9}$$

If the two samples would coincide, $\hat{\delta}$ would be the usual TSLS estimator.

In Klevmarken (1982) it was shown that the estimator $\hat{\delta}$ is biased but consistent. The following asymptotic properties can also be proved.[1]

If $n_B = cn_A$, where $c>0$ is an arbitrary finite constant, if $(1/n_A)(X^{A'}X^A)$ and $(1/n_B)(X^{B'}X^B)$ both tend to finite non-singular limits when n_A and n_B tend to infinity, and if the rows of the error matrix U are stochastically independent, then $\sqrt{n_A}(\hat{\delta}-\delta)$ asymptotically follows a normal distribution with zero mean vector and covariance matrix
$(\sigma_{11} + 2\sigma_1'B^*\beta + 2\beta'B^{*\prime}\Sigma B^*\beta)Q^{-1}$.

σ_1 is the first column of the matrix Σ. The first element of this vector is σ_{11} and the other elements are the covariances between the error term in the first equation and those in the other equations. B^* is a matrix of g columns from $(B')^{-1}$ such that

$$V_1 = UB^*.$$

Q, a finite non-singular matrix, is the limit to which $(1/n_A)(Z'Z)$ tends in probability when n_A tends to infinity.

The asymptotic moment matrix of the ordinary TSLS estimator based on a complete A-

[1] This corrects results given in Klevmarken (1982). A proof is parallel to that given in Klevmarken (1983b).

sample is $\sigma_{11}\ Q^{-1}$. As shown in Klevmarken (1982), it is possible to find cases for which the asymptotic variance of $\hat{\delta}$ is smaller than the asymptotic variance of the ordinary TSLS estimator.

As a preliminary to the next section, suppose that sample A does not only lack the Y_1-observations but also all observations on one or more of the X_1-variables. Could we then use the information in sample B to predict X_1 in A? Since X_1 by definition is exogenous, there is no theoretical justification for predicting X_1 within the present model. Additional assumptions about X_1 are needed. This problem and the nature of the new assumptions are, however, not particular to a model of interdependent equations, and could more conveniently be discussed within the framework of an ordinary regression model.

3. Estimation of a linear regression model

As above, we assume that there are two independently drawn samples, A and B. These now include the following variables and observations,

Sample A: $y^A_{(n_A \cdot 1)'}$ $X^A_{2(n_A \cdot k_2)}$

Sample B: $X^B_{1(n_B \cdot k_1)'}$ $X^B_{2\,(n_B \cdot k_2)}$.

The two samples will be used to estimate the model,

$$y = X_1\gamma + u;\ u \sim N(0, \sigma_u^2 I), \qquad (3.1\ \text{a,b})$$

where γ is an unknown parameter vector. X_1 is assumed to be a matrix of k_1 stochastic variables, which depends linearly on X_2.

$$X_1 = X_2 R + \varepsilon; \qquad (3.2\ \text{a})$$

$$\varepsilon = \{\varepsilon_1, \ldots, \varepsilon_{k_1}\}; \qquad (3.2\ \text{b})$$

$$\varepsilon_j \sim N(0, \sigma_j^2 I); \qquad (3.2\ \text{c})$$

$$E(\varepsilon_i \varepsilon_j') = 0\ \forall i \neq j. \qquad (3.2\ \text{d})$$

R is a parameter matrix. X_2 is treated as a matrix of non-stochastic variables. It is also assumed that there are at least as many predictors as predictands, i.e. $k_2 \geqq k_1$. For the two samples we thus obtain the following three relations,

$$y^A = X_1^A\gamma + u^A; \qquad (3.3\ \text{a})$$

$$X_1^A = X_2^A R + \varepsilon^A; \qquad (3.3\ \text{b})$$

$$X_1^B = X_2^B R + \varepsilon^B. \qquad (3.3\ \text{c})$$

Since X_1^A is not observed, (3.3 b) is inserted into (3.3 a) to give the following reduced system of observed variables

$$y^A = X_2^A R\gamma + \varepsilon^A\gamma + u^A \qquad (3.4\ \text{a})$$

$$X_1^B = X_2^B R + \varepsilon^B. \qquad (3.4\ \text{b})$$

This model shows great similarities with an errors-in-variables model analysed in Goldberger (1972). Goldberger's starting point was a regression model containing a single explanatory variable observed with a random error. This model also assumed that the true unobserved variable was a non-stochastic linear function of a number of independent variables. Zellner (1970) had previously considered this model and developed a generalized least-squares estimator and also presented a Bayesian analysis of the model. Goldberger developed the corresponding maximum-likelihood theory and extended it to a model in which the true unobserved variable is a stochastic function of the independent variables rather than an exact function. The observable equivalent of Goldberger's model has the same form as eq:s (3.4 a, b). The only difference is that the two equations refer to two different samples, while Goldberger's observa-

tions were assumed to come from a single sample.

It is convenient to rewrite eq:s (3.4 a, b) in the following way,

$$y^A = X_2^A\pi+v, \tag{3.5 a}$$

$$x_{1j}^B = X_2^B r_j + \varepsilon_j^B;\ j = 1,\ldots,k_1, \tag{3.5 b}$$

where x_{1j}^B is the j:th column of X_1^B, r_j is the j:th column of R, and

$$\pi = R\gamma; \tag{3.5 c}$$

$$v = \mathcal{E}\,\gamma+u. \tag{3.5 d}$$

If it is assumed that u and ε_j are uncorrelated for all j, it follows that

$$E(VV') = (\gamma'\Sigma\gamma+\sigma_u^2)I; \tag{3.6 a}$$

where

$$\Sigma = \text{diag}\,\{\sigma_1^2,\ldots,\sigma_{k_1}^2\}. \tag{3.6 b}$$

Since the two samples are independent, the joint distribution of the observable variables is,

$$F(y^A,\ X_1^B) = F_y(y^A)\cdot F_{X_1}(X_1^B). \tag{3.7}$$

The likelihood function becomes (disregarding irrelevant constants),

$$L(\gamma,R,\sigma_u^2,\Sigma) = -n_A\log(\gamma'\Sigma\gamma+\sigma_u^2) -$$

$$(\gamma'\Sigma\gamma+\sigma_u^2)^{-1}(y-X_2^AR\gamma)'(y-X_2^AR\gamma) -$$

$$n_B\sum_j\log(\sigma_j^2) -$$

$$\sum_j(\sigma_j^2)^{-1}(x_{1j}^B-X_2^Br_j)'(x_{1j}^B-X_2^Br_j). \tag{3.8}$$

Differentiating with respect to the unknown parameters and putting the derivatives equal to zero gives,

$$\frac{\partial L}{\partial\sigma_u^2} = -n_A(\gamma'\Sigma\gamma+\sigma_u^2)^{-1} +$$

$$(\gamma'\Sigma\gamma+\sigma_u^2)^{-2}(y-X_2^AR\gamma)'(y-X_2^AR\gamma) = 0. \tag{3.9}$$

If we define $\xi = \gamma'\Sigma\gamma+\sigma_u^2$, it thus follows that

$$\hat{\xi} = n_A^{-1}(y-X^AR\gamma)'(y-X_2^AR\gamma), \tag{3.10}$$

$$\frac{\partial L}{\partial\sigma_j^2} = -n_A(\gamma'\Sigma\gamma+\sigma_u^2)^{-1}\gamma_j^2 +$$

$$\gamma_j^2(\gamma'\Sigma\gamma+\sigma_u^2)^{-2}(y-X_2^AR\gamma)'(y-X_2^AR\gamma) -$$

$$n_B(\sigma_j^2)^{-1} +$$

$$(\sigma_j^2)^{-2}(x_{1j}^B-X_2^Br_j)'(x_{1j}^B-X_2^Br_j) = 0. \tag{3.11}$$

From eq:s (3.9) and (3.11) it follows that

$$\hat{\sigma}_j^2 = n_B^{-1}(x_{1j}^B-X_2^Br_j)'\ (x_{1j}^B-X_2^Br_j). \tag{3.12}$$

The concentrated likelihood function then becomes

$$L^* = -n_A\log\{(y^A-X_2^AR\gamma)'(y^A-X_2^AR\gamma)\} -$$

$$n_B\sum_j\log\{(x_{1j}^B-X_2^Br_j)'(x_{1j}^B-X_2^Br_j)\}. \tag{3.13}$$

$$\frac{\partial L^*}{\partial\gamma} = -n_A[(y^A-X_2^AR\gamma)'(y^A-X_2^AR\gamma)]^{-1}.$$

$$R'X_2^{A\prime}(y^A-X_2^AR\gamma) = 0. \tag{3.14}$$

$$\because\ \hat{\gamma} = (R'X_2^{A\prime}X_2^AR)^{-1}R'X_2^{A\prime}y^A. \tag{3.15}$$

If eq. (3.15) is premultiplied by $R'X_2^{A\prime}X_2^AR$ one easily sees that

$$\hat{\pi} = \hat{R\gamma} = (X_2^{A\prime}X_2^A)^{-1}X_2^{A\prime}y^A. \tag{3.16}$$

Eq. (3.16) inserted into eq. (3.13) gives a new concentrated likelihood function,

$$L^{**} = -n_B \sum_j \log\{(x^B_{1j} - X^B_2 r_j)' (x^B_{1j} - X^B_2 r_j)\}. \tag{3.17}$$

If the derivative of this function is put equal to zero, we obtain the ordinary least-squares solution,

$$\hat{r}_j = (X^{B\prime}_2 X^B_2)^{-1} X^{B\prime}_2 x_{1j}; \tag{3.18 a}$$

or in matrix form

$$\hat{R} = (X^{B\prime}_2 X^B_2)^{-1} X^{B\prime}_2 X^B_1. \tag{3.18 b}$$

From (3.15) and (3.18 b) we thus find that the maximum-likelihood solution is equivalent to a TSLS procedure analogous to the estimation method suggested for the interdependent model in the previous section.

This result no longer holds if the X_1-variables are correlated, i.e. if Σ is not diagonal, or if the errors are heteroscedastic. With such changes in the model specification we would have to find the maximum likelihood estimates by numerical maximization of the likelihood function.

If we allow both for contemporaneously correlated errors and for heteroscedasticity, the model can be written in the following way,

$$y^A = X^A_2 R\gamma + \varepsilon^A_\gamma + u^A; \tag{3.19}$$

$$x^B = (I \otimes X^B_2) r + \varepsilon^B; \tag{3.20}$$

where $x^{B\prime} = \{x'_{11}, x'_{12}, \ldots, x'_{1k_1}\}$,
$r' = \{r'_1, r'_2, \ldots, r'_{k_1}\}$,
$\varepsilon' = \{\varepsilon'_1, \varepsilon'_2, \ldots, \varepsilon'_{k_1}\}$,
and
$\varepsilon'_j = \{\varepsilon_{j1}, \varepsilon_{j2}, \ldots, \varepsilon_{jn_B}\}$.

Now let

$$E(\varepsilon_{is}\varepsilon_{jt}) = \begin{cases} \sigma_{ij(t)} & \text{if } s = t \\ 0 & \text{if } s \neq t \end{cases} \tag{3.21}$$

We define the contemporaneous moment matrices for observation t,

$$E(\varepsilon_{1t}\, \varepsilon_{2t} \ldots \varepsilon_{k_1 t})\,(\varepsilon_{1t}\, \varepsilon_{2t} \ldots \varepsilon_{k_1 t})' = \Sigma_{(t)} = \{\sigma_{ij(t)}\}_{k_1 \cdot k_1}, \tag{3.22}$$

and the individual specific moment matrices,

$$E(\varepsilon_i \varepsilon_j') = \Omega_{ij} = \operatorname{diag}\{\sigma_{ij(1)}\, \sigma_{ij(2)} \ldots \sigma_{ij(n_B)}\}. \tag{3.23}$$

These two types of matrices are thus functions of the same parameters.

We also allow u to be heteroscedastic, i.e.

$$E(uu') = \Omega_u = \operatorname{diag}\{\sigma^2_{u(1)}\, \sigma^2_{u(2)} \ldots \sigma^2_{u(n_A)}\}. \tag{3.24}$$

With these assumptions it follows that

$$E(\varepsilon\gamma + u)(\varepsilon\gamma + u)' = E(\varepsilon\gamma\gamma'\varepsilon + uu') =$$

$$= \begin{Bmatrix} \gamma'\Sigma_{(1)}\gamma & 0 & \ldots & 0 \\ 0 & \gamma'\Sigma_{(2)}\gamma & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & & \gamma'\Sigma_{(n_A)}\gamma \end{Bmatrix} + \Omega_u, \tag{3.25}$$

and

$$E(\varepsilon\varepsilon') = \begin{Bmatrix} \Omega_{11} & \Omega_{12} \ldots\ldots & \Omega_{1k_1} \\ \Omega_{21} & \Omega_{22} \ldots\ldots & \Omega_{2k_1} \\ \ldots\ldots\ldots\ldots & & \\ \Omega_{k_1 1} & \Omega_{k_1 2} \ldots & \Omega_{k_1 k_1} \end{Bmatrix}. \tag{3.26}$$

The matrix (3.25) is denoted V and the matrix (3.26) Ω. Disregarding irrelevant constants the likelihood function then becomes,

$$L = -\log\det(V) + tr\{V^{-1}(y^A - X^A_2 R\gamma)(y^A - X^A_2 R\gamma)'\} - \log\det(\Omega) + tr\{\Omega^{-1}(x - (I \otimes X^B_2) r)(x - (I \otimes X^B_2) r)'\}. \tag{3.27}$$

The model will not be identified unless additional assumptions are made about the nature of the heteroscedasticity.

The TSLS estimate of γ is consistent. This result follows, because it is a function of consistent estimates of the parameters R of the auxiliary relations and it is in itself consistent conditional on R. This is true also for non-

normal errors.

$\hat{\gamma}$, however, is not unbiased. This is easily seen from the simple model with $k_1 = k_2 = 1$ and only one γ-parameter:

$$\hat{\gamma} = \left(\frac{\Sigma x_1^B x_2^B}{\Sigma (x_2^B)^2} \right)^{-1} \left(\frac{\Sigma x_2^A y}{\Sigma (x_2^A)^2} \right),$$

$$E(\hat{\gamma}) = E_{x_1}(E(\hat{\gamma}|x_1)) =$$

$$E_{x_1} \left\{ \left(\frac{\Sigma x_1^B x_2^B}{\Sigma (x_2^B)^2} \right)^{-1} \gamma r \right\} \neq r^{-1} \gamma r.$$

The critical assumptions of the present model and of the whole approach are those about the auxiliary relations (3.3 b) and (3.3 c). It might at first seem very constraining to assume that there exist linear relations which explain the unobserved variables and that these relations are the same for both samples. However, we do not necessarily have to give them a causal interpretation but could rather look upon them as predictive relations. In practical applications of this approach we will then face the problem of finding good predictors. Linearity is not a binding assumption. At the minor cost of discontinuities, non-linear relations can be transformed into linear relations with dummy variables. If the samples are sufficiently large, we could use dummy variables only and still obtain a good precision of the estimates. That would be equivalent to grouping the two samples by the same groups, merging them at the group level and then estimating the relation between y and X_1 from group means.

The auxiliary relations must be formulated in such a way that no parameters are needed to predict X_1 in sample A, which cannot be estimated from sample B. This means that the two samples cannot differ too much with respect to observational units, variable definitions etc. However, the two samples do not have to be of the same size, and the number of observations with a particular configuration of X_2-values can differ between the samples. Other differences might also be acceptable. For instance, if the model explains household behavior and sample A includes household data, it might be possible to use data on individual persons in sample B, viz. if household behavior (data) can be predicted from individual behavior (data).

4. A simulation experiment

A simulation experiment was made to illustrate the estimation procedure numerically and to get an idea about its small sample properties. The experiment was based on a demand equation, which relates expenditures on food, beverages and other everyday commodities to household disposable income, age of household head and household size. The parameters of this function were estimated from two samples of household data. The first sample included all variables except disposable income. There was, however, in this sample information about the assessed income for each household member. The second sample included both household disposable income and assessed income, as well as additional variables which could be used to predict disposable income, the most important one being the ownership of owner-occupied houses[1]. The model was specified as follows,

$$\begin{aligned} \text{LEXP} = {} & 3.594 + .4\,\text{LDISP} + .04\,\text{AGE} - \\ & .00043\,\text{AGE}^2 + .644\,\text{HS} - \\ & .0723\,\text{HS}^2 + u; \end{aligned}$$

$$u \sim N(0, \sqrt{.45}\,);$$

$$\begin{aligned} \text{LDISP} = {} & 2.665 + .7445\,\text{LTAXINC} + \\ & .5626\,\text{DH} - \\ & .00000176\,\text{DHTAXINC} + \varepsilon; \end{aligned}$$

[1] Deduction of interest payments reduces the assessed income for owners of owner-occupied houses, as compared with non-owners.

$\varepsilon \sim N(0, \sqrt{.053})$;

$E(u\varepsilon) = 0$;

where LEXP = The logarithm of expenditures on food, beverages and other everyday commodities,

LDISP = The logarithm of disposable income,

AGE = The age of the household head,

HS = Household size,

LTAXINC = The logarithm of the total of the assessed incomes of all household members,

DH = A dummy variable for owners of owner-occupied houses,

and

DHTAXINC = Assessed income for owners of owner-occupied houses; zero for non-owners.

With minor adjustments the parameter values were obtained as the least squares estimates of the model from a small sample of 144 households, which included survey and register information of all variables. The observations on the exogenous variables of this sample were then used in the simulations. First, the sample was randomly divided into two samples, *A* and *B*. Then for each observation, LDISP was simulated either four times or ten times. In sample *A* the simulated LDISP values were used to simulate LEXP values. In this way we obtained a simulated sample four or ten times larger than the original sample but with the same covariance structure of the exogenous variables. Finally, the parameters of the demand function were estimated by the TSLS procedure using the two subsamples. The whole simulation and estimation procedure was then replicated.

There are altogether seven simulations. In all but the last two of these, the two subsamples were of the same size, 288 in simulations 1 and 2 and 788 in 3 to 5. In simulations 7 and 8 sample *A* included 360 observations, while sample *B* was three times as large.

The results are shown in Table 1. They indicate that, when the true predictive relation is used, there is virtually no bias. The bias estimates given in the table are so small that they are dominated by the random fluctuations of the simulation experiment. In simulation 5 the predictive relation used in the estimation was misspecified. The house-owner variable and the interaction variable were both deleted and the logarithm of assessed income was thus the only predictor.[1] The result is a small bias. The income elasticity is underestimated by 5 per cent, and there is also an 8–9 per cent bias in the estimates of the household size parameters.

Since the LDISP variable is predicted with an error, it might be expected that the efficiency of the TSLS estimator would be less than for the LS estimator based on a complete sample. In Table 1 we can compare the relative root mean-square errors for these two cases.[2] In both cases this measure includes the variability caused by the drawing of a new LDISP vector for each sample. We find that the estimated relative efficiency, defined as the ratio of the two root mean-square errors, is approximately 80 per cent for the income elasticity, i.e. for parameter γ_2. It is between 80 and 90 per cent for the intercept and above 90 per cent for the other parameters. The number of replications are not large enough to justify any conclusions about the dependence of the efficiency on the sample size. When the

[1] The full model was, of course, used to simulate data.

[2] The relative root mean-square error for the estimator $\hat{\gamma}_i$ is defined as $\sqrt{\text{MSE}_i}/\gamma_i$.

Table 1. Results from seven simulation experiments

	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6
Parameter values	3.594	.400	.040	– .00043	.644	– .0723
Simulation 1: TSLS						
n_A 288						
n_B 288						
Replications 800						
Relative bias (%)	.42	– .82	2.42	1.72	.79	1.36
Relative $\sqrt{\text{MSE}}$ (%)	32.82	29.55	57.50	53.49	27.00	36.66
Relative sq. root mean variance estimate (%)	33.46	30.30	56.00	53.49	27.83	38.59
Simulation 2: LS, x_2 = LDISP known						
n_A 288						
Replications 400						
Relative $\sqrt{\text{MSE}}$ (%)	29.54	24.60	56.25	53.49	24.78	34.16
Relative sq. root mean variance estimate (%)	28.54	24.75	54.75	51.16	26.89	37.48
Simulation 3: TSLS						
n_A 720						
n_B 720						
Replications 600						
Relative bias (%)	–1.00	.61	1.83	.28	– .14	–1.14
Relative $\sqrt{\text{MSE}}$ (%)	21.82	19.16	35.30	33.23	17.23	24.17
Relative sq. root mean variance estimate (%)	21.22	19.21	35.54	33.53	17.64	24.45
Simulation 4: LS, x_2 = LDISP known						
n_A 720						
Replications 1000						
Relative $\sqrt{\text{MSE}}$ (%)	17.57	15.05	34.83	32.94	16.62	23.49
Relative sq. root mean variance estimate (%)	17.99	15.61	34.57	32.66	16.99	23.69
Simulation 5: TSLS, wrong predictors						
n_A 720						
n_B 720						
Replications 600						
Relative bias (%)	1.62	–4.86	2.18	– .22	9.13	–7.70
Relative $\sqrt{\text{MSE}}$ (%)	24.16	22.57	37.55	35.52	19.66	25.84
Relative sq. root mean variance estimate (%)	24.02	21.56	36.22	34.28	17.38	24.51

(cont.)

Table 1 (cont.). Results from seven simulation experiments

	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6
Simulation 6: TSLS						
n_A 360						
n_B 1080						
Replications 400						
Relative bias (%)	–1.37	2.24	–5.30	5.47	–1.65	3.02
Relative $\sqrt{\text{MSE}}$ (%)	31.11	31.31	59.01	56.01	28.78	42.99
Relative sq. root mean variance estimate (%)	29.74	29.85	60.06	57.20	27.48	40.72
Simulation 7: LS, x_2 = LDISP known						
n_A 360						
Replications 400						
Relative $\sqrt{\text{MSE}}$ (%)	25.37	23.90	54.13	51.74	27.62	41.92
Relative sq. root mean variance estimate (%)	24.91	23.56	56.28	53.79	26.59	39.77

predictor relation is incorrectly specified there is an additional, but in this particular case, modest loss in efficiency. This is at least partly caused by the bias component.

Finally, Table 1 also shows that the variance formula for the ordinary least-squares estimates applied to the second step of the TSLS procedure, i.e. the diagonal elements of $s^2(\hat{X}_1'\hat{X}_1)^{-1}$, gives good estimates of the true variance of $\hat{\gamma}$. s^2 is the residual variance in the second estimation step. The last row of each panel of Table 1 shows the square-root of the mean of the replications of each diagonal element relative to the true parameter value. They differ very little from their corresponding relative root mean-square errors.

5. Concluding remarks

It is not unusual in nonexperimental studies that variables are missing or replaced by proxies. Even when a new survey is to be made it might not be feasible to collect all items of information from every respondent, either because of the risk for invasion of privacy or because the respondent burden might become so high that the response rate would drop below an acceptable level. One approach to solve this problem has been suggested in this paper. When the statistical problem is to estimate a linear relation between y and x or an equation in an interdependent linear system, one sample including all variables is not necessarily needed, but two or more samples, each with missing variables, can be used instead. A sampling scheme like matrix sampling could thus be used in combination with the estimation method suggested. A condition is that it is possible to predict the missing variables.

If "true" predictors are used, the two-stage least-squares estimator suggested has good properties. It is consistent asymptotically normal. For a proof see Klevmarken (1983b).

What has been shown here is that the estimator is a maximum-likelihood estimator, if the moment matrices of the regression model and the predictive relations are scalar. When they are not scalar matrices, the maximum-likelihood estimates can be obtained by numerical optimization of the likelihood function.

A sampling experiment has indicated that the two-stage least-squares estimator also has favorable small sample properties. We found virtually no bias, and the decrease in efficiency, relative to the case with one complete sample, was small.

In practice it might be difficult to know the "true" predictors. In the case of the interdependent model, the predictors are given by the model. If all of them cannot be used, e.g. because of a shortage of data, the estimates will no longer be consistent (see Klevmarken (1982) pp. 159–160). For the regression model the same conclusion does not necessarily follow. In this case the predictive relations do not necessarily have the same status of theory. To some extent we can choose these relations at our convenience. If x_1 is a stochastic variable, it is always possible to define distributions conditional on some x_2. If these distributions do not have the same mean, $E(x_1|x_2)$ could, in principle, be used as a predictive relation. The efficiency of the resulting estimate would, however, depend on how well this relation predicts x_1. If the residual variance is high, the estimates are likely to have a high variance as well. In practical applications it would thus be desirable to use at least part of sample B to find good predictors. This search process is stochastic, and the variability introduced by the search should, in principle, be taken into account when the properties of the estimator are evaluated.

With X_1 unobserved and with no supplementary information, the regression model is unidentified. The identification is achieved by the predictive relation, which adds the necessary *a priori* information and bridges the two samples. *A priori* information might also come in other forms. For instance, if the unobserved X_1 variables do not only explain one y-variable but two or more variables, we could look upon the y-variables as indicators of X_1 and arrive at a model which is similar to a factor analysis model. All these indicators do not necessarily have to come from the same sample. Although the details have to be worked out, it should be possible to combine two or more independent samples in this case as well.

There are common features of the approach suggested in this paper and the general principles of statistical matching. In statistical matching similar observations in the two samples are matched. Similarity is defined either by a grouping principle or by a distance measure defined on a set of variables. These variables basically serve the same purpose as the predictor variables in the TSLS approach. There are, however, also important differences. In most applications of statistical matching there is the implicit assumption of independence between y and X_1 conditional on X_2. The models discussed in this paper do not involve this assumption. The conditional covariance between y and X_1 in the regression model is a function of the structural parameters γ and the moment matrix of the errors in the predictor relations.

Finally, a remark on functional form. Linear models have been assumed throughout this paper. Extending the same approach to non-linear models would be quite conceivable, and there is no reason why this would not be feasible. The properties of the estimates would, however, be more difficult to derive, and this remains to be done.

6. Acknowledgements

Paul Olovsson very efficiently helped with the programming, and Claes Cassel contributed useful comments on a previous draft. A grant from The Bank of Sweden Tercentenary Foundation is gratefully acknowledged.

7. References

Barr, R.S., Stewart, W.H. and Turner, J.S. (1982): An Empirical Evaluation of Statistical Matching Methodologies. Report prepared for the Office of Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Resources.

Brown, C. (1980): Equalizing Differences in the Labor Market. The Quarterly Journal of Economics, pp. 113–134.

Dalenius, T. (1982): A Sample of Ideas for Research and Development in the Theory and Methods of Sample Surveys. Utilitas Mathematica, Vol. 21 A, pp. 59–74.

Goldberger, A.S. (1972): Maximum-Likelihood Estimation of Regressions Containing Unobservable Independent Variables. International Economic Review, Vol. 13, No. 1, pp. 1–15.

Klevmarken, N.A. (1982): Missing Variables and Two-Stage Least-Squares Estimation from more than One Data Set. American Statistical Association, Proceedings of the Business and Economic Statistics Section, pp. 156–161.

Klevmarken, N.A. (1983a): There is a Need for Applied Micro Econometrics. Micro Econometrics, the IUI Yearbook 1982/83, The Industrial Institute for Economics and Social Research, Stockholm, Sweden.

Klevmarken, N.A. (1983b): Asymptotic Properties of a Least-Squares Estimator Using Incomplete Data. Research Report 1983:3, Department of Statistics, University of Gothenburg, Sweden.

Paass, G. (1982): Statistical Match with Additional Information. Report IPES.82.0204, Gesellschaft für Matematik und Datenverarbeitung MBH, Bonn.

Rodgers, W.L. and DeVol, E.B. (1982): An Evaluation of Statistical Matching. Report from ISR, University of Michigan.

U.S. Department of Commerce (1980): Report on Exact and Statistical Matching Techniques. Statistical Policy Working Paper 5.

Zellner, A. (1970): Estimation of Regression Relationships Containing Unobservable Variables. International Economic Review, Vol. XI, pp. 441–454.

Balancing Bias in Vital Rates Due to an Informative Sampling Plan

by Jan M. Hoem[1]

0. Abstract

Members of a target population are sometimes stratified into population segments according to status on a closing date of an individual level process which is to be investigated, or a closely related process. Population level occurrences and exposures can be estimated from a random sample using these segments as strata, by means of the Horvitz-Thompson estimator, which makes use of reciprocal inclusion probabilities to weight individual contributions from the sample. A population vital rate can then be computed as the ratio between estimated occurrences and estimated exposures. This paper demonstrates that such a vital rate is a consistent estimator of a corresponding theoretical transition intensity, provided all members of the population have a positive inclusion probability which is known up to a constant proportionality factor. When there is selective nonresponse, or when some segment(s) of the population are not sampled, trial calculations on resulting biases may be based on formulas given in this paper.

1. Introduction

Population data will sometimes be organised into separate segments according to the outcome of an individual level population process. In a population register, for instance, data for those who have died or outmigrated during a given period may be kept in separate files, and the current population may perhaps be filed by region of residence. Some analytic investigation may focus on the process by which the data are organised, perhaps mortality, or on some other process related to it, for instance fertility. Frequently, it will then be convenient to work with a separate sample from each segment of data, much in the same way as separate samples are drawn from each stratum of a sample survey. The effect will be of the same character as when the sampling plan of a retrospective interview survey, say, of a given regional population has separate strata for the resident population and for recent outmigrants.

When vital rates are computed from complete data of a total population register, ideally the rates will be calculated as the ratio between two population totals, viz., the ratio between a numerator consisting of the total count of individual occurrences of the event in question in the relevant population group, and a denominator consisting of a corresponding total count of individual exposures. It is natural to compute corresponding vital rates from sample data as ratios between estimates (or predictors) of population total occurrences and total exposures. Such estimators or predictors can be weighted sums of count of individual occurrences and exposures in the sample, with weights computed as reciprocals of individual sample inclusion probabilities, by straightforward application of procedures from sampling theory. Intuitively, this ought to give good estimates of the population totals, hence of their ratio, and, a fortiori, of the theoretical population parameter

[1] Professor of Demometry, University of Stockholm, Department of Sociology, Section for Demography.

estimated by the total population vital rate. It is the purpose of this paper to investigate the basis of these intuitive notions and to demonstrate that they are sound.

At first blush, such a demonstration may seem superfluous. On the surface, the computations involved look like algorithms which have been given any amount of attention in the (design-based) sampling survey literature and whose properties should need no further investigation in the present context. The following consideration suggests that a further check may be useful, however, if only to put to rest the suspicious conscience of statistical purists.

First, while design-based sampling theory is concerned with matters like the estimation of population totals of *nonstochastic* individual level variables, the individual occurrences and exposures which make up the population totals in the numerators and denominators of vital rates are random outcomes of individual level *stochastic processes*. The latter call for a superpopulation approach to sampling theory.

Secondly, most current sampling theory, whether superpopulationist or otherwise, assumes that the sampling plan is noninformative, i.e., essentially stochastically independent of the phenomenon investigated in the survey. The sampling plans indicated at the beginning of this section, on which we now focus, obviously are *informative,* however, in that the "strata" from which the sample is selected (by whichever method) are defined by the outcome of a relevant stochastic process.

Thirdly, well-known theory of purged processes (Hoem and Funck Jensen (1982), Section 5.2) as well as general theory of social science statistics (e.g., Blalock (1964)) stress the biases inherent in observational plans defined in terms of outcomes of the variable under investigation.

These items, taken together, should be sufficient to make any method user uneasy. It is a nice feature of the procedures suggested that they balance out potential biases and produce estimators for theoretical transition intensities which are statistically consistent under general assumptions. Our method of proof permits us to indicate when these assumptions break down and biases may occur. We also draw the connection to existing theory of purged processes and throw some new light over the latter.

2. Simple mortality estimation

To set the stage, we start by considering the relatively uncomplicated problem of estimating a force of mortality in a simple life table for a closed cohort of size N of stochastically independent males all aged z at the time of investigation. The data come from a population register, organised into a "current" file for the members of the cohort who are alive at age z and a separate "death" file for those who have died before that age. For the latter, age at death has been recorded. A simple random sample of records is drawn from each file, with a sampling fraction of f_0 from the "current" file and a fraction of f_1 from the "death" file. For individual No. i among the N original members of the cohort, we define $Y_i = 0$ if he is alive at age z, and $Y_i = 1$ otherwise. Thus, the number of members of the "death" file is $Y = \Sigma Y_i$ at the time when the sample is drawn, and the corresponding number of records actually selected for inspection is, say, $[f_1 Y]$, where $[a]$ is the integer value of the real number a. Similarly, the number of records actually sampled from the "current" file is $[f_0(N-Y)]$. For individual No. i in the original cohort, therefore, the chance of getting included in the sample is $\Pi(0)=[f_0(N-Y)]/(N-Y)$ if he is alive at age z, and it is $\Pi(1)=[f_1 Y]/Y$ if he has died before that age. (We define $\Pi(1)=1$ if $Y=0$ and $\Pi(0)=1$ if $Y=N$.) Note that $\Pi(0)$ and $\Pi(1)$ are random probabilities because they depend on Y. As $N\to\infty$, $\Pi(0)\to f_0$ and $\Pi(1)\to f_1$ in probability. (We disregard the

managable complication that the sampling fractions themselves may actually be determined on the basis of Y, for instance from cost considerations which also involve the total sample size $n=[f_0(N-Y)]+[f_1 Y]$.) We let $I_i=1$ if individual i does become a sample member, and $I_i=0$ otherwise.

Assume that the force of mortality is constant and equal to μ_x over some interval $[x,x+h>$, and that the investigator wants to estimate μ_x, presumably as part of a scheme for estimating the force of mortality for all relevant ages up to age z. (See Hoem and Funck Jensen (1982), Section 1.3, for a discussion of piecewise constant intensities.) Let D_{xi} be the number of times (0 or 1) which individual No. i has died in age interval $[x,x+h>$, and let R_{xi} be the time this individual has been alive in the given age interval. (If he died before age x, then $R_{xi}=0$; if he survives age $x+h$, then $R_{xi}=h$; and so on.)

The samples actually constitute complete counts if $f_0=f_1=1$, in which case D_{xi} and R_{xi} will be known to the investigator for all i in the original cohort. In this case, the natural estimator for the force μ_x of mortality is the usual vital rate

$$\hat{\mu}_x = D_x / R_x ,$$

for $D_x = \Sigma_i D_{xi}$ and $R_x = \Sigma_i R_{xi}$.

If the samples are not complete counts, then usual practice in sampling theory would lead one to estimate the total counts D_x and R_x by

$$\hat{D}_x = \Sigma_i I_i D_{xi} / \Pi(Y_i) = D_x^{(1)}/\Pi(1),$$

and

$$\hat{R}_x = \Sigma_i I_i R_{xi}/\Pi(Y_i) =$$

$$= R_x^{(0)}/\Pi(0) + R_x^{(1)}/\Pi(1),$$

respectively, where

$$D_x^{(k)} = \sum_{\{i:I_i=1 \& Y_i=k\}} D_{xi} \text{ and}$$

$$R_x^{(k)} = \sum_{\{i:I_i=1 \& Y_i=k\}} R_{xi}$$

are the "stratum"-specific sample subtotals, for $k=0,1$. Note that $D_x^{(0)}=0$, for $D_{xi}=0$ when $Y_i=0$. The mortality intensity μ_x is then estimated by

$$\hat{\hat{\mu}}_x = \hat{D}_x/\hat{R}_x .$$

We will now prove that $\hat{\hat{\mu}}_x$ converges to μ_x in probability as $N\to\infty$ by comparing the probability limits of $\hat{D}_x/N$ and $\hat{R}_x/N$. We will not go into asymptotic normality results for $\hat{\hat{\mu}}_x$ here.

Since both $\hat{D}_x$ and $\hat{R}_x$ are sums of identically distributed random variables, it suffices to consider the mean of each individual term. We start by noting that

$$E\{I_i D_{xi}/\Pi(Y_i)\} =$$

$$= E\{\Pi(Y_i)^{-1} E[I_i D_{xi} \mid \underline{\underline{Y}}]\}$$

$$= E\{\Pi(Y_i)^{-1} E(I_i \mid Y_i, \underline{\underline{Y}}) E(D_{xi} \mid Y_i)\},$$

for $\underline{\underline{Y}}=(Y_1,\ldots,Y_N)$. Note that I_i and D_{xi} are stochastically independent when $\underline{\underline{Y}}$ is given. Because $D_{xi}=0$ when $Y_i=0$, $E(D_{xi}|Y_i=0)=0$. Furthermore,

$$E(I_i \mid Y_i=k, \underline{\underline{Y}}) = \Pi(k), \text{ for } k=0,1,$$

and

$$E(D_{xi} \mid Y_i=1) = \int_0^h \ell_{x+t}\, \mu(x+t)dt/(1-\ell_z),$$

where $\mu(x+t)$ is the force of mortality function and ℓ_{x+t} is the usual decrement series function from life table theory, with $\ell_0=1$. Since $\mu(\cdot)$ is constant over $[x,x+h>$, we get

$$E(D_{xi} \mid Y_i=1) = \mu_x \int_0^h \ell_{x+t}\, dt/(1-\ell_z),$$

and

$$E\{I_i\, D_{xi}\,/\Pi(Y_i)\} = \mu_x \int_0^h \ell_{x+t}\, dt. \qquad (1)$$

Similarly,

$$E\{I_i R_{xi}\,/\Pi(Y_i)\} =$$
$$= E\{\Pi(Y_i)^{-1} E[I_i\, R_{xi}\,|\underline{\underline{Y}}]\};$$

$$E[I_i\, R_{xi}\,|Y_i = 0, \underline{\underline{Y}}] =$$
$$= E[I_i\,|Y_i = 0, \underline{\underline{Y}}]\, E[R_{xi}\,|Y_i = 0, \underline{\underline{Y}}] =$$
$$= \Pi(0)h;$$

and

$$E[I_i\, R_{xi}\,|Y_i = 1, \underline{\underline{Y}}] = \Pi(1)\,[\int_0^h \ell_{x+t}\,\mu(x+t)\,dt$$
$$+ h(\ell_{x+h} - \ell_z)](1-\ell_z)$$
$$= \Pi(1)\,[\int_0^h \ell_{x+t}\,dt - h\ell_z\,]\,/\,(1-\ell_z)$$

by partial integration; which makes

$$E\{I_i\, R_{xi}\,/\Pi(Y_i)\} = h\ell_z + [\int_0^h \ell_{x+t}\,dt - h\ell_z]$$
$$= \int_0^h \ell_{x+t}\,dt. \qquad (2)$$

Since the probability limit of $\hat{\hat{\mu}}_x$ is the ratio between the expressions in (1) and (2), $\hat{\hat{\mu}}_x$ converges in probability to μ_x, as suggested.

For large N, $\Pi(0)$ and $\Pi(1)$ will be practically indistinguishable from f_0 and f_1, respectively. For $f_0 = f_1$, then, $\hat{\hat{\mu}}_x$ will practically be the same as $D_x^{(1)}/(R_x^{(0)} + R_x^{(1)})$, which means that the estimator is computed *as if* the sample were the whole population.

3. General theory

We will now generalise the above reasoning to the estimation of a piecewise constant transition intensity in a general Markov chain with a finite state space and a continuous time parameter. We consider a closed cohort of stochastically independent males and think of them as organised into *closing date segments* according to the status of the individual cohort member as of a particular closing date. Let individual No. i have age z_i and status Y_i (in the state space) as of this date. Members with the same value of Y_i constitute a closing date segment. Note that we permit different ages for different individuals as of the closing date, and that "dead" may be a segment defining status.

Assume that the investigator is interested in particular events (deaths, first births, marriages) represented here as transitions from a particular state j to another particular state $k \neq j$ at ages in the interval $[x, x+h>$. Let the number of such transitions made by individual No. i be D_{xi}^{jk}, and let his total sojourn time in state j in this age interval be R_{xi}^{j}. We assume that D_{xi}^{jk}, R_{xi}^{j}, z_i, and Y_i are included in a record for this individual, possibly along with other information about him, and that $x < z_i$ for all i.

In addition to the segmentation according to the value of Y_i, we will permit the cohort to be partitioned into (real) strata according to background variables whose values are fixed before age x and cannot change during the ages investigated. Typical stratification variables for adult cohort members are sex, social status of parents, and family circumstances during adolescence. Let θ_i denote the stratum to which individual No. i belongs.

Data records are made available to the investigator for a sample of individuals from the cohorts. The sample is selected by some random mechanism which provides record No. i with probability $\pi_i(\underline{\underline{y}}, \theta_i)$ when $\underline{\underline{Y}} = (Y_1, \ldots, Y_N)$ has the value $\underline{\underline{y}} = (y_1, \ldots, y_N)$. As in Section 2, we will assume that $\pi_i(\underline{\underline{y}}, \theta_i)$ depends directly on y_i as well as possibly on the other $\{y_j\}$ for $j \neq i$. Let $I_i = 1$ if the record actually is made available, and $I_i = 0$ otherwise. Note that

$$\pi_i(\underline{\underline{y}}, \theta_i) = E(I_i\,|\,\underline{\underline{Y}} = \underline{\underline{y}}). \qquad (3)$$

We will not yet restrict ourselves to any particular specification of $\pi_i(\cdot)$, but we will mention one which is of particular interest in the presence of nonresponse in survey data. Assume that the stratum partition of the cohort is not available when the sample is selected, and that the stratum membership of any member must be ascertained from his data record, perhaps in the sample interview. The $\{\theta_i\}$ can then only be used to form post strata. By contrast, the segmentation according to the $\{Y_i\}$ *is* available and is used as a basis for drawing a stratified random sample of individuals. Assume that there is some known probability $\alpha_i(\underline{\underline{Y}})$ that individual No. i is included in the drawn sample. Typically, $\alpha_i(\underline{\underline{Y}})$ depends on the segment Y_i to which individual i belongs, as well as on the total number of individuals in that segment, much as in Section 2. Given that individual i is included in the (gross) sample, there is some probability $\beta(Y_i, \theta_i)$ that a record really is collected, $1-\beta(Y_i, \theta_i)$ being a nonresponse probability. Then

$$\pi_i(\underline{\underline{y}}, \theta_i) = \alpha_i(\underline{\underline{y}})\,\beta(y_i, \theta_i). \qquad (4)$$

If the data come from a complete population register, then presumably $\beta(Y_i, \theta_i) \equiv 1$, and $\pi_i(\underline{\underline{y}}, \theta_i)$ is known. If they are collected through an interview survey, nonresponse may be important and $\beta(Y_i, \theta_i)$ may be less than 1. If nonresponse is selective, $\beta(Y_i, \theta_i)$ really depends on Y_i and/or θ_i, and this probability may be completely or partly unknown. If nonresponse is nonselective (which rarely will be realistic), β is some constant less than 1.

The transition intensities of the Markov process, which governs the moves of an individual between the states of the system investigated, may depend on the post stratum to which he belongs, i.e., on the value of θ_i. Let us focus on a particular post stratum, and for convenience of presentation let $\theta_i = 1$ if individual i belongs to this stratum, and let $\theta_i = 0$ otherwise. (This slightly changes the meaning of θ_i.) Assume that individuals with $\theta_i = 1$ have the transition intensity $\mu_{jk}(\cdot)$ for moves from state j to state k, and let this function be constant and equal to the parameter μ_x^{jk} over the age interval $[x, x+h>$. To estimate this parameter, we use

$$\hat{\mu}_x^{jk} = \hat{D}_x^{jk} / \hat{R}_x^j,$$

where by analogy with Section 2 we let

$$\hat{D}_x^{jk} = \Sigma_i I_i \theta_i D_{xi}^{jk} w_i(\underline{\underline{Y}})$$

and

$$\hat{R}_x^j = \Sigma_i I_i \theta_i R_{xi}^j w_i(\underline{\underline{Y}}).$$

While the function of I_i in each term is to pick up terms only for individuals in the sample, the function of θ_i is to select terms only for individuals in the post stratum of interest. The factor $w_i(\underline{\underline{Y}})$ is a possibly random individual weight. If $\pi_i(\underline{\underline{Y}}, \theta_i)$ is known and positive, then $w_i(\underline{\underline{Y}}) = 1/\pi_i(\underline{\underline{Y}}, \theta_i)$. Note that we need to consider the case of $\theta_i = 1$ only, and we let $\pi_i(\underline{\underline{Y}}) = \pi_i(\underline{\underline{Y}}, 1)$. If $\pi_i(\underline{\underline{Y}})$ is positive but otherwise unknown, then $w_i(\underline{\underline{Y}})$ could be a sensible guess for its reciprocal. Under (4), $w_i(\underline{\underline{Y}}) = 1/\alpha_i(\underline{\underline{Y}})$ is a natural possibility, particularly suitable if β is (nearly) constant. (As is well-known among survey practioners, guessing nonresponse probabilities is a chancy game which may lead to biases *worse* than if β is "guessed" to equal 1 and $w_i(\underline{\underline{Y}})$ is set equal to the known $1/\alpha_i(\underline{\underline{Y}})$. See Lyberg (1983) for a study of a situation like ours). If $\pi_i(\underline{\underline{Y}}, \theta_i)$ is known to be 0, let $w_i(\underline{\underline{Y}}) = 0$. In this case, $I_i = 0$ anyway, and individual i will not contribute to $\hat{D}_x^{jk}$ and $\hat{R}_x^j$.

As is evident from (3), the sampling plan is *informative*, i.e., inclusion probabilities depend on the values of random variables connected with individual cohort members. We will assume that the sampling plan is *condi-*

tionally non-informative, however, in the sense that for given $\underline{\underline{Y}}$, I_i is stochastically independent of $(D_{xi}^{jk}, R_{xi}^{j})$.

To investigate the probability limit of $\hat{\mu}_x^{jk}$ as $N \to \infty$, we note that

$$E(I_i D_{xi}^{jk} | \underline{\underline{Y}} = \underline{y}) =$$

$$E(I_i | \underline{\underline{Y}} = \underline{y})\, E(D_{xi}^{jk} | \underline{\underline{Y}} = \underline{y}). \qquad (5)$$

The first factor on the right here appears in (3). For $\theta_i = 1$, the second factor equals

$$\int_0^h P_j(x+t)\, \mu_{jk}(x+t)\, P_{ky_i}(x+t, z_i)\, dt / P_{y_i}(z_i),$$

where $P_j(s)$ is the probability that a member of the post stratum with $\theta_i = 1$ is in state j at age s, and $P_{kl}(s,t)$ is a corresponding transition probability. Of course, y_i is the ith member of $\underline{y}$. Since $\mu_{jk}(x+t) = \mu_x^{jk}$ for $0 < t < h$, we get

$$E(D_{xi}^{jk} | \underline{\underline{Y}} = \underline{y}) =$$

$$= \mu_x^{jk}\, \gamma_x^{jk}(y_i, z_i) \,/\, P_{y_i}(z_i) \quad \text{if } \theta_i = 1,$$

where

$$\gamma_x^{jk}(y, z) = \int_0^h P_j(x+t) P_{ky}(x+t, z)\, dt. \qquad (6)$$

Combining these results and using the fact that

$$P\{\underline{\underline{Y}} = \underline{y}\} = \prod_{l=1}^{N} P_{y_l}(z_l),$$

we get

$$E\, \hat{D}_x^{jk} =$$

$$\mu_x^{jk} \sum_i \theta_i \sum_{y_i} \cdots \sum_{y_N} w_i(\underline{y}) \pi_i(\underline{y}) \gamma_x^{jk}(y_i, z_i) \prod_{l \neq i} P_{y_l}(z_l).$$

If we introduce $\underline{\underline{Y}}^{(i)} =$ $(Y_1, \ldots, Y_{i-1}, y_i, Y_{i+1}, \ldots, Y_N)$ and let

$$\zeta_x^{jk}(N) =$$

$$\sum_i \theta_i \sum_{y_i} \gamma_x^{jk}(y_i, z_i)\, E\{w_i(\underline{\underline{Y}}^{(i)}) \pi_i(\underline{\underline{Y}}^{(i)})\} / N,$$

we get

$$E\, \hat{D}_x^{jk} = \mu_x^{jk}\, N\, \zeta_x^{jk}(N).$$

By similar reasoning,

$$E(R_{xi}^{j} | \underline{\underline{Y}} = \underline{y}) =$$

$$\int_0^h P_j(x+t) P_{jy_i}(x+t, z_i) dt / P_{y_i}(z_i)$$

$$= \gamma_x^{jj}(y_i, z_i) / P_{y_i}(z_i),$$

and

$$E\, \hat{R}_x^{j} = N\, \zeta_x^{jj}(N).$$

Now assume that $\zeta_x^{jl}(N) \to$ some ζ_x^{jl} for $l = j, k$, and that $\hat{D}_x^{jk}/N$ and $\hat{R}_x^{j}/N$ converge in probability to $\mu_x^{jk}\, \zeta_x^{jk}$ and to ζ_x^{jj}, respectively, as in Section 2. Then $\hat{\mu}_x^{jk}$ converges in probability to

$$\mu_x^{jk}\, \zeta_x^{jk} / \zeta_x^{jj}, \qquad (7)$$

i.e., as an estimator of μ_x^{jk} it is subject to an asymptotic distortion factor of $\zeta_x^{jk}/\zeta_x^{jj}$.

The possibility of progressing further and getting useful assessments of this distortion factor depends on what we can assume about $E\{w_i(\underline{\underline{Y}}^{(i)})\, \pi_i(\underline{\underline{Y}}^{(i)})\}$. If (4) holds for an unknown nonconstant β, then (7) mainly has a value as a basis for trial calculations based on trial assumptions about $\beta(\cdot)$ and $w_i(\cdot)$ and knowledge of $\alpha_i(\cdot)$.

To get further general results, we must make more assumptions about $\pi_i(\cdot)$, as in the two subsequent sections here.

4. The case of known, positive inclusion probabilities

Assume first that (4) holds, that $\pi_i(\underline{y})>0$ for all i and $\underline{y}$, and that $w_i(\underline{y})=1/\alpha_i(\underline{y})$, as will frequently be the case. Then

$$\zeta_x^{jl}(N) = \sum_i \theta_i \sum_{y_i} \gamma_x^{jl}(y_i,z_i)\,\beta(y_i)/N, \tag{8}$$

where $\beta(y)=\beta(y,1)$. Evidently, selective nonresponse may lead to bias in the estimation of μ_x^{jk}. If, however, $\beta(\cdot)$ is constant, say $\beta(y) \equiv b>0$, then

$$\zeta_x^{jl}(N) = b \sum_i \theta_i \sum_{y_i} \gamma_x^{jl}(y_i,z_i)/N$$

and by (6),

$$\sum_y \gamma_x^{jl}(y,z) = \int_0^h P_j(x+t)dt \text{ for } l=k,j,$$

which means that

$$\zeta_x^{jl}(N) = b \int_0^h P_j(x+t)dt \,\Sigma\,\theta_i/N \text{ for } l=k,j. \tag{9}$$

Here, $\Sigma\,\theta_i/N$ is the proportion of individuals in the post stratum on which we focus, among the members of the target population. Assuming that this proportion converges to some positive limit, (9) shows that $\zeta_x^{jk}/\zeta_x^{jj}=1$, and the distortion factor in (7) does not really distort μ_x^{jk}.

Note that nonexistent nonresponse, as in a study of records of a complete population register, corresponds to the case of $\beta(y)\equiv 1$, i.e. $b=1$. We have proved the general result that *the estimated vital rate $\hat{\mu}_x^{jk}$ is a consistent estimator for the transition intensity μ_x^{jk} if the weights $w_i(\underline{Y})$ are chosen as reciprocal inclusion probabilities and nonresponse is nonexistent or nonselective.*

5. The case with some zero inclusion probabilities

In most cases, investigators surely will make a strong effort to sample data from all relevant population segments and make all $\pi_i(\underline{y})>0$. In analyses of records from complete population registers, presumably this is the most sensible thing to do, in which case Section 4 shows that the natural subsequent estimation procedure is safe (asymptotically, as $N\to\infty$). There exist some important situations where no data are (or can be) collected from certain segments, however. In particular, segments of individuals who are no longer members of the target population as of the target date (because of death, outmigration, and similar features) are liable to be left out, for instance in interview surveys. As is well-known from the theory of purged processes (Hoem and Funck Jensen (1982), Section 5.2), estimation biases will then tend to occur.

To see how they appear in the present context, assume that $\pi_i(\underline{y})\equiv 0$ if $y_i\in R$, $\pi_i(\underline{y})>0$ if $y_i\in K$, where R and K are disjoint subsets of the state space, which equals $R\cup K$, and where R is absorbing. Assume that (4) holds. At the cost of some slight notational complication, we could account for the possibility of nonresponse ($\beta(y_i,\theta_i)<1$). To bring out the current main argument more clearly, we disregard nonresponse, however, and let $\beta\equiv 1$.

We assume that $w_i(\underline{y})=1/\pi_i(\underline{y})$ for $y_i\in K$, $w_i(\underline{y})=0$ for $y_i\in R$. (The value of $w_i(\underline{y})$ for $y_i\in R$ is not important, for $I_i=0$ and individual i does not contribute to the numerator and denominator of $\hat{\mu}_x^{jk}$ if $y_i\in R$.)

In this case, for $l=k,j$,

$$\zeta_x^{jl}(N) = \sum_i \theta_i \sum_{y_i\in K} \gamma_x^{jl}(y_i,z_i)/N$$

$$= N^{-1}\sum_i \theta_i \int_0^h P_j(x+t)\,P_{lK}(x+t,z_i)\,dt,$$

where $P_{kK}=\sum_{y\in K} P_{ky}$. By the mean value theorem, there exists an age $z_l(N)$ such that

$$\zeta_x^{jl}(N) = \int_0^h P_j(x+t)\,P_{lK}(x+t,z_l(N))\,dt.$$

Assuming that $z_l(N) \to$ some z'_l as $N \to \infty$, $\hat{\hat{\mu}}_x^{jk}$ can be regarded as an estimator of

$$\mu_x^{jk} \frac{\int_0^h P_j(x+t)P_{kK}(x+t,z'_k)dt}{\int_0^h P_j(x+t)\,P_{jK}(x+t,z'_j)\,dt}. \qquad (10)$$

In general, the ratio in (10) will differ from 1, and $\hat{\hat{\mu}}_x^{jk}$ *will estimate a distorted value for the transition intensity* μ_x^{jk}.

If the data come from a real cohort, there will be restricted variation in $\{z_i\}$, and the distortion caused by the difference between z'_j and z'_k in (10) will be unimportant. Most of the distortion will be due to the difference between the probability measures P_{jK} and P_{kK}. As is known from the theory of purged processes, these measures generally coincide if and only if transition into R is nondifferential, i.e., if $\sum_{\lambda \in R} \mu_{l\lambda}(\cdot)$ is the same for all $l \in K$, where $\mu_{l\lambda}(\cdot)$ is the intensity for transitions from state l to state λ.

We conclude that $\hat{\hat{\mu}}_x^{jk}$ *may be largely consistent for* μ_x^{jk} *if the variation in* $\{z_i\}$ *is small and if transition into R is largely nondifferential. Otherwise, there is risk of distortion.*

6. Acknowledgement

I am grateful to Professor Robert Erikson for a conversation which made me realise the need to look into the matters of this paper.

7. References

Blalock, H. M. (1964): Causal Inferences in Nonexperimental Research. Chapel Hill: University of North Carolina Press.

Hoem, J. M. and Funck Jensen, U. (1982): Multistate Life Table Methodology: A Probabilist Critique. Pp. 155–264 in Multidimensional Mathematical Demography, edited by Land, K.C. and Rogers, A. Academic Press.

Lyberg, I. (1983): Nonresponse Effects on Survey Estimates in the Analysis of Competing Exponential Risks. Stockholm Research Reports in Demography, No. 12, Department of Statistics, University of Stockholm.

On the Practical Use of Subsampling Among Nonrespondents

by Ingrid Lyberg[1]

As a student of Tore Dalenius I have become interested in nonsampling errors and the problem of allocation of resources in respect to both sampling and nonsampling errors. In this paper I have tried to illustrate the application of sampling theory and methods in practice (which differs from theory), as Tore did in his well-known and still relevant Ph.D. thesis (Dalenius (1957)).

1. Introduction

Most surveys suffer from nonresponse, and an unknown bias is involved in any assumption that the respondents are representative of the combined total of respondents and nonrespondents. Imputation, reweighting, poststratification or even more sophisticated techniques are often used to adjust for this unknown nonresponse bias. However, the only sure way to reduce the nonresponse bias is to reduce the nonresponse rate. This is often possible by using more effective data collection procedures. It is well-known that data collection by personal interviews generally results in a higher response rate than mail questionnaires and that the response rate increases by the number of call-backs or reminders.

Usually we have to pay for a higher response rate by a higher cost per unit selected. This means that we have either to decrease the sample size and accept a lower precision at a fixed budget, or accept a higher total cost to achieve the desired precision.

By means of the technique of subsampling among the nonrespondents we can combine the advantages of two collection procedures; the economies of a low-cost procedure and the high response rate involved in a more expensive one. The technique was originally mentioned by Cornfield (1942), and later theoretically confirmed and developed by Hansen and Hurwitz (1946).

The principle adopted is to make a first attempt to collect data with a low-cost method from an initial sample of units and then make a second attempt using a more expensive method among a random subsample of those who did not respond in the first attempt. It is then possible to obtain unbiased estimates of population means (or other population quantities), provided all units in the subsample respond. Hansen and Hurwitz showed how the optimal initial sample size and subsampling fraction could be determined if the (expected) response rate (first attempt) and the variances in the two response "strata" (see below) were known.

Hansen and Hurwitz assumed that the second attempt was to "elicit a substantially complete response", and their results are based on this assumption, which is rather unrealistic in most survey situations.

At Statistics Sweden we use this technique more and more in our surveys. The main data collection method is the mail questionnaire followed-up by telephone interviews. When designing these surveys we take into account the occurrence of some nonresponse at the follow-up. We have used the technique in connection with stratified sampling and one-stage cluster sampling based on the theory de-

[1] Senior statistician, Statistics Sweden.

scribed in Swensson (1983) and Lyberg (1979), respectively.

This paper deals with our experiences at Statistics Sweden of subsampling among nonrespondents. In Section 2 we describe the theoretical framework, which involves the assumption of response behavior strata. In Sections 3 and 4 we show how the sample should be allocated between the first and the second attempt in order to obtain an effective design with respect to variances and costs. We also illustrate how much we lose in cost efficiency if our allocation differs from the optimal one. In fact, the loss is usually small for moderate deviations from the optimal allocation. In Section 5 we show the advantages of using a combination of two data collection methods as compared with using one method only. In Section 6 we propose a model to estimate the bias due to nonresponse in the second attempt. Finally, we discuss some problems inherent in combining two data collection methods on a subsampling basis.

2. General assumptions and notations

Suppose we have two collection methods, denoted α and β, where the latter can (and perhaps can only) be used to follow-up the former. For instance, method β may be callbacks or reminders to method α, or β may be personal interviews if method α is a mail questionnaire.

By simple random sampling (SRS) we select an initial sample of $n = fN$ units, and $n_\alpha = w_\alpha n$ among these units respond to method α. Among the $n_{\bar{\alpha}} = n - n_\alpha$ initial nonrespondents we select by SRS a subsample of $r = f_2 n_{\bar{\alpha}}$ units, and $r_{\bar{\alpha}\beta} = w'_{\bar{\alpha}\beta} r$ among these respond to method β. Then, for instance, the population mean, $\bar{Y}$, is estimated by:

$$\hat{\bar{Y}} = w_\alpha \bar{y}_\alpha + w_{\bar{\alpha}} \bar{y}'_{\bar{\alpha}\beta}, \qquad (2.1)$$

where $w_\alpha = 1 - w_{\bar{\alpha}}$ is the response rate obtained in the first attempt, and $\bar{y}_\alpha$ and $\bar{y}'_{\bar{\alpha}\beta}$ are the means for the respondents in the initial sample and the subsample, respectively.

Suppose that the data collection cost per selected unit is C_α and C_β for methods α and β, respectively, and that the processing cost per respondent unit is C_p. The total expected cost for data collection and processing then becomes:

$$C = n\{C_1 + f_2 W_{\bar{\alpha}} C_2\}, \qquad (2.2)$$

where

$$C_1 = C_\alpha + W_\alpha C_p \quad \text{and}$$

$$C_2 = C_\beta + W_{\bar{\alpha}\beta} C_p.$$

$W_\alpha = 1 - W_{\bar{\alpha}}$ and $W_{\bar{\alpha}\beta}$ are the *expected* response rates in the first and second attempt, respectively, n is the initial sample size, and f_2 is the subsampling fraction.

In accordance with Cochran (1977, pp. 370–374), we assume a non-probabilistic response behavior and consider the population divided into "strata" defined by whether or not a unit would respond (in theory), given a specific collection procedure. The definitions and notations of these theoretical strata are illustrated in Diagram 1.

The assumption of response behavior strata makes it possible to express the variance and bias of the estimator $\hat{\bar{Y}}$ (2.1) in terms of the variances and means of those strata. Provided the samples are large enough to permit Taylor expansions, Swensson (1983) shows that the variance is approximately:

$$V = V\{\hat{\bar{Y}}\} \approx \frac{1}{n}\{V_1 + V_2/f_2\} - \frac{1}{N}\{V_1 + V_2\}, \qquad (2.3)$$

where

$$V_1 = W_\alpha \{S_\alpha^2 + W_{\bar{\alpha}}(\bar{Y}_\alpha - \bar{Y}_{\bar{\alpha}\beta})^2\},$$

$$V_2 = \frac{W_{\bar{\alpha}}}{W_{\bar{\alpha}\beta}} S_{\bar{\alpha}\beta}^2,$$

<table>
<tr><td rowspan="2">Would respond to method α</td><td colspan="2">Would not respond to method α</td></tr>
<tr><td>Would respond to method β</td><td>Would not respond to method β</td></tr>
<tr><td rowspan="2">α = initial response stratum

$N_\alpha = W_\alpha N$</td><td colspan="2">$\bar{\alpha}$ = initial nonresponse stratum

$N_{\bar{\alpha}} = W_{\bar{\alpha}} N$</td></tr>
<tr><td>$\bar{\alpha}\beta$ = second attempt response stratum

$N_{\bar{\alpha}\beta} = W_{\bar{\alpha}\beta} N_{\bar{\alpha}}$</td><td>$\bar{\alpha}\bar{\beta} = \bar{\gamma}$ = final nonresponse stratum

$N_{\bar{\alpha}\bar{\beta}} = W_{\bar{\alpha}\bar{\beta}} N_{\bar{\alpha}}$</td></tr>
<tr><td colspan="2">γ = final response stratum

$N_\gamma = W_\gamma N = (W_\alpha + W_{\bar{\alpha}} W_{\bar{\alpha}\beta}) N$</td><td>$= N_{\bar{\gamma}} = W_{\bar{\gamma}} N$</td></tr>
</table>

Diagram 1. Number of units in theoretical response behavior strata

and where S_α^2, $S_{\bar{\alpha}\beta}^2$, $\bar{Y}_\alpha$ and $\bar{Y}_{\bar{\alpha}\beta}$ are the variances and means of the initial and second attempt response stratum, respectively.

The bias of the estimator is approximately

$$B = E\{\hat{\bar{Y}} - \bar{Y}\} \approx W_{\bar{\gamma}} \{\bar{Y}_{\bar{\alpha}\beta} - \bar{Y}_{\bar{\alpha}\bar{\beta}}\},$$

where $W_{\bar{\gamma}} = 1 - W_\gamma = W_{\bar{\alpha}} W_{\bar{\alpha}\bar{\beta}}$ is the relative size of the final nonresponse stratum, and $\bar{Y}_{\bar{\alpha}\beta}$ and $\bar{Y}_{\bar{\alpha}\bar{\beta}}$ are the means for the respondents and nonrespondents, respectively, within the initial nonresponse stratum (see Diagram 1).

3. Sample allocation

Let us express the cost efficiency, CEFF, of the design described above by the product of the expected cost, C in (2.2), and variance, V in (2.3) (omitting the last term, which represents the finite population correction (fpc)). We want to allocate our sample in such a way as to make CEFF$=V\cdot C$ as small as possible.

By means of Cauchy's inequality we find the minimum of CEFF:

$$\text{CEFF} = V \cdot C \geq \{\sqrt{V_1 C_1} + \sqrt{V_2 W_{\bar{\alpha}} C_2}\}^2 =$$

$$= \tilde{S}_\alpha^2 Z^2,$$

where the minimum is obtained for

$$f_2 = f_{2OPT} = \sqrt{\frac{V_2}{V_1} \frac{C_1}{W_{\bar{\alpha}} C_2}} \tag{3.1}$$

$$= \sqrt{Q} \sqrt{\frac{C_1}{W_\alpha W_{\bar{\alpha}\beta} C_2}}, \tag{3.2}$$

provided $f_{2OPT} \leq 1$, and where

$$Q = S_{\bar{\alpha}\beta}^2 / \tilde{S}_\alpha^2, \tag{3.3}$$

$$\tilde{S}_\alpha^2 = S_\alpha^2 + W_{\bar{\alpha}} (\bar{Y}_\alpha - \bar{Y}_{\bar{\alpha}\beta})^2, \tag{3.4}$$

and

$$Z = W_\alpha \sqrt{\frac{C_1}{W_\alpha}} + W_{\bar{\alpha}} \sqrt{\frac{C_2}{W_{\bar{\alpha}\beta}}} \sqrt{Q}. \tag{3.5}$$

With $W_{\bar{\alpha}\beta}=1$, i.e., complete response in the second attempt, (3.1) is equal to the corresponding expression given by Hansen and Hurwitz.

With $f_2 = f_{2OPT}$ the expected subsample size, cost, and variance become:

$$E(r_{OPT}) = n\, W_{\bar{\alpha}} f_{2OPT},$$

$$C_{OPT} = n \sqrt{\frac{C_1}{W_\alpha}}\, Z, \text{ and}$$

$$V_{OPT} = \frac{1}{n} \sqrt{\frac{W_\alpha}{C_1}}\, Z\, \tilde{S}_\alpha^2, \text{ respectively.}$$

The initial sample size, n, may be chosen either to minimize V for a specified $C = C^*$, or to minimize C for a specified $V = V^*$, or to obtain a specified expected subsample size. The latter choice may be necessary if the administrative resources for the follow-up are limited, or if time is lacking.

Q in (3.2) reflects the relation between the variances in the response strata α and $\bar{\alpha}\beta$. In practice it is usually difficult to predict Q; we may also want to account for several variables with different variance relations when allocating our sample. It may then be convenient to assume that $Q \approx 1$ to obtain a "proximal" allocation:

$$f_{2PROX} = \sqrt{\frac{C_1}{W_\alpha W_{\bar{\alpha}\beta} C_2}}. \tag{3.6}$$

If this is done the expected cost and variance become:

$$C_{PROX} = n \sqrt{\frac{C_1}{W_\alpha}} \left\{ W_\alpha \sqrt{\frac{C_1}{W_\alpha}} + W_{\bar{\alpha}} \sqrt{\frac{C_2}{W_{\bar{\alpha}\beta}}} \right\}$$

and

$$V_{PROX} = \frac{1}{n} \sqrt{\frac{W_\alpha}{C_1}} \left\{ W_\alpha \sqrt{\frac{C_1}{W_\alpha}} + W_{\bar{\alpha}} \sqrt{\frac{C_2}{W_{\bar{\alpha}\beta}}}\; Q \right\} \tilde{S}_{\bar{\alpha}}^2,$$

respectively.

However, we might expect the variance for the second attempt respondents, $S_{\bar{\alpha}\beta}^2$, to be less than the variance for the initial respondents, S_α^2, and the second term in (3.4) to be small compared with the first. In this case $Q < 1$ and

$$\tilde{S}_\alpha^2 Z^2 < V_{PROX} C_{PROX} < \tilde{S}_\alpha^2 \tilde{Z}^2, \tag{3.7}$$

where $\tilde{Z}$ is identical to Z in (3.5) for $Q = 1$. Thus, using $\tilde{S}_\alpha^2 \tilde{Z}^2$ to predict the efficiency of the proximal allocation, as we shall do from now on, will probably be too pessimistic.

In practice it may also be difficult to guess the response rates (and even the cost components) in advance. However, when the first attempt has been accomplished we know the initial response rates. We can then choose the subsampling fraction to obtain a specified subsample size, cost or variance. In the last situation we choose the subsampling fraction according to the variance V' conditional on the known values of $w_\alpha = 1 - w_{\bar{\alpha}}$ and n. Given $n_\alpha = w_\alpha\, n > 0$, $n_{\bar{\alpha}} = w_{\bar{\alpha}} n > 0$ and $r = f_2 n_{\bar{\alpha}}$, the sample design may be regarded as a stratified sample with a selection by SRS of n_α and r units in stratum α and $\bar{\alpha}$, respectively. The conditional variance then approximately becomes:

$$V' = V\{\hat{\bar{Y}} | w_\alpha\}$$

$$= w_\alpha^2 \frac{1-f}{n_\alpha} S_\alpha^2 + w_{\bar{\alpha}}^2 \frac{1-ff_2}{f_2 n_{\bar{\alpha}} W_{\bar{\alpha}\beta}} S_{\bar{\alpha}\beta}^2$$

$$= \frac{1}{n}\{V_1' + \frac{1}{f} V_2'\} - \frac{1}{N}\{V_1' + V_2'\},$$

where

Table 1. Optimal allocation and cost efficiency rates, $VC/\{VC\}_{min}$, for different subsampling fractions assuming equal variance in the response strata

Cost relations		Expected response rates. Per cent			Optimal allocation		Cost efficiency rates $f_2 =$			
		Initial	Second	Final						
$\frac{C_\beta}{C_\alpha}$	$\frac{C_p}{C_\alpha}$	W_α	$W_{\bar{\alpha}\beta}$	W_γ	$\tilde{Z}^2/C_\alpha$	f_{2OPT}	.25	.50	.75	1.00
5	1	50	60	80	5.7	.94	1.47	1.10	1.01	1.00
		50	100	100	4.4	.71	1.29	1.03	1.00	1.03
		80	25	85	4.6	1.31[1]	1.81	1.22	1.06	1.00
		80	75	95	3.1	.72	1.27	1.03	1.00	1.02
		80	100	100	2.9	.61	1.18	1.01	1.01	1.05
20	2	50	60	80	15.8	.56	1.13	1.00	1.02	1.06
		50	100	100	11.2	.43	1.06	1.01	1.07	1.16
		80	25	85	10.6	.80	1.37	1.05	1.00	1.01
		80	75	95	6.3	.45	1.09	1.00	1.07	1.17
		80	100	100	5.7	.38	1.04	1.02	1.11	1.24
40	6	50	60	80	32.2	.55	1.12	1.00	1.02	1.07
		50	100	100	23.1	.42	1.06	1.01	1.07	1.17
		80	25	85	22.4	.84	1.41	1.07	1.00	1.01
		80	75	95	13.7	.47	1.10	1.00	1.06	1.15
		80	100	100	12.3	.40	1.05	1.01	1.10	1.22
40	1	50	60	80	24.8	.35	1.02	1.02	1.09	1.17
		50	100	100	16.5	.27	1.00	1.07	1.19	1.33
		80	25	85	14.0	.47	1.09	1.00	1.05	1.13
		80	75	95	7.2	.27	1.00	1.10	1.28	1.48
		80	100	100	6.2	.23	1.00	1.15	1.38	1.63

[1] In this case we choose $f_{2OPT} = 1$.

$$V_1' = w_\alpha S_\alpha^2 \text{ and } V_2' = \frac{w_{\bar{\alpha}}}{W_{\bar{\alpha}\beta}} S_{\bar{\alpha}\beta}^2.$$

Of course, the possibility to choose f_2 when the first attempt has been completed does not change our choice of n and the efficiency of the design. Table 1 shows how f_{2OPT} and $\{VC\}_{\min}$ depend on the cost components and response rates, and the loss of efficiency when f_2 differs from f_{2OPT}. The tabulated values of cost components and response rates may reflect the situation where α is a mail questionnaire and β is a telephone interview or a face-to-face interview. According to our experience at Statistics Sweden, the final (weighted) response rate, $w_\gamma = w_\alpha + w_{\bar{\alpha}} w_{\bar{\alpha}\beta}$, usually becomes 85–95 % (see Fig. 1 and 2).

Table 1, as expected, shows that f_{2OPT} is small when method β is expensive compared with method α and vice versa. As Durbin (1954) has pointed out, the loss of efficiency by choosing $f_2 = 1$ is small, unless method β is much more expensive than method α. (However, even with $f_2 = 1$ we should use the weighted estimator in (2.1) to adjust for nonresponse bias.)

In practice, we may be tempted to choose $f_2 < f_{2OPT}$ and $n > n_{OPT}$ to get a larger number of respondents. However, this may cause a substantial decrease of the efficiency if the second attempt response rate, $W_{\bar{\alpha}\beta}$, is small.

If we are not sure of our prior knowledge about the expected response rates, $f_2 = .5$ seems to be a good choice.

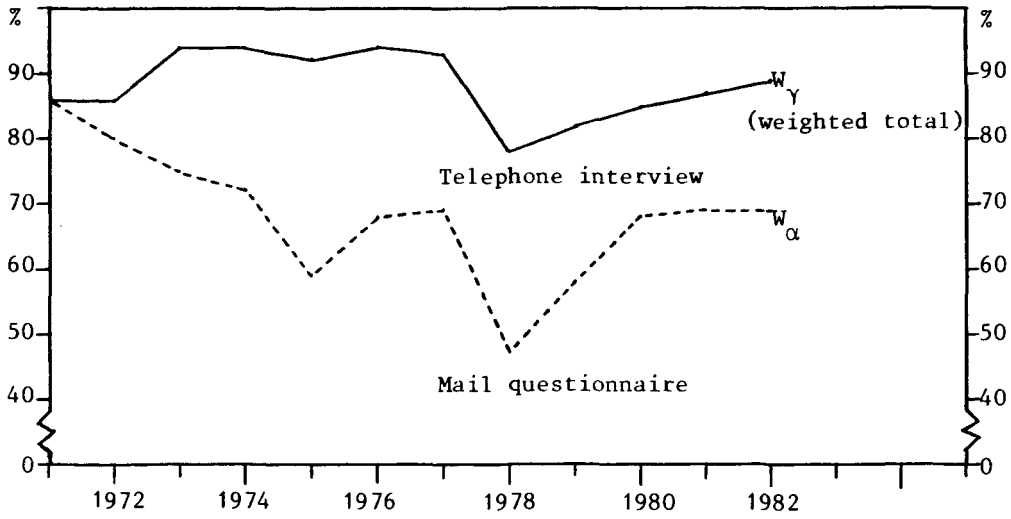


Fig. 1. Response rates for pupils who have left the 9-year comprehensive school and have not been admitted to the integrated upper secondary school. (One year after leaving, except in 1978 and 1982, which years refer to pupils leaving the comprehensive school in 1971 and 1980, respectively.)

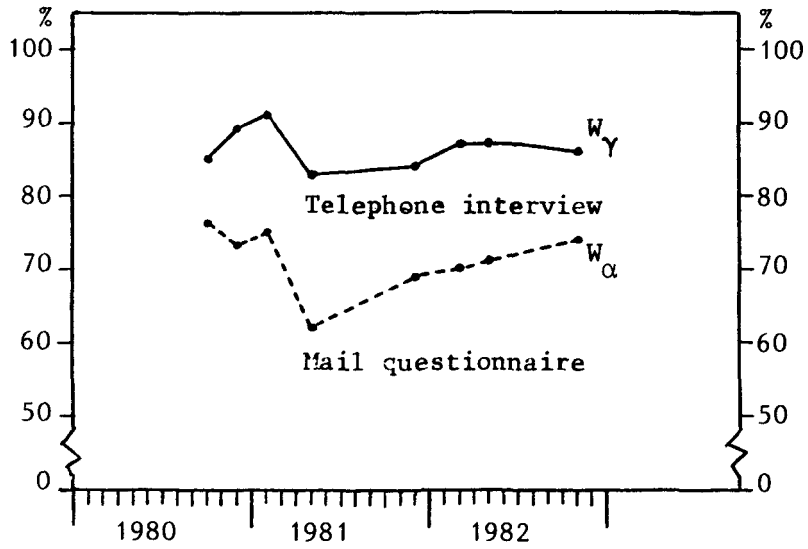


Fig. 2. Response rates in the Omnibus surveys at Statistics Sweden

Table 2. Response rates and optimal subsampling fractions (assuming equal variance) in the 1980 Survey of Pupils Leaving the 9-Year Comprehensive School in 1979. ($C_{h\beta}/C_{h\alpha} = 20$, $C_{hp}/C_{h\alpha} = 2$)

Population group	Admitted to integrated upper secondary school	Response rate Per cent			Optimal subsampling fraction f_{2PROX}
		Mail W_α	Telephone $W_{\bar{\alpha}\beta}$	Final W_γ	
All pupils		81	60	92	.50
	no	68	54	85	.55
	yes	86	64	95	.48
Immigrants		72	68	91	.48
	no	57	61	83	.54
	yes	78	74	94	.45

4. Subsampling among the nonrespondents in connection with stratified sampling

At Statistics Sweden we often use subsampling among the nonrespondents in connection with stratified sampling (with SRS within strata). The population mean is then estimated by

$$\hat{\bar{Y}} = \sum_h W_h \hat{\bar{Y}}_h,$$

where $W_h = N_h/N$ is the relative size of stratum h, and $\hat{\bar{Y}}_h$ is the weighted estimate of the stratum mean given by (2.1). According to the results in Section 2, the cost and variance become

$$C = \sum C_h \quad \text{and}$$

$$V = \sum W_h^2 V_h, \text{ respectively,}$$

where C_h and V_h are given by (2.2) and (2.3) (with the index "h" inserted to denote a specific stratum).

In order to obtain an efficient allocation of our resources between the two collection methods as well as between strata, we must account for the differences between the strata as regards expected response rates, variances, and costs per unit. Table 2 shows how the response rates may differ between population groups.

By means of Cauchy's inequality, Swensson (1983) shows how the sample should be allocated to obtain the minimum of VC. The optimal subsampling fraction within the strata is given by (3.1), and the optimal initial sample size by:

$$n_h = n_{hOPT} = kW_h \sqrt{\frac{V_{h1} + V_{h2}/f_{h2}}{C_{h1} + W_{h\bar{\alpha}} f_{h2} C_{h2}}},$$

which for f_2 given by (3.1) becomes:

$$n_h = kW_h \sqrt{V_{h1}/C_{h1}} \qquad (4.1)$$

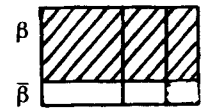
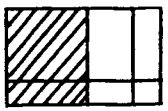
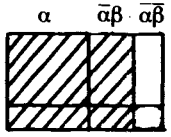
$$= kW_h \sqrt{W_{h\alpha} \tilde{S}^2_{h\alpha}/C_{h1}},$$

where k is a constant which may be chosen either to obtain a specified expected cost, subsample size, or variance.

Note that the solution given by (4.1) is analogous to the well-known solution, $n_h \propto W_h \sqrt{S_h^2/C_h}$, for the one-phase sampling situation (see e.g. Cochran (1977) pp. 96–98). Thus, in practice we can determine the initial sample sizes on the basis of $V_{h1} = W_{h\alpha} \tilde{S}^2_{h\alpha}$ and C_{h1} only.

If the strata coincide with the domains of study, we may wish to ascertain that the variances of the estimated stratum means are equal, i.e., $V_h = \bar{V} =$ constant. Even in this situa-

Diagram 2. Response stratum, estimator, nonresponse bias, variance and expected cost for three data collection procedures, assuming equal variance in response strata and optimal allocation for the combined procedure

	Data collection procedure		
	Method β only	Method α only	Method α followed-up by method β
Response stratum			
Estimator $\hat{\bar{Y}}$	$\bar{y}_\beta$	$\bar{y}_\alpha$	$\bar{y}_\gamma = w_\alpha \bar{y}_\alpha + w_{\bar{\alpha}} \bar{y}'_{\bar{\alpha}\beta}$
Bias B	$W_{\bar{\beta}}(\bar{Y}_\beta - \bar{Y}_{\bar{\beta}})$	$W_{\bar{\alpha}}(\bar{Y}_\alpha - \bar{Y}_{\bar{\alpha}})$	$W_{\bar{\gamma}}(\bar{Y}_{\bar{\alpha}\beta} - \bar{Y}_{\bar{\alpha}\bar{\beta}})$
Variance V (omitting *fpc*)	S_β^2/nW_β	S_α^2/nW_α	$S_\gamma^2\{W_\alpha + \frac{W_{\bar{\alpha}}}{W_{\bar{\alpha}\beta}} \frac{1}{f_2}\}$
Cost C	nC'_2 $C'_2 = C_\beta + W_\beta C_p$	nC_1 $C_1 = C_\alpha + W_\alpha C_p$	$n\{C_1 + f_2 W_{\bar{\alpha}} C_2\}$ $C_2 = C_\beta + W_{\bar{\alpha}\beta} C_p$
CEFF = VC (assuming optimal allocation)	$S_\beta^2 C'_2/W_\beta$	$S_\alpha^2 C_1/W_\alpha$	$S_\gamma^2 \{\sqrt{C_1 W_\alpha} + \sqrt{\frac{W_{\bar{\alpha}}^2}{W_{\bar{\alpha}\beta}} C_2}\}^2$

tion it is advisible to choose subsampling fractions according to (3.1), since such a choice will minimize the within-stratum product of cost and variance. In order to obtain equal variances, the initial sample sizes should then be:

$$n_h = \frac{1}{\bar{V}} \sqrt{\frac{V_{h1}}{C_{h1}}}\, Z_h,$$

where $\bar{V}$ is the constant variance and Z_h is defined by (3.5), with "h" inserted. $\bar{V}$ may be chosen in advance to achieve a specified total variance, expected total cost, or total subsample size, which become:

$$V = \bar{V} \Sigma W_h^2,$$

$$C = \Sigma Z_h^2/\bar{V} \text{ and}$$

$$E(r) = \frac{1}{\bar{V}} \Sigma Z_h \sqrt{\frac{V_{h2} W_{h\bar{\alpha}}}{C_{h2}}},$$

respectively.

Comment. We have omitted the fpc as represented by the last term in (2.3). This, however, will not substantially change the solutions with respect to allocation, as long as the fpc does not vary too much between strata.

5. Combining two data collection methods vs using one method only

By combining the two data collection methods, α and β, we may derive the advantages of both. By our using the low-cost method α the design becomes effective with respect to cost and variance, and by applying method β we achieve more accurate results as

Table 3. Efficiency rates, $VC/VC_{\gamma min}$, for data collection method α only, and method β only, and the combination of α and β on a subsample basis, γ. (Assuming equal variance in response strata)

Cost relations		Expected response rates Per cent				Efficiency, $VC/VC_{\gamma min}$				f_{2OPT}
		Method β	Methods α and γ			Method β	Method α	Method γ		
			Initial	Second	Final			$f_2=.5$	$f_2=1$	
$\frac{C_\beta}{C_\alpha}$	$\frac{C_p}{C_\alpha}$	W_β	W_α	$W_{\overline{\alpha\beta}}$	W_γ					
5	1	80	50	60	80	1.27	.52	1.10	1.00	.95
		90	50	80	90	1.34	.61	1.06	1.01	.80
		100	50	100	100	1.37	.69	1.03	1.03	.71
		85	80	25	85	1.51	.49	1.22	1.00	1.31[1]
		95	80	75	95	2.04	.73	1.03	1.02	.72
		100	80	100	100	2.10	.79	1.10	1.05	.61
20	2	80	50	60	80	1.71	.25	1.00	1.07	.56
		90	50	80	90	1.87	.31	1.00	1.11	.48
		100	50	100	100	1.97	.36	1.01	1.16	.43
		85	80	25	85	2.41	.31	1.05	1.00	.80
		95	80	75	95	3.65	.52	1.00	1.17	.45
		100	80	100	100	3.88	.57	1.02	1.24	.38

[1] In this case we choose $f_{2OPT}=1$.

regards nonresponse bias. However, using method β results in increased costs, and using method α before method β may cause a reduction of the response stratum, as compared with using method β only. The latter may be true if there are units that would respond to β only if they were not approached by means of method α first.

Thus, using one method only may be better than combining them in the subsample technique. Diagram 2 shows the nonresponse bias, variance and cost for one collection method (α and β, respectively) and for the subsampling technique (γ).

Obviously, the least expensive collection procedure is the best one if we regard cost and variance only. Table 3 shows how much the product $\{VC\}$ increases if we use method β only and decreases if we use method α only, as compared with a combination of the methods. The cost to achieve a specified variance may be doubled by using method β only and reduced to half by using method α only, as compared with combining them.

However, we must also consider the nonresponse bias when comparing the data collection procedures. Let us express the total error of an estimate by the mean square error (MSE):

$$\mathrm{MSE}(\hat{\bar{Y}}) = V\{\hat{\bar{Y}}\} + E\{\hat{\bar{Y}} - \bar{Y}\}^2$$

$$= \frac{\{VC\}}{C^*} + B^2,$$

where $\{VC\}$ is the product of variance and cost, C^* is a specified cost, and B is the bias. In a low-budget survey, the bias component may be small compared with the variance, and we should use a low-cost method (or choose to cancel the survey). In a large-budget survey the bias component may dominate, and it might be worthwhile to use a more expensive method aiming at decreasing the bias (due to

Table 4. Accuracy rates, $\sqrt{MSE/MSE_{\gamma min}}$, for data collection method α and method β only, assuming equal variances (S^2) in response strata and optimal allocation using the combined procedure (γ)

Cost relations $\frac{C_\beta}{C_\alpha}$ $\frac{C_p}{C_\alpha}$	Bias rates B/S Per cent Method α	β	γ	Expected response rates Per cent W_β	W_α	$W_{\bar{\alpha}\beta}$	W_γ	Budget $C^*/C_1 = n_\alpha$	$\sqrt{MSE/MSE_{\gamma min}}$ Method α	β
5 1	1	.5	.1	90	80	50	90	500	.8	1.4
								10 000	1.1	1.4
	8	6	.6	85	75	60	88	500	1.4	1.6
								10 000	4.9	3.8
	50	5	5	95	50	80	90	500	5.3	1.1
								10 000	9.4	1.0
20 1	1	.5	.1	90	80	50	90	500	.7	1.8
								10 000	.9	1.8
	8	6	.6	85	75	60	88	500	1.1	1.8
								10 000	3.9	3.3
	50	5	5	95	50	80	90	500	4.1	1.3
								10 000	8.9	1.1

nonresponse and measurement errors). This is illustrated by Table 4.

In Table 4 the bias rate, B/S, ranges from 1 to 50 per cent for method α, from .5 to 6 per cent for method β, and from .1 to 5 per cent for the combined procedure γ. In practice, we almost never know the size of the bias. However, according to the results in Table 4, the combined procedure γ seems superior to using one data collection method only.

In order to compare different data collection procedures, an experiment was carried out at Statistics Sweden in connection with the 1977 Survey of Pupils Leaving the 9-Year Comprehensive School in 1976. For a sample of about 3 600 pupils we used the regular data collection procedure (γ), i.e. a mail questionnaire (α) followed-up by telephone interviews (β) for a subsample of the initial nonrespondents.

For another sample of about 410 pupils we tried to collect data by telephone interviews (β) only. Thus, we were able to estimate the size of the response strata obtained by using method α or β only, or by combining the two.

By means of register data, accessible for respondents as well as nonrespondents, it was also possible to estimate the means of the different response behavior strata. The results are shown in Table 5.

The results in Table 5 indicate that we may increase the size of the (final) response stratum considerably by combining two data collection methods. (The comparatively small response rate for method β is explained by difficulties to get in touch with the pupils within the stipulated time of a week or two.)

The combined data collection technique reduces the nonresponse bias in two ways. Firstly, the size of the (final) nonresponse stratum decreases. Secondly, the bias is reduced by using reweighting. The reweighted estimate becomes $p'_\gamma = .448$, which is closer to the

Table 5. Results obtained by three different data collection procedures in the 1977 Survey of Pupils Leaving the 9-Year Comprehensive School in 1976 but Not Admitted to Upper Secondary School. Parameter estimated by register data: p = the proportion of pupils with a special course in mathematics or English

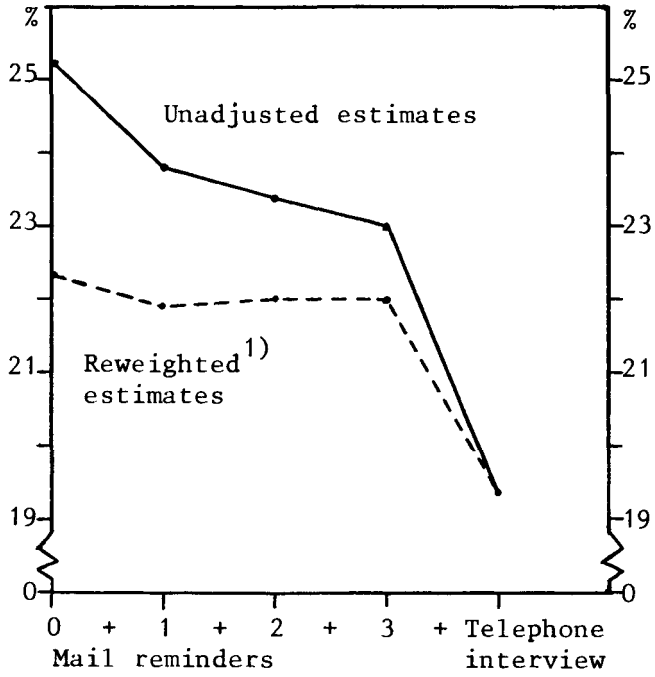
Mail questionnaire (α) followed-up by telephone interview (β)			Telephone interview (β) only[1]	
Initial respondents	Initial nonrespondents		Respondents	Nonrespondents
	Second attempt respondents	Final nonrespondents		
α	$\bar{\alpha}\beta$	$\alpha\bar{\beta}$	β	$\bar{\beta}$
$w = .681$ $p = .486$	$w = .319$ $p = .359$			
	$w = .783$ $p = .367$	$w = .217$ $p = .330$	$w = .798$ $p = .477$	$w = .202$ $p = .317$
γ		$\bar{\gamma}$		
$w = .931$ $p = .454$		$w = .069$ $p = .330$		
$w = 1$ $p = .445$				

[1] The results have been adjusted for sampling errors (without changing the nonresponse bias) to obtain the same value of p for the two samples.

Table 6. Nonresponse bias, standard error and total error for estimates of p (as in Table 5) based on register data for those responding by mail (α), telephone interview (β), or a combination of the two (γ). Per cent

	Data collection method		
	Mail questionnaire α	Telephone interview β	Combined $f_2 = .5$ γ
p	44.5	44.5	44.5
$\hat{p}$	48.6	47.7	44.8
Bias	+4.1	+3.2	+ .3
Standard error			
$C^*/C_1 = n_\alpha =$ 500	2.7	7.6	4.2
5 000	.9	2.4	1.3
10 000	.6	1.7	.9
Accuracy, $\sqrt{\text{MSE}}$			
$C^*/C_1 = n_\alpha =$ 500	4.9	8.2	4.2
5 000	4.2	4.0	1.4
10 000	4.1	3.6	1.0

$C_\alpha = 5$, $C_\beta = 100$, $C_p = 10$,
$W_\alpha = .68$, $W_{\bar{\alpha}\beta} = .78$, $W_\gamma = .93$, $W_\beta = .80$



[1] Reweighted by sex and type of course in English and mathematics.

Fig. 3. Estimates of number of pupils still studying (one year after leaving the 9-year comprehensive school) based on responses obtained before reminders, after each of the three reminders and the telephone interview, respectively, in the 1977 Survey of Pupils Leaving the 9-Year Comprehensive School in 1976

"true" proportion, $p = .445$, than the proportion in the final response stratum, $p_\gamma = .454$. (The bias may be reduced even more, see Section 6.)

Table 6 shows the standard error and mean square error that would have been obtained by using method α, β, and γ, respectively . If, with method α, we could afford a sample of only 500 pupils, the standard error would become large compared with the nonresponse bias. Nevertheless, the combined procedure, γ, would generate a smaller mean square error. With a larger budget, the combined procedure will become even more favorable. Using method β only should never be recommended in a survey like this, since it involves high costs and a comparatively large nonresponse bias.

Unfortunately we have no estimates of nonresponse bias for real survey variables. Fig. 3 shows, however, that the follow-up by telephone interviews changes the estimate of "number of pupils who study one year after they have left the 9-year comprehensive school". Furthermore, reweighting by means of sex and type of course in English and mathematics does not adjust the estimate to the same degree as does the follow-up.

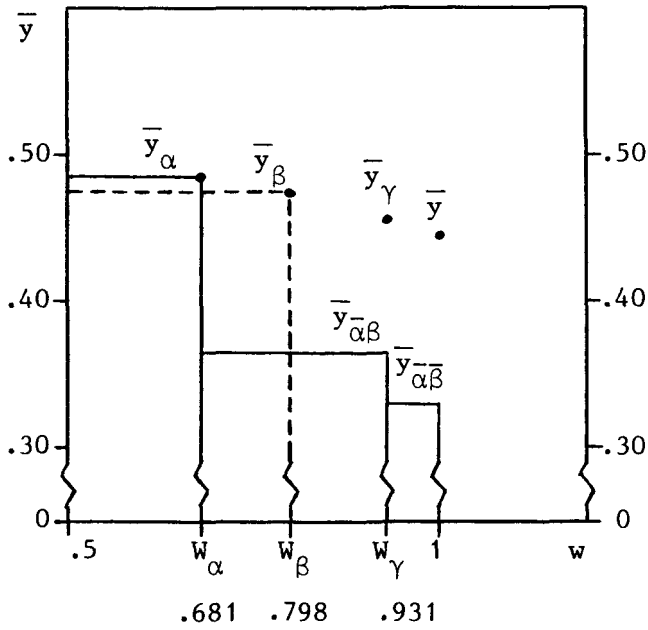


Fig. 4. Relation between means and size of response strata in the 1977 Survey of Pupils

6. A model to estimate and adjust for nonresponse bias

The results in Table 5 indicate that there is a linear relation between the means (proportions) and relative sizes of the response strata obtained by different procedures (see Fig. 3 and 4), i.e.

$$\begin{aligned}\bar{Y}_\alpha &= a + bW_\alpha,\\ \bar{Y}_\gamma &= a + bW_\gamma,\\ \bar{Y} &= a + b,\\ \bar{Y}_\beta &= a + bW_\beta,\end{aligned}$$

and, as a consequence:

$$\begin{aligned}&\bar{Y}_{\bar{\alpha}\beta} = \bar{Y}_{\bar{\gamma}} = a + b(1+W_\gamma) \text{ and}\\ &\bar{Y}_{\bar{\alpha}\bar{\beta}} = a + b(W_\alpha + W_\gamma).\end{aligned}$$

If the model is true, the nonresponse bias of the three estimators in Diagram 2 becomes

$$\begin{aligned}B_\beta &= W_{\bar{\beta}}(\bar{Y}_\beta - \bar{Y}_{\bar{\beta}}) = -W_{\bar{\beta}}b,\\ B_\alpha &= W_{\bar{\alpha}}(\bar{Y}_\alpha - \bar{Y}_{\bar{\alpha}}) = -W_{\bar{\alpha}}b, \text{ and}\\ B_\gamma &= W_{\bar{\gamma}}(\bar{Y}_{\bar{\alpha}\beta} - \bar{Y}_{\bar{\alpha}\bar{\beta}}) = -W_{\bar{\alpha}}W_{\bar{\gamma}}b, \qquad (6.1)\end{aligned}$$

respectively.

We can also estimate b and adjust for the nonresponse bias by means of the observed values of $\bar{y}_\alpha$, $\bar{y}'_{\bar{\alpha}\beta}$, w_α, and $w'_\gamma = w_\alpha + w_{\bar{\alpha}}w'_{\bar{\alpha}\beta}$:

$$\hat{b} = \frac{1}{w'_\gamma}\{\bar{y}'_{\bar{\alpha}\beta} - \bar{y}_\alpha\}, \qquad (6.2)$$

which yields

$$\begin{aligned}\hat{\bar{Y}}^* &= \bar{y}'_\gamma + w_{\bar{\alpha}}w_{\bar{\gamma}}\hat{b}\\ &= \bar{y}_\alpha + \frac{w_{\bar{\alpha}}}{w_\gamma}\{\bar{y}'_{\bar{\alpha}\beta} - \bar{y}_\alpha\}. \qquad (6.3)\end{aligned}$$

For the results in Table 5 we get $\hat{b} = -.128$ and $\hat{\bar{Y}}^* = .445$.

The bias of the adjusted estimator (6.3) becomes:

$$B(\hat{\bar{Y}}^*) = E(\hat{\bar{Y}}^*) - \bar{Y}$$

$$= W_{\bar{\gamma}}\{(\bar{Y}_{\bar{\alpha}\beta} - \bar{Y}_{\bar{\alpha}\bar{\beta}}) - \frac{W_{\bar{\alpha}}}{W_\gamma}(\bar{Y}_\alpha - \bar{Y}_{\bar{\alpha}\beta})\}.$$

Thus, the nonresponse bias is really reduced by the estimator (6.3), provided the differences $(\bar{Y}_{\bar{\alpha}\beta} - \bar{Y}_{\bar{\alpha}\bar{\beta}})$ and $(\bar{Y}_\alpha - \bar{y}_{\bar{\alpha}\beta})$ have the same sign, and $(\bar{Y}_\alpha - \bar{Y}_{\bar{\alpha}\beta})/(\bar{Y}_{\bar{\alpha}\beta} - \bar{Y}_{\bar{\alpha}\bar{\beta}}) < 2W_\gamma/W_{\bar{\alpha}}$. Furthermore, the adjusted estimator is unbiased if the linear model is true.

If we derive the variance of $\hat{\bar{Y}}^*$ by means of Taylor expansions, we obtain an expression similar to, but more complicated than the variance expression in (2.3) for the reweighted estimator. If the means and variances for the initial and second attempt respondents are equal, the variance of $\hat{\bar{Y}}^*$ is approximately

$$V(\hat{\bar{Y}}^*) \approx \frac{S^2}{n}\left\{\left(\frac{1 - W_{\bar{\alpha}}/W_\gamma}{W_\alpha}\right)^2 W_\alpha + \left(\frac{1}{W_\gamma}\right)^2 \frac{W_{\bar{\alpha}}}{f_2 W_{\bar{\alpha}\beta}}\right\}. \qquad (6.4)$$

By means of Cauchy's inequality it can be shown that the minimum of (6.4) is obtained for

$$f_2 = f_{2PROX}^* = f_{2PROX}\frac{W_\alpha}{W_\gamma - W_{\bar{\alpha}}} > f_{2PROX},$$

where f_{2PROX} is given in (3.6). It can also be shown that the minimum of (6.4) is larger than $\tilde{S}^2\tilde{Z}^2$ in (3.7), provided $C_\beta/C_\alpha > W_{\bar{\alpha}\beta}/W_\alpha$. Thus, in most situations we should expect the variance of $\hat{\bar{Y}}^*$ to be larger than the variance of $\hat{\bar{Y}}$. We have to pay for the possible bias reduction in $\hat{\bar{Y}}^*$ by an increased variance. The linear model may be used to estimate the nonresponse bias in $\bar{Y}$ by means of (6.1) and (6.2), but it should not be used to adjust for the bias, unless it is very large.

7. Drawbacks involved in the combination of two data collection methods on a subsampling basis

The advantages associated with combining two data collection methods have been dealt with in Section 5. However, combined data collection methods are not without problems.

When mail data collection is followed-up by interviews on a subsampling basis, we do not use the questionnaires from delayed mail respondents who are not subsampled. In fact we do not know how to make use of this information without introducing an unknown estimation bias.

To minimize the useless number of mail questionnaires received too late, the follow-up is put off until 7–8 weeks after the questionnaires have been mailed. This of course affects the quality of the data if the survey concerns conditions during a specific period.

In connection with the 1977 Survey of Pupils, we also conducted an experiment concerning measurement errors. For a sample of about 370 pupils we made telephone interviews at the same time as the questionnaires were sent out by mail for the regular data collection. One part of the sample was then reinterviewed after three weeks (at the same time as the third reminder was sent out in the regular data collection). Another part of the sample was reinterviewed after seven weeks (at the same time as the follow-up by telephone was carried out in the regular data collection). The results from this experiment are shown in Table 7.

The results in Table 7 show that we can expect a larger number of (random) measurement (recall) errors the longer we wait before we make the interview. Estimates based on responses from the three interview occasions, however, did not differ significantly. Thus, the discrepancies seem to be random rather than systematic. The "*" in Table 7 shows for which items the proportion of discrepancies was significantly larger for the 7-week reinter-

Table 7. Number of discrepancies between the first interview and a reinterview made after three weeks and seven weeks, respectively, in the 1977 Survey of Pupils Leaving the 9-Year Comprehensive School in 1976. Per cent

Item (number of alternatives)	Reinterview made after		
	3 weeks	7 weeks	Significance = *
Date of leaving school (4)	4	5	–
Present employment status (9)	3	12	*
Reason for not working or studying (8)	14	59	*
Unemployed (2)	6	13	*
Total length of unemployment	42	63	*
Present occupation (3-digit code)	24	26	–
Planned occupation (3-digit code)	42	48	–

view than for the 3-week reinterview (5 per cent significance level).

The results in Table 7 indicate that we should not wait too long to make a follow-up. However, if we start too early we might increase the useless number of mail questionnaires received too late.

Another problem with combining two data collection procedures is that we may obtain data of different quality (in a broad sense).

In our theory we have disregarded the measurement aspects and assumed that we measure the same item with different data collection methods. This may be true for variables concerning factual conditions, but surely not for variables related to attitudes, plans, expectations, etc. Furthermore, it is difficult to develop questionnaires suitable for two different data collection methods.

8. Acknowledgements

I am grateful to Ms Ulla-Britt Engström, Statistics Sweden, for drawing the pictures. I also want to thank Mr Lars-Gunnar Holmström, Statistics Sweden, for providing realistic values of the cost components.

9. References

Cochran, W.G. (1977): Sampling Techniques, 3rd edition. New York, Wiley & Sons.

Cornfield, J. (1942): On Selection Biases in Sampling of Human Populations. Journal of the American Statistical Association, pp. 63–68.

Dalenius, T. (1957): Sampling in Sweden. Stockholm, Almqvist & Wiksell.

Durbin, J. (1954): Non-Response and Call-Backs in Surveys. Bulletin of International Statistical Institute, 34, pp. 72–86.

Hansen, M.H. and Hurwitz, W.N. (1946): The Problem of Non-Response in Sample Surveys. Journal of the American Statistical Association, pp. 517–529.

Lyberg, I. (1979): Variansformler och urvalsallokering vid enstegs klusterurval med bortfallsuppföljning enligt Hansen-Hurwitz-planen. Statistics Sweden, unpublished report, 1979–01–12. (In Swedish.)

Swensson, B. (1983): On Double Sampling for Stratification with Non-Response in the Second Phase. Statistical Review, 1983:5, pp. 111–116.

Hot Deck Imputation Techniques that Permit Standard Methods for Assessing Precision of Estimates

by Jan Lanke[1]

To Tore Dalenius in gratitude for all inspiration, encouragement and friendship he has given me over the years.

1. Introduction

In survey investigations and censuses alike, a common experience is that some units fail to provide information on all variables asked for. Thus the problem of missing values crops up and various techniques have been devised to cope with that problem. In the present note we shall discuss only one of these, viz. hot deck imputation. This consists in substituting for each missing value a value determined in some way, deterministic or random, by values given by respondents who in some sense are similar to the non-responding ones.

In the sequel we shall assume that from a set of N units we have chosen n and asked each of these to provide his value on a certain variable x; the mean m of x for the N units is to be estimated. Further, we shall assume that n' of the units have actually responded, giving the values

$$x_1, x_2, \ldots, x_{n'},$$

while the remaining n'' $(=n-n')$ have failed to do so. Then imputation amounts to defining n'' values to be used as substitutes for the x-values of the n'' non-responding units; hot deck imputation means that only information present in the sample under consideration is used when creating the values to be imputed.

[1] Assistant professor of Statistics, University of Lund, Sweden.

Using the phrase 'domain of imputation' for a group of individuals among whom the non-response can be assumed to be random, i.e. a group such that the non-respondents in that group can be considered as a random sample from it, we shall assume that the group for which we want an estimate of the mean, is one single domain of imputation; cf., however, Section 6.

Here it should be stressed that hot deck imputation can never create any new information as a substitute for the information lost because of non-response; rather, the main point is to produce a data set which, although composed of a mixture of true response values and imputed ones, can be treated in exactly the same way as if all the values had been true responses. Thus the purpose of imputation is to simplify the data processing.

In Sections 2 and 3 we briefly review the two most common hot deck imputation techniques, at the same time pointing out a serious drawback inherent in both of them; for more details about these methods (and others) we refer to Elofsson (1973) and Lindström et al. (1979). In Section 4 we suggest a new technique, closely related to the one described in Section 2, but without the drawback just hinted at. Similarly Section 5 modifies the standard technique described in Section 3. Finally Section 6 points out an area for future research.

2. Standard two-pass technique

Using the notation introduced in the introduction, the most commonly employed technique

amounts to using the average

$$\bar{x}' = \frac{1}{n'} \sum_{1}^{n'} x_i$$

of all the values provided by the responding units as a substitute for each missing value.

Admittedly this does not seem very profound and, as mentioned above, the main point is just to fill out the data set to a seemingly complete one, thus permitting use of standard programs for analyzing the data. The rationale for choosing $\bar{x}'$ as a substitute is that the mean of the values

$$x_1, x_2, \ldots, x_{n'}, \bar{x}', \bar{x}', \ldots, \bar{x}'$$

in the completed data set then equals the mean of the true response values.

The method just described is clearly a two-pass method, i.e. it requires us to go through the data set twice: first in order to compute $\bar{x}'$, then once more in order to perform the substitutions. If that is considered too troublesome, there is a variant to be described in Section 3, which calls for just one pass through the data.

Another drawback of the technique described is that, although standard programs will give a reasonable point estimate of the mean of x, the precision of the estimate will be considerably overestimated unless the non-response fraction f'' $(=n''/n)$ is negligible. More precisely, the variance of the estimator m^* $(=\bar{x}')$ is

$$V[m^*] = \frac{N-n'}{N-1} \cdot \frac{\sigma^2}{n'} \quad , \qquad (2.1)$$

where σ^2 is the population variance among the x-values in the entire group of N units. Furthermore, letting s'^2 denote the sample variance among the true responses,

$$s'^2 = \frac{1}{n'-1} \sum_{1}^{n'} (x_i - \bar{x}')^2 \ ,$$

we know that s'^2 is an unbiased estimator of $N\sigma^2/(N-1)$. Thus, the correct form of the standard error of m^* is

$$\sqrt{\frac{N-n'}{N}} \cdot \frac{s'}{\sqrt{n'}} \ .$$

However, a standard program is likely to produce s^2, the sample variance in the completed data set, as an estimator of $N\sigma^2/(N-1)$ and, furthermore, to give

$$\sqrt{\frac{N-n}{N}} \cdot \frac{s}{\sqrt{n}}$$

as the standard error of m^*. The program, not being able to know any better, thus makes two mistakes: it underestimates σ^2 (since $s^2 < s'^2$; cf. below) and it overestimates the number of units observed. It is easy to see that

$$s = \sqrt{\frac{n'-1}{n-1}}\, s'$$

and using the notation

$$\varphi = \sqrt{\frac{N-n}{N-n'} \cdot \frac{n'(n'-1)}{n(n-1)}}$$

the identity

$$\sqrt{\frac{N-n}{N}} \cdot \frac{s}{\sqrt{n}} = \varphi \sqrt{\frac{N-n'}{N}} \cdot \frac{s'}{\sqrt{n'}}$$

tells us that the program inadvertently attaches the factor φ to the correct standard error. Now

$$\varphi \approx (1-f'')(1-\tfrac{1}{2}ff''), \qquad (2.2)$$

where $f = n/N$, $f'' = n''/n$, and thus e.g. a 20 % non-response fraction will cause the standard error to be underestimated by about 20 %, turning an alleged 95 % confidence interval into an 88 % one.

The success of the imputation technique described is thus restricted to the point estimate; as soon as we turn to assessing the precision of the estimate, over-optimistic results will emerge.

3. Standard one-pass technique

If it is considered essential to go through the data set just once when performing the imputation, the following version of hot deck imputation is sometimes employed: for each missing value, substitute the true response value last encountered. When analyzing this technique it is most often assumed that the imputed values are drawn by simple random sampling without replacement from the set of true response values. (For a discussion of the realism of that assumption, see Elofsson

(1973); for an alternative treatment, see Bailar and Bailar (1978) or Ernst (1980).) This in particular means that we must assume that the non-response fraction is at most 50 %.

In this method, as opposed to the one described in the preceding section, the imputation is random in the sense that the set of imputed values is not a deterministic function of the set of true responses. Thus we have two types of random variation present; the sources are

(a) the sampling and

(b) the imputation.

In this section (and in Section 5) we shall use the symbols

$E_{\text{samp}}, E_{\text{imp}}, V_{\text{samp}}, V_{\text{imp}}$

to denote expectations and variances with respect to these two sources of randomness; further we shall use E for the combined expectation, i.e.

$E = E_{\text{samp}} E_{\text{imp}}$,

and V for the corresponding variance.

The completed data set can be denoted

$x_1, x_2, \ldots, x_{n'}, x_{I_1}, x_{I_2}, \ldots, x_{I_{n''}}$,

where $I_1, I_2, \ldots, I_{n''}$ are chosen with SRSWOR from $\{1, 2, \ldots, n'\}$. Letting $\bar{x}$ denote the mean in the completed data set we of course put $m^* = \bar{x}$ and hence

$$E_{\text{imp}}[m^*] = \bar{x}' \tag{3.1}$$

$$E[m^*] = E_{\text{samp}}[\bar{x}'] = m. \tag{3.2}$$

After heavy algebra the standard decomposition

$$V = E_{\text{samp}} V_{\text{imp}} + V_{\text{samp}} E_{\text{imp}} \tag{3.3}$$

gives the result

$$V[m^*] = \frac{N\sigma^2}{N-1} \cdot \frac{1}{n} \left(1 - \frac{n}{N} + 2\frac{n''}{n}\right); \tag{3.4}$$

for details we refer to Elofsson (1973).

The ratio of the variance (3.4) to the one obtained when using the two-pass technique (given in (2.1)) is approximately

$$(1-f'')(1+2f''-f)(1-ff''+f), \tag{3.5}$$

where $f = n/N$ and $f'' = n''/n$; thus the factor (3.5) is the price we have to pay for the benefit of avoiding a second pass through the data set.

More relevant to the theme of the present paper is the fact that the sample variance s^2 in the completed data set has

$$E[s^2] = \frac{N\sigma^2}{N-1} \left(1 - \frac{2n''}{n(n-1)}\right);$$

cf. again Elofsson (1973). Some algebra shows that the standard error $\sqrt{1-f}\,s/\sqrt{n}$ that a standard program is likely to produce, if not told otherwise, is in error by a factor approximately equal to

$$\left(1 + \frac{2f''}{1-f}\right)^{-1/2}. \tag{3.6}$$

Comparison with (2.2) shows that for normally occurring values of f and f'' the standard error is underestimated somewhat less in the present case than in the two-pass case.

4. A new two-pass technique

If we want an imputation technique such that standard formulae, applied to the completed data set, give not only a reasonable point estimate m^* of m but also a reasonable impression of the precision of m^*, a natural idea is to try to introduce a proper amount of variation among the n'' imputed values. The simplest way to do that is to use two different values, say x''_- and x''_+, alternately for substitution. Since the amount of variation desired must be proportional to the variation s' among the true responses, the choice

$x''_- = \bar{x}' - cs'$

$x''_+ = \bar{x}' + cs'$

is close at hand; here c is a constant to be determined.

To facilitate the determination of c, assume that n'' is even. Then the mean in the completed data set equals $\bar{x}'$; furthermore the sample variance s^2 in the completed data set is given by

$$s^2 = \frac{n' - 1 + c^2 n''}{n-1} s'^2.$$

Since we want

$$\sqrt{\frac{N-n}{N}}\cdot\frac{s}{\sqrt{n}}$$

to be a valid standard error for m^* $(=\bar{x}=\bar{x}')$, we get the equation

$$\sqrt{\frac{N-n}{N}}\cdot\frac{s}{\sqrt{n}}=\sqrt{\frac{N-n'}{N}}\cdot\frac{s'}{\sqrt{n'}}$$

giving

$$\frac{n'-1+c^2n''}{n-1}=\frac{n}{n'}\cdot\frac{N-n'}{N-n}$$

from which

$$c=\sqrt{2+\frac{n''-1}{n'}+\frac{n(n-1)}{(N-n)n'}}\,.$$

As a first approximation, assume that n' is not too small; then

$$c\approx\sqrt{2+\frac{f''}{1-f''}+\frac{f}{1-f}\cdot\frac{1}{1-f''}}\,.$$

Next, assume that f is moderately large; then

$$c\approx\sqrt{2+\frac{f''}{1-f''}}\left(1+\frac{f}{2(2-f'')}\right)\,. \qquad (4.1)$$

and if f is negligible we of course get

$$c\approx\sqrt{2+\frac{f''}{1-f''}}\,. \qquad (4.2)$$

If f'' as well is moderate, (4.1) and (4.2) turn into

$$c\approx\sqrt{2}\left(1+\frac{f}{4}+\frac{f''}{4}\right)$$

and

$$c\approx\sqrt{2}\left(1+\frac{f''}{4}\right),$$

respectively. Finally it is seen that

$$c=\sqrt{2}$$

may be a good choice. If e.g. $f<.10$ and $f''<.20$ we shall be at most 8 % in error concerning c if we choose $c=\sqrt{2}$ and, more important, the standard error of m^* is at most 3 % too small in those cases.

5. A new one-pass technique

Now we shall show how it is possible to combine the one-pass feature with the feature of properly increased variation among the imputed values. Thereby we want the variation among the imputed values to be, at least on the average, proportional to the variation among the true response values; at the same time we have no desire whatever to increase the variability of the resulting estimate m^*. Using the notation

$$x_1, x_2, \ldots, x_{n'}, x_1'', x_2'', \ldots, x_{n''}''$$

for the completed data set, our task is to describe the imputed values x_1'', x_2'', ..., $x_{n''}''$ as functions, presumably random, of

$$x_1, x_2, \ldots, x_{n'}.$$

After some thought one arrives at a procedure, the formal description of which is as follows. Let $I_1, I_2, \ldots, I_{n''+1}$ be chosen by simple random sampling without replacement from $\{1, 2, \ldots, n'\}$ and define

$$x_j''=x_{I_j}+c(x_{I_{j+1}}-x_{I_j}), \qquad (5.1)$$

where c is a constant to be determined. Here the first term on the right hand side is intended to give, on the average, a sensible mean, while the purpose of the second is to introduce a convenient amount of variation. For the sum of the imputed values we have

$$\sum_1^{n''}x_j''=\sum_1^{n''}x_{I_j}+c(x_{I_{n''+1}}-x_{I_1})$$

and it is seen that the variation-introducing terms contribute very little to the variability of the estimator m^* $(=\bar{x}=$ average in the completed data set); that fact will also be obvious from the formula (5.2) for $V[m^*]$, given be-

low.

The relations (3.1) and (3.2), given in Section 3 for the standard one-pass technique, clearly hold in the present case as well.

Turning to the variance of m^* we obtain after some algebra

$$V_{\text{imp}}[m^*] = \frac{s'^2}{n^2}\left(\frac{(n'-n'')n''}{n'} - 2c(1-c)\right)$$

and the first term in the decomposition (3.3) of $V[m^*]$ is given by

$$E_{\text{samp}}[V_{\text{imp}}[m^*]] = \frac{N\sigma^2}{N-1}\cdot\frac{1}{n^2}\left(\frac{(n'-n'')n''}{n'} - 2c(1-c)\right).$$

Since

$$V_{\text{samp}}[E_{\text{imp}}[m^*]] = V_{\text{samp}}[\bar{x}'] = \frac{N-n'}{N-1}\cdot\frac{\sigma^2}{n'},$$

we get

$$V[m^*] = \frac{N\sigma^2}{N-1}\cdot\frac{1}{n'}\left(f''(1-2f'') - 2c(1-c)\frac{f'}{n} + 1 - ff'\right) \tag{5.2}$$

where, as before, $f = n/N$, $f' = n'/n$, and $f'' = n''/n = 1 - f'$. Unless n is very small we can clearly drop the term $2c(1-c)f'/n$ and in the sequel we shall assume that to be permitted. Skipping the factor $N/(N-1)$ as well we have the approximate formula

$$V[m^*] \approx \frac{\sigma^2}{n'}\left(1 - f + (1+f)f'' - 2f''^2\right), \tag{5.3}$$

which is seen to agree with a similarly approximate version of the formula (3.4) for $V[m^*]$ in the standard one-pass technique.

Now we address the problem of choosing c in order to introduce the proper amount of variation. Letting, as usual, s^2 denote the sample variance in the completed data set, one arrives after considerable work at

$$E_{\text{imp}}[s^2] = \frac{s'^2}{n-1}\left(n - 2c(1-c)n'' - 1 - 2f'' + 2c(1-c)\frac{1}{n}\right)$$

which, for reasonable values of the parameters involved, can be approximated by

$$\frac{s'^2}{n-1}(n - 2c(1-c)n'').$$

Thus we obtain

$$E[s^2] \approx \frac{N\sigma^2}{N-1}\cdot\frac{1}{n-1}(n - 2c(1-c)n'') \approx \sigma^2(1 - 2c(1-c)f''). \tag{5.4}$$

A combination of (5.3) and (5.4) shows that the correct standard error is

$$\frac{s}{\sqrt{1-2c(1-c)f''}}\cdot\frac{1}{\sqrt{n'}}\sqrt{1-f+(1+f)f''-2f''^2} \tag{5.5}$$

and thus our goal is to choose c in such a way that (5.5) equals the usual expression $\sqrt{1-f}\,s/\sqrt{n}$. This gives an equation which miraculously simplifies into

$$c(1-c) = -\frac{1}{1-f}. \tag{5.6}$$

This equation clearly has two solutions:

$$c = \frac{1}{2} \pm \frac{1}{2}\sqrt{\frac{5-f}{1-f}}. \tag{5.7}$$

Writing the definition (5.1) of x_j'' on the form

$$x_j'' = (1-c)x_{l_j} + c\,x_{l_{j+1}},$$

it is clear that interchange of c and $1-c$ will give a procedure with unchanged statistical properties and hence the choice between the two solutions (5.7) is irrelevant.

A much more essential property of the equation (5.6) is that it is independent of the non-response fraction f''. In fact, a moment's reflection shows that without this property the whole procedure would have been almost pointless: at the moment when c is to be determined, viz. when the sweep through the data set is about to start, the value of f'' is unknown.

For moderate values of f an approximation to the (larger) root of (5.6) is given by

$$c = \frac{1+\sqrt{5}}{2} + \frac{f}{\sqrt{5}}$$

and for small values of f one can of course use

$$c = \frac{1+\sqrt{5}}{2} = 1.62.$$

Finally a word of warning should be given. In the usual one-pass technique, described in Section 3, the formal description is that one should use as substitutes true values drawn by simple random sampling without replacement while in practice one uses, for each substitution, the last true value seen. In the present case, however, it is not permitted to use simply a linear combination of the two true values last seen; rather, it is of paramount importance that the same true value $x_{I_{j+1}}$ should occur in x'_j and x''_{j+1} since this makes these two imputed values strongly negatively correlated, a fact which has obvious pleasant consequences for the variance of m^*. The present imputation technique thus requires a set of temporary registers for storing true values already seen.

6. A problem to be investigated

In the present paper we have so far assumed that the group within which imputations are performed, i.e. the set of units where each can act as a substitute for any other, is the same as the group for which an estimate is desired. Often this is not the case in practice: a common situation is that the population is more finely divided during the imputation than during the subsequent estimation. This of course raises the question as to how our two new methods work under such circumstances.

The author intends to return to that question on a later occasion.

7. References

Bailar, J.C. and Bailar, B.A. (1978): Comparison of Two Procedures for Imputing Missing Survey Values. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 462–467.

Elofsson, G. (1973): Teoretisk jämförelse mellan tre metoder att korrigera för bortfall, speciellt avseende skattningar av varianser. Statistisk tidskrift 1973:2, pp. 93–101. (In Swedish.)

Ernst, L.R. (1980): Variance of the Estimated Mean for Several Imputation Procedures. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 716–720.

Lindström, H., Wretman, J., Forsman, G. and Cassel, C. (1979): Standard Methods for Non-Response Treatment in Statistical Estimation. Memo, Statistics Sweden, Stockholm.

On Double Sampling for Stratification with Non-Response in the Second Phase

by Bengt Swensson [1]

0. Foreword

I was fortunate enough to be introduced to survey sampling by Tore Dalenius, and even more fortunate to have Tore as my teacher during my post-graduate studies. As a part of these studies I participated in his two large research projects during the seventies – "Errors in Surveys" and "Confidentiality in Surveys". I am now very happy to have the opportunity to contribute to this volume in honor of Tore.

As is well known, and as even a cursory look at the compilation of his works in this volume tells, Tore has devoted much work in the field of non-sampling errors. The purpose of this paper is to give some results concerning one specific source of error: non-response.

1. Introduction

In Rao (1973) the following survey situation is considered. Suppose that we wish to stratify a population of size N according to the values of an auxiliary variable ξ, but the frequency distribution of ξ is unknown. We take a large sample $s(n')$ of size n' by simple random sampling without replacement and observe ξ alone. The selected units are classified into H strata according to ξ. Let n'_h denote the number of units in $s(n')$ falling in stratum h ($h = 1, \ldots, H$; $\Sigma n'_h = n'$). A subsample $s(n_h)$ of size n_h is drawn from $s(n'_h)$ by simple random sampling without replacement independently for each h, and the character of interest u is observed. Let all of the following symbols refer to stratum h: $W_h = N_h/N$, $w_h = n'_h/n'$, $\bar{u}_h$ is the mean for $s(n_h)$, $\bar{U}_h$ is the true mean, S^2_{uh} is the true variance and s^2_{uh} is the unbiased estimator of S^2_{uh} based on $s(n_h)$.

[1] Director, Statistical Research Unit, Statistics Sweden.

Assuming that n' is large Rao proves the following two theorems when subsampling according to the subsampling procedure

$$n_h = \nu_h n'_h, \tag{1.1}$$

where ν_h is a constant, chosen in advance, such that $0 < \nu_h \leq 1$.

Theorem 1.1. The estimator $\bar{u}_d = \Sigma\, w_h \bar{u}_h$ is unbiased for $\bar{U} = \Sigma\, W_h \bar{U}_h$ with variance

$$V(\bar{u}_d) = \{\frac{1}{n'} - \frac{1}{N}\} S_u^2 + \Sigma\, \frac{W_h S^2_{uh}}{n'}\ \{\frac{1}{\nu_h} - 1\}\ , \tag{1.2}$$

where S_u^2 is the population variance, provided n' is so large that $\mathrm{pr}(n'_h = 0)$ can be put equal to 0 for all h.

Theorem 1.2. A non-negative unbiased estimator of $V(\bar{u}_d)$ is

$$v(\bar{u}_d) = \frac{N-1}{N}\ \Sigma\Big(\frac{n'_h - 1}{n' - 1} - \frac{n_h - 1}{N-1}\Big)\ \frac{w_h s^2_{uh}}{n_h} + \\ + \frac{N-n'}{N(n'-1)}\ \Sigma w_h (\bar{u}_h - \bar{u}_d)^2\ , \tag{1.3}$$

provided n' is so large that $\mathrm{pr}(n_h \geq 2)$ can be put equal to 1 for all h.

2. Non-response in the second phase

In many applications of the sample design in Section 1.1 we will fail to measure all units in the chosen subsamples, that is we will have non-response. Following Cochran (1977) we will "think of the population as divided into two 'strata', the first consisting of all units for

which measurements would be obtained if the units happened to fall in the sample, the second of the units for which no measurements would be obtained". The first stratum is the response stratum, and the second is the non-response stratum.

Let the character of interest be y (instead of u). When non-response is present we will have measurements on y for a subset $s(n_{hr})$ only, of size n_{hr}, of the subsample $s(n_h)$. We shall assume that n' is so large that $\text{pr}(n_{hr} = 0)$ can be put equal to 0, and we introduce the following symbols all of which refer to the response stratum of stratum h: $W_{hr} = N_{hr}/N_h$, $w_{hr} = n_{hr}/n_h$, $\bar{y}_{hr}$ is the mean for $s(n_{hr})$, $\bar{Y}_{hr}$ is the corresponding population mean, while s^2_{yhr} – based on $s(n_{hr})$ – is the unbiased estimator of S^2_{yhr}.

As an estimator for $\bar{Y}$ we will take

$$\bar{y}^*_d = \Sigma\, w_h \bar{y}_{hr}. \tag{2.1}$$

We will now determine the bias and approximate variance of $\bar{y}^*_d$.

Since $s(n_{hr})$ for fixed n_{hr} is a simple random sample drawn without replacement of size n_{hr} from N_{hr}

$$E(\bar{y}^*_d) = E\, E(\bar{y}^*_d \mid \underline{n}', \underline{n}_r) = E(\Sigma\, w_h \bar{Y}_{hr}) = \\ = \Sigma\, W_h \bar{Y}_{hr}, \tag{2.2}$$

where $\underline{n}' = (n'_1, \ldots, n'_H)$ and $\underline{n}_r = (n_{1r}, \ldots, n_{Hr})$. Thus the bias of $\bar{y}^*_d$ is given by

$$B(\bar{y}^*_d) = E(\bar{y}^*_d) - \bar{Y} = \Sigma\, W_h(\bar{Y}_{hr} - \bar{Y}_h). \tag{2.3}$$

To determine the approximate variance we observe that $\bar{y}^*_d$ can be written

$$\bar{y}^*_d = \frac{1}{\hat{N}} \Sigma\, \hat{N}_h \frac{\hat{Y}_{hr}}{\hat{N}_{hr}}, \tag{2.4}$$

where

$$\hat{N}_h = \frac{N}{n'} \frac{1}{\nu_h} \sum_1^{n_h} z_{hi} \quad (z_{hi} \equiv 1), \tag{2.5}$$

$$\hat{Y}_{hr} = \frac{N}{n'} \frac{1}{\nu_h} \sum_1^{n_h} y^*_{hi}, \tag{2.6}$$

$$y^*_{hi} = \begin{cases} y_{hi} \text{ if the unit is in } s(n_{hr}) \\ 0 \quad \text{otherwise} \end{cases}, \tag{2.7}$$

$$\hat{N}_{hr} = \frac{N}{n'} \frac{1}{\nu_h} \sum_1^{n_h} t_{hi}, \tag{2.8}$$

and

$$t_{hi} = \begin{cases} 1 \text{ if the unit is in } s(n_{hr}) \\ 0 \text{ otherwise} \end{cases}. \tag{2.9}$$

Following Woodruff (1971) we take the partial derivatives of $\bar{y}^*_d$ with respect to $\hat{N}_h$, $\hat{Y}_{hr}$ and $\hat{N}_{hr}$, evaluate the partial derivatives at expected values and form

$$\bar{y}^{*\prime}_d = \Sigma\Big(\frac{\partial \bar{y}^*_d}{\partial \hat{N}_h} \hat{N}_h + \frac{\partial \bar{y}^*_d}{\partial \hat{Y}_{hr}} \hat{Y}_{hr} + \\ + \frac{\partial \bar{y}^*_d}{\partial \hat{N}_{hr}} \hat{N}_{hr}\Big) = \\ = \Sigma\Big(\frac{1}{N} \frac{Y_{hr}}{N_{hr}} \hat{N}_h + \frac{1}{N} \frac{N_h}{N_{hr}} \hat{Y}_{hr} \\ - \frac{1}{N} \frac{N_h Y_{hr}}{N^2_{hr}} \hat{N}_{hr}\Big) = \\ = \Sigma\, w_h \bar{u}_h = \bar{u}_d, \tag{2.10}$$

where $\bar{u}_h$ is the mean of the variable

$$u_{hi} = \bar{Y}_{hr}(1 - t_{hi}/W_{hr}) + y^*_{hi}/W_{hr} \tag{2.11}$$

for $s(n_h)$.

Since $V(\bar{y}^*_d) \cong V(\bar{y}^{*\prime}_d) = V(\bar{u}_d)$ for large n' we have, using theorem 1.1, the following large-sample approximation for the variance of $\bar{y}^*_d$:

$$V(\bar{y}^*_d) = \Big(\frac{1}{n'} - \frac{1}{N}\Big) S^2_u + \Sigma \frac{W_h S^2_{uh}}{n'} \Big(\frac{1}{\nu_h} - 1\Big). \tag{2.12}$$

As an estimator for this variance approximation we may take, using theorem 1.2 and defining the variable u by

$$u_{hi} = \bar{y}_{hr}(1 - t_{hi}/w_{hr}) + y^*_{hi}/w_{hr} \tag{2.13}$$

assuming large samples,

$$v(\bar{y}_d^*) = \frac{N-1}{N}\Sigma\left(\frac{n_h'-1}{n'-1} - \frac{n_h-1}{N-1}\right)\frac{w_h s_{uh}^2}{n_h} + \frac{N-n'}{N(n'-1)}\Sigma w_h(\bar{u}_h - \bar{u}_d)^2, \quad (2.14)$$

where $\bar{u}_h$ now is the mean of the variable (2.13) and $\bar{u}_d = \Sigma w_h \bar{u}_h$.

Thus, if a computer program is already available for double sampling for stratification without non-response, it can still be used in a non-response situation after performing the transformation (2.13).

Since

$$\bar{U}_h = \bar{Y}_{hr},\ \bar{U} = \bar{Y}^* = \Sigma W_h \bar{Y}_{hr},\ S_{uh}^2 \cong S_{yhr}^2/W_{hr},$$

and

$$S_u^2 \cong \Sigma W_h S_{yhr}^2/W_{hr} + \Sigma W_h(\bar{Y}_{hr} - \bar{Y}^*)^2,$$

an alternative approximation for the variance of $\bar{y}_d^*$ is given by

$$V(\bar{y}_d^*) = \left(\frac{1}{n'} - \frac{1}{N}\right)\{\Sigma W_h S_{yhr}^2/W_{hr} + \Sigma W_h(\bar{Y}_{hr} - \bar{Y}^*)^2\} + \frac{1}{n'}\Sigma\frac{W_h}{W_{hr}} S_{yhr}^2\left(\frac{1}{\nu_h} - 1\right). \quad (2.15)$$

Analogously, an alternative variance estimator is given by

$$v(\bar{y}_d^*) = \Sigma\left(\frac{n_h'-1}{n'-1} - \frac{n_h-1}{N-1}\right)\frac{w_h s_{yhr}^2}{n_{hr}} + \frac{N-n'}{N(n'-1)}\Sigma w_h(\bar{y}_{hr} - \bar{y}_d^*)^2, \quad (2.16)$$

which may be preferred to (2.14), especially in situations where no computer program for the complete response case is available.

3. Weighting subclass means to reduce non-response bias

A special case of the situation described in Sections 1 and 2 is the following: A sample $s(n)$ of size n is drawn by simple random sampling without replacement from a population of size N. Measurements on the variable y are obtained for a subset $s(n_r)$ of size n_r. The simple mean $\bar{y}_r$ for $s(n_r)$ might be seriously biased. In order to try to reduce this non-response bias, the units in $s(n)$ are classified into H subclasses according to an auxiliary variable. Let n_h denote the number of units falling in subclass h ($h = 1, \ldots, H$). Let n_{hr} denote the number of units in the response stratum of subclass h and $\bar{y}_{hr}$ the corresponding response mean. As an estimator for the population mean $\bar{Y}$ we take

$$\bar{y}^* = \Sigma\frac{n_h}{n}\bar{y}_{hr}. \quad (3.1)$$

This is a special case of the survey situation described in Section 2, with $\nu_h = 1$, that is $n_h' = n_h$, for all h. Thus the bias of $\bar{y}^*$ is given by

$$B(\bar{y}^*) = \Sigma W_h(\bar{Y}_{hr} - \bar{Y}_h), \quad (3.2)$$

which will be small if we have been able to find a classification such that $(\bar{Y}_{hr} - \bar{Y}_h)$ is small.

The variance of $\bar{y}^*$ is for large samples approximately given by

$$V(\bar{y}^*) = \left(\frac{1}{n} - \frac{1}{N}\right)\{\Sigma W_h S_{hr}^2/W_{hr} + \Sigma W_h(\bar{Y}_{hr} - \bar{Y}^*)^2\}. \quad (3.3)$$

Ignoring the finite population correction, this is the same variance expression as in Thomsen (1978).

As a variance estimator we can take (2.16) with $n' = n$ and $n_h' = n_h$. Ignoring finite population corrections we then arrive at approximately the same result as in Thomsen (1978).

4. The Hansen-Hurwitz design for mail questionnaire and personal interview

In the classical Hansen and Hurwitz (1946) design for the combined use of mail questionnaire and personal interview an initial simple random sample of size n' is drawn and mail questionnaires are sent out. Let n_1' be the number of respondents and n_2' the number of non-respondents. A subsample of size $n_2 = n_2'/k$ ($k \geqq 1$) is selected by simple random sampling without replacement. The subsample units are followed up by personal interview. As pointed out in Rao (1973) this is a special case of the design described in Section 1 with $L = 2$, $\nu_1 = 1$ and $\nu_2 = 1/k$. By the results given in Section 2, extensions to the case with non-response in the subsample immediately follow.

5. Stratified sampling and double sampling within strata

In many survey situations we will first stratify the N units of the population into L strata with N_g units in stratum g ($g = 1, \ldots, L$). After this initial stratification we may wish to stratify the N_g units in stratum g according to the values of an auxiliary variable ξ_g with unknown frequency distribution, e.g. we may wish to apply the sampling scheme given in earlier sections.

This means that we in stratum g ($g = 1, \ldots, L$) will take a large sample $s(n_g')$ of size n_g' by simple random sampling without replacement and observe ξ_g alone. The selected units are classified into L_g strata according to ξ_g. Let n_{gh}' denote the number of units in $s(n_g')$ falling in stratum gh where $h = 1, \ldots, L_g$; $g = 1, \ldots, L$; $\sum_h n_{gh}' = n_g'$. A subsample $s(n_{gh})$ of size n_{gh} is drawn from $s(n_{gh}')$ by simple random sampling without replacement and the character of interest y is observed. Due to non-response we will have measurements on y for a subset $s(n_{ghr})$ only, of size n_{ghr}, of the subsample $s(n_{gh})$.

In stratum gh the subsample size n_{gh} is determined from

$$n_{gh} = \nu_{gh} n_{gh}', \tag{5.1}$$

where ν_{gh} is a constant, chosen in advance, such that $0 < \nu_{gh} \leqq 1$.

With obvious notational extensions we will as an estimator of the population mean

$$\bar{Y} = \sum_g W_g \bar{Y}_g = \sum_g W_g \sum_h W_{gh} \bar{Y}_{gh} \tag{5.2}$$

take

$$\bar{y}^* = \sum_g W_g \bar{y}_g^* = \sum_g W_g \sum_h w_{gh} \bar{y}_{ghr}. \tag{5.3}$$

From the results given in Section 2 it now easily follows that $\bar{y}^*$ is biased for $\bar{Y}$ with bias

$$B(\bar{y}^*) = \sum_g W_g \sum_h W_{gh} (\bar{Y}_{ghr} - \bar{Y}_{gh}) \tag{5.4}$$

and approximate variance

$$V(\bar{y}^*) = \sum_g W_g^2 V(\bar{y}_g^*) = \sum_g W_g^2 V(\bar{u}_g) \tag{5.5}$$

with

$$V(\bar{u}_g) = \left(\frac{1}{n_g'} - \frac{1}{N_g}\right) S_{ug}^2 + \sum_h \frac{W_{gh} S_{ugh}^2}{n_g'} \left(\frac{1}{\nu_{gh}} - 1\right), \tag{5.6}$$

where the variable u is defined by

$$u_{ghi} = y_{ghi}^*/W_{ghr} + \bar{Y}_{ghr}(1 - t_{ghi}/W_{ghr}) \tag{5.7}$$

with

$$y_{ghi}^* = \begin{cases} y_{ghi} \text{ if unit } ghi \text{ belongs} \\ \quad \text{to the response stratum} \\ 0 \quad \text{otherwise,} \end{cases} \tag{5.8}$$

and

$$t_{ghi} = \begin{cases} 1 \text{ if unit } ghi \text{ belongs} \\ \quad \text{to the response stratum} \\ 0 \text{ otherwise,} \end{cases} \tag{5.9}$$

($i = 1, \ldots, N_{gh}$; $h = 1, \ldots, L_g$; $g = 1, \ldots, L$).

As an estimator of $V(\bar{y}^*)$ we can take

$$v(\bar{y}^*) = \sum_g W_g^2 v(\bar{y}_g^*) = \sum_g W_g^2 v(\bar{u}_g) \tag{5.10}$$

with

$$v(\bar{u}_g) = \frac{N_g-1}{N_g} \sum_h \left(\frac{n'_{gh}-1}{n'_g-1} - \frac{n_{gh}-1}{N_g-1}\right) \frac{w_{gh} s^2_{ugh}}{n_{gh}} +$$

$$+ \frac{N_g-n'_g}{N_g(n'_g-1)} \sum_h w_{gh}(\bar{u}_{gh} - \bar{u}_g)^2, \quad (5.11)$$

where the variable u now is defined by

$$u_{ghi} = y^*_{ghi}/w_{ghr} + \bar{y}_{ghr}(1 - t_{ghi}/w_{ghr}). \quad (5.12)$$

Results for the case when we do not have non-response follow directly, since we then have $\bar{Y}_{ghr} = \bar{Y}_{gh}$, $W_{ghr} = w_{ghr} = 1$, $t_{ghi} = 1$, and $y^*_{ghi} = y_{ghi} = u_{ghi}$.

If the purpose is to estimate the population total $Y = N\bar{Y}$, we use the estimator $\hat{Y}^* = N\bar{y}^*$. Corresponding results for this estimator easily follow.

For the estimation of the population ratio $R = Y/X$ we use

$$\hat{R}^* = \hat{Y}^*/\hat{X}^* \quad (5.13)$$

which, since $\mathrm{Cov}(\hat{R}^*, \hat{X}^*) = Y^* - E(\hat{R}^*)X^*$, is biased for R with bias

$$B(\hat{R}^*) = E(\hat{R}^*) - R^* + R^* - R =$$
$$- \mathrm{Cov}(\hat{R}^*, \hat{X}^*)/X^* + R^* - R, \quad (5.14)$$

where

$$R^* = Y^*/X^* = \left(\sum_g \sum_h N_{gh} \bar{Y}_{ghr}\right)/\left(\sum_g \sum_h N_{gh} \bar{X}_{ghr}\right). \quad (5.15)$$

For large samples we will expect the first term in (5.14) to be small compared to the standard error of $\hat{R}^*$, that is

$$B(\hat{R}^*) \cong R^* - R. \quad (5.16)$$

Since we are dealing with large samples the variance of $\hat{R}^*$ is, using first-order Taylor-expansion, approximately equal to the variance of

$$\hat{R}^{*\prime} = (\hat{Y}^* - R^*\hat{X}^*)/X^* = \hat{U} = N\bar{u} =$$
$$= N \sum_g W_g \bar{u}_g \quad (5.17)$$

with the variable u defined by

$$u_{ghi} = \{y^*_{ghi} - R^* x^*_{ghi})/W_{ghr} +$$
$$+ (\bar{Y}_{ghr} - R^* \bar{X}_{ghr})(1 - t_{ghi}/W_{ghr})\}/X^*, \quad (5.18)$$

where y^*_{ghi} and t_{ghi} are defined by (5.8) and (5.9) respectively, and x^*_{ghi} is defined in the same way as y^*_{ghi}.

Thus the variance of $\hat{R}^*$ is approximately given by

$$V(\hat{R}^*) = N^2 \sum_g W_g^2 V(\bar{u}_g), \quad (5.19)$$

where $V(\bar{u}_g)$ is given by expression (5.6).

As an estimator of $V(\hat{R}^*)$ we can take

$$v(\hat{R}^*) = N^2 \sum_g W_g^2 v(\bar{u}_g), \quad (5.20)$$

where $v(\bar{u}_g)$ is given by expression (5.11) if the variable u is defined by

$$u_{ghi} = \{(y^*_{ghi} - \hat{R}^* x^*_{ghi})/w_{ghr} +$$
$$+ (\bar{y}_{ghr} - \hat{R}^* \bar{x}_{ghr})(1 - t_{ghi}/w_{ghr})\}/\hat{X}^*. \quad (5.21)$$

6. Optimum allocation

To determine the optimal values of $\underline{n}' = (n'_1, \ldots, n'_L)$ and $\underline{\mathrm{v}} = (\mathrm{v}_{11}, \ldots, \mathrm{v}_{LL_L})$ suppose the cost function is

$$C = \sum_g (n'_g c'_g + \sum_h n_{gh} c_{gh}), \quad (6.1)$$

where c'_g is the cost of classification per unit in stratum g and c_{gh} is the cost of measuring a unit in stratum gh. Since the n_{gh} are random variables we take expected costs

$$E(C) = \sum_g n'_g (c'_g + \sum_h W_{gh} c_{gh} \mathrm{v}_{gh}). \quad (6.2)$$

Now the (approximate) variances for the estimators of $\bar{Y}$, Y and R can all be written on the same form, i.e.

$$V = \sum_g \frac{A_g}{n'_g} + \sum_g \sum_h \frac{A_{gh}}{n'_g \mathrm{v}_{gh}} - B. \quad (6.3)$$

Taking $\hat{R}^*$ as an example, we have

$$A_g = N_g^2(S_{ug}^2 - \sum_h W_{gh}S_{ugh}^2), \tag{6.4}$$

$$A_{gh} = N_g^2 W_{gh} S_{ugh}^2, \text{ and} \tag{6.5}$$

$$B = \sum_g N_g S_{ug}^2. \tag{6.6}$$

Under realistic survey conditions every A_g and A_{gh} will be > 0.

To minimize V for a given $E(C)$ – or to minimize $E(C)$ for a given V – we form the product $(V+B)E(C)$, which by the Cauchy inequality is seen to be minimized if, for every h and g,

$$n_g'^2 \frac{c_g'}{A_g} = n_g'^2 \nu_{gh}^2 \frac{W_{gh}c_{gh}}{A_{gh}} = \gamma, \tag{6.7}$$

that is, optimal $\underline{\nu}$ is given by

$$\nu_{gh} = \{ \frac{1}{W_{gh}} \frac{c_g'}{c_{gh}} \frac{A_{gh}}{A_g} \}^{\frac{1}{2}} \tag{6.8}$$

provided $\nu_{gh} \leqq 1$ for all g and h. Optimal $\underline{n}'$ is given by (6.7), (6.8) and the relevant side condition provided $n_g' \leqq N_g$ for every g.

7. An alternative subsampling procedure

In Srinath (1971) and Rao (1973) an alternative subsampling procedure is suggested, to be used in situations where the stratum weights are not accurately known. In the present stratified context the suggestion is: In stratum gh choose

$$n_{gh} = n_{gh}'^2/(\alpha_{gh}n_g' + n_{gh}'), \tag{7.1}$$

where $\alpha_{gh} \geqq 0$ is a constant chosen in advance.

The approximate variance of $\bar{y}^*$ is still given by (5.5), but with

$$V(\bar{u}_g) = \{ \frac{1}{n_g'} - \frac{1}{N_g} \} S_{ug}^2 + \frac{1}{n_g'} \sum_h \alpha_{gh} S_{ugh}^2, \tag{7.2}$$

which is independent of the weights W_{gh}.

As an estimator of $V(\bar{y}^*)$ we can still take (5.10).

Expected costs are now given by

$$E(C) \cong \sum_g n_g'(c_g' + \sum_h c_{gh} W_{gh}^2/(\alpha_{gh} + W_{gh})). \tag{7.3}$$

Since the variance of $\bar{y}^*$ using the present subsampling scheme and the variance of $\bar{y}^*$ using the original subsampling scheme are equal under the transformations

$$\alpha_{gh} = W_{gh}(\frac{1}{\nu_{gh}} - 1) \tag{7.4}$$

and since the same is true for the expected costs, we have

$$\alpha_{gh(\text{opt})} = W_{gh}(\frac{1}{\nu_{gh(\text{opt})}} - 1), \tag{7.5}$$

while $\underline{n}'_{(\text{opt})}$ will be equal for the two subsampling procedures.

8. References

Cochran, W.G. (1977): Sampling Techniques, 3rd edition, New York: Wiley.

Hansen, M.H. and Hurwitz, W.N. (1946): The Problem of Nonresponse in Sample Surveys. Journal of the American Statistical Association 41, pp. 517–529.

Rao, J.N.K. (1973): On Double Sampling for Stratification and Analytical Surveys. Biometrika 60, pp. 125–133.

Srinath, K.P. (1971): Multiphase Sampling in Nonresponse Problems. Journal of the American Statistical Association 66, pp. 583–586.

Thomsen, I. (1978): A Second Note on the Efficiency of Weighting Subclass Means to Reduce the Effects of Non-Response When Analyzing Survey Data. Statistisk tidskrift 3, pp. 191–196.

Woodruff, R.S. (1971): A Simple Method for Approximating the Variance of a Complicated Estimate. Journal of the American Statistical Association 66, pp. 411–414.

On Variance Estimation in Survey Sampling with Measurement Errors

by Jan Håkan Wretman[1]

1. Introduction

Sampling, measurement, and data processing contribute to the total error of a sample survey. Tore Dalenius has repeatedly stressed the importance of total survey design, aiming at an overall economic balance between sampling and nonsampling errors. I want to pay tribute to Tore by presenting the following note on a problem connected with the assessment of the total survey error.

The importance of measurement errors in survey sampling has long been recognized, as evidenced by, e.g., Mahalanobis (1946), Sukhatme and Seth (1952), Hansen, Hurwitz and Bershad (1961), Hansen, Hurwitz and Pritzker (1964), and Bailar and Dalenius (1969). Interest has mainly focused on estimating certain components of the total mean square error of an estimator by utilizing such tools as, e.g., interpenetrating samples, reinterviews, and randomized assignment of interviewers and coders.

Most textbooks on survey sampling, however, still do not provide an integrated treatment of sampling and nonsampling errors. When it comes to current survey practice, it seems that the above-mentioned tools for estimating error components are usually lacking. It also seems that the total variance of an estimator is usually estimated by means of standard textbook formulas, derived under the assumption that observations are free from errors.

With the last-mentioned fact in mind, Hartley and Rao (1978) stated that "the application of standard textbook formulas for the estimation of the variances of survey estimates may lead to serious underestimates of the real variability which should incorporate the nonsampling errors." The purpose of the present note is to study what happens if we take "standard textbook formulas," originally derived under the assumption of no measurement errors, and apply them mechanically in a situation where measurement errors are present. We generalize earlier results in that we consider, for an arbitrary sampling design, the case when a linear population function L is estimated by a linear estimator $\hat{L}$, and the variance $V = V(\hat{L})$ is estimated by a quadratic variance estimator $\hat{V}$.

2. Sampling design

We consider a finite population U consisting of N units labeled $1, 2, \ldots, k, \ldots, N$, i.e., $U = \{1, \ldots, N\}$. A sample (a subset of U) is selected by a given probability sampling design p, which gives any sample $s \subseteq U$ a known probability $p(s)$ of being obtained.

We define the following sets of samples from U:

$S_p = \{s: s \subseteq U \text{ and } p(s)>0\}$, $A_k = \{s: s \in S_p \text{ and } k \in s\}$, and $A_{kl} = \{s: s \in S_p, k \in s, \text{ and } l \in s\}$ $(k = 1, \ldots, N; l = 1, \ldots, N)$.

The inclusion probabilities determined by the design p are $\pi_k = \Sigma_{s:\, A_k} p(s)$ and $\pi_{kl} = \Sigma_{s:\, A_{kl}} p(s)$ $(k = 1, \ldots, N; l = 1, \ldots, N)$.

Note that $A_{kk} = A_k$, and $\pi_{kk} = \pi_k$.

3. The aim of the survey

With each unit $k \in U$ we associate an unknown constant θ_k, which is the true value of unit k

[1] Director, Statistical Research Unit, Statistics Sweden.

with respect to the variable under study. The aim of the survey is to estimate the linear function

$$L = \Sigma_{k=1}^{N} c_k \theta_k \,. \tag{3.1}$$

Examples of such linear functions L are: Population totals (all $c_k = 1$); population means (all $c_k = 1/N$); domain totals ($c_k = 1$ for k in the domain in question, $c_k = 0$ otherwise); and a difference between two domain totals.

4. Measurement error model

After the sample s has been selected, attempts are made to obtain the true values θ_k for all units $k \in s$. We assume, however, that measurements are imperfect. Hence, instead of obtaining the true values we obtain observed values Y_k for all $k \in s$, where

$$Y_k = \theta_k + \text{measurement error.}$$

Remark 1: The terms "observed value" and "measurement error" are used here in a wide sense. The observed value is the value that enters into the final calculation of an estimate of L. Thus, the measurement error of an observation is the net result of all nonsampling errors generated in previous stages of the survey process: data collection errors, coding errors, editing errors etc. □

The measurement errors are assumed to be random variables obeying a probability model, called the measurement error model, and denoted by m. It follows that the observed values Y_k $(k \in s)$ from a given sample s are random variables with a joint distribution determined by the measurement error model m.

The survey is thus considered as a two-stage process, where randomness is contributed from each stage:

(i) First stage: The sample selection, generating a sample s.
Probability model: The sampling design p.

(ii) Second stage: The measurement procedure, generating observed values Y_k for $k \in s$.
Probability model: The measurement error model m.

By "the measurement error model" we mean, strictly speaking, a set of conditional models for the second stage above: For each given $s \in S_p$, there is a conditional measurement error model describing the behavior of exactly those Y_k for which $k \in s$. We assume that these conditional models satisfy the following conditions for any given sample $s \in S_p$, and for any $k \in s$ and $l \in s$:

$$E_m(Y_k|s) = \mu_k, \tag{4.1}$$

$$V_m(Y_k|s) = \sigma_k^2, \text{ and} \tag{4.2}$$

$$C_m(Y_k, Y_l|s) = \sigma_{kl}, \tag{4.3}$$

where $E_m(Y_k|s)$, $V_m(Y_k|s)$ and $C_m(Y_k, Y_l|s)$ denote conditional expectation, variance and covariance, respectively, for any given sample s, with respect to the measurement error model m. (Note that $\sigma_{kk} = \sigma_k^2$.)

Remark 2: The frequency interpretation of properties (4.1) – (4.3) of the measurement error model m refers to a hypothetical series of independent replications of the whole process which generates observed values Y_k for all $k \in s$, with a fixed sample s, and with the same staff, the same organization, and the same techniques. In a long series of repetitions of this kind it is assumed that the mean value of the observed Y_k-values for unit $k \in s$ would be equal to μ_k, the variance of these Y_k-values would be σ_k^2, and the covariance between observed values from two units $k \in s$ and $l \in s$ would be σ_{kl}. Although hard to imagine in reality, this hypothetical setup will serve as a frequentist frame of reference for probability statements under the measurement error model. For the following presentation it is not

essential that a series of independent repetitions under identical conditions is physically realizable, since we will not assume replicated observations to be available. □

Remark 3: Note that, by assumption, the model parameters μ_k, σ_k^2 and σ_{kl} do not depend on the given sample s. Hence, it is assumed for any two samples $s_1 \in A_k$ and $s_2 \in A_k$ that $E_m(Y_k|s_1) = E_m(Y_k|s_2) = \mu_k$, and $V_m(Y_k|s_1) = V_m(Y_k|s_2) = \sigma_k^2$. In the same way, for any two samples $s_1 \in A_{kl}$ and $s_2 \in A_{kl}$, $C_m(Y_k, Y_l|s_1) = C_m(Y_k, Y_l|s_2) = \sigma_{kl}$. In other words, no interaction between sample and measurement error is assumed, as would have been done in a less simplified model. □

Remark 4: By (4.1) – (4.3), the model parameters μ_k, σ_k^2, and σ_{kl} were assumed not dependent on s. But there is another kind of dependence that we do not wish to rule out, namely, the dependence on p, e.g., through expected sample size. If, e.g., p_1 is a small size design, and p_2 is a census (i.e., $S_{p_2} = \{U\}$ with $p_2(U) = 1$), then the measurement errors arising under p_1 would certainly not behave in the same way as those arising under p_2, because the measurement conditions could never be identical in these two cases. Hence, the model parameters defined with respect to one specific sampling design may not be generalized to any other design. Especially, we are not allowed to generalize from a small size design to a census. □

For future use we define the quantities

$$\mu = \Sigma_{k=1}^N \mu_k/N, \tag{4.4}$$

$$\sigma^2 = \Sigma_{k=1}^N \sigma_k^2/N, \tag{4.5}$$

$$\gamma = \Sigma_{k=1}^N \Sigma_{l=1,\, l\neq k}^N \sigma_{kl}/N(N-1), \text{ and} \tag{4.6}$$

$$\varrho = \gamma/\sigma^2. \tag{4.7}$$

The following notation will be used to distinguish different types of mathematical expectation: $E_p(.)$ denotes expectation with respect to the sampling design p only (i.e., with respect to the "first stage" as described earlier). $E_m(.|s)$ denotes expectation with respect to the measurement error model m only, for a given sample s (i.e., conditional expectation with respect to the "second stage", for given outcome of the "first stage"). $E_{pm}(.) = E_p\{E_m(.|s)\}$ denotes expectation with respect to sampling design and measurement error model jointly (i.e., with respect to both "stages" jointly). A similar notation will also be used for variances and covariances.

5. Estimation formulas

We assume that the population characteristic $L = \Sigma_{k=1}^N c_k \theta_k$ is estimated by the linear homogeneous estimator

$$\hat{L} = \Sigma_{k \in s,\, s \in S_p} w_{ks} Y_k \tag{5.1}$$

with weights w_{ks} ($k \in s$, $s \in S_p$) satisfying the conditions

$$\Sigma_{s:\, k \in s} p(s) w_{ks} = c_k \quad (k=1, \ldots, N). \tag{5.2}$$

Remark 1: Assumption (5.2) entails design-unbiasedness, in the sense that

$$E_p(\Sigma_{k \in s,\, s \in S_p} w_{ks} a_k) = \Sigma_{k=1}^N c_k a_k$$

for any fixed $(a_1, \ldots, a_N) \in \mathbf{R}^N$. □

It follows from (5.2) that

$$E_{pm}(\hat{L}) = E_p\{E_m(\hat{L}|s)\} = \\ = E_p(\Sigma_{k \in s,\, s \in S_p} w_{ks} \mu_k) = \Sigma_{k=1}^N c_k \mu_k.$$

By consequence, the total bias of $\hat{L}$ (with respect to sampling design and measurement error model jointly) is given by

$$B = E_{pm}(\hat{L}) - L = \Sigma_{k=1}^{N} c_k(\mu_k - \theta_k). \quad (5.3)$$

The total variance of $\hat{L}$ (with respect to sampling design and measurement error model jointly) can be expressed as follows:

$$V = V_{pm}(\hat{L}) = V_p\{E_m(\hat{L}|s)\} + \\ + E_p\{V_m(\hat{L}|s)\} = V_1 + V_2, \quad (5.4)$$

where

V_1 = sampling variance component
and
V_2 = measurement variance component.

(The measurement variance component can be decomposed into simple measurement variance and correlated measurement variance, in conformity with the variance decomposition given by, e.g., Hansen, Hurwitz and Bershad (1961)).

We assume that the total variance V is estimated by the variance estimator

$$\hat{V} = \Sigma_{k\in s}\Sigma_{l\in s} u_{kls} Y_k Y_l \quad (5.5)$$

with weights u_{kls} ($k \in s$, $l \in s$, $s \in S_p$) satisfying

$$\Sigma_{s\supset A_{kl}} p(s) u_{kls} = \psi_{kl} \\ (k=1, \ldots, N; l=1, \ldots, N), \quad (5.6)$$

where

$$\psi_{kl} = \Sigma_{s\supset A_{kl}} p(s) w_{ks} w_{ls} - c_k c_l. \quad (5.7)$$

This variance estimator is of a form general enough to cover most "standard textbook formulas" for variance estimation.

Remark 2: Assumption (5.6) implies design-unbiasedness in the same sense as described in Remark 1 above, namely that

$$E_p(\Sigma_{k\in s}\Sigma_{l\in s} u_{kls} a_k a_l) = V_p(\Sigma_{k\in s} w_{ks} a_k)$$

for any fixed $(a_1, \ldots, a_N) \in \mathbf{R}^N$. □

Remark 3: As an example of estimators $\hat{L}$ and $\hat{V}$ satisfying (5.2) and (5.6) we consider Horvitz-Thompson estimation of the population total $L = \Sigma_{k=1}^{N} \theta_k$. Let p be a sampling design with $\pi_k > 0$ and $\pi_{kl} > 0$ ($k=1, \ldots, N$; $l=1, \ldots, N$). The Horvitz-Thompson formula for estimating a population total gives

$$\hat{L}_{HT} = \Sigma_{k\in s} Y_k/\pi_k.$$

The total variance given by (5.4) now becomes

$$V = \Sigma_{k=1}^{N}\Sigma_{l=1}^{N}\left(\frac{\pi_{kl}}{\pi_k\pi_l} - 1\right)\mu_k\mu_l +$$

$$\Sigma_{k=1}^{N}\Sigma_{l=1}^{N}\frac{\pi_{kl}}{\pi_k\pi_l}\sigma_{kl}.$$

Two well-known variance estimators, derived under the assumption of no measurement errors, are given by, respectively, Horvitz-Thompson's formula

$$\hat{V}_{HT} = \Sigma_{k\in s}\Sigma_{l\in s}\left(1 - \frac{\pi_k\pi_l}{\pi_{kl}}\right)\frac{Y_k}{\pi_k}\frac{Y_l}{\pi_l}$$

and Yates-Grundy's formula

$$\hat{V}_{YG} = \\ \Sigma_{k\in s}\Sigma_{l\in s,\, l>k}\left(\frac{\pi_k\pi_l}{\pi_{kl}} - 1\right)\left(\frac{Y_k}{\pi_k} - \frac{Y_l}{\pi_l}\right)^2.$$

We finally consider the special case when the sampling design p is simple random sampling without replacement of n units. In this case $\hat{L}_{HT}$, V, $\hat{V}_{HT}$ and $\hat{V}_{YG}$ reduce to

$$\hat{L}_{HT} = N\bar{Y}_s,$$

$$V = N^2(1-f)\Sigma_{k=1}^{N}(\mu_k - \mu)^2/n(N-1) + \\ + N^2(\sigma^2/n)\{1 + (n-1)\varrho\}, \quad (5.8)$$

and

$$\hat{V}_{HT} = \hat{V}_{YG} = N^2(1-f)\Sigma_{k\in s}(Y_k - \bar{Y}_s)^2/n(n-1),$$

where $f = n/N$, $\bar{Y}_s = \Sigma_{k \in s} Y_k / n$, and μ, σ^2 and ϱ are defined by (4.4) – (4.7). □

Conditions (5.2) and (5.6) ensure that in the case with no measurement errors, i.e., when $Y_k = \theta_k$ ($k \in s$, $s \in S_p$), $\hat{L}$ is design-unbiased for L, and $\hat{V}$ is design-unbiased for V (which would then consist of only a sampling variance component). Our question now is: If the "standard textbook formula" $\hat{V}$ is used when measurement errors are present, what is the bias of this variance estimator?

6. The bias of the variance estimator

In the presence of measurement errors, $\hat{V}$ as defined in the previous section is not in general unbiased for V, if expectation is with respect to sampling design and measurement error model jointly. Nor is $\hat{V}$ unbiased for the sampling variance component of V. The following theorem gives an expression for the bias of $\hat{V}$.

Theorem: Consider an arbitrary sampling design p with $\pi_k > 0$ and $\pi_{kl} > 0$ ($k=1, \ldots, N$; $l=1, \ldots, N$), and an arbitrary measurement error model m satisfying (4.1) – (4.3). Let $\hat{L}$ be as defined by (5.1), satisfying (5.2), and let $\hat{V}$ be as defined by (5.5), satisfying (5.6). Then,

$$E_{pm}(\hat{V}) - V = -\Sigma_{k=1}^{N} \Sigma_{l=1}^{N} c_k c_l \sigma_{kl} \qquad (6.1)$$

and

$$E_{pm}(\hat{V}) - V_1 = \Sigma_{k=1}^{N} \Sigma_{l=1}^{N} \psi_{kl} \sigma_{kl}. \square \qquad (6.2)$$

A proof of the theorem is given in the next section.

Remark 1: Note the remarkable simplicity of the bias (6.1). It does not depend on the weights w_{ks} and u_{kls} used for defining $\hat{L}$ and $\hat{V}$. Nor does it depend explicitly on any sampling design probabilities. However, it may depend on the sampling design in the weak sense that the σ_{kl}'s may be specific for the design under consideration. □

Remark 2: When all $\sigma_{kl} \geq 0$ and all $c_k \geq 0$, the total variance V is underestimated by $\hat{V}$:

$$E_{pm}(\hat{V}) - V = -\Sigma_{k=1}^{N} \Sigma_{l=1}^{N} c_k c_l \sigma_{kl} \leq 0.$$

This seems to be the situation most usually met with in practice. □

Remark 3: If $\sigma_{kl} = 0$ whenever $k \neq l$, i.e., if measurements on different units are always uncorrelated, the total variance V is also underestimated by $\hat{V}$:

$$E_{pm}(\hat{V}) - V = -\Sigma_{k=1}^{N} c_k^2 \sigma_k^2 \leq 0. \quad \square$$

Remark 4: In the case when $c_k = 1$ ($k = 1, \ldots, N$), i.e., when L is a population total, the bias (6.1) can be expressed, using (4.4) – (4.7), as

$$E_{pm}(\hat{V}) - V = -\Sigma_{k=1}^{N} \Sigma_{l=1}^{N} \sigma_{kl} = \\ = -N\sigma^2 \{1 + (N-1)\varrho\}. \qquad (6.3)$$

In this case, the total variance V is underestimated by $\hat{V}$ if and only if $\sigma^2 > 0$ and $\varrho > -1/(N-1)$. □

Remark 5: Theoretically, the total variance V can be overestimated by $\hat{V}$, as demonstrated by the following numerical example, where large negative correlations between measurements on different units cause the bias (6.1) to be positive. These correlations are certainly unrealistic, but the example can not be rejected on logical grounds. Additional constraints must be put on the covariances σ_{kl} in order to ensure a strictly negative bias (6.1). The example is as follows.

Suppose we want to estimate the total of a population of size $N = 3$: $U = \{1, 2, 3\}$, $c_k = 1$ ($k = 1, 2, 3$). The sampling design p is simple random sampling without replacement of $n = 2$ units: $S_p = \{\{1,2\}, \{1,3\}, \{2,3\}\}$, and $p(s) = 1/3$ for all $s \in S_p$. Measurement variances and covariances are put together in the $N \times N$ matrix

$$\underset{\sim}{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix} = \begin{bmatrix} 1.0 & -.5 & -.6 \\ -.5 & 1.0 & -.6 \\ -.6 & -.6 & 1.0 \end{bmatrix}.$$

It is easily verified that this choice of $\underset{\sim}{\Sigma}$ makes $\Sigma_{k=1}^N \Sigma_{l=1}^N \sigma_{kl} < 0$, and hence $E_{pm}(\hat{V}) < V$.

Note that the matrix $\underset{\sim}{\Sigma}$, given above, is not positive semi-definite, and hence may not be interpreted as the variance matrix of any three-dimensional random vector. But this choice of $\underset{\sim}{\Sigma}$ is nevertheless formally permitted, because the measurement model assumptions in Section 4 are still satisfied. As is easily checked, for any $s \in S_p$, the 2x2 variance matrix pertaining to those Y_k for which $k \in s$ is positive semi-definite. Hence, for any given $s \in S_p$ we have a proper conditional measurement error model for those Y_k for which $k \in s$. (Unlike us, Koop (1974) assumed the NxN matrix $\underset{\sim}{\Sigma}$ to be always positive semi-definite and, as a consequence, obtained a different result.) □

Remark 6: It is seen from (6.2) that $\hat{V}$ is not *pm*-unbiased for the sampling variance V_1. Probably, $\hat{V}$ will often overestimate V_1, as in the following example. Consider estimation of a population total under simple random sampling without replacement of size n. Using (5.8) and (6.3) we then have

$$\begin{aligned} &E_{pm}(\hat{V}) - V_1 = \{E_{pm}(\hat{V}) - V\} + V_2 \\ &= -N\sigma^2\{1 + (N-1)\varrho\} + N^2(\sigma^2/n)\{1 + (n-1)\varrho\} \\ &= N(N-n)(\sigma^2/n)(1-\varrho) > 0 \end{aligned}$$

provided $\sigma^2 > 0$ and $\varrho < 1$. □

Remark 7: The theorem above relates as follows to earlier results on the bias of a variance estimator when measurement errors are present. Cochran (1953, and subsequent editions) considered p = simple random sampling, L = the population mean, $\hat{L}$ = the sample mean, and $\hat{V}$ = the traditional variance estimator used with simple random sampling. Koop (1974) and also Koch, Freeman and Freeman (1975) considered p = arbitrary design, L = the population total, $\hat{L} = \hat{L}_{HT}$, and $\hat{V} = \hat{V}_{HT}$. The present theorem extends these results, in that it considers p = arbitrary design, L = arbitrary linear function, $\hat{L}$ = arbitrary linear estimator, and $\hat{V}$ = arbitrary quadratic variance estimator. □

7. Proof of the theorem

For any given $s \in S_p$ the weights w_{ks} and u_{kls} were defined only for $k \in s$ and $l \in s$. For the purpose of this proof, we introduce new weights w'_{ks} and u'_{kls} defined for all $k = 1, \ldots, N$ and $l = 1, \ldots, N$:

$$w'_{ks} = \begin{cases} w_{ks} & \text{when } k \in s \\ 0 & \text{otherwise} \end{cases}$$

and

$$u'_{kls} = \begin{cases} u_{kls} & \text{when both } k \in s \text{ and } l \in s \\ 0 & \text{otherwise.} \end{cases}$$

From (5.2), (5.6) and (5.7) we now have

$$E_p(w'_{ks}) = c_k$$

and

$$E_p(u'_{kls}) = \psi_{kl} = C_p(w'_{ks}, w'_{ls}).$$

Noting that

$$\begin{aligned} V_p\{E_m(\hat{L}|s)\} &= V_p\{E_m(\Sigma_{k\in s} w_{ks} Y_k|s)\} \\ &= V_p(\Sigma_{k\in s} w_{ks} \mu_k) \\ &= V_p(\Sigma_{k=1}^N w'_{ks} \mu_k) \\ &= \Sigma_{k=1}^N \Sigma_{l=1}^N \psi_{kl} \mu_k \mu_l \end{aligned}$$

and

$$\begin{aligned} E_p\{V_m(\hat{L}|s)\} &= E_p\{V_m(\Sigma_{k\in s} w_{ks} Y_k|s)\} \\ &= E_p(\Sigma_{k\in s} \Sigma_{l\in s} w_{ks} w_{ls} \sigma_{kl}) \\ &= E_p(\Sigma_{k=1}^N \Sigma_{l=1}^N w'_{ks} w'_{ls} \sigma_{kl}) \\ &= \Sigma_{k=1}^N \Sigma_{l=1}^N (\psi_{kl} + c_k c_l) \sigma_{kl}, \end{aligned}$$

we can express the total variance V as

$$\begin{aligned} V &= V_p\{E_m(\hat{L}|s)\} + E_p\{V_m(\hat{L}|s)\} \\ &= \Sigma_{k=1}^N \Sigma_{l=1}^N \psi_{kl} \mu_k \mu_l + \\ &+ \Sigma_{k=1}^N \Sigma_{l=1}^N (\psi_{kl} + c_k c_l) \sigma_{kl} \\ &= V_1 + V_2. \end{aligned} \tag{7.1}$$

The expected value of $\hat{V}$ is

$$\begin{aligned} E_{pm}(\hat{V}) &= E_p\{E_m(\Sigma_{k\in s}\Sigma_{l\in s} u_{kls} Y_k Y_l|s)\} \\ &= E_p\{\Sigma_{k\in s}\Sigma_{l\in s} u_{kls}(\sigma_{kl} + \mu_k\mu_l)\} \\ &= E_p\{\Sigma_{k=1}^N \Sigma_{l=1}^N u'_{kls}(\sigma_{kl} + \mu_k\mu_l)\} \\ &= \Sigma_{k=1}^N \Sigma_{l=1}^N \psi_{kl}(\sigma_{kl} + \mu_k\mu_l). \quad (7.2) \end{aligned}$$

Comparing (7.1) and (7.2) we find that

$$E_{pm}(\hat{V}) - V = -\Sigma_{k=1}^N \Sigma_{l=1}^N c_k c_l \sigma_{kl}$$

and

$$E_{pm}(\hat{V}) - V_1 = \Sigma_{k=1}^N \Sigma_{l=1}^N \psi_{kl} \sigma_{kl}$$

and the theorem is proved.

8. Further developments

Survey results are often presented in form of confidence intervals

$$\hat{L} \pm z_{1-\alpha/2}\sqrt{\hat{V}} \qquad (8.1)$$

where α is the desired confidence level, and $z_{1-\alpha/2}$ is the quantile of order $1-\alpha/2$ of the standard normal distribution. The fact that $\hat{V}$ is not unbiased for V will make the real confidence level differ from the desired one and may thus lead to erroneous inference statements. More work needs to be done to study how serious these effects are. (Of course, the real confidence level of the interval (8.1) also depends on how well the distribution of $\hat{L}$ is approximated by a normal distribution.)

The results of this paper were derived assuming a simple measurement error model with no interaction between sample and measurement error. If such interaction is allowed, things become more difficult, and it remains to find a simple expression for the bias of $\hat{V}$ in that case.

9. References

Bailar, B.A., and Dalenius, T. (1969): Estimating the Response Variance Components of the U.S. Bureau of the Census' Survey Model. Sankhyā, B, 31, pp. 341–360.

Cochran, W.G. (1953): Sampling Techniques. New York: Wiley.

Hansen, M.H., Hurwitz, W.N., and Bershad, M.A. (1961): Measurement Errors in Censuses and Surveys. Bulletin of the International Statistical Institute, 38, pp. 359–374.

Hansen, M.H., Hurwitz, W.N., and Pritzker, L. (1964): The Estimation and Interpretation of Gross Differences and the Simple Response Variance. In C.R. Rao, Ed., Contributions to Statistics. London: Pergamon Press, and Calcutta: Statistical Publishing Society.

Hartley, H.O., and Rao, J.N.K. (1978): Estimation of Nonsampling Variance Components in Sample Surveys. In N.K. Namboodiri, Ed., Survey Sampling and Measurement. New York: Academic Press.

Koch, G.C., Freeman Jr., D.H., and Freeman, J.L. (1975): Strategies in the Multivariate Analysis of Data from Complex Surveys. International Statistical Review, 43, pp. 59–78.

Koop, J.C. (1974): Notes for a Unified Theory of Estimation for Sample Surveys Taking into Account Response Errors. Metrika, 21, pp. 19–39.

Mahalanobis, P.C. (1946): Recent Experiments in Statistical Sampling in the Indian Statistical Institute. Journal of the Royal Statistical Society, Series A, 109, pp. 325–370.

Sukhatme, P.V., and Seth, G.R. (1952): Non-Sampling Errors in Surveys. Journal of the Indian Society of Agricultural Statistics, 4, pp. 5–41.

Randomized Response Designs With Attribute-Based Randomization

by Patrick D. Bourke[1]

1. Introduction

It is a great pleasure to contribute to this special issue of the Statistical Review in honor of Professor Tore Dalenius. Apart from his great contribution to Survey Sampling, Professor Dalenius is well-known for his generosity in putting forward ideas for further research and development, and he may remember that it was he who introduced me to the Randomized Response Technique (and to many other areas of Survey Sampling). In view of Professor Dalenius' continued interest in Randomized Response and his many contributions to its development, it is appropriate that my paper for this special issue should focus on some further development of Randomized Response methods.

Most of the Randomized Response (RR) designs described in the literature protect the privacy of the respondent through the use of *artificial* randomization (such as that provided by a deck of cards, a box of beads or a coin-toss) which directs the respondents to one of a number of alternative questions. (For a recent review of RR developments, see Boruch and Cecil (1979)). An alternative to artificial randomization is the use of *attribute-based* randomization by which the respondent is asked to use some personal attribute (such as the telephone-number of a close friend) to make the selection of the question to be answered. The Combined Questions approach of Swensson (1974, 1977) may be viewed as randomized response using attribute-based randomization. Other designs in which a respondent-attribute is used as a basis for question-selection are presented in Bourke (1974a), DeLacy (1975) and Hochberg (1975). Some applications of the RR technique with attribute-based randomization are described in Goodstadt and Gruson (1975), Lavrakas (1975), and Takahasi and Sakasegawa (1977).

All the RR designs presented here use attribute-based randomization, either as the only source of randomization or in conjunction with artificial randomization. In Section 2, a design for estimating a single proportion is presented and a novel feature is that the distribution of the randomizing attribute need not be known. Designs for multi-proportions estimation are presented in Section 3 while Section 4 indicates a possible direction for multivariate RR designs utilizing attribute-based randomization.

2. Designs for estimation of a single proportion

Suppose we wish to estimate the population proportion π having a stigmatizing attribute S. We begin this section with a design in which the artificial randomization of the Simmons unrelated question design is replaced with attribute-based randomization. We need to choose some non-stigmatizing attribute, A say, which we shall assume for the moment is independent of S, although we shall see later that this requirement is not essential.

Each respondent in a simple random sample (of size n, selected with replacement) is presented with a card such as the following:

[1] Lecturer, Department of Statistics, University College, Cork, Ireland.

If you have attribute A answer from here:	If you do not have attribute A answer from here:
(1) S	(1) U
(2) $\bar{S}$	(2) $\bar{U}$

Here, S denotes a suitably-worded statement admitting to the possession of attribute S, while $\bar{S}$ refers to a statement denying possession of attribute S. The symbols U and $\bar{U}$ denote statements admitting to having, and not having, an "unrelated" attribute U. We shall assume for the moment that attribute U is independent of the attribute A. For obvious reasons of encouraging respondent co-operation it is desirable that U should be independent of S, but this is not essential in deriving the form of an estimator for π.

Let n_1 be the number of respondents giving the reply '1', and let μ and π_A be the proportions of the population having attributes U and A respectively. If λ denotes the probability that a randomly selected respondent gives the reply '1', then we have the familiar equation:

$$\lambda = \pi_A \pi + (1-\pi_A)\mu . \qquad (2.1)$$

If π_A is known, or can be determined by a separate investigation, then the form of equation (2.1) suggests the following estimator for π:

$$\hat{\pi} = \{\hat{\lambda} - (1-\pi_A)\mu\} / \pi_A , \qquad (2.2)$$

where $\hat{\lambda} = n_1/n$. (We are assuming that μ is known.) The above estimator can produce estimates outside the range (0,1), although the chances of this occurring are quite small for the large sample sizes that are likely to be used in an RR investigation.

Up to now in any applications of this approach it has been assumed that π_A is known (or can be established separately). Boruch and Cecil (1979, p. 154) mention a number of RR surveys that have used attribute-based randomization. In one survey the randomization was based on a respondent's preference for color, while in another it was based on the last digit of a telephone number. Boruch and Cecil warn that difficulties can arise from what may appear to be plausible assumptions about the distribution of such attributes as telephone-number digits.

Since π_A appears as a divisor in the expression for $\hat{\pi}$, then an error in the value of π_A can seriously affect our estimate for π. We now present a procedure which eliminates the need for π_A to be known. We split our sample into two groups, and the respondents in the first group are presented with the card described above (card 1), while the respondents in the second group are presented with the following card (card 2):

If you have attribute A answer from here:	If you do not have attribute A answer from here:
(1) U	(1) S
(2) $\bar{U}$	(2) $\bar{S}$

We need to modify our earlier notation:

Let n_i be the size of group i and let n_{i1} be the number of respondents replying '1' in group i, while λ_i is the probability that a randomly selected respondent will reply '1' when faced with card i, $i = 1, 2$. We have

$$\lambda_1 = \pi_A \pi + (1-\pi_A)\mu ,$$

$$\lambda_2 = (1-\pi_A)\pi + \pi_A \mu$$

so that

$$\lambda_1 + \lambda_2 = \pi + \mu . \qquad (2.3)$$

This equation suggests that π be estimated using

$$\hat{\pi} = \hat{\lambda}_1 + \hat{\lambda}_2 - \mu\,, \tag{2.4}$$

where

$$\hat{\lambda}_i = n_{i1}/n_i\,.$$

The estimator $\hat{\pi}$ can produce estimates outside the range (0,1), and the ML estimator for π is given by

$$\begin{aligned}\hat{\pi}_{ML} &= \hat{\pi} \text{ provided } 0 \leqslant \hat{\pi} \leqslant 1\\ &= 0 \text{ if } \hat{\pi} < 0\,,\\ &= 1 \text{ if } \hat{\pi} > 1\,.\end{aligned} \tag{2.5}$$

Asymptotically, this estimator is unbiased, and its asymptotic variance may be estimated by

$$\hat{V}(\hat{\pi}_{ML}) = \frac{\hat{\lambda}_1(1-\hat{\lambda}_1)}{n_1-1} + \frac{\hat{\lambda}_2(1-\hat{\lambda}_2)}{n_2-1}. \tag{2.6}$$

The procedure described above may also be applied when the Simmons unrelated-question design is used with *artificial* randomization, so that it is unnecessary to know the probability that the randomizing device will choose the "sensitive" question.

The final result presented in this section is concerned with the removal of the requirements that attribute A be independent of attribute S and that attribute A be independent of attribute U. We again use the procedure of using two groups or samples, with the first card above being presented to the respondents of the first group, while the second card is used in the second group. The data then consists of $(n_{11}, n_{21}, n_1, n_2)$, as defined earlier.

Let π_{11} be the proportion of the population having the attributes A and S, while μ_{11} denotes the proportion having the attributes A and U. Let π_{21} denote the proportion having attribute S but not attribute A, and similarly for μ_{21}.

We have

$$\lambda_1 = \pi_{11} + \mu_{21}$$

and

$$\lambda_2 = \pi_{21} + \mu_{11}$$

so that

$$\lambda_1 + \lambda_2 = \pi + \mu\,,$$

which is the same as equation (2.3) above. Thus the form of the estimator for π given above in equation (2.5), and the subsequent expressions, will apply whether or not S is independent of A and U is independent of A.

It is easy to show that the allocation of sample sizes to the two groups which will minimize $V(\hat{\pi})$ is such that

$$\frac{n_1}{n_2} = \sqrt{\frac{\lambda_1(1-\lambda_1)}{\lambda_2(1-\lambda_2)}}\,, \tag{2.7}$$

but preliminary estimates of λ_1 and λ_2 are needed to apply this.

3. Designs for multi-proportions estimation

3.1. A design without artificial randomization

There is an obvious extension of the approach in Section 2. Suppose that the population of interest is categorized into 3 categories, two of which are stigmatizing and we wish to estimate the proportion π_i in category i, $i = 1, 2, 3$. The RR design to be presented makes use of attribute-based randomization rather than artificial randomization and the randomizing attribute (A, say) need not be independent of the stigmatizing attribute. Furthermore, we can extend the approach in Section 2 so that the distribution of attribute A in the population need not be known.

The sample is divided into three groups of sizes n_1, n_2, n_3 (it seems sensible that these should be approximately equal but further investigation is required on the sample allocation). Each respondent in each of the groups is presented with a card as illustrated in Figure 1.

Fig. 1. Form of Inquiry Cards for a Multi-Proportions Design

If you are in A_1 answer from here:	If you are in A_2 answer from here:	If you are in A_3 answer from here:
(1) S_1	(1) S_2	(1) U_1
(2) S_2	(2) S_3	(2) U_2
(3) S_3	(3) S_1	(3) U_3

(a): Form of Card for Group 1

If you are in A_1 answer from here:	If you are in A_2 answer from here:	If you are in A_3 answer from here:
(1) S_2	(1) U_1	(1) S_1
(2) S_3	(2) U_2	(2) S_2
(3) S_1	(3) U_3	(3) S_3

(b): Form of Card for Group 2

If you are in A_1 answer from here:	If you are in A_2 answer from here:	If you are in A_3 answer from here:
(1) U_1	(1) S_1	(1) S_2
(2) U_2	(2) S_2	(2) S_3
(3) U_3	(3) S_3	(3) S_1

(c): Form of Card for Group 3

We let A_1, A_2 and A_3 denote three mutually exclusive and collectively exhaustive categories of the randomizing attribute A, while U_i and S_j denote suitably-worded statements admitting membership in category i of the "unrelated" attribute U and category j of the sensitive attribute S, respectively. (The statements S_j, $j = 1, 2, 3$, are of course mutually exclusive and collectively exhaustive, as are U_i, $i = 1, 2, 3$).

Let n_{ij} denote the number of respondents giving the reply j in group i, and let λ_{ij} denote the probability that a randomly selected respondent will give the reply j when faced with card i, $(i = 1, 2, 3)$, $(j = 1, 2, 3)$. Let π_{ij} denote the proportion of the population who are in category i of attribute A and category j of attribute S, while μ_{ij} is the proportion of the population in category i of attribute A and category j of attribute U.

Then, for group 1:

$$\begin{aligned}\lambda_{11} &= \pi_{11} + \pi_{22} + \mu_{31}\,,\\ \lambda_{12} &= \pi_{12} + \pi_{23} + \mu_{32}\,,\\ \lambda_{13} &= \pi_{13} + \pi_{21} + \mu_{33}\,.\end{aligned} \tag{3.1}$$

For group 2:

$$\begin{aligned}\lambda_{21} &= \pi_{12} + \mu_{21} + \pi_{31}\,,\\ \lambda_{22} &= \pi_{13} + \mu_{22} + \pi_{32}\,,\\ \lambda_{23} &= \pi_{11} + \mu_{23} + \pi_{33}\,.\end{aligned} \tag{3.2}$$

For group 3:

$$\begin{aligned}\lambda_{31} &= \mu_{11} + \pi_{21} + \pi_{32}\,,\\ \lambda_{32} &= \mu_{12} + \pi_{22} + \pi_{33}\,,\\ \lambda_{33} &= \mu_{13} + \pi_{23} + \pi_{31}\,.\end{aligned} \tag{3.3}$$

Notice that

$$\begin{aligned}\sum_{i=1}^{3}\lambda_{i1} &= \pi_1 + \pi_2 + \mu_1\,,\\ \sum_{i=1}^{3}\lambda_{i2} &= \pi_2 + \pi_3 + \mu_2\,,\\ \sum_{i=1}^{3}\lambda_{i3} &= \pi_1 + \pi_3 + \mu_3\,,\end{aligned} \tag{3.4}$$

where we have written $\sum_{i=1}^{3}\pi_{ij}$ as π_j and $\sum_{i=1}^{3}\mu_{ij}$ as μ_j.

These equations indicate that π_1, π_2 and π_3 may be estimated (assuming that μ_1, μ_2 and μ_3 are known). For example, an estimator for π_2 is

$$\hat{\pi}_2 = 1 - \sum_{i=1}^{3}\hat{\lambda}_{i3} + \mu_3 \tag{3.5}$$

where $\hat{\lambda}_{ij} = n_{ij}/n_i$.

Asymptotically, this is the ML estimator, with asymptotic variance:

$$V(\hat{\pi}_2) = \sum_{i=1}^{3}\lambda_{i3}(1-\lambda_{i3})/n_i \tag{3.6}$$

which may be estimated by

$$\hat{V}(\hat{\pi}_2) = \sum_{i=1}^{3}\hat{\lambda}_{i3}(1-\hat{\lambda}_{i3})/(n_i-1). \tag{3.7}$$

It may be noticed that this design does not have completely symmetric response, which is a property of the original Warner (1965) design, the unrelated question design of Bourke (1974b) and the multi-proportions design of Bourke and Dalenius (1973). In fact the procedure presented above will break down for completely symmetric designs. However, the design does have what might be termed *partially symmetric response*, since each of the design responses may be associated with *either* at least one sensitive category *or* at least one non-sensitive category.

3.2. A design with both artificial and attribute-based randomization

The design presented here makes use of the *sensitive* attribute to provide the attribute-based randomization. The design may be used to estimate k proportions, but as in Section 3.1, the case of $k = 3$ will be presented.

Each respondent (in a random sample of size n, selected with replacement) undergoes two separate RR trials. The first RR trial makes use of the Simmons unrelated question design, while the second RR trial incorporates attribute-based randomization. In the first RR trial, the respondent operates a randomizing device with two outcomes (such as a coin-toss) which directs him to answer from one section of the following card, by giving the number (1 or 2) of the statement describing his state.

If outcome 1 answer from here:	If outcome 2 answer from here:
(1) S	(1) U
(2) $\bar{S}$	(2) $\bar{U}$

(The symbols S, $\bar{S}$, U, $\bar{U}$ have the same meaning as in Section 2).

In the second RR trial the respondent *does not* operate the randomizing device again but rather uses the outcome from the first trial to choose from which section of the following card to respond:

If outcome 1 answer from here:	If outcome 2 answer from here:
If you have attribute *S*:	If you don't have attribute *S*:
(1) S_2 (2) S_3	Say '2'
If you don't have attribute *S*:	If you have attribute *S*:
Say '1'	(1) S_3 (2) S_2

(The symbols S_2 and S_3 denote statements admitting membership in the two stigmatizing categories of attribute *S*.) Again, the possible responses are '1' and '2'.

Let n_{ij} denote the number of respondents giving the response vector (i,j), $i = 1, 2; j = 1, 2$, and let λ_{ij} denote the probability that a randomly selected respondent will give this response vector. If p is the probability that the randomizing device selects outcome 1 and if μ is the proportion having attribute U, then we have

$$\begin{aligned}
\lambda_{11} &= p\pi_2 + (1-p)\,\mu\,\pi_3\,, \\
\lambda_{12} &= p\pi_3 + (1-p)\,\mu\,(\pi_1 + \pi_2)\,, \\
\lambda_{21} &= p\,\pi_1 + (1-p)\,(1-\mu)\,\pi_3\,, \\
\lambda_{22} &= (1-p)\,(1-\mu)\,(\pi_1 + \pi_2)\,.
\end{aligned} \tag{3.8}$$

The ML estimator for (π_1, π_2, π_3) must be found by numerical methods, i.e. solving the non-linear programming problem of maximizing the log-likelihood:

$$\sum_{i,j} n_{ij} \log \lambda_{ij}$$

subject to $(\pi_1 + \pi_2 + \pi_3) = 1$ and $\pi_i \geq 0$.

It may be noted that this design is partially symmetric rather than fully symmetric. Work is in progress on the comparison of this design with some of the earlier multi-proportions designs.

4. A design for bivariate estimation

The option of using attribute-based randomization has greatly increased the range of possible RR designs for multivariate estimation. The design presented here is for estimation of a $(2 \times k)$ table where not all cells are of interest. The first variate (S, say) is dichotomous, while the second variate (T, say) has k categories (in the analysis below, we take $k = 3$). Frequently one wishes to estimate the distribution of T when S is at level 2 (the stigmatizing level), but for S at level 1 (the innocuous level) one may have no interest in the distribution of T. Thus, one wishes to estimate π_{2j}, $j = 1, 2, \ldots, k$, but one is not interested in π_{1j}. Two cards are drawn up, one for each variate, as in Figure 2.

As in the earlier sections T_1, T_2, T_3 denote statements admitting membership in each of the three categories of attribute T (two of which are stigmatizing), and similarly for S_1, S_2 and U_1, U_2. Attribute U is non-stigmatizing and "unrelated", the proportions in each category being μ_1 and μ_2, assumed known.

Each respondent operates a randomizing device that can produce one of three possible outcomes. This single randomization is used by the respondent to give a reply when the card for variate S is presented, and again later when the card for variate T is presented.

The information gathered by the interviewer is n_{ij}, the number of respondents giving the reply vector (i, j), $i = 1, 2; j = 1, 2, 3$. If λ_{ij} is used to denote the probability that a randomly selected respondent gives the reply (i, j) and if p_i is the probability that the randomizing

Fig. 2. Form of Inquiry Cards for a Bivariate Design

If outcome 1 answer from here:	If outcome 2 answer from here:	If outcome 3 answer from here:
(1) S_1 (2) S_2	(1) S_2 (2) S_1	(1) U_1 (2) U_2

(a): Card for variate S

If outcome 1 answer from here:	If outcome 2 answer from here:	If outcome 3 answer from here:
If in S at level 2 (1) T_1 (2) T_2 (3) T_3	If in S at level 1 Say '2'	If in S at level 2 (1) T_3 (2) T_1 (3) T_2
If in S at level 1 Say '1'	If in S at level 2 (1) T_2 (2) T_3 (3) T_1	If in S at level 1 Say '3'

(b): Card for variate T

device gives outcome i, then we have the following equations:

$$\begin{aligned}
\lambda_{11} &= p_1\,\pi_{1\cdot} + p_2\,\pi_{22} + p_3\,\mu_1\pi_{23},\\
\lambda_{12} &= \quad p_2\,\pi_{23} + p_3\,\mu_1\,\pi_{21},\\
\lambda_{13} &= \quad p_2\,\pi_{21} + p_3\,\mu_1\,(\pi_{22} + \pi_{1\cdot}),\\
\lambda_{21} &= p_1\,\pi_{21} \quad + p_3\,\mu_2\,\pi_{23},\\
\lambda_{22} &= p_1\,\pi_{22} + p_2\,\pi_{1\cdot} + p_3\,\mu_2\,\pi_{21},\\
\lambda_{23} &= p_1\,\pi_{23} \quad + p_3\,\mu_2\,(\pi_{22} + \pi_{1\cdot})
\end{aligned} \tag{4.1}$$

where $\pi_{1\cdot} = \sum_j \pi_{1j}$.

As in Section 3.2, the ML estimator for the vector $\underline{\pi} = (\pi_{1\cdot}, \pi_{21}, \pi_{22}, \pi_{23})$ must be found by a numerical procedure in which the log likelihood

$$\sum_{i,j} n_{ij} \log \lambda_{ij}$$

is maximized, subject to $(\pi_{1\cdot} + \pi_{21} + \pi_{22} + \pi_{23}) = 1$ and each element of the $\underline{\pi}$ vector being non-negative.

5. References

Boruch, R.F. and Cecil, J.S. (1979): Assuring the Confidentiality of Social Research Data. University of Pennsylvania Press, pp. 127–173.

Bourke, P.D. and Dalenius, T. (1973): Multi-Proportions Randomized Response Using a Single Sample. Report No. 68 of the Research Project "Errors in Surveys", Department of Statistics, University of Stockholm.

Bourke, P.D. (1974a): Multi-Proportions Randomized Response Using the Unrelated Question Technique. Report No. 74 of the Research Project "Errors in Surveys", Department of Statistics, University of Stockholm.

Bourke, P.D. (1974b): Symmetry of Response in Randomized Response Designs. Report No. 75 of the Research Project "Errors in Surveys", Department of Statistics, University of Stockholm.

DeLacy, P. (1975): Randomized Conditional Response. American Statistical Association, Proceedings of the Social Statistics Section, pp. 383–386.

Goodstadt, M.S. and Gruson, V. (1975): The Randomized Response Technique: A Test on Drug Use. Journal of the American Statistical Association, 70, pp. 814–818.

Hochberg, Y. (1975): Two-Stage Randomized Response Schemes for Estimating a Multinomial. Communications in Statistics, 4, pp. 1021–1032.

Lavrakas, P.J. (1975): A Randomized Response Technique for Assuring Confidentiality of Data in a Group Survey Situation. Research Memo, Psychology Department, Loyola University, Chicago.

Swensson, B. (1974): Combined Questions: A New Survey Technique for Eliminating Evasive Answer Bias. Report No. 70 of the Research Project "Errors in Surveys", Department of Statistics, University of Stockholm.

Swensson, B. (1977): Survey Measurement of Sensitive Attributes: Some Contributions. Unpublished Ph.D. Thesis, University of Stockholm.

Takahasi, K. and Sakasegawa, H. (1977): A Randomized Response Technique Without Making Use of a Randomizing Device. Annals of the Institute of Statistical Mathematics, 29(a), pp. 1–8.

Warner, S.L. (1965): Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. Journal of the American Statistical Association, 60, pp. 63–69.

Balancing Recall Bias Against Sampling Error in the Swedish TV Audience Surveys

by Anders Christianson[1]

1. Introduction

Research into the field of integrated error models in the realm of survey theory and practice has increased in interest during the last three decades. Although in the beginning of the 1950's the research was still focused on specific sources of errors, such as sampling errors and non-response errors, more and more efforts were devoted to work with integrated error models. This development can be measured by the increasing number of scientific papers and conferences devoted to this topic.

One of the highlights during this period was the joint ISI-IASS session on the balance between sampling and non-sampling errors, especially in official statistics, which was held in Vienna in 1973, organized and chaired by Tore Dalenius.

In one of the papers discussed, Nathan (1973), an analysis of recall bias and sampling errors in the 1971 Israeli Radio and Television Survey was presented. This analysis led to the choice of single day reporting over whole week reporting.

At present the reporting period of the Swedish Television Audience Survey is three days. However, a decision is to be made as to whether or not it should be changed. The alternatives considered here will be one, two, four and five program days. The solution to this problem is mainly a question of achieving a proper balance between the recall bias and the sampling variance.

The purpose of this paper is to gain some insight into the problem as regards the specific type of survey design used by the Audience and Program Research Department of the Swedish Broadcasting Company.

2. The survey design

The TV audience surveys are performed three or four periods every year in order to measure the program audience sizes. About 140 days are covered annually.

Although different types of parameters are estimated, like audience sizes for particular programs or series of programs and the proportion of non-viewers a certain day, this paper concerns only the parameter

U, the average number of minutes spent watching TV per person for a given day, regardless of channel.

The survey is based upon simple random samples of the Swedish population 9–79 years old. Each day during a survey period the interview work is begun with one such sample or sample package, where

n is the sample package size; at present $n = 175$.

The interviews for a particular sample package are carried out during three *interview days*. They are made by telephone and follow a questionnaire, listing all the TV-programs in the two channels in their broadcast order for each *program day* of the reporting period. In general,

k will equal the number of program days in the interview.

[1] Statistical Research Director, Audience and Program Research Department of the Swedish Broadcasting Corporation.

The program days are always those immediately preceding the first interview day and in the current design: $k = 3$. However, the data analysed in this paper come from an experiment performed in the fall of 1981, when a fourth program day was added to the reporting period in ten consecutive sample packages.

The parameter U is estimated by

y, the average viewing time for the k sample packages covering a given day.

3. The problem

The combination ($k = 3$; $n = 175$) that is used in the current survey design gives a total sample size, $nk = 525$, for a normal day, since every day is covered by three sample packages except for the two first and the two last days of the survey period. The age of the observations varies between 1 and 5 days, since there are both 3 program days and 3 interview days.

The age distribution of the observations will depend upon the proportion of the interviews that are completed during the first, second and third interview day and on the number of program days, k.

Changing k therefore means changing the sample size as well as the age distribution of the observations. Decreasing k implies a larger sampling variance and probably a smaller recall bias whereas increasing k gives a smaller sampling variance and an enlarged recall bias.

The problem under consideration is to determine the optimum choice of k and n, that is the one that gives the smallest total error per cost unit.

4. The general approach

The total error of the survey is measured by the mean square error of the estimator y:

$$\text{MSE}y = E\,(y-U)^2.$$

Denoting by

Y, the expected value of y, the mean square error can be decomposed into

$$\text{MSE}y = E(y-Y)^2 + (Y-U)^2 = Vy + B^2y,$$

which is the sum of the sampling variance and the squared bias of y.

The choice of n and k will be made subject to the optimization criterion:

"Minimize the cost for a given mean square error."

The mean square error given will be the result of the current survey design with $k = 3$ and $n = 175$. The sample package sizes, n, that are necessary to produce the same MSE will be computed for $k = 1, 2, 4$ and 5 and the cost will be calculated for the five alternatives.

These computations require, however, a mathematical model describing the way in which recall and other errors are introduced into the survey results. Such a model will be presented in the next paragraph.

5. The model

Let the age of an observation be defined as the number of days, t, between the program day and the interview day. Since the number of interview days is three,

$$t = 1, 2, \ldots, (k + 2) \text{ days.}$$

Let further

$f(t)$ denote the distribution of the observations over age such that $\Sigma f(t) = 1$, and

$b(t)$ denote the recall bias of the observations, which is assumed to be a function of their age.

The final bias of y may then be expressed as a weighted mean of the $b(t)$'s with the $f(t)$'s as weights:

$$By = \sum_{t=1}^{k+2} f(t)b(t).$$

As regards the recall bias function $b(t)$, the following two postulates are introduced:

$b(t) = 0$ for $t = 1$, and
$b(t) = \alpha$ for $t = 2, 3, \ldots, (k+2)$.

This implies that day-old observations are unbiased, whereas older observations are affected by a constant bias α, regardless of how much older they are.

Using these two postulates the final bias of y may be rewritten as

$$By = \sum_{t=2}^{k+2} f(t)\alpha = (1-\pi)\alpha, \text{ where}$$

π denotes the fraction of the observations that are one day old.

$\pi = f(1) = P/k$, where

P denotes the proportion of interviews in the first sample package that are completed on the first interview day.

Then the expression

$$By = (1-P/k)\alpha$$

demonstrates the influence of the number k of program days on the final bias.

The sampling variance of y may be expressed as

$Vy = S^2/kn$, where

S^2 is the population variance.

Using these expressions for By and Vy the mean square error of the estimator y may be expressed as

$$\text{MSE}y = (1-P/k)^2\alpha^2 + S^2/kn.$$

Thus, with estimates of the parameters P, α and S^2, the mean square error of y may be computed for a given combination $(k; n)$.

6. Estimating the parameters

The data used for estimating the parameters originate from an experiment during 10 days in October and November 1981, in which an extra program day was added in the interviews. The overall mean for the 10 days was

$y = 123.4$ minutes of viewing per day.

This mean was decomposed into 12 estimates, one for each combination of the four program days and three interview days:

		Interview day		
		1	2	3
Program day	1	$y_{11} = 123.5$	$y_{12} = 111.6$	$y_{13} = 96.2$
	2	$y_{21} = 129.2$	$y_{22} = 112.0$	$y_{23} = 111.6$
	3	$y_{31} = 128.8$	$y_{32} = 115.7$	$y_{33} = 101.9$
	4	$y_{41} = 130.7$	$y_{42} = 112.8$	$y_{43} = 103.6$

Here, the first index refers to program day and the second to interview day. For instance, y_{32} is based only on third program day estimates for interviews completed on the second interview day. The first program day is always the one preceding the first interview day, the second program day is the day before first program day and so on. This is the order in which the program days are taken at the interview.

The second and third interview days estimates seem to be affected by a selection effect. Therefore, only the left column of the above table will be used to estimate α. Since, under the recall bias function

$Ey_{11} = W$, and

$Ey_{21} = Ey_{31} = Ey_{41} = W + \alpha$, where

W is the expected value of first program day estimates based only upon interviews completed during the first interview day,

α may be estimated by

$$\hat{\alpha} = (y_{21} + y_{31} + y_{41})/3 - y_{11} = 6.1.$$

P is estimated by the average proportion, p, of the interviews completed during the first interview day during the experiment period, where

$$p = .75.$$

For S^2 the average sample variance s^2 will be used as an estimate. During the experiment period it was calculated to be

$$s^2 = 9490.$$

7. The mean square error

The estimates that were computed in the preceding paragraph will be used here to demonstrate how the model may by used to calculate the mean square error for alternative combinations of n and k.

In the next table, the results for $n = 175$ and $k = 1, 2, \ldots, 5$ are presented in terms of final mean square error as well as the variance and the squared bias components.

No. of program days (k)	Sample size nk	Vy	+ B^2y =	MSE
1	175	54.2	2.3	56.6
2	350	27.1	14.5	41.7
3	525	18.1	20.9	39.0
4	700	13.6	24.6	38.1
5	875	10.8	26.9	37.7

The table indicates that the substantial fall in MSE is obtained when the number of program days is increased from one to two.

After that, the decrease of MSE is rather slow since the decreased variance is almost balanced by an increased recall bias.

The MSE of the present system with 3 program days and $n = 175$ is estimated at 39.0 minutes, of which 18.1 is variance and 20.9 is squared bias.

The sample package sizes necessary to produce approximately the same mean square error for $k = 1, 2, 4$ and 5 are derived by means of the formula

$$n = \frac{S^2}{(\mathrm{MSE}_y - B^2y)\,k} = \frac{9490}{(39 - B^2y)\,k},$$

where the B^2y's of the preceding table may be used, since they do not depend upon the sample package size.

The combinations of k and n producing the same MSE (39.0) are tabulated below for $k =$ 1 to 5, together with their variance and squared bias components:

k	n^*	Vy	B^2y	MSEy
1	259	36.7	2.3	39.0
2	194	24.4	14.5	39.0
3	175	18.1	20.9	39.0
4	164	14.4	24.6	39.0
5	157	12.1	26.9	39.0

* rounded to the nearest integer.

The five combinations of k and n are equivalent in terms of mean square error. The difference between them lies in the proportion of the variance and squared bias components adding to the same mean square error.

Although, from the viewpoint of mean square error, any one of the combinations might be chosen, the natural choice is the combination that gives the lowest cost.

8. Cost evaluation

The cost for an interview is assumed to be the sum of two terms,

$C = C_0 + kC_1$, where
C_0 = the initial cost for an interview and
C_1 = the additional cost per program day.

The cost for a whole sample package may then be expressed as

$$C(y) = n(C_0 + kC_1).$$

Although this is the cost for a sample package it will be used as the cost for an estimate y of a program day. This may create a difference for a particular day, which however evens out in the long run.

Thus, the expression for $C(y)$ provides a means of evaluating the cost for y for a given combination $(k; n)$, when estimates of C_0 and C_1 are available.

The following estimates will be used in the cost function:

C_0 = 45:30 Skr (Swedish Crowns), and
C_1 = 7:20 Skr.

These figures mirror the fact that most of the cost for an interview is accumulated in the preparatory work and a relatively small part in telephoning and interviewing the respondent. Using the above figures, the following daily costs for the five alternatives are calculated:

k	n	Cy (Skr)
1	259	13 597:50
2	194	11 581:80
3	175	11 707:50
4	164	12 152:40
5	157	12 764:10

The results favor the $(k=2; n=194)$ system with a daily estimated cost of 11 581:80 Skr. However, the combination $(k=3; n=175)$ is only slightly more expensive, and a small change in the estimate of the parameter α may change the picture substantially.

9. Final remarks

The results of the preceding paragraph suggest that a reporting period of two program days – one less than in the current design – is the best choice for achieving a rational balance between recall bias and sampling variance. It must be stressed, however, that these calculations were made to illustrate the potential use of the model. The restrictions listed below demonstrate that further research is still necessary before a final decision can be made.

i. The choice of the type of recall bias function is, to a large extent, subjective. The sudden increase in recall bias after one day is suggested by the variation between the observations y_{11} to y_{41} demonstrated in Paragraph 6. Some preliminary analyses of data from other periods seem to indicate, at least in one case, both a later introduction and another magnitude of the bias.

ii. A good standard for validating the observations has not been available. In this study, day-old observations have been assumed to be unbiased. Of course, the results of the survey may be affected by an overall bias, for instance caused by non-response. If this is the case, which is plausible, the parameter α should be considered as measuring the change in bias rather than the bias itself.

iii. The analysis is restricted to a particular parameter, e.g. average time spent viewing television per individual on a given day. The optimum for this parameter may be far from optimum for other parameters. For instance, if the parameter is the mean daily viewing time for a period, the observations will be accumulated and the sampling variance thus reduced. If the recall bias is approximately constant, the bias term will then be too large for the design to be in balance with respect to this parameter.

On the other hand, the design which is balanced with respect to the average daily viewing time for the whole population may provide too large a variance for domains, such as age groups in which the sample sizes are smaller.

Again, if recall biases are approximately constant over time, they will tend to cancel

out when differences between time periods are estimated.

Finally it should be mentioned that the approach may be used to evaluate only a limited class of survey designs. Alternatives like those based on audiometer panels could not be evaluated by this method since the error picture is quite different.

With all these restrictions in mind, the analysis might however still provide some understanding of how to balance the different sources of error.

10. Reference

Nathan, G. (1973): Utilization of Information on Sampling and Non-Sampling Errors for Survey Design – Experiences in Israel. International Statistical Institute, Proceedings of the 39th session, book 3, pp. 393–406.

The Development of Procedures for Industry and Occupation Coding at Statistics Sweden

by Lars Lyberg[1]

1. Introduction

Coding is a major operation in such statistical studies as censuses of population, censuses of business and labor force surveys.

The coding operation has three components:

(1) Each element in, for instance, a population is to be coded with respect to a specific variable by means of verbal descriptions.

(2) There exists a code for this variable, i.e. a set of code numbers in which each code number denotes a specific category of the variable under study.

(3) There is a coding function relating (1) and (2), i.e. a set of coding instructions relating verbal descriptions with code numbers.

The problems with coding are of different kinds. As with most other survey operations, coding is susceptible to errors. The errors occur because the coding function is not always properly applied and because either the coding function itself or the code is improper. In fact, in some statistical studies coding is the most error-prone operation next to data collection. For some variables error frequencies at the 10 per cent level are not unusual. Another problem is that coding is difficult to control. Accurate coding requires a lot of judgement on the part of the coder, and it can be extremely hard to decide upon the correct code number. Even experienced coders display a great deal of variation in their coding. Thus there are problems in finding efficient designs for controlling the coding operation. A third problem is that many coding operations are difficult to administer. Coding is generally time-consuming and costly: for instance, in the 1970 Swedish Census of Population carrying out the coding took more than 300 man-years. In many countries coders in large-scale operations must be hired on a temporary basis, and the consequences for maintaining good quality are obvious. There are even reasons to believe that in the future it might be difficult to obtain even temporary coders for this kind of relatively monotonous work. So there is certainly room for new ideas on the effectiveness of the coding operation.

Much development work on coding has been carried out at Statistics Sweden during the last 20 years. Professor Tore Dalenius has played a very important part in this work. He has supervised various studies and he has continuously suggested new ideas. In other words, he has clearly been at the forefront in this field.

This paper reviews the development work at Statistics Sweden from 1965 to this day.

2. Different kinds of coding

Coding can be carried out manually outside an agency, manually within an agency, or automatically by a computer.

Manual coding outside an agency is more common than one might think. In the continuous Swedish Labor Force Survey the coding of occupation is done by interviewers. In the 1975 Swedish Census of Population local authorities coded some of the census variables

[1] Head of the Survey Research Unit, Statistics Sweden.

in order to make it possible to produce some employment statistics without the usual time lag. A third example is when the respondents themselves code different variables. At one extreme we have the case when respondents receive a copy of the code and are asked to use it for coding purposes. At the other extreme we have the case called "self-coding" where the respondent is presented with a number of fixed alternatives and asked to choose one of them.

Coding outside an agency shifts the burden of coding. Generally the procedure generates low coding costs for the agency but the control of the agency over the coding is reduced as well.

The literature on evaluation of this kind of decentralized coding is not very extensive.

Manual coding inside an agency (centralized coding) is very common. At Statistics Sweden such coding is used in over 100 surveys each year. However, in many of these cases, the coding operation is small and presents no serious problems. In other and more interesting cases, we have large-scale coding on a continuous basis. Examples are the censuses, in which Statistics Sweden during the past decades has hired around 300 coders on a temporary basis for each census. The coding operation is complex and many variables must be coded for each element. Such operations put tremendous pressures on the central staff and the organization. Examples of such operations are continuous surveys where variables such as occupation, industry, education and employment status are coded. Such surveys are the Labor Force Survey (both centralized and decentralized coding), different Pupil Surveys, Income Distribution Surveys, etc. The coding in these surveys is done by a regular coding staff.

Large-scale coding is costly and time-consuming and computer coding could be an attractive alternative. Automated coding is used as a complement to manual coding: manual coders take care of cases rejected by the computer. The methods for automated coding have been used experimentally for some years. Quite recently it was applied for the first time in production, in the 1978 Household Expenditure Survey, where the variable "goods" was coded by computer. Immediate successors were the Swedish Author's Fund where authors and book titles were coded, and the 1980 Census of Population where occupation and socio-economic classification were coded.

These different types of coding operations generate specific problems. In a decentralized coding situation there are problems involved in supervising and controlling the operation. In a centralized situation one often faces complex coding and has to make compromises between quality and cost. In computerized coding one must administer parallel manual systems and be very alert with respect to the performance of the computer programs.

Thus the main dimensions of the coding operations are:

(a) manual vs. computerized;

(b) centralized vs. decentralized.

In practice, combinations of these dimensions may be used.

3. Coding errors

3.1. The meaning of "coding error"

In this paper it is assumed that a true code number exists for each element with respect to the variable under study. A coding error occurs if an element is assigned to a code number other than the true one. This seemingly simple definition needs some further elaboration. Three different aspects will be considered.

First, it is often difficult to decide upon a true code number. The basic assumption is that every element belongs to one and only one category. In practice there are difficult situations where a specific description is such that it can be assigned different code numbers depending on interpretation.

Second, even if the description is detailed, problems might arise in assigning true code numbers. Who are the experts to decide them? Studies show that the variation between "experts" can be considerable and as a consequence true code numbers often have to be defined operationally. For instance, three or more experts code the same element. By means of a majority rule a true code number is assigned; i.e. the code number assigned by a majority of the experts is considered the true one.

Third, consider the following example. A dentist fills out a mail questionnaire. One of the questions is "What is your occupation?" For some reason or another the dentist answers "brain surgeon". Thus the information available to the coder is "brain surgeon"; if the code number of "brain surgeon" is, say, 411 and the coder assigns 411 then the coding is correct. If any other code number is assigned, including the one for dentist, a coding error has occurred. Thus we say that the coding operation starts with the available element description, whether or not it is proper, and ends with the assignment of a code number. This limitation is practical from the standpoint of control. The obvious response error in the example must be dealt with by other means.

When the code has two or more digits, the notion of a coding error must take this fact into account. Assume for example that six coders code a specific element with respect to the variable industry using a six-digit code. The outcome may be as follows:

Table 1.

True code number		3	6	9	9	2	2
Coder	1	3	6	9	9	2	1
	2	3	6	9	9	1	1
	3	3	6	9	1	1	0
	4	3	6	2	0	9	0
	5	3	5	5	1	1	0
	6	2	1	0	0	0	0

According to the definition of "coding error" all six code numbers assigned are wrong since none of them coincides with the true one. However, the errors are of different kinds. The first code number, 3 6 9 9 2 1, differs from the true one with respect to the last digit. The second coder has been able to code correctly down to the fourth level. The third, fourth and fifth coders have been able to code correctly down to the third, second and first level, respectively. The sixth coder has not been able to assign even the first digit correctly, and as a result the element has been coded to the main group "mining" instead of "manufacture". The point is that as soon as an error occurs on a specific level all subsequent levels are erroneously coded as well. Furthermore, as a consequence of the way an x-digit code is constructed, an error in the first position is more serious than an error in the second position which, in turn, is more serious than an error in the third, etc. For instance, an error in the sixth position only does not affect the quality of a presentation of results on the fifth level, but an error in the first position affects the presentation on results on any level.

The error rate is the number of incorrectly coded elements divided by the total number of coded elements. The error rate can be calculated for a specific material, for a specific time period, for individual coders and for different levels of the code.

Such error rates are *gross* errors. Normally the results of the coding operation are displayed in statistical tables. The coding errors which remain in the table are *net* errors and can sometimes be much less than the gross errors, since the errors tend to some extent to cancel out.

3.2. *The magnitude of the coding error*

3.2.1. The 1965 Swedish Census of Population

In 1967 an evaluation study of coding errors in the 1965 Swedish Census of Population was

conducted (see Dalenius and Lyberg (n.d.)). From a population of census material comprising about 70 per cent of the 1965 population a two-stage sample of verified census schedules was selected. The population was partitioned into four strata and four subsamples were obtained. The evaluation study was confined to industry among other variables. The code used for industry was a three-digit code.

Since we were dealing with four subsamples we obtained four different estimates of error rates. These are given in Table 2a below. The four strata reflected the organization of the census operations. Thus subsample 1 consisted of large city schedules subjected to total verification. Subsample 2 consisted of large city schedules subjected to sample verification. Subsample 3 consisted of totally verified schedules from other areas and subsample 4 consisted of sample verified schedules from these other areas.

Table 2a. Estimates of error rates (%) in production coding of industry in the 1965 Census of Population

Subsample	Error rate
1	14.5
2	8.2
3	14.5
4	8.7

Most of the total verification was done for still inexperienced production coders; this explains the differences in error rates between total and sampling verification.

In Lyberg (n.d.) studies concerning within-coder replication in the evaluation of the 1965 Swedish Census are presented. Each of the three coders X, Y and Z in the experiment made one original coding (trial 1). After three weeks the material was coded once again by the same coders (trial 2). These independent trials gave the estimates shown in Table 2b of the within-coder variability $\hat{P} = m/n$, where n is the total number of coded elements for the specific variable and m is the number of elements differently coded when comparing the two trials.

Table 2b. Estimates of within-coder variability, $\hat{P}$ (%), in the 1965 Swedish Census of Population

Variable	$\hat{P}$ for coder			Subsample
	X	Y	Z	
Industry	7.4	5.7	4.5	1
	6.5	5.2	3.2	2
	6.7	1.8	1.8	3
	5.8	5.3	2.9	4

Industry variability ranges from 1.8 to 7.4 per cent; this illuminates the difficulty of coding more complex variables.

When we are dealing with variables such as occupation and industry we must use hundreds of categories. The codes must therefore be multi-digit and they are built upon the principle of Chinese boxes; i.e. the first digit in the code associates the element with a major group, the second digit in the code associates the element with a subgroup, and so on. Statistical tables often present only the major group data. Data for subgroups and minor groups, i.e. those represented by the last digits in the code, are available on request for special purposes. As pointed out above the conclusion must be that errors in major group coding are more serious than errors in the last digits. In addition, an error in a major group automatically results in errors in subgroups and decimals. For most purposes errors in, say, the third or fourth digit (or sometimes the fifth or sixth) are of little vital importance. A self-evident question arises: are most of the errors in occupation coding, for instance, concentrated in the last digits? If so, an estimated total error rate of, say, 12 per cent is very misleading.

In the evaluation of the 1965 Swedish Census the error distribution for the three-digit variable industry was studied. The three digits

represent a major group, a subgroup, and a minor group within the subgroup (decimal). The error distributions are displayed in Tables 2c and 2d. The results are given separately for the four different subsamples and for the regular coder and the verifier.

Table 2c. Error distribution for industry coding in the 1965 Swedish Census of Population, regular coder

Sample	Major group	Sub-group	Decimal	Total number of errors
1	41	17	22	80
2	22	8	7	37
3	37	17	20	74
4	28	8	16	52
Unweighted total	128	50	65	243

Table 2d. Error distribution for industry coding in the 1965 Swedish Census of Population, verifier

Sample	Major group	Sub-group	Decimal	Total number of errors
1	28	9	16	53
2	18	7	8	33
3	27	10	15	52
4	23	4	15	42
Unweighted total	96	30	54	180

3.2.2. The 1970 Swedish Census of Population

In the 1970 Swedish Census of Population the number of variables to be coded increased over that in 1965. For evaluation purposes a sample was drawn from the population of census schedules.

A pool of expert coders was used to generate a set of "true" evaluation code numbers for each schedule in this sample. These code numbers were compared with the production code numbers after verification, and this led to estimates of error rates for the different variables on economic activity.

Estimates of error rates for industry and occupation are given in Table 2e.

Table 2e. Estimated error rates in coding economic activity in the 1970 Swedish Census of Population

Variable	Code	Per cent error rate (total population)
Occupation	3-digit	13.5
Industry	4-digit	9.9

The table shows that these variables are difficult to code.

In the evaluation of the 1970 Census coding, the experiments for investigating the within-coder variability were repeated. This time five expert coders were used. The results are given in Table 2f.

Table 2f. Estimates of within-coder variability (%) in the 1970 Swedish Census of Population

Variable	Expert coders				
	A	B	C	D	E
Occupation	8.0	10.6	10.9	9.2	7.1
Industry	3.7	8.8	11.6	6.9	5.4

The estimates are based on sample sizes ranging from 300 to 1 000; this range reflects the fact that the economic variables are coded only for economically active persons. There were also some differences in expert coder work load.

As Table 2f shows, the within-coder variability is considerable. This is disturbing when we remember that these coders were used as producers of "true" code numbers to evaluate the coding operation.

The experiment with error distribution was repeated during the evaluation of the 1970 Swedish Census coding. This time occupation (a three-digit code) was also investigated, and the industry variable had become four-digit, the first two digits representing the major group and the last two the subgroup. The results are given in Tables 2g and 2h. The results here are given for different subsamples after verification.

Table 2g. Error distribution for occupation coding in the 1970 Swedish Census of Population, after verification

Sample	Major group	Sub-group	Decimal	Total number of errors
1	67	36	56	159
2	84	39	52	175
3	68	13	35	116
Unweighted total	219	88	143	450

Table 2h. Error distribution for industry coding in the 1970 Swedish Census of Population, after verification

Sample	Major group	Sub-group	Total number of errors
1	77	22	99
2	88	35	123
3	56	18	74
Unweighted total	221	75	296

3.2.3. The 1975 Swedish Census of Population

The number of variables was smaller in the 1975 Census of Population than in the 1970 Census. Evaluation studies show that the error rates also were smaller in this census than in the 1970 Census. As for industry and occupation estimated error rates are given in Table 2i.

Table 2i. Estimated error rates in coding economic activity in the 1975 Swedish Census of Population

Variable	Code	Per cent error rate (total population)
Occupation	3-digit	7.8
Industry	4-digit	3.5

The results given in this table differ strikingly from those obtained in the 1970 evaluation study. The error rates have dropped for both variables. The occupation error rate of almost 8 per cent is still very serious, but compared to the 13.5 per cent rate in 1970 it is a good result. Even better is the estimate for industry.

The error distribution experiments gave the results shown in Tables 2j and 2k.

Table 2j. Error distribution for occupation coding in the 1975 Swedish Census of Population, after verification

Major group	Sub-group	Decimal	Total number of errors
138	68	116	322

Table 2k. Error distribution for industry coding in the 1975 Swedish Census of Population, after verification

Major group	Subgroup	Total number of errors
83	61	144

The tables show a pattern which is quite clear: serious errors are most frequent. This might seem remarkable at first glance since intuitively it does not appear to be so difficult to choose between, say, ten major occupation groups. One would expect that this would be carried out correctly and that then some difficulties might perhaps appear in the details of the code. On the contrary, things seem to work the other way; provided the major group is identified correctly it is relatively easy to code the details. This indicates that we have problems with the code dictionary and that the dividing-lines between major categories may be diffuse.

In Lyberg (1974 b) and in Andersson (1977) the evaluation major group code numbers for occupation were compared with the outgoing production code numbers in the 1970 and 1975 Censuses.

It was found that the marginal distributions were rather similar; this illustrates the fact that gross errors cancel out to some extent in statistical tables.

4. Control of coding operations

4.1. The need for control

We have noted that error frequencies can be substantial; i.e. the gross errors can be large. Such errors may or may not have a significant impact on survey results (for instance, estimates in statistical tables). The general experience seems to be that coding errors do not affect tables of over-all statistics very much, since gross errors have a tendency to cancel out and become rather small net errors.

Statistical tables on over-all statistics are seldom, however, the single and final output from surveys. Statistics in breakdowns may be seriously in error even if over-all statistics are not. Besides, most surveys are multi-purpose and coded materials are often saved for future known or unknown analyses. It is common that a coded material with large gross errors is presented as a frequency distribution, say N_1, N_2, ..., N_k for k categories, where the net effects of coding errors are small. After some time it is decided that a new survey or a special analysis should be carried out for individuals belonging to one or a few specific categories. Now, the gross errors may become serious as discussed above.

Other difficult situations occur when the material is used in cross-classification or in prediction.

These problems make it imperative in most statistical series that coding control be part of the over-all program for producing the statistics. However, knowledge of the error rate is not enough if we want to be far-sighted. We need to know about the error structure, the reliability of the coding process, different types of errors, the seriousness of different errors and the effects of errors, in order to take suitable corrective measures with respect to the code or the coder.

It should be pointed out, though, that the control called for must be designed so as to take it into account that there are also other sources of errors. What is called for is a rational balance of various control efforts, as discussed in Dalenius (1974).

4.2. Schemes for statistical quality control

Manual coding can be characterized as an endless sequence of operations, and it thus seems rather well suited for the application of statistical quality control schemes as originally developed for industrial applications. More specifically, control of coding could be based on various quality control sampling plans. However, coding differs somewhat from, say, car manufacturing. Often there is a problem in finding the true code number, and this therefore forms a sharp contrast to the situation where the diameter of a screw nut is to be checked. As a consequence, errors of the first and second kind are usually much more common in administrative applications than in other quality control areas. Furthermore, it is often impractical or even impossible to establish risk functions for producers and consumers, since coding is only one part of the statistical production process. Nevertheless statistical quality control has been used for several decades as a means for keeping the desired quality level of coding.

A sampling inspection plan can assure quality for any pre-specified level. The literature discussing such plans is extensive; an early example is Dodge and Romig (1944).

The statistical quality control of coding aims at controlling the code numbers assigned (a control oriented towards the user of data) and at controlling the coder (a control oriented towards the producer of data). The major instruments available are acceptance sampling, process control, and combined procedures which utilize both acceptance sampling and process control. Applications of these techniques may be found in Fasteau et al. (1964), Minton (1969), US Bureau of the Census (1965) and Minton (1970).

4.3. Specific coding control schemes

There are certain control schemes designed specifically for coding. These schemes are applicable in three different areas, namely

- Training of coders
- Dependent and independent verification
- Evaluation

4.3.1. Training of coders

The training and education of coders is indeed valuable since the error rate often decreases with time. If it is possible to "cut" error rates at the beginning of a coding process one will probably end up with a higher average outgoing quality.

The literature on the training and education of clerks is not very extensive. However, the subject is discussed in Minton (1969) and in Dalenius and Frank (1968).

4.3.2. Verification

There are two main schemes for verification of coding, i.e. for deciding whether or not a code number is correctly assigned: dependent and independent verification. In dependent verification the verifier has access to the code number assigned by the production coder. In independent verification the verifier has no such access and the decision on the outgoing code number must be based on different rules such as majority or modal rules. Within these major schemes several realistic subschemes can be defined. Dependent and independent verification is dealt with in Lyberg (n.d.), Lyberg (1969) and Minton (1969).

Let us start with a definition of dependent verification.

An element is coded by production coder A. The code number is then reviewed by verifier B. B inspects the code number assigned by A and decides if it is correct or not. If it is considered correct it remains unchanged; otherwise B changes the code number.

With this type of verification, experience tells us that the verifier has a tendency to let erroneous code numbers remain unchanged: his judgement is influenced by (*depends* on) the code number assigned by A. Various studies show that the proportion of incorrect code numbers which remain unchanged can be substantial. At Statistics Sweden we often use 50 per cent as a rule of thumb.

Independent verification is defined as follows:

An element is production-coded by a coder A. The code number is denoted x_A. The element is also coded by N other coders, where $N \geq 1$. Their code numbers are denoted $x_B, x_C, \ldots, x_{N+1}$. Each coder works independently and as a consequence does not know how the others have coded. Thus we end up with a set of code numbers $x_A, x_B, \ldots, x_{N+1}$. These code numbers are matched and a decision rule defines the outgoing code number.

This definition gives rise to at least two questions. How do we create a situation in which each coder works independently of the other coders? What decision rules are possible?

The first question is a matter of administrative resources. One option is to code directly on the schedule and then mask the code numbers after each coding. This is a rather clumsy procedure, however, and is seldom used. Another option is to copy the schedules to be verified. This procedure is costly, of course, but it has been used in some studies. A third alternative is to code on special forms. This is the best alternative so far, and could be very smooth in a computerized environment, where the matching is done by a computer program.

There are various possible decision rules. One rather natural one is the majority rule: if a majority of the $N+1$ coders involved agrees upon a specific code number, then this code number is the outgoing one. (If a majority is not reached, then special measures are

taken.) One early example of this rule is the use of the three-way independent verification system in the coding process of the 1960 US Censuses of Population and Housing. With three coders involved in each decision we have three possible outcomes, which can be denoted 3-0, 2-1 and 1-1-1; a majority is reached in all cases except the last.

A very natural way to improve the efficiency is to use a sequential procedure. In the three-way system this means that we start with two coders. After the matching of their code numbers, it is decided whether a third coder is needed or not. Obviously if the outcome of the first matching is 2-0, then the code number of the third coder is completely unnecessary to reach a majority. The third coder's contribution could only lead to either 3-0 or 2-1, and we have probably wasted some money. However, if the outcome is 1-1 the third coder must enter the scene and his contribution leads to a majority 2-1 case or a 1-1-1 case. This sequential system, sometimes called the two-way independent system, was used in the 1970 US and 1970 Swedish Censuses of Population and Housing.

4.3.3. Evaluation

Evaluation of coding results provides a basis for the allocation of quality control efforts. We have already given examples of results from different evaluation studies. The results of such studies give suggestions concerning the size and emphasis of the necessary quality control program.

Evaluation studies assume the existence of "true" code numbers, which usually are those generated by more skilled coders or expert coders. By comparing these true code numbers to those assigned by the production coders, an estimate of the gross error rate for the production coding can be obtained.

It is obvious that many of the coding errors do not depend on the ability of coders: the codes and the coding manuals may be insufficient and thus cause great variability in the coding process. Improvements of these tools therefore seems to be an urgent task.

4.4. Automated coding as a control technique

Manual coding is error-prone, and within- and between-coder variabilities are often considerable. An evaluation of manual coding from the viewpoint of control reveals that the error-detection process is complicated by the difficulty of defining the true code numbers and by difficulties of supervising the coding staff. There are problems in implementing the coding instructions together with all the changes and code supplements which always show up with probability one in coding operations of at least moderate size. Furthermore, variabilities exist at all levels and all times in the process. For instance, a single coder has a tendency to look upon the same coding problem differently at different times. Two or more different coders have an even stronger tendency to look at the same coding problem differently at the same time. Different teams have a tendency to look at the coding problems from different points of view. This pattern applies to experienced coders too.

Generally, manual coding and the necessary control efforts consume lots of time and money. This fact, and the fact that the process generates many errors call for some new solution. A possible alternative is to let a computer take over the process, at least to a large extent.

Computerized or automated coding can be carried out in the following way. The code is stored in the computer. (This code is usually referred to as a "dictionary".) The (verbal) information to be coded is input into the computer, which is programmed to "match" the input with the dictionary. For those cases with a "match", the corresponding code number is assigned; other cases are referred to manual coding.

The variability in manual coding is not present in computerized coding (except, naturally, for those elements which are referred to manual coding). However, the method presents new problems.

5. Verification schemes

5.1. Dependent verification

In 1967 a study, presented in Lyberg (n.d.), was conducted in order to illuminate, in a concrete fashion, the performance of the dependent verification used in the 1965 Swedish Census of Population. A general evaluation of the coding was obtained as a byproduct.

In our Swedish study verified schedules from the 1965 census material were sampled from four strata covering about 70 per cent of the population. The stratification was based on geographic area and on whether the schedules were verified totally or on a sampling basis. The study was confined to the industry variable among others.

The samples were coded by a team of three experimental coders who were considered especially skilled. Each coder coded independently of the others and then the code numbers were matched. Three cases were possible. First, all three coders could agree; we call that case 3-0. Second, two could agree but not the third; we call that case 2-1. Finally, we have the case when no two coders agree; we call that case 1-1-1. In the first and second cases, clearly, we were able to define a majority code number; that number was used as an evaluation code number. In the 1-1-1 cases we let an expert decide the evaluation code number. After that the evaluation code number was compared both to the unverified production code number (P) and to the dependent verification code number (V). Tables 3a–b show the results for industry for one of the subsamples.

Table 3a. A comparison between the majority code number and the production code number (P) (number of cases)

Industry. Subsample 1

Experimental coder combinations	P agrees with ... experimental coders 3	2	1	0	Expert	Total cases
3-0	427	–	–	48	–	475
2-1	–	41	24	8	–	73
1-1-1	–	–	–	–	5	5
						553

Table 3b. A comparison between the majority code number and the dependent verification code number (V) (number of cases)

Industry. Subsample 1

Experimental coder combinations	V agrees with ... experimental coders 3	2	1	0	Expert	Total cases
3-0	451	–	–	24	–	475
2-1	–	44	23	6	–	73
1-1-1	–	–	–	–	5	5
						553

This is a difficult variable for the coders: they agreed in only 475 of the 553 cases (86 per cent). The error rate in production coding, estimated by (48 + 8 + 24)/548, is a striking 14.6 per cent.

The 80 cases within the triangle in Table 3a were reduced to 53 by means of the verification system. This time the system took care of only 34 per cent of the errors.

As pointed out in Linebarger et al. (1976), dependent verification has certain advantages, though. The operation is quick, fairly non-disruptive to handle, requires little work if handled clerically, and is rather inexpensive. The serious disadvantages are to be found in the quality dimension. But if the disadvantages of a verification system are found in the quality dimension we are in big trouble.

5.2. *Independent verification*

In Dalenius and Frank (1968) and in Lyberg (1969) several alternative schemes for independent verification are discussed. The schemes differ in complexity and in cost and probability characteristics.

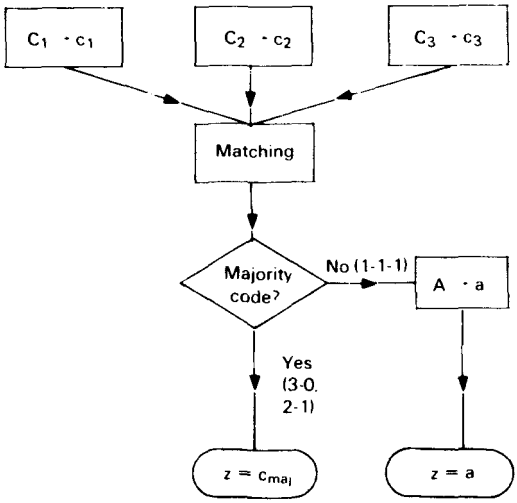
Here, we give two examples of such schemes.

Scheme No. 1 (S1)

Team formation

The coding team consists of three production coders chosen at random and a referee A. The code numbers assigned by these coders are denoted c_1, c_2, c_3, and a respectively. z is the outgoing code number of the scheme.

The coding decision rule is diagrammed below:



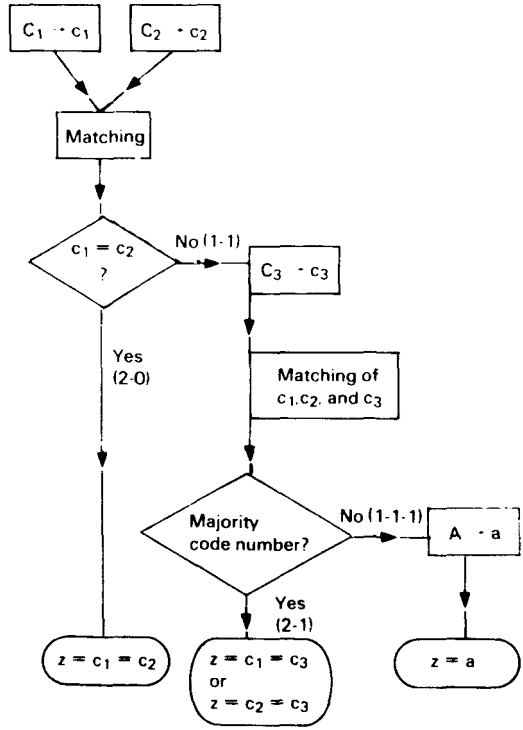
This scheme is called three-way independent verification.

Scheme No. 2 (S2)

Team formation

The coding team is the same as in S1.

The coding decision rule is diagrammed below:



It can be shown that S1 and S2 have the same probability characteristics. Thus S2 should be preferred in most situations.

Independent verification is a more reliable process than dependent verification: it generally allows more accurate estimation of errors and the coding results are more credible. However, the advantages of dependent verification, listed in Linebarger et al. (1976), simultaneously constitute the drawbacks of independent verification.

The basic idea in all independent schemes is that the outgoing code number is very likely to be the correct one. Thus very seldom will a minority coder be unjustly charged with an error. However, it is important that the coders whose code numbers are compared be of approximately equal ability. In Harris (1974) it is pointed out that two "poor" coders could overrule one "good" coder simply because the former have not read the instructions properly.

In Boston (1977) some doubts are thrown on the basic idea of independent verification. An expert evaluation of cases of coding agreement shows that the error rate in occupation coding cannot be neglected: a series of tests showed that the majority code number was incorrect in .7 to 1.8 per cent of the cases. However, everything depends on what is meant by "highly likely to be correct".

Despite what has been said about error rates in agreement cases, independent verification is the best system we have today in terms of quality. However, the administrative apparatus required and the costs can be large.

6. Automated coding: the challenge

As illustrated, manual coding has various drawbacks. Especially it is time-consuming, costly, error-prone and boring.

To cope with these drawbacks, it appears inevitable to focus on the very basis of manual coding and to consider the possibilities offered by access to a computer of developing a basically new approach. This idea is, of course, not in principle new: for instance, at the US Bureau of the Census geographic coding has been conducted by means of computer since 1963. What is new is the suggestion in that agency that the computer be used extensively in the coding of such complex variables as occupation and industry. This suggestion may be viewed as a natural extension of earlier uses of computers in the editing operations.

During the past decade we have conducted a series of experiments at Statistics Sweden in order to find out whether or not it is possible to automate the coding process.

We distinguish four operations:

(i) Construction of a computer-stored dictionary;
(ii) Entering element descriptions into the computer;
(iii) Matching and coding;
(iv) Evaluation,

and will describe them briefly in turn.

6.1. *Construction of a computer-stored dictionary*

In automated coding a dictionary stored in the computer takes the place of the coding instructions used in manual coding. Obviously the construction of such a dictionary is a very important task. The construction work could be carried out manually but, when dealing with multi-digit variables, using the computer seems to be a better alternative. The resulting dictionary should consist of a number of verbal descriptions with associated code numbers. The descriptions could be a sample from the population to be coded or a sample from an earlier survey of the same kind. Of course an important problem is the size of the sample underlying the dictionary construction. Whether the dictionary is constructed manually or by computer the code numbers appearing in it should be those assigned by the best of the available coders with special independent verification procedures involved.

6.2. *Entering element descriptions into the computer*

Descriptions are to be entered into the computer. One possible method is to punch the descriptions in a more or less free format on cards or magnetic tape. However, this method has some serious drawbacks: first it consumes a lot of "space", and second, the errors involved in large-scale keypunching of alphabetic information are relatively unknown; more-

over such keypunching is rather costly.

There are reasons to believe that at present the entering of verbal descriptions to the computer is the most important practical problem in designing systems for automated coding.

6.3. *Matching and coding*

Each element description now put into the computer is compared with the list of occupation descriptions in the dictionary. If an element description agrees with an occupation description (is a "match"), it is assigned the corresponding code number, otherwise it is referred to manual coding.

In an automated coding system we will obtain exact matching for a fraction of all elements only. A primary task in developing such a system is to design criteria for the degree of similarity between input words and dictionary words necessary for them to be considered to match.

6.4. *Evaluation*

The system must include continuous evaluation studies. Such studies aim at

(i) controlling the quality of computerized coding;
(ii) improving the dictionary; and
(iii) controlling the cost.

6.5. *Some experiments and applications*

6.5.1. Industry

The very first experiment with automated coding (see Lyberg (1972, 1974a, 1977a, b)) concerned the industry variable.

This experiment gave the following results. Out of 419 input verbal descriptions 213 were coded or about 50 per cent. Of these, 170 or 80 per cent were correctly coded when compared to verified manual census coding.

Automated coding of industry has also been carried out in an experiment with data from the Labor Force Survey (LFS). We used a PLEX[1] dictionary containing 531 descriptions and a SLEX[2] dictionary containing 400 descriptions. A total of 712 verbal descriptions of industry were entered into the computer. The computer was able to code 65 per cent of these, but only 69 per cent were "correctly" coded (in agreement with the LFS manually assigned code numbers). This result was worse than those of the previous census experiments. However, the reason is obvious: in the LFS the interviewers collecting the data strive for detailed descriptions, and as a consequence the descriptions are sometimes composed of one or two whole sentences. In the censuses the respondents themselves fill in the answers and this usually results in short descriptions. The experiment showed that out of the 315 coded descriptions only 50 per cent were coded by means of PLEX. Usually a large majority is coded by that dictionary, but this time many SLEX words fitted the long input descriptions. For census data we have also tested a computerized dictionary based on 6 000 descriptions in the basic file. This time the descriptions came from the 1970 Census, the computer coded 61 per cent, and 83 per cent of these were correctly coded.

The main experiments are summarized in Table 4a.

Perhaps one can accept the low coding degree, but the errors are too frequent. However, we have not been working with this variable very long. In fact, almost all the "trial and error" work, in our opinion the very essence in developing methods for automated coding, is still waiting to be done for the industry variable.

[1] PLEX is a primary dictionary which is scanned first. PLEX consists of unambiguous descriptions, each of them supposed to fit one input description only.

[2] SLEX is a secondary dictionary which is scanned if PLEX fails to code. SLEX consists of discriminating word strings that can fit several input descriptions.

Table 4a. Automated coding of industry

Experiment	Type of dictionary	Survey	Coding degree (%)	Quality (% agreement)
1	Manual	1965 Census	50	80
2	Manual	Labor Force Survey 1974	65	69
3	Computerized	1970 Census	61	83

6.5.2. Occupation

When we first started to deal with the occupation variable we had the results from the first industry experiments at hand. Thus we were convinced that the size of the dictionary must be quite large to start with. The basic file consisted of 14 000 verbal descriptions from census material which were coded using an independent verification scheme. The final dictionary, constructed manually, consisted of 900 descriptions. This first dictionary was a type of PLEX dictionary, but different sophisticated matching methods (scanning, rank correlation coefficient) could be used if necessary. The first experiment, which was carried out with an independent set of 3 800 occupation descriptions from the 1965 Census, gave some encouraging results: the coding degree was 62 per cent and 95 per cent of these were "correctly" coded. This was considered to be very satisfactory.

Later on several trials were carried out and we introduced both PLEX and SLEX. With PLEX and SLEX together we sometimes obtained coding degrees around 80 per cent and qualities around 90 per cent.

We also tested our program for computerized dictionary construction on census descriptions. An experiment with this resulted in a coding degree of 69 per cent; 87 per cent of these were correctly coded.

Furthermore, we have constructed a dictionary for LFS occupations. When coding occupations in LFS we do not use the same dictionary as in census coding (as we do when coding industry). We were merely curious to investigate how well the census dictionary would work in LFS coding, and thus, the census three-digit code numbers were simply translated into LFS two-digit code numbers. This was done manually and resulted in a PLEX with 1 131 descriptions and a SLEX with 393 descriptions.

Thus, 711 independent descriptions were entered into the computer. Of these 81 per cent were coded and 81 per cent of these were in accordance with the LFS coding. This result is worse than those obtained in the census coding experiments. The reason is obvious. A direct translation is not to be recommended due to the different data collection methods: occupation in censuses and occupation in the LFS are not really the same variables. A special LFS dictionary should be constructed by means of a basic material consisting of LFS descriptions.

During 1978 a new set of experiments on LFS material was conducted. This time the computerized construction program was used on a basic file consisting of 6 000 descriptions. From these descriptions the program created a PLEX containing 1 637 descriptions and a SLEX containing 1 230 discriminating words. Three sets of independent new records were coded by means of PLEX or a combination of PLEX and SLEX. Between each run certain changes were made in the PLEX dictionary. In a fourth run a greatly extended dictionary was used.

First run

In all, 1 002 verbal descriptions were coded. The coding degree was 84.3 per cent. Of these 84.6 per cent were in agreement with the manual code number assigned in the LFS. The dictionary was PLEX and SLEX. When only PLEX was used the coding degree decreased to 69 per cent while the agreement rate increased to 92.2 per cent. Obviously SLEX was not very accurate.

Second run

Prior to the second run some minor changes were made in PLEX: 22 descriptions were removed and 162 were added. Then a new set of 1 000 records was coded by the modified PLEX; the resulting coding degree was 69.1 per cent with an agreement rate of 93.3 per cent.

The cost of coding 1 000 occupations was 18 kronor (approximately US$ 2).

Third run

Prior to the third run some further minor changes were made in PLEX: 19 descriptions were removed and 145 were added. Again only PLEX was used for coding of 989 new records. The resulting coding degree was 69.2 per cent with an agreement rate of 91.8 per cent. An evaluation of the disagreements revealed that the absolute automated coding error rate was 4.7 per cent, which we think is a good result.

Fourth run

The minor modifications of PLEX had no effect on the coding degree. However, the manual coders had an alphabetical list of approximately 11 000 occupations at their disposal. This list, provided by Statistics Sweden, is not based on knowledge of empirical response patterns. Rather, it is a desk product. We were anxious to know what would happen if that list, already available on tape, was used as a PLEX. Said and done. We made new runs of the materials used in runs 1 and 2 and obtained the coding degrees 40.2 per cent and 36.0 per cent respectively. When we merged the alphabetical list with our computerized constructed PLEX the coding degree increased to 75.8 per cent and 74.2 per cent respectively and the agreement rates increased to 93.0 per cent and 93.8 per cent respectively.

The cost of coding 1 000 descriptions by means of this large dictionary was 127 kronor (approximately US$ 16).

The main experiments are summarized in Table 4b.

We think that it is possible to code occupation in our surveys by means of the program package together with some manual efforts in the dictionary construction phase. However, for an automated system to be considered, the cost situation must be carefully evaluated.

Our experiments with occupation coding are described in, for instance, Lyberg (1972, 1974a, 1977a, b, 1979).

Table 4b. Automated coding of occupation

Experiment	Type of dictionary	Survey	Coding degree (%)	Quality or agreement rate (%)
1	Manual	1965 Census	62	95
2	Manual	1970 Census	66	92
3	Manual	1970 Census	74	84
4	Manual	1970 Census	80	90
5	Manual	Labor Force, 1974	81	81
6	Computerized (PLEX+SLEX)	1970 Census	69	87
7	Computerized (PLEX+SLEX)	Labor Force, 1976	84	85
8	Computerized (PLEX)	Labor Force, 1976	69	93
9	Computerized (PLEX)	Labor Force, 1976	69	92
10	Computerized and manual combined	Labor Force, 1976	74–76	93–94

Automated coding of occupation is now used in the processing of the 1980 Swedish Census of Population. As this is being written the production results are still incomplete but the operation has been real smooth. Let us just give one example.

For the County of Gothenburg comprising 341 529 economically active individuals the following results are available:

The dictionary consisted of 10 291 PLEX descriptions and 513 SLEX descriptions. The computer managed to code 74.7 per cent or 254 991 individuals. The coding percentage for PLEX was 72.2 per cent and for SLEX 2.5 per cent only. The computer cost for this automated coding was 303 kronor (approximately US$ 38). Some extra costs for sortings and listings are unavoidable but, in practice, the cost for the automated part is negligible.

The automated coding in the 1980 Census is described in more detail in Andersson and Lyberg (1983).

7. Epilogue

The background: increasing use of administrative records and development of computers

The future of coding control, like that of other fields, is hard to predict. However, we know a few things for sure. We know that administrative records will be increasingly used for statistical purposes.

An efficient link between administrative records and statistics production implies the necessity of uniform coding. Uniform coding means uniform codes, i.e. for the variable under study a specific verbal description should be assigned the same code number irrespective of the system. This is not always the case even within statistics production. Uniform coding is the necessary basis for effective coding control and is therefore one of the most urgent tasks.

Coding within administrative systems is of course very sensitive to errors since each error can have important effects on the individuals for whom the system is designed. The coding control must therefore be tighter than in statistical systems, where the errors sometimes tend to cancel out. However, special care must be taken in using administrative systems in statistics production. Suppose that data in an administrative system are used to construct sampling frames for statistics production. Then, again, each individual error might constitute a serious flaw.

Thus, it seems obvious that one can expect in the future more intensive use of efficient control methods such as independent verification.

The development of computers has been important over the years. We have seen, for instance, that the replacement of manual matching of code numbers by computer matching facilitates things a great deal. Recent studies show that data processing can be made more efficient by letting a single clerk carry out several processing tasks more or less simultaneously. For instance, in Rustemayer (1977) it is shown how interviewing can be computer-assisted by means of a system in which telephone interviewing is carried out at a computer terminal. No wild fantasy is needed to imagine that such a procedure can be extended to cover such operations as editing, coding and punching as well. Concentration of several operations, including coding, in a centralized facility makes things a lot easier. Coding instructions can be displayed on terminals and quick changes in codes and instructions can be entered on terminals. Verification and feedback can be conducted more easily in a computerized environment. Coding in such an environment should be a cheaper and more effective operation.

At Statistics Sweden we have been working with computer-assisted data collection since 1979. Rapid progress is now being made along two interacting lines:

- tests of available, handheld computers in survey data collection,
- specifications of requirements for a prototype equipment for survey data collection to be available on the market within a couple of years.

We aim at constructing a handheld microcomputer to be used in many different kinds of data collection. Our endeavours so far are described in Danielsson and Mårstad (1982a, b) and Lyberg et al. (1983).

The future of automated coding

Obviously the success of automated coding is a function of language complexity. It seems that the Swedish language is more forgiving than English in this respect. Our experiments have shown that automated coding might be a possible option when designing the coding operation.

Our experiments with industry and occupation show that occupation is much easier to code than industry. It may well be possible to use automated coding in the Labor Force Survey. Currently we are very close to the payoff point in that survey. Automated coding of census occupations is also possible and is used in the 1980 Swedish Census of Population. Our experiments with industry have been very limited since this variable is semantically much more complex than occupation. (Curiously, overseas it is the other way around.) Our strategy has been to put the easier variables to a test first, and then, if possible, proceed to the more difficult ones.

8. Special note

Statistics Sweden is indebted to the US Bureau of the Census. That agency has always been in the pole position when it comes to dealing with coding problems. Methods such as independent verification and automated coding were invented there and much of the work carried out by Statistics Sweden in this field over the years would have been impossible without the support and encouragement given by our colleagues at the Census Bureau. References to work in the field of coding carried out by the U.S. Census Bureau are given in Lyberg (1981).

9. References

Andersson, R. (1977): Kodningskvaliteten i FoB 75 – ekonomisk aktivitet. Memo, Statistics Sweden. (In Swedish.)

Andersson, R. and Lyberg, L. (1983): Automated Coding at Statistics Sweden. Paper presented at the ASA meeting in Toronto, Canada, August 1983.

Boston, G.F.P. (1977): Quality Control of Occupation Coding. Unpublished draft, Office of Population Censuses and Surveys, U.K.

Dalenius, T. (1974): Ends and Means of Total Survey Design. Forskningsprojektet FEL I UNDERSÖKNINGAR, Stockholms universitet, Statistiska Institutionen.

Dalenius, T. and Frank, O. (1968): Control of Classification. Review of the International Statistical Institute, vol. 36:6, pp. 279–295.

Dalenius, T. and Lyberg, L. (n.d.): An Experimental Comparison of Dependent and Independent Verification of Coding. Memo from Tore Dalenius to Leon Pritzker.

Danielsson, L. and Mårstad, P.-A. (1982a): Towards Development of a Computer Assisted Survey Data Collection System. DBM-PM 82:35. Memo, Statistics Sweden.

Danielsson, L. and Mårstad, P.-A. (1982b): Statistical Data Collection with Handheld Microcomputers – A Test in Consumer Price Index. DBM-PM 82:41. Memo, Statistics Sweden.

Dodge, H.F. and Romig, H.G. (1944): Sampling Inspection Tables. Wiley.

Fasteau, H.H., Ingram, J.J. and Minton, G. (1964): Control of Quality of Coding in the 1960 Censuses. Journal of the American Statistical Association, Vol. 59, No. 305, pp. 120–132.

Harris, K. (1974): Analysis of the Independent Three-Way Verification System in Mortality Medical Coding. Memo, US Department of Health, Education and Welfare.

Linebarger, J.S., Jablin, C. and Davie, W.C. (1976): Dependent versus Independent Verification. Memo, US Bureau of the Census.

Lyberg, L. (n.d.): Beroende och oberoende kontroll av kodning. Rapport nr 4, Forskningsprojektet FEL I UNDERSÖKNINGAR, Stockholms universitet, Statistiska institutionen. (In Swedish.)

Lyberg, L. (1969): On the Formation of Cod-

ing Teams in the Case of Independent Verification Under Cost Considerations. Rapport nr 18, Forskningsprojektet FEL I UNDERSÖKNINGAR, Stockholms universitet, Statistiska institutionen.

Lyberg, L. (1972): A Note on Automatic Coding. Statistisk tidskrift, 1972:2, pp. 137–142.

Lyberg, L. (1974a): Automatic Classification of Industry and Occupation – Some Experiments. Memo, Statistics Sweden.

Lyberg, L. (1974b): Kodningskvaliteten i FoB 70 – ekonomisk aktivitet. Memo, Statistics Sweden (In Swedish.).

Lyberg, L. (1977a): Coding of Verbal Information. American Statistical Association, Proceedings of the Social Statistics Section, Part I, pp. 354–359.

Lyberg, L. (1977b): Coding of Verbal Information. Memo, Statistics Sweden.

Lyberg, L. (1979): Några experiment med automatisk kodning i AKU. Memo, Statistics Sweden (In Swedish.).

Lyberg, L. (1981): Control of the Coding Operation in Statistical Investigations – Some Contributions. PhD thesis, Statistics Sweden, Urval No. 13.

Lyberg, L., Swensson, B., and Wretman, J. (1983): Current Survey Research at Statistics Sweden. Paper presented at the ASA meeting in Toronto, Canada, August 1983. 1983.

Minton, G. (1969): Inspection and Correction Error in Data Processing. Journal of the American Statistical Association, Vol. 64, pp. 1256–1275.

Minton, G. (1970): Comments on Quality Control and Research in Data Processing Programs. Paper presented at the American Society for Quality Control, March 12–13, Arlington, Virginia.

Rustemeyer, A. (1977): Toward Development of a Computer-Assisted Telephone Interviewing System. Memo, US Bureau of the Census.

US Bureau of the Census (1965): United States Censuses of Population and Housing, 1960: Quality Control of Preparatory Operations, Microfilming, and Coding. Washington, DC.

Teaching Questionnaire Design

by Thomas B. Jabine[1]

1. Introduction

Among the great strengths of Professor Tore Dalenius, in whose honor this article is presented, are his wide-ranging intellectual curiosity and his facility for recognizing diverse dimensions of an activity or problem and organizing them to form a coherent, integrated structure.

It is not surprising, therefore, that Dalenius was an early advocate of the concept of *total survey design*. About 25 years ago, in *Sampling in Sweden* (Dalenius (1957)), he showed a clear understanding of the importance of non-sampling errors in surveys, saying that (p. 142):

"The impact of this choice [of measurement method] on the overall adequacy of sample survey results has often been neglected. In fact the problem of 'non-sampling errors' is, in our opinion, a most serious problem in today's Swedish sample survey practice."

He also pointed to the need to work on

". . . the integration of existing theories and methods made use of in survey practice into a general theory of statistical investigations . . ." (p. 247).

Dalenius and others pursued this goal. A response error model capable of dealing with all sources of errors in surveys was developed at the United States Bureau of the Census (Hansen, Hurwitz and Bershad (1959)), an organization with which Dalenius has had frequent and productive (for all concerned) associations throughout his career. The Census Bureau response error model provided a suitable tool for an approach to survey design which takes all sources of error into account and tries to strike an economic balance in the allocation of resources to the measurement and control of different kinds of survey errors.

In the mid-1970's, this approach came to be called "total survey design." It appears that Dalenius himself was the first to use the term in print. In June 1974 he published a monograph with the title *Ends and Means of Total Survey Design*. The author has seen the term used often in publications issued since that date, but never earlier. In any case, the work of all competent survey researchers today is guided by the concept of total survey design, and there is no question that Dalenius has had a major role as a developer and advocate of this idea.

What does all of this have to do with the teaching of questionnaire design? Fellegi and Sunter (1974, p.124) put the basic idea of total survey design in very simple and practical terms: ". . . concentrate additional resources where they do the most good." The author has felt for many years that, in those terms, questionnaire design in surveys has often not received its proper share of attention and resources, considering that substantial reductions in error can often be had from fairly modest investments in questionnaire improvement. Hansen, Hurwitz and Madow made this point nearly 30 years ago (1953, pp. 87–88):

". . . significant decreases in response errors can sometimes be achieved at slight increase in expenditure, perhaps by a change in wording or by one or two additional questions on a questionnaire that may add little to the time of each interview."

Consequently, when an opportunity arose to teach a course in Questionnaire Design for the Statistics Department of George Washing-

[1] Professorial Lecturer, Statistics Department, George Washington University, USA.

ton University in the Spring of 1982, the challenge was readily accepted. This paper describes the nature and content of the course. Evaluation of the results by the professor is likely to be biased in one way or another; however, some subjective views on what worked well and what improvements can be made are included.

2. Need for the course

Courses in survey research methods are given in many universities, frequently in sociology departments. Such courses generally try to cover the sample survey from start to finish, including specification of objectives, sample design and selection, questionnaire design, data collection, data processing, and analysis. Frequently, as a class project, a small survey is undertaken, with students doing most of the work, including interviewing, under the guidance of the instructor.

Students who require only a general acquaintance with survey research techniques may be well served by a one-semester course of this kind. For those who expect to or are already making a career in survey research, each major aspect of survey work needs to be examined in more depth. Thus, one or more courses in sampling, data processing, and data analysis are needed. Questionnaire design and the associated issues of content development and data collection methods likewise deserve separate treatment.

The American Statistical Association's Section on Statistical Education Committee on Training Statisticians for Government recently published a report (1982) which included a list of courses which it felt were necessary background for government statisticians working at various levels. Heavy emphasis was placed on courses in applied survey techniques. Included in the course list for junior professional statisticians (GS-9/11 levels) were: sampling theory, questionnaire design principles, analysis and interpretation of survey data, estimation and control of non-sampling errors in survey data, imputation/incomplete data techniques and a computer processing course. In regard to the availability of courses on these topics, the report stated that:

"Most graduate statistical curricula are so theoretically based that the elementary principles of applied survey methods are not covered at all. With the exception of many social science statistics courses, the principles of questionnaire design and the analysis and interpretation of survey data are rarely touched."

The Statistics Department of George Washington University offers an unusually wide range of survey methods courses, covering most of the topics mentioned above. The course in Questionnaire Design (STAT 193), first given in the Spring of 1982, replaced a somewhat broader course in Survey Design. With all or nearly all of the appropriate courses now in place, and in view of the wide use of and interest in survey methods in the Washington area, the Department now plans to offer a Masters program in applied survey sampling.

3. Can questionnaire design be taught?

Many practitioners have referred to questionnaire design as an art rather than a science. Stanley Payne's classic work of 1951 was titled *The Art of Asking Questions*. A recent text on questionnaire design (Labaw (1980)) states that ". . . questionnaire design is an art and takes an artist's sensitivity." A recently revised version of the United Nations *Handbook of Household Surveys* (1981) says ". . . the preparation of questionnaires is still more of an art than a science."

It is certainly true that there is an element of what might be called art, imagination or creativity that is essential to the production of "good" questionnaires. However, there are also objective standards for good questionnaires, and there are several useful tools, such

as flow charting, focus group interviewing and, above all, experimental research techniques (especially "split-panel testing") that can be presented to the student. There is also an ever-increasing body of literature describing empirical findings from carefully designed experiments with alternate versions of questionnaires. Some of these experimental results contradict conventional wisdom about questionnaire design, e.g., the finding that increasing the length of sensitive questions tends to decrease response error (Bradburn and Sudman (1979)). The survey researcher who is familiar with these standards, tools and research findings is more likely to succeed in designing good questionnaires than the one who is not. Considering the extent to which survey research is used as a tool in the social *sciences* it would be discouraging indeed to think that its main vehicle, the questionnaire, could only be judged by artistic standards.

4. Setting

Questionnaire Design (STAT 193) was a one-semester, 3 credit hour course, which could be taken both by upper-level undergraduates and by graduate students. A course in Theory of Sampling (STAT 187) was listed as a prerequisite; it could be waived by the instructor and was waived for a few applicants who had already had some survey research experience.

Prior to registration, a flyer describing the course was sent to key people in several major Federal statistical agencies. The final enrollment was 19, of whom 4 took the course for graduate credit. Over half of the class were full-time employees of Federal statistical agencies, probably a good turnout considering that budget cuts at the time had substantially reduced funds available for agency funding of training for employees. Several of the students had already had substantial experience in survey work, and were able to make important contributions to class discussions of various issues.

The course was presented in 13 weekly two and one-half hour sessions. Like most courses in the Department it met in the evenings for the convenience of part-time students.

5. Objectives

The objectives of the course, as stated at the first session, were:

- To establish a broad framework, namely that of total survey design, for setting objectives and evaluating results in the design of questionnaires.
- To become familiar with and to get some practical experience in the steps essential to the development of a new questionnaire.
- To learn what to look for and to obtain some practice in reviewing existing questionnaires or drafts.
- To become familiar with past and current literature on questionnaire design, and to learn how to distinguish findings based on adequate experimental or observational data from those based on faulty or no quantitative evidence.
- To identify specific principles and guidelines for use in designing questionnaires and forms.

As may be noted from the last objective, although the course focussed primarily on survey questionnaires, some attention was given to forms used for non-statistical purposes, such as application forms. Most relevant methodological research to date appears to have been on survey questionnaires as opposed to administrative forms, but it would seem that much of what has been learned could profitably be applied in the latter area. There has been some interesting experimental work on administrative forms, for example studies reported by Wright (1980) and colleagues in the United Kingdom, and by Scherr (1980) in connection with the redesign of the social security number application form in the United States.

Within the category of survey questionnaires, most of the course materials related to questionnaires for surveys of households and individuals, as opposed to businesses, simply because that is where more of the methodological work on questionnaires has been done.

6. Content

The subject of questionnaire design, with its close linkage to all other aspects of survey design, is not an easy one to organize and present, or even to define. The solution adopted was to present various aspects more or less as they would (or should) appear naturally in the course of a survey. Following is a list of the major topics covered, with brief comments on each:

(1) Introduction. This section covered the course objectives and structure. The basic ideas of total survey error, components of non-sampling error, and total survey design were presented and related to the question of appropriate goals to be established for the design of questionnaires. The literature on questionnaire design and survey research methods was discussed, with emphasis on convenient sources, such as Dalenius' (1977) bibliography on non-sampling errors, the U.S. Census Bureau's Survey Methodology Information System (unfortunately a victim of recent budget cuts in U.S. statistical programs), and the publications of the American Statistical Association, including especially the *Proceedings of the Section on Survey Research Methods* published annually since 1978.

(2) Purposes of surveys and definition of survey variables. In many surveys, the drafting of a structured questionnaire is begun with only the vaguest notions of specific data requirements. The need to start with at least a tentative list of clearly defined survey variables was given strong emphasis. For this section, Chapter 3 of the text chosen for the course (Lininger and Warwick (1975)) provided an excellent discussion of the different purposes for which surveys are undertaken. A detailed discussion of the process of defining survey variables was available in a document prepared by the author for the Inter-American Statistical Institute (Jabine (1981, Chapter B)).

(3) Unstructured design activities. This section covered techniques that are used by survey designers to gather background information needed to design an effective questionnaire, especially when the survey relates to a new topic or a new target population. This information relates primarily to communication with survey respondents and to their ability and willingness to provide particular kinds of information. Techniques discussed included participant observer studies, focus group interviews, and individual depth interviews. An especially useful discussion of the latter two is found in Chapter 2 of Hoinville, Jowell and associates (1978).

(4) Strategic decisions, questionnaire structure and format. This section covered the survey design decisions that determine or place constraints on the specific type of questionnaire or questionnaires to be designed in order to meet the data requirements. These decisions include collection method, choice of units of observation, rules for associating units of observation with sampling units, respondent rules, number and timing of interviews, number of separate questionnaires needed, constraints on overall length, and data entry method to be used. Each of these decisions was discussed in terms of its effects on the questionnaire(s). The use of flow charting as a preliminary to design of a structured

questionnaire was also introduced at this stage.

(5) Questionnaires and questions. This section could be said to deal with questionnaire design proper. The extent of the preliminaries needed to reach this stage reflects the author's view of the proper way to approach questionnaire development. This section was divided into 3 parts:

I. Identification and "factual" items
II. Subjective items
III. Sensitive items and other topics.

The paper by Kalton and Schuman (1980) on "The effects of the question on survey responses: a review" provided an excellent guide for the topics to be covered in this section.

(6) Alternative collection methods. The nature of a questionnaire is, of course, strongly affected by the data collection method chosen. The main options are interview and self-administration; each of these has various sub-options, e.g., face-to-face vs. telephone interviewing. While these effects had been discussed to a considerable extent in earlier sections, it seemed necessary at this point to present them in a more systematic way. In addition, the use of computer-assisted telephone interviewing (CATI) techniques introduces a radically different concept of what is meant by a "questionnaire" and how it is used as an instrument in conveying information from respondents to the data processing stage of the survey.

(7) Pretesting. The purpose of this section was to emphasize the need to define pretest objectives clearly and to plan for the capture of information that will be useful in spotting problems and making improvements in the questionnaire.

(8) Clearance, burden and fair information practices. For the benefit of students working for U.S. Government agencies, a section on the formal procedures and requirements for clearance of government-administered or government-sponsored questionnaires was included. In addition, the procedures for the "information collection budget," recently introduced by the U.S. Office of Management and Budget (OMB) as a means of exercising better control over total respondent burden, were covered. Both of these topics were presented by Mr. Richard Eisinger from OMB's Office of Information and Regulatory Affairs, the office responsible for these functions. Under "fair information practices", issues relating to privacy, confidentiality and informed consent were treated. Throughout the course, students were urged to put themselves in the respondent's place and to recognize that the success of survey research depends in the long run on equitable and considerate treatment of respondents.

(9) Methodological research on questionnaires. Results of methodological research had been introduced in connection with many of the earlier topics. This section covered the design of experiments relating to questionnaires, with emphasis on split-panel tests of alternate versions and on methods of measuring various components of response error, such as record checks and reinterviews.

7. Teaching methods and materials

Major elements of the course were (1) lectures and class discussions on the topics described in the previous section, (2) a group questionnaire development project, and (3) individual literature research projects.

Prior to each lecture-discussion period readings were assigned from the text (Lininger and Warwick (1975)) or other items included in a reading list made available at the first

session. The author's study on experimental and pilot survey activities (Jabine (1981)) was used in connection with several topics. For the lectures and discussions on questionnaires and questions, the author was fortunate to have available several draft chapters of a book on questionnaire design in preparation by Sudman and Bradburn (1981). Chapter summaries from this draft covering questionnaire format, order of questions, and the uses of knowledge questions served as excellent outlines for discussion of these topics.

A substantial part of class time, especially in the first half of the course, was devoted to a *group project* on questionnaire development. The students were divided into 4 groups, each led by one of the 4 graduate students. Each group developed and pretested a questionnaire, all on the same topic but, for most phases of development, working independently. This project went through several stages:

(1) Choice of topic. Students were asked to suggest questionnaire topics meeting the following criteria:
   - Exclude topics for which more or less standard question sets already exist.
   - The subject should be one for which at least some members of the class are potential respondents, and with which some of them are at least moderately familiar.
   - The topic should be one for which both "hard" and "soft" survey items are relevant.
   - The data requirements should not be too complex. For example, it probably would have been too ambitious to try to design a detailed household expenditure survey questionnaire.

   Proposed topics meeting these criteria were narrowed down to 5, and then voted on. The one selected was volunteer work. This was a topic of considerable current interest because it is the philosophy of the Reagan administration to encourage private organizations to provide many of the services and types of assistance that have been coming from Federal programs.

(2) Focus group interview. Using a topic outline provided by the instructor, one of the graduate students acted as a moderator for a group interview to explore the participants' knowledge of and attitudes toward volunteer work, and to try to identify salient dimensions of the topic. Half the students participated in the interview; the other half acted as observers.

(3) Development of a list of survey variables. Starting at this stage, each of the 4 groups worked separately. They were asked to prepare a brief statement of purpose for a survey of volunteer work, and to develop a list of variables to be covered, following the steps and specifications given by the author (Jabine (1981, Chapter B)).

(4) Individual depth interviews. After review of each group's variables list by the instructor, each student was asked to conduct two individual depth interviews, in order to determine the kinds of questions needed to obtain the desired information and to identify problems such as imprecise definition of concepts, memory problems and sensitive items. A written report was prepared for each interview.

(5) Draft of structured questionnaire. After assimilating the results of the individual depth interviews and revising variables lists as necessary, the students in each group collaborated in preparing a first draft of a structured questionnaire.

(6) Pretest of draft questionnaire. Each student conducted two interviews with his or her group's draft questionnaire and prepared a written report describing problems encountered and suggesting

changes.

(7) Redraft of questionnaire. This was the final product of the group project. While it probably would have been profitable to go through one or two more cycles of pretesting and revision, time did not permit this.

Each student was required to undertake an *individual literature research project.* The assignment was:

"To do a literature search for information relating to some specific aspect of questionnaire design, and to synthesize and make a presentation of the findings. Explain the implications of the findings for questionnaire design practice."

Early in the course, the individual topics were proposed and reviewed by the instructor. Presentations were made during class sessions in the latter part of the course. At the time of presentation, the student was required to turn in: (1) an outline of the presentation, (2) a list of references (for most topics, at least 10 references were expected), and (3) an abstract of the presentation. In addition, graduate students were required to turn in a written report of their findings.

The flavor of these reports can be determined from the following partial list of topics:

- Memory aids: when to use and how effective are they?
- Effects of varying the order of response categories.
- Differences between telephone and face-to-face interviewing.
- The use of advance and cover letters in survey research.
- The influence of pollster identification on political poll results.
- The effect of using a proxy respondent for "soft" items.
- How colors and size of print affect questionnaire results.

In addition to these 3 major elements, other class features included: review and discussion of specific questionnaires, a flow charting exercise (using the flow chart as diagnostic tool for an existing questionnaire), and, from time to time, discussion of news items relating to surveys. Students were asked to bring in and present news items relevant to questionnaires, forms, and surveys. It was remarkable how many such items turned up during the 3 month duration of the course, including not just reports of survey results, but stories where methodological issues, such as the suitability of the particular questions used or inconsistencies between results for similar questions appearing on different surveys, came into play.

Finally, the course included mid-term and final examinations. These emphasized practical applications of the course content, such as:

- Identifying flaws in particular questions and recommending alternative versions.
- Evaluating a particular questionnaire or administrative form (for the mid-term examination, the University's standard employee data questionnaire was used for this purpose).
- Preparing a short list of variables for a particular topic.
- Drafting questions for topic variables.

Many of the text, reading and reference materials have already been identified; however, there are some additional items that were especially useful and deserve mention:

(1) The Social Science Research Council (1975) publication *Basic Background Items for Household Surveys.* This publication contains "a model set of basic background questions and coding procedures for general population household interview surveys in the United States". It was developed by a working group of survey researchers interested in encouraging standardization and thereby improving the comparability of data collected by different survey organizations. Students were encouraged to use this resource in developing questions for

standard demographic variables for their questionnaires on volunteer services.

(2) *Survey Methods in Social Investigation*, by Moser and Kalton (1972). The chapter on "Scaling Methods" was used as a primary source for the presentation of that important and somewhat complex subject.

(3) The section on "Design of questionnaires and administrative forms" from the Census Bureau's (1979) *Popstan: A Case Study for the 1980 Censuses of Population and Housing*. This unit includes many useful examples, and a fairly comprehensive coverage of the practical aspects of questionnaire design from start to finish.

(4) Sirken's (1972) manual, *Designing Forms for Demographic Surveys*. This manual is especially useful for its discussion of data requirements and strategic decisions affecting questionnaire design, and for illustrations of the use of flowcharting techniques.

(5) "Split-panel tests of census and survey questionnaires" by Jabine and Rothwell (1970). A useful feature of this paper was a standard format developed by the authors for preparing summary descriptions of split-panel tests of questionnaires. Describing the individual elements of the standard format to the class was a convenient and effective way of presenting the design and interpretation of controlled experiments with questionnaires.

(6) The body of work by Schuman and Presser, culminating in their recent book *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context (Schuman and Presser (1981))*. Some examples from their work, presented early in the course, were very effective in raising the consciousness of students to the sensitivity of response to small variations in question wording and format. Their findings were used extensively in some of the individual literature research projects.

8. Discussion

The comments which follow are based largely on the author's personal evaluation, supplemented by informal discussions with some of the students and review of various exercises, assignments and examination papers.

Both the group and individual projects proved to an important and valuable part of the course. The group project, developing a questionnaire step by step from the beginning, demonstrated to the participants, as no amount of lecturing or reading could, that the process of questionnaire development requires patience, careful planning, a clear definition of data requirements, the participation of persons with some understanding of all major phases of survey operations, and willingness to work directly with members of the survey target population to understand their reactions to what is being asked of them.

The individual literature review projects were valuable for several reasons. They introduced students to a body of literature which is widely scattered in a variety of professional journals, books, and reports of national and international statistical organizations. They required students to organize and evaluate often conflicting findings and recommendations on various aspects of questionnaire design. Allowing the students to select their topics permitted them to focus on the areas of most interest to them. The class presentations, which were generally well done, provided the class with substantive information on a broad range of topics that might not otherwise have received such thorough treatment. Finally, the specialized bibliographies that were prepared and in most cases distributed

to the class provided, collectively, good coverage of most of the major issues in questionnaire design.

Looking at the first presentation of this course as a pilot test, the author sees a need for improvements along the following lines:

(1) The concept of a list of survey variables needs to be sharpened, and more actual practice in preparing such lists would be helpful.

(2) The students should be given more exercises involving the drafting or redrafting of individual questions or question sets.

(3) The standard format for describing split-panel experiments should be presented and used earlier in the course, as it will be useful to students in pursuing their individual literature research projects.

(4) An attempt should be made to introduce more illustrative material from business (as opposed to household) surveys.

(5) As a teaching tool, it would be useful to have a "check-list" for use in reviewing draft questionnaires. The author has seen one or two versions of such a check-list and indeed has himself made an attempt to develop one, but finds none of these efforts completely satisfactory as yet. The effort will continue.

(6) To accommodate the greater emphasis on practical exercises implied by some of the preceding comments, it may be necessary to cut back some on lecturing and rely more on reading assignments to cover some of the basic ideas. Fortunately, the amount of good material for this purpose is increasing year by year.

In closing, the author expresses the hope that Professor Dalenius will review this paper and, as he has invariably done in the past, react with perceptive and constructive comments, based on his broad knowledge of all aspects of survey research and his remarkable ability to apply that knowledge to whatever project comes to his attention.

9. References

American Statistical Association (1978–): Proceedings of the Section on Survey Research Methods. Annual. Washington, D.C.

American Statistical Association (1982): Preparing Statisticians for Careers in the Federal Government: Report of the ASA Section on Statistical Education Committee on Training of Statisticians for Government. The American Statistician, Vol. 36, No. 2, pp. 69–81.

Bradburn, N. and Sudman, S. (1979): Improving Interview Method and Questionnaire Design. San Fransisco: Jossey-Bass.

Dalenius, T. (1957): Sampling in Sweden. Stockholm: Almquist and Wiksell.

Dalenius, T. (1974): Ends and Means of Total Survey Design. Stockholm: University of Stockholm.

Dalenius, T. (1977): Bibliography on Non-Sampling Errors in Surveys. International Statistical Review 45: pp. 71–89 (A to G), pp. 187–197 (H to Q), pp. 303–317 (R to Z).

Fellegi, I. and Sunter, A. (1974): Balance Between Different Sources of Survey Errors – Some Canadian Experiences. Sankhyā 36 (Series C), pp. 119–142.

Hansen, M., Hurwitz, W. and Bershad, M. (1959): Measurement Errors in Censuses and Surveys. Bulletin of the International Statistical Institute 38, pp. 359–374.

Hansen, M., Hurwitz, W. and Madow, W. (1953): Sample Survey Methods and Theory. Volume I, Methods and Applications. New York: Wiley.

Hoinville, G., Jowell, R. and associates (1978): Survey Research Practice. London: Heinemann Educational Books.

Jabine, T. (1981): Guidelines and Recommendations for Experimental and Pilot Survey Activities in Connection with the Inter-American Household Survey Program. Document 7679a. Washington: Inter-

American Statistical Institute.

Jabine, T. and Rothwell, N. (1970): Split-Panel Tests of Census and Survey Questionnaires. American Statistical Association, Proceedings of the Social Statistics Section, pp. 4–13.

Kalton, G. and Schuman, H. (1980): The Effect of the Question on Survey Responses. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 30–39.

Labaw, P. (1980): Advanced Questionnaire Design. Cambridge, Mass: Abt Books.

Lininger, C. and Warwick, D. (1975): The Sample Survey, Theory and Practice. New York: McGraw-Hill.

Moser, C. and Kalton, G. (1972): Survey Methods in Social Investigation (second edition). New York: Basic Books.

Payne, S. (1951): The Art of Asking Questions. Princeton: Princeton University Press.

Scherr, M. (1980): The Use of Focus Group Interviews to Improve the Design of an Administrative Form: A Case Study at the Social Security Administration. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 347–352.

Schuman, H. and Presser, S. (1981): Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context. New York: Academic Press.

Sirken, M. (1972): Designing Forms for Demographic Surveys. Laboratory for Population Statistics Manual Series, No. 3. Chapel Hill: University of North Carolina.

Social Science Research Council (1975): Basic Background Items for U.S. Household Surveys. R. van Dusen and N. Zill, editors. Washington: Center for Coordination of Research on Social Indicators.

Sudman, S. and Bradburn, N. (1981): Asking Questions. Draft manuscript, selected chapters.

United Nations (1981): Handbook of Household Surveys (draft revision) DP/UN (INT-79-020/2) with two addenda. New York: United Nations.

U.S. Bureau of the Census (1974): Indexes to Survey Methodology Literature. Technical Paper 34. Washington, D.C.: U.S. Department of Commerce.

U.S. Bureau of the Census (1979): Popstan: A Case Study for the 1980 Census of Population and Housing. Part B, Planning and preparation for Popstan Census. Washington, D.C.: U.S. Department of Commerce.

Wright, P. (1980): Strategy and Tactics in the Design of Forms. Visible Language XIV, 2, pp. 151–191.

Organization of the U.S. Federal Committee on Statistical Methodology

by Maria Elena Gonzalez[1]

1. Introduction

The Federal Committee on Statistical Methodology (FCSM), which has a coordinating role within the U.S. statistical system, held its first meeting in July of 1975. The FCSM was initially organized by the Statistical Policy Division of the Office of Management and Budget (OMB). The U.S. Federal statistical system is decentralized. The major statistical agencies are located in various departments such as the Bureau of the Census and the Bureau of Economic Analysis in Commerce, the Bureau of Labor Statistics in Labor and the Statistical Reporting Service in Agriculture.[2] The location of the FCSM has moved twice because of reorganizations of the Federal statistical system. In October 1977, the FCSM moved to the Office of Federal Statistical Policy and Standards (OFSPS) in the Department of Commerce. In August 1981, the FCSM returned to OMB to the Office of Information and Regulatory Affairs (OIRA).

The main purpose of the FCSM is to identify and investigate technical and methodological problems which affect the quality of Federal statistical data and to suggest ways for improving the statistical methods used by Federal agencies for designing, gathering, processing, analyzing and disseminating statistical data. Another purpose is to study the cost effectiveness of various statistical methods.

[1] Office of Management and Budget, Washington, D.C., USA.

[2] A comprehensive description of the agencies within the U.S. Federal Statistical System is given in *Principal Federal Statistical Programs*, Statistical Policy Branch, Office of Information and Regulatory Affairs, Office of Management and Budget, March 1982.

Members of the FCSM are selected by the Office of Information and Regulatory Affairs of OMB. These members cover a broad spectrum of Federal agencies; however, they do not represent the agencies, but rather are individuals with recognized expertise in statistical methodology who are concerned with the quality of Federal statistical data.

The FCSM selects topics of methodological concern to be studied by subcommittees. The topics are proposed by individual members and then selected based on the amount of interest expressed by all members. When topics are selected for subcommittee work, the FCSM usually prepares a very general description of the area to be investigated by the subcommittee. This description serves as a starting point for the work of the subcommittees.

2. Organization of the subcommittees of the FCSM

After a topic has been selected for a subcommittee a chair is identified; the chair may be a member of the FCSM. The membership of the subcommittees comes from various Federal agencies and is open to any Federal employee interested in the selected area. Members of the FCSM are asked to suggest possible participants for the subcommittee work. After an initial group has been invited, a first meeting is convened. During this first meeting the objectives of the subcommittee are reviewed and initial tasks are defined. The subcommittees provide an opportunity for collaboration by persons from different agencies with expertise and interest in a common subject. This collaboration provides an opportunity for fruitful

exchange of experiences. The kind and extent of work of the subcommittee members varies depending on the expertise and interest of the individuals.

An element common to the work of nearly all subcommittees is the process of gathering information from Federal agencies (and in one case from contractors) on relevent procedures and methods. Frequently, this is done through a fairly structured survey process. In addition to providing the data needed to address the topic, this process is in itself useful to those who participate because it often encourages them to reexamine what they are doing.

There is no fixed time period for the work of the subcommittee, nor is there a set periodicity for the meetings. In general, however, the subcommittees have prepared their reports within 18 to 30 months and meetings are scheduled every 4 to 6 weeks. In one case, a follow-up subcommittee has been set up to consider the technical aspects of the implementation of the recommendations made by the original subcommittee; this committee was on statistical uses of administrative records.

After studying the area of concern, the subcommittee prepares a report which is published by the office in which the FCSM is located. These reports have formed a series called *Statistical Policy Working Papers*. After publication of the reports, FCSM-sponsored sessions are organized to discuss the findings and recommendations with other Federal employees. Reports on the work of the FCSM and its subcommittees have also been organized at meetings of the Washington Statistical Society and at the annual meetings of the American Statistical Association. Participants in these sessions include both Federal and non-Federal statisticians.

The *Statistical Policy Working Papers* published to date are:

(1) Report on Statistics for Allocation of Funds (1978).
(2) Report on Statistical Disclosure and Disclosure-Avoidance Techniques (1978).
(3) An Error Profile: Employment as Measured by the Current Population Survey (1978).
(4) Glossary of Nonsampling Error Terms: An Illustration of a Semantic Problem in Statistics (1978).
(5) Report on Exact and Statistical Matching Techniques (1980).
(6) Report on Statistical Uses of Administrative Records (1980).

The book review section of JASA (Journal of the American Statistical Association) has published quite favorable reviews of the *Report on Statistics for Allocation of Funds* (June 1980) and on the *Report on Statistical Disclosure and Disclosure-Avoidance Techniques* (December 1979).

The next sections of this paper discuss the history and the work of four subcommittees:

- Contracting for Statistical Surveys--Approaching the Problem.
- Statistical Interagency Agreements.
- An Interagency Review of Time-Series Revision Policies.
- Data Collection Methodology in Federal Government.

Other topics being addressed by subcommittees are the design and evaluation of questionnaires and the implementation of statistical uses of administrative records.

3. Contracting for statistical surveys

The Subcommittee on Contracting for Statistical Surveys was first organized in the Fall of 1977. An outline of the initial task was defined by the FCSM as follows.[3]

1. The task of the subcommittee should be developed sequentially comprising the following major steps (in chronological order):

[3] Memorandum from Maria E. Gonzalez to FCSM on "Work Plan for a Possible Subcommittee on Standards for Contracting Statistical Surveys", September 9, 1977.

A. Prepare a comprehensive list of topics in the realm of contracts for Federal statistical surveys. The life-cycle of the contract should be covered.
B. Determine priorities with respect to topics to investigate.
C. Investigation of topics selected.

2. Various topics were discussed as possibly within the scope of the task:
   A. In terms of types of contractual arrangements:
      i. Competitive Federal contracts for statistical surveys.
      ii. Sole source Federal contracts for statistical surveys.
      iii. Statistical interagency agreements.
      iv. Cooperative agreements with universities, etc.
      v. International contracts under PL-480.
      vi. Contracts in Federal-State cooperative agreements.
   B. In terms of types of statistical services:
      i. Developmental work on concepts and methods.
      ii. Studies of statistical methodology.
      iii. Data collection in statistical surveys.
      iv. Program evaluation.
      v. Analysis of data.
   C. Expertise of persons to include (or consult) in the subcommittee:
      i. Survey statisticians.
      ii. Sponsors of surveys with specific subject-matter knowledge.
         a. Policy/evaluation groups.
         b. Statistics and research groups.
      iii. Contracting officers.
      iv. Procurement experts.
      v. GAO auditors of statistical surveys.
      vi. Agency counsel with responsibility for reviewing contracts.
   D. How was agreement for contract carried out?
      i. Study of "Requests for Proposals" (RFPs) issued for competitive contracts.
         a. Select and study a sample of RFPs to determine characteristics.
         b. What firms get invited to bid?
      ii. How are statistical interagency (reimbursable) agreements achieved?
      iii. Investigation of methods used for sole source Federal contracts.
   E. Study of procedures used for selection of contractor; review of proposals:
      i. Membership of review committee.
      ii. How are points assigned?
      iii. Is there an acceptable range for evaluating proposals?
   F. Post-award activities:
      i. Modification of contracts.
      ii. Monitoring of progress of contract.
      iii. Need for termination of contract.
   G. Evaluation of final product:
      i. Have the needs of the sponsor been met adequately?
      ii. Performance evaluation of contractor for future reference (need for legal advice).

The subcommittee's initial discussions considered the use of contracts vs. in-house work, as well as the use of contracts vs. statistical interagency agreements. After many hours of deliberation a working group was set up to investigate statistical interagency agreements (that is, contracts between Federal agencies). This working group later became a separate subcommittee.

Over the years, there has been much turnover in the membership of this subcommittee as well as refocusing of its objectives. However, in the past year the membership has been stable, the outline for the final report has changed little and significant progress has been made in drafting the report. The main topics covered are:

1. Introduction (focus and scope of report);
2. Technical skills needed in contracting for statistical surveys;
3. Considerations in the preparation of requests for proposal (RFPs);
4. Contractor selection process;

5. Contract management and evaluation: post-award activities;

Appendix A. Informal survey of agencies and contractors concerning contracting for statistical surveys;

Appendix B. Office of Management and Budget Requirements: Clearance and Information Collection Budget.

One of the early activities of the subcommittee was to interview nine agencies and nine contractors to identify practices and pitfalls of the contracting process. Appendix A of the report will summarize these interviews.

The report being prepared should be useful to Federal employees involved in specifying statistical contracts, selecting contractors or monitoring statistical surveys. Based on the OMB clearances (August 1981) about 25 percent of the approved dockets (excluding application and regulatory requests) involve private contractors; for clearances involving telephone data collection 67 percent involve private contractors; for data collected by personal interviews 47 percent involve private contractors; and for those involving both telephone and personal interviews 61 percent involve private contractors. Therefore, the contracting process is an important factor in Federal data collection.

Even though focusing the work of this subcommittee and keeping a stable membership were difficult tasks, the initial draft report is now almost complete. A first report of this material is given in the paper entitled "Contracting for Statistical Surveys – – Approaching the problem."[4]

4. Statistical interagency agreements

A working group of the subcommittee on Contracting for Statistical Surveys was formed to study the issue of statistical interagency agreements. Statistical interagency agreements are contractual agreements between Federal agencies where the sponsor agency agrees to pay for the services done by the contractor agency. A work plan written in 1978 states that a description of how interagency agreements are arranged, as well as identification of the various circumstances under which interagency agreements are suitable would be the foci of this group's work.

To illustrate the process of changing the focus of groups an example will be cited. A memorandum from Monroe Sirken (chair of the subcommittee) on "Plans for the Working Group on Interagency Agreements," June 30, 1978 states:

> "The goal of the Working Group (WG) has been modified. Originally, the WG planned to compare the cost, error, and administrative effects of three alternative arrangements available to a Federal agency for compiling and analyzing statistics: (1) in-house, (2) interagency agreements, and (3) contracts with non-Federal agencies. The WG soon recognized that the original objective was too ambitious. The WG is planning to limit its scope to a study of interagency agreements for statistical services."

This group carried out a small scale study of contractor and sponsor agencies of statistical interagency agreements to investigate (1) interagency agreement procedures and policies and how they vary from one Federal agency to another, and (2) the extent to which Federal agencies are involved in interagency agreements.[5] Based on the results of these interviews a final report has been prepared.[6] The

[4] Pearce, N. and Brehm, H.P. (1982): Contracting for Statistical Surveys– –Approaching the Problem. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 165–170. A final report was issued as Statistical Policy Working Paper 9, *Contracting for Surveys*, OMB, Washington, DC, 1983.

[5] Copies of the sponsor and contractor questionnaire are available upon request from M. Gonzalez, OMB, NEOB, Rm. 3019, Washington, DC 20503.

[6] Statistical Policy Working Paper 8, *Statistical Interagency Agreements*, OMB, Washington, DC, 1982.

draft report was circulated to the agencies that participated in the feasibility study and their extensive comments have been incorporated into the final report. The topic of statistical interagency agreements is of interest basically to the Federal agencies involved in these studies. Their importance is, however, much broader since important statistical data are collected through interagency agreements. The Current Population Survey, providing monthly estimates of employment and unemployment, and the Health Interview Survey, which is the source for health prevalence data, are examples of major surveys conducted through statistical interagency agreements.

5. An interagency review of time-series-revisions policies

In the Fall of 1980, a subcommittee was formed to address the issue of making and publishing revisions and corrections to time series data. The initial specification of the problem pointed out some issues which might be addressed:

- Frequency of revisions.
- Labeling of preliminary estimates and revisions in publications and in public use tapes.
- File maintenance procedures. Establishment of audit trails at the micro and macro levels showing amount and reasons for change.
- How large a change is needed to justify a revision?

After a few meetings, it was decided to prepare a questionnaire to collect information about revision of time series. The topics covered in the questionnaire were: identification of time series, description of series, timing of release of statistics, reasons for revisions, magnitude of revisions, methods of dissemination and methodologies used in preparing original and revised data, and policies for revisions.[7] The agencies represented in the subcommittee were asked to select a few series published by their agencies and to complete the questionnaire with respect to these series. The results were summarized and a report including eight recommendations was prepared. These recommendations used input from the survey, as well as existing Federal practices in this area. "An Interagency Review of Time-Series Revision Policies" represents the report of this subcommittee.[8]

6. Telephone data collection in Federal statistics

In the Fall of 1980 a subcommittee was started to study the role of telephone, mail and personal interview surveys in Federal statistics. One of the first decisions of this subcommittee was to concentrate only on telephone surveys. This decision was based on the fact that the methodology of telephone surveys has changed dramatically in recent years and that several Federal agencies are beginning to take advantage of these new developments.

To determine the prevalence of the use of the telephone for Federal data collection, a request was made for a listing of the projects approved in the clearance process of the Office of Management and Budget (OMB). The various characteristics given on the request form (SF–83, 3–81 version) were included in this listing. Based on the listing, tabulations were prepared of the clearances for surveys which report collecting data by telephone.

The next planned step of this subcommittee is to gather further information on the characteristics of projects which report using the telephone for data collection and to identify issues which will improve the use of telephone data collection in Federal Government. This subcommittee is still active and a progress report of their findings is given in the paper on

[7] A copy of the questionnaire is available upon request from M. Gonzalez, OMB, NEOB, Rm. 3019, Washington, DC 20503.

[8] Statistical Policy Working Paper 7, *An Interagency Review of Time-Series Revision Policies*, OMB, Washington, DC, 1982.

"Data Collection Methods in Federal Statistical Surveys."[9]

7. Organization of the FCSM

The FCSM has provided the leadership in the selection of topics to be studied and has also served to review the reports of the subcommittees. Further the FCSM has encouraged the work of Federal interagency groups in investigating crosscutting methodological issues of importance to the Federal statistical system. The interagency group setting has provided a fruitful opportunity for persons with similar interests from different agencies to meet and obtain a good exchange of ideas. This type of forum serves to strengthen the Federal statistical system.

As chair of the FCSM, one of the author's duties is to work with the subcommittees to ensure results which are both useful and timely. Work with the subcommittees has many different aspects. For example, one needs to consider how to maintain sufficient interest of the members in the task in order to achieve the defined objectives. The task of focusing the work of the subcommittee takes ingenuity and much effort. Most of the subcommittees have modified their objectives several times before settling on the task of preparing a final report. The necessary process of focusing and refocusing a report can be both time consuming and frustrating to the members of the group. However, the final product is likely to be improved by sharpening the objectives of the task to be accomplished.

Another important function is the selection of members for the subcommittees. Since subcommittees are open to any Federal person interested in working on the selected topic, the selection is done by assigning tasks and determining whether the person handles the assigned task effectively.

Experts from outside Federal agencies have sometimes been used to provide technical support to subcommittees. For example, Tore Dalenius worked very effectively in support of the Subcommittee on Statistical Disclosure and Disclosure-Avoidance Techniques; Tom Jabine is now working with the Subcommittee on Implementation of Statistical Uses of Administrative Records.

Once a final report is prepared, its publication and dissemination are also important aspects of the work. In 1982 reports were issued on "Time-Series Revision Policies" and on "Statistical Interagency Agreements." In 1983 "Contracting for Surveys" was issued. An additional step which may sometimes be necessary is to determine the implications of the report for existing statistical standards or those to be developed. It is essential to develop operational recommendations to implement within the Federal statistical system the work of the subcommittees.

[9] Thornberry, O., Nicholls, W. and Kulpinski, S. (1982): Data Collection Methods in Federal Statistical Surveys. American Statistical Association, Proceedings of the Section on Survey Research Methods, pp. 185–190.

Statistical Disclosure Control

by Ove Frank[1]

1. Introduction

In recent years there has been an increased interest in individual privacy in connection with release of statistical data. See Dalenius (1974) for an overview and Cassel (1976), Cox (1980, 1981), Dalenius (1977), Duncan and Lambert (1982) and Frank (1976, 1978, 1979) for various approaches to disclosure control.

The purpose of this article is to review two different approaches to disclosure control and discuss and illustrate some fundamental problems. The first approach introduces a probabilistic model for individual data and controls the potential disclosure possibilities in the summary statistics that are considered for release. This control is based on the conditional distribution of individual data for fixed summary statistics. The summary statistics are not supposed to be released if the conditional distribution of individual data is too concentrated. This approach is discussed in Sections 2 and 3.

The second approach is also based on probability models but not for individual data. Rather a disclosure is specified as a piece of exact knowledge of the value of some sensitive variable for a particular individual, and probability models are used to describe the uncertainty of such knowledge. This idea is developed in Sections 4 and 5. Finally Section 6 gives some concluding comments on the applicability of the two approaches and on the need for further development.

[1] Professor of Statistics, University of Lund, Sweden.

2. Conditional information about individual data in summary statistics

Various authors have used an approach to disclosure control that can be described generally in the following way. Let n individuals have data given by random variables $X_1,\ldots,X_n$. Each random variable X_v might be a uni- or multidimensional variable containing some sensitive component. Let $T = T(X_1,\ldots,X_n)$ be some summary statistic obtained from $X_1,\ldots,X_n$: for instance, frequency counts, moments, or other statistics giving collective information. Generally, T is multidimensional and represents statistics considered for release. The distribution of $(X_1,\ldots,X_n)$ conditioned by T gives all the information considered to be relevant in order to decide whether or not T should be released. A peaked conditional distribution implies that T reveals too much individual information, while a more spread out conditional distribution indicates that T better protects individual information.

If individual data refer to a random sample or to uncertain measurements or observations, there may be no particular difficulties with this approach other than those that are generally present in statistical model building. However, if individual data refer to a census or to some other non-random set, then the probability distribution of individual data has to be defined in some other way. One possibility is a Bayesian approach with superpopulations. Another approach is predictive distributions obtained by some exchangeability condition that seems natural to the situation at hand as discussed by Duncan and Lambert

(1982). A similar approach discussed by Frank (1979) is to specify the probability distribution of individual data by considering their likeliness due to available background information. Prediction or guessing of individual data is also the motivation used by Cassel (1976) and Frank (1976).

As soon as a probability distribution for individual data $(X_1,\ldots,X_n)$ has been defined, it is conceptually straightforward to introduce a conditional distribution for individual data, given some statistic T. Technically there might be problems in deriving this conditional distribution, and it might therefore be hard to perform disclosure control. Another general problem is that the specified probability distribution for $(X_1,\ldots,X_n)$ can have unknown parameters in it that have to be estimated by using T only. This can introduce further technical problems before the disclosure control can be affected. It should be noted that if T is a sufficient statistic for the parameters in the initial distribution, then the conditional distribution does not contain any parameters, and it is immediately available for disclosure control.

As a simple example, consider the case of independent identically distributed Bernoulli(p) variables $X_1,\ldots,X_n$. Let the total frequency $T=X_1+\ldots+X_n$ be considered for release. The distribution of $(X_1,\ldots,X_n)$ conditional on $T=k$ is a uniform distribution over $\binom{n}{k}$ outcomes. Obviously, individual data are disclosed if $T=0$ or if $T=n$, and otherwise a reasonable disclosure control rule releases $T=k$ if $\binom{n}{k}$ is sufficiently large. Technical problems that occur for insufficient statistics can be illustrated in this example by considering the statistic $T'=\max(X_1,\ldots,X_n)$. The distribution of $(X_1,\ldots,X_n)$ conditional on $T'=1$ is a distribution over 2^n-1 outcomes that has probability $p^k(1-p)^{n-k}/(1-(1-p)^n)$ given to each outcome with $X_1+\ldots+X_n=k$ for $k=1,\ldots,n$. This distribution involves the unknown parameter p, and p is poorly estimated by using T'. In fact, the maximum likelihood estimator of p based on the likelihood of $T'=1$ is $\hat{p}=1$. This illustrates that disclosure control is not always immediate from the conditional distribution.

A more realistic example will be discussed in the next section.

3. Conditional information about individual data in contingency tables

Assume that individual data $X_1,\ldots,X_N$ consists of categorical information of some sensitive kind. There are c categories labeled $1,\ldots,c$ and $X_v=i$ if individual v belongs to the ith category. The categorical frequencies are given in a contingency table with c cells. Assume that this contingency table is considered for release or not.

Let $Y_1,\ldots,Y_c$ be the categorical frequencies. If X_v are independent identically distributed random variables with categorical probabilities

$$P(X_v = i) = p_i \qquad i=1,\ldots,c,$$

then $(Y_1,\ldots,Y_c)$ is multinomially distributed with parameters N and $p_1,\ldots,p_c$, and it follows that the conditional distribution of $(X_1,\ldots,X_N|Y_1,\ldots,Y_c)$ is uniform over $\binom{N}{Y_1,\ldots,Y_c}$ outcomes. Consequently, a reasonable disclosure control rule releases the contingency table if $\binom{N}{Y_1,\ldots,Y_c}$ is sufficiently large. The logarithm of this multinomial coefficient can be approximated according to Stirling's formula by

$$-\sum_{i=1}^{c} Y_i \log (Y_i/N),$$

and therefore the rule essentially releases the contingency table if the entropy

$$-\sum_{i=1}^{c} (Y_i/N) \log(Y_i/N)$$

is sufficiently large. Disclosure rules based on entropies have been discussed by several authors. See, for instance, Frank (1978) and

Duncan and Lambert (1982) for other rules based on entropies.

If focus is on a single individual and not on the total set of N individuals, then the conditional distribution of interest is

$$P(X_v = i|Y_1,\ldots,Y_c) = Y_i/N \qquad i=1,\ldots,c$$

and the entropy above is a measure of how spread-out this distribution is. Consequently the same disclosure control rule seems reasonable both for protecting a single individual and for protecting the whole group of individuals in the data set.

If a contingency table considered for release does not report frequencies for all N individuals of interest but only for a simple random sample of n individuals, then the appropriate conditional distribution might be

$$P(X_v = i|Y_1 \geq n_1,\ldots,Y_c \geq n_c) \qquad i=1,\ldots,c,$$

where $n_1,\ldots,n_c$ are the reported frequencies for the sample. By considering separately the cases when v belongs and does not belong to the sample, this conditional distribution is found to be given by

$$-[n_i + (N-n)p_i]/N \qquad i=1,\ldots,c.$$

Frank (1979) has shown that the entropy of this distribution is at least equal to the product of the sampling fraction n/N and the entropy of the sample frequencies, i.e.

$$(n/N)\sum_{i=1}^{c}(n_i/n)\log(n_i/n).$$

Consequently, in this case when there are unknown parameters $p_1,\ldots,p_c$ in the conditional distribution of interest, the entropy of this distribution can be bounded below by an expression that does not contain unknown parameters, and the disclosure control rule can be based on whether or not this lower bound is sufficiently large. This rule never releases a frequency table that should not be released according to the rule based on the entropy of the conditional distribution.

4. Prior knowledge needed together with data considered for release in order to make additional disclosures

An approach to disclosure control that has been discussed by Frank (1976, 1978, 1979) uses a strict sense concept of disclosure meaning that knowledge is available to find the exact value of a sensitive variable for an individual. Thus, the disclosure of an individual's value is based on available background information, previously released data, prior knowledge, etc. Since the information about what kind of such knowledge is available necessarily involves uncertainty, a probability model is introduced for the disclosures in the following way.

Consider n individuals and a sensitive variable with c values or categories labeled $1,\ldots,c$. Let $X_1,\ldots,X_n$ be the categories of the individuals. Let $\alpha_1,\ldots,\alpha_n$ be variables indicating disclosure possibilities prior to the release of a statistic T, i.e., α_v is 1 or 0 according to whether or not X_v can be deduced before the release of T. Let $\beta_1,\ldots,\beta_n$ be variables indicating disclosure possibilities after the release of T.

A very simple model for prior and posterior disclosure uncertainty can be specified as follows. Assume that $\alpha_1,\ldots,\alpha_n$ are independent Bernoulli variables with a common probability p of prior disclosure. Let T be the frequencies of individuals in the c categories, i.e., $T=(n_1,\ldots,n_c)$ where n_i is the number of individuals v with $X_v = i$. The adjusted frequencies of not disclosed individuals in the categories are given by

$$T(\alpha) = (n_1-m_1,\ldots,n_c-m_c),$$

where $\alpha = (\alpha_1,\ldots,\alpha_n)$ and m_i is the number of individuals v having $\alpha_v = 1$ and $X_v = i$. Assume that $\beta_v = 1$ if $\alpha_v = 1$ or if $\alpha_v = 0$ and $T(\alpha)$ is a degenerate frequency distribution with only one positive frequency. This is an as-

sumption that means that only the disclosures specified by α are combined with T in order to obtain further disclosures, and there is no way to use previous partial knowledge about X_v. This is certainly an oversimplification, and this model serves merely to illustrate that disclosure rules can be obtained with this kind of approach. More complicated models with coalitions and other disclosure possibilities are considered by Frank (1976). Now we can calculate the probability distribution of $\beta = (\beta_1,\ldots,\beta_n)$ and, for instance, find the probability P of at least one additional disclosure caused by release of T. According to Frank (1979, Formula (31)) it follows that

$$P = \sum_{i=1}^{c} (p^{n-n_i} - p^n) ,$$

which for small values of p can be approximated by

$$P \approx p^{n-\max n_i} .$$

Consequently a reasonable disclosure control rule releases T if P is sufficiently small, i.e. if $\max n_i$ is sufficiently small. Duncan and Lambert (1982) have motivated this and similar rules by other arguments.

The reasoning above assumes that a single user of T is making disclosures. In the next section the model will be extended to cover the situation with internal disclosures among the n individuals in the data.

5. Disclosure graphs

The concept of a disclosure graph is useful to describe internal disclosures. In order to introduce a disclosure graph, we consider the case of n individuals with categorical data $X_1,\ldots,X_n$ and a table of frequency counts $T = (n_1,\ldots,n_c)$ for c categories. The disclosures prior to the release of T are indicated by variables

$$\alpha_{uv} = \begin{cases} 1 \text{ if } X_u \text{ is known by } v \text{ prior to the release of } T \\ 0 \text{ otherwise,} \end{cases}$$

and the disclosures posterior to the release of T are indicated by variables

$$\beta_{uv} = \begin{cases} 1 \text{ if } X_u \text{ is known by } v \text{ posterior to the release of } T \\ 0 \text{ otherwise.} \end{cases}$$

The prior and posterior disclosures can be represented by directed graphs on the set of individuals. There is an arc from u to v if X_u is known by v. Let α and β denote the prior and the posterior disclosure graphs. The difference $\delta = \beta - \alpha$ defined by $\delta_{uv} = \beta_{uv} - \alpha_{uv}$ is the graph of additional disclosures made possible by the release of T.

Now, by arguing as in the preceding section, we can adjust the frequencies in T separately for each individual and subtract the numbers of prior disclosures in order to obtain individual frequency tables $T_v(\alpha)$. If we use the oversimplified model with α_{uv} for $u \neq v$ being independent Bernoulli(p) variables, then it follows (see Frank (1976)) that the probability of at least one additional disclosure caused by the release of T is given by

$$P = 1 - [1 - \sum_{i=1}^{c} (p^{n-n_i} - p^n)]^n ,$$

which is approximated

$$P \approx np^{n-\max n_i}$$

for small values of p. This formula shows how P varies with the prior disclosure probability p. If two different tables $T = (n_1,\ldots,n_c)$ and $T' = (m_1,\ldots,m_c)$ with totals n and m ($n \leq m$) are considered for release, then T has a smaller P for all p if

n-max $n_i > m$-max m_i ,

and if this inequality does not hold, then it depends on p which one of the two tables has the smaller P.

6. Comments

Both the approaches to disclosure control discussed above have obvious disadvantages. The first approach does not include background information or give any suggestions concerning how to take into consideration that it is usually not known what kind of prior knowledge is available for making disclosures.

The second approach has a restricted concept of disclosure that does not distinguish between various degrees of partial knowledge about sensitive variables. Also there is no general way to tell how posterior disclosures are obtained from the prior disclosures and the kind of data considered for release. The model considered in Section 4 is an oversimplified attempt to handle data in the form of contingency tables, and it does not work at all for other kinds of data.

The concept of a disclosure graph can be useful if extended to cover more general kinds of internal disclosure. This is achieved by assigning a conditional distribution $(X_u|\theta_v)$ or some characteristic of this distribution to the arc from u to v; here X_u is the data of u and θ_v is the prior information available to v. The prior information θ_v can be considered as a parameter or as a random variable. For instance the particular case $\theta_v = X_v$ with arc values given by expected conditional entropies $EH(X_u|X_v)$ might be of interest to study complicated dependence structures among random variables $X_1,\ldots,X_n$.

An interesting observation can be made in connection with the conditional distributions discussed in Section 2. When data $(X_1,\ldots,X_n)$ has a distribution depending on some parameters for which T is a sufficient statistic, then the conditional distribution of $(X_1,\ldots,X_n|T)$ does not contain the parameters, and it is usually considered to be of no interest, since all the information required for statistical inference purposes is contained in the distribution of T. Now, when we are interested in predictions of individual data from a statistic T considered for release, also the conditional distribution appears to be of importance.

7. References

Cassel, C. M. (1976): Probability Based Disclosures. In T. Dalenius and A. Klevmarken (Eds.). Personal Integrity and the Need for Data in the Social Sciences. Swedish Council for Social Science Research, Stockholm, pp. 189–193.

Cox, L. H. (1980): Suppression Methodology and Statistical Disclosure Control. Journal of the American Statistical Association, 75, pp. 377–385.

Cox, L. H. (1981): Linear Sensitivity Measures in Statistical Disclosure Control. Journal of Statistical Planning and Inference, 5, pp. 153–164.

Dalenius, T. (1974): The Invasion of Privacy Problem and Statistics Production – An Overview. Statistisk tidskrift 3, pp. 213–225.

Dalenius, T. (1977): Towards a Methodology for Statistical Disclosure Control. Statistisk tidskrift 5, pp. 429–444.

Duncan, G. T. and Lambert, D. (1982): A Model for Statistical Disclosure Control Based on Predictive Distributions and Uncertainty Functions. Department of Statistics, Carnegie-Mellon University, Pittsburgh, PA.

Frank, O. (1976): Individual Disclosures from Frequency Tables. In T. Dalenius and A. Klevmarken (Eds.). Personal Integrity and the Need for Data in the Social Sciences. Swedish Council for Social Science Research, Stockholm, pp. 175–187.

Frank, O. (1978): An Application of Information Theory to the Problem of Statistical Disclosure. Journal of Statistical Planning and Inference, 2, pp. 143–152.

Frank, O. (1979): Inferring Individual Information from Released Statistics. Invited paper for the 42nd Session of the International Statistical Institute, Manila, Philippines. To appear in the Bulletin of the International Statistical Institute.

Some Mathematical Problems Arising from Confidentiality Concerns

by Lawrence H. Cox[1]

0. Abstract

Selected mathematical formulations to define and control statistical disclosure in two-way tables are discussed.

Keywords: statistical disclosure, two-way tables, mathematical formulations.

AMS 1970 Subject Classification: 62P99.

1. Introduction

I am very pleased by the opportunity to contribute an article in the area of statistical disclosure to this special issue of *Statistical Review* dedicated to Tore Dalenius. Tore and I have shared strong interests, stimulating discussions and professional activities in the privacy and confidentiality field over several years.

This paper is prepared under the auspices of the Committee on Privacy and Confidentiality of the American Statistical Association, which I currently chair. The Committee was formed as a result of the activities and Report (American Statistical Association (1977)) of its precursor, the American Statistical Association (ASA) Ad Hoc Committee on Privacy and Confidentiality, with which Dalenius was active. At about the same time, Tore also was a member of the Subcommittee on Disclosure-Avoidance Techniques of the U.S. Federal Committee on Statistical Methodology. He made significant contributions to that Committee's *Statistical Policy Working Paper 2: Report on Statistical Disclosure and Disclosure-Avoidance Techniques. Working Paper 2* established a comprehensive framework for policy discussions of statistical disclosure and has become required reading in this field. For purposes of its discussion, *Working Paper 2* adopted a definition of disclosure developed by Dalenius (U.S. Department of Commerce (1977, p. 10)).

Tore has published extensively in the privacy and confidentiality field, most notably Dalenius (1974), entitled "The Invasion of Privacy Problem and Statistics Production – An Overview," and Dalenius (1977), entitled "Towards a Methodology for Statistical Disclosure Control." The latter article was one among over thirty reports comprising the *Confidentiality in Surveys Report Series* of the University of Stockholm established by Dalenius. The *Series* was particularly successful in disseminating research findings among practitioners in the field. I rate the *Series* high among Tore's many contributions in the privacy and confidentiality area. U.S. Bureau of the Census (1978), entitled *Information Privacy and Statistics: A Topical Bibliography*, provided a needed access to the privacy and confidentiality literature. This bibliography was the work of Tore Dalenius.

The purpose of this present article is to illustrate the kinds of mathematical problems and methods which arise from the problem of statistical disclosure control. The article is expository and is written to inform the reader interested in learning about disclosure methodology, but stops short of reporting specific results in detail. Appropriate references to

[1] Senior Mathematical Statistician, U.S. Bureau of the Census, Washington, D.C., USA.

such results are provided. To achieve concreteness and brevity, the discussion is limited to the problem of statistical disclosure control in two-way tables. Therefore, we do not deal with other interesting topics such as disclosure in microdata or statistical database security. For these topics, the reader is referred to various reports in the *Confidentiality in Surveys Report Series*, to Cox (1981a) and Dalenius and Reiss (1978) or to Denning (1982), respectively.

A broad technical discussion of the disclosure problem is provided in Cox and Sande (1979). The reader unfamiliar with the concerns of statistical confidentiality is referred to the papers by Dalenius previously cited or Chapters I–III of *Working Paper 2* for the requisite background and motivation.

2. Approaching Disclosure Mathematically: Disclosure Definition

The broadest interpretation of the Dalenius definition of disclosure (U.S. Department of Commerce (1977, p. 10)) is that published statistics cause disclosure of individual respondent data if, on the basis of the published data, more is known (or knowable) about the respondent by others than was known before publication.

The methodological problem at hand is to develop techniques to quantify and refine this definition and to develop techniques to control statistical disclosure in two-way tables. The essence of the problem can be stated simply: proceeding from intuitive notions of what constitutes statistical disclosure, frame a (mathematical/statistical) definition of disclosure which preserves these notions and, next, develop disclosure control mechanisms which are both sufficient and consistent with this definition.

For two-way tables of frequency counts, the widely accepted intuitive notion of statistical disclosure is that disclosure occurs if, on the basis of the released cell counts and publicly available non-confidential information (such as residence, occupation, age group, etc.), an individual respondent can be associated with a small subgroup of the population which is defined by at least one confidential characteristic. In other words, disclosure occurs if a respondent can be *identified* as possessing a confidential or sensitive characteristic. In this paper, we do not address the issue of distinguishing between confidential and non-confidential characteristics, but rather assume that such determination already has been made and that at least one confidential characteristic is present. However, it should be noted that, where confidentiality concerns are addressed completely, this determination is a key part of the analysis and often affects the definition of what constitutes a "small" subgroup of the population. For example, if the characteristic were truly non-confidential, then no subsets would be defined to be "small" (i.e., only subsets of "negative size"); subsets of size zero, one, two or three would by "small" in the case of confidential characteristics such as income class. The disclosure control problem for frequency count data in two-way tables is to prevent the derivation of "small" cell counts from the released table.

In the case of two-way tables of aggregate magnitude data such as economic data, usually it is not the presence of a respondent in particular cells of the table which is confidential but rather the values and sums of values of respondent contributions to cells. For example, assume that the table presents total sales receipts over sets of retail establishments in a state cross-classified by location (county within state) and type of retail activity, designated by *kind of business* category such as clothing, jewellery, sporting goods, general merchandise, etc. Assume further that the term "respondent" refers to legal entities such as individual companies, any one of which may consist of multiple establishments active in different counties and kinds of business. In this example, the places and kinds of business of these establishments and their affiliation with their

parent companies (respondents) typically would by well-known or readily discovered, so that this information is not confidential (or sensitive). However, what is sensitive are the values of the contributions of the individual respondent companies to cells and unions of cells in the table. The disclosure control problem in this context is to prevent close estimation of individual company data from the published cell values.

Because the fundamental unit of data release in a (two-way) table is the tabulation cell, disclosure in (two-way) tables is analyzed at the cell level and is defined in terms of cell values and their constituent parts. As discussed above, disclosure in frequency count tables is defined by *threshold rules* as the release of "small" cell values. More precisely, a cell $\underline{X}$ in a two-way table of frequency counts is defined to be a *sensitive* (or *disclosure*) *cell* if $V(X)$, the value of $\underline{X}$, satisfies $V(X) < n$. The parameter $\underline{n}$, the operational definition of "smallness", is called the *threshold value* and is fixed. Threshold rules are well-behaved mathematically and, in particular, enjoy the important property that the union of two non-disclosure cells necessarily is a non-disclosure cell. Disclosure control in frequency count tables amounts to reducing to an acceptable level the likelihood that an individual respondent can be identified as belonging to a particular cell in the table.

Disclosure definition in two-way tables of aggregate magnitude data, such as economic data, is more complex. We refer to the preceding example of a two-way table of sales receipts data. Each establishment contributes a non-negative value $\underline{x}$ to precisely one cell in the table and $\underline{0}$ to all other cells. This is analogous to the frequency count case wherein each respondent represented in the table contributes the value $\underline{1}$ to precisely one cell in the table and $\underline{0}$ to all other cells. However, because the unit of response (*respondent*) in the retail sales example is a company, a legal entity which often consists of several establishments doing business in different counties and/or different kinds of business categories, disclosure control in tables of aggregate magnitude data must assure that estimates of both company and sub-company (including establishment) data are rendered suitably imprecise. For example, if the ABC Merchandising Company consists of the DEF Clothing Store Chain and the GHI Waterbed Company, then aggregates by county for DEF or GHI or their totals (i.e., aggregates for ABC) all must pass the test of non-disclosure.

There are three general principles which must be observed when framing disclosure definitions and tests for aggregate magnitude data such as economic data. First, the test must be defined on a company (respondent) basis, so that all meaningful aggregates of establishment values are afforded disclosure protection. Second, the test must be defined so that it can be easily applied to meaningful unions of cells within the table. For example, if three counties within the table comprise a Standard Metropolitan Statistical Area (SMSA), then the test should be applicable to the union across any fixed kind of business activity of the corresponding three counties by kind of business cells. Analogously, if a subset of the kind of business categories forms a meaningful set, such as several clothing, shoes, etc., classifications forming an "apparel" category, then unions of corresponding cells also must be tested for disclosure. Third, disclosure risk must not increase through aggregation. As we soon shall see, this condition is equivalent to requiring that the disclosure/non-disclosure rule on which the test is based is *subadditive*. A consequence of subadditivity is that the collection of non-disclosure cells and cell unions in a table is closed under the operation "cell union". Because a *sufficient* disclosure methodology must provide disclosure protection to meaningful cell unions as well as individual tabulation cells, we sometimes use the term "cell" to refer to both individual cells and cell unions.

Familiar examples of disclosure rules for aggregate magnitude data are the "n respondent, k%" rules (*(n,k)-rules*) and *(p,q)-prior-posterior ambiguity rules*. (n,k)-rules define a disclosure cell as a cell for which the sum of the n largest respondent contributions to the cell exceeds k% of the cell value. (n,k)-rules are part of the larger class of *cell concentration rules*. These rules attempt to provide disclosure protection to respondents under the following assumptions: (1) the identities and number of respondents in a cell X are known and (2) coalitions of respondents within X consist of at most $(n-1)$ respondents. Such coalitions presumably pool their own contributions to form P, the sum of their contributions, and then estimate the contribution of another respondent (the *target respondent*) by $V(X)$-P. In cells for which the total contribution of all non-coalition respondents represents a close estimate of the contribution of the target respondent, disclosure occurs.

A somewhat more direct and perhaps more intuitive disclosure rule for coalitions requires that a coalition be unable to estimate from above the contribution of another respondent to within p% of its value, where the fixed value of the parameter p serves as the operational definition of "close" estimate. It can be shown that, given n, this so-called p%-rule is equivalent to an (n,k)-rule with $k=100\,(100/(100+p))$ (Cox 1981b.)

(p,q)-prior-posterior ambiguity rules are designed to prevent p% estimation of individual respondent data in a cell assuming that q% estimation can be achieved from prior knowledge. More precisely, under the assumption that, from general knowledge, anyone (another respondent in the cell, in particular) can estimate the contribution of the target respondent to within q% of its value, the problem is to develop a subadditive disclosure rule which prevents estimation of this respondent's data to within p% of its value, where $0 < p < q$.

Mathematics comes into play in defining appropriate disclosure rules and studying their properties. Cox (1981b) investigates this problem for disclosure rules S which are defined on a cell basis and which (1) are subadditive and (2) can be defined as linear functions of the contributions of individual respondents to the cell. Such rules are called *subadditive linear sensitivity measures* because they measure disclosure (*cell sensitivity*) on a continuous scale, in addition to their ability to distinguish between disclosure and non-disclosure cells. We sketch the salient results of Cox (1981b) below.

The following notation is used: X, Y and Z denote cells (or cell unions); x_i, y_i and z_i denote the ith largest respondent contributions to X, Y and Z, respectively (zero for respondents who do not contribute to the particular cell); and N denotes the total number of distinct respondents in the table. For notational simplicity, we assume that within each cell X, the respondents are enumerated in descending order of size, i.e., $x_i \leq x_j$ whenever $1 \leq j \leq i \leq N$. Therefore, the ordering of respondents typically is different in different cells. Also, whenever $Z = X \cup Y$ (the cell union), the contribution to Z of each respondent is the sum of its contributions to X and Y. The subadditivity condition guarantees that if X and Y are non-disclosure cells, then Z also will be a non-disclosure cell. If X is a disclosure cell, then X cannot be released but certain cells $Z \supset X$ can. The results on linear sensitivity measures in Cox (1981b) provide a mechanism for determining which cells $Z \supset X$ to release and which not to release. These measures may be applied with equal ease to categorical or magnitude data and their use is not restricted to two-way tables.

As the name implies, linear sensitivity measures are linear functions S designed to measure the extent of disclosure (sensitivity) in cells and cell unions X with regard to a specific disclosure rule. Although both disclosure

due to estimation of respondent and cell data from above (subadditive, upper linear sensitivity measures) and disclosure due to estimation from below (superadditive, lower linear sensitivity measures) must be considered, all important properties of linear sensitivity measures can be derived in the subadditive case. We therefore restrict our attention to subadditive linear sensitivity measures and the problem of disclosure due to estimation from above.

With the notation established above, $S(X) = w_o + \sum_{i=1}^{\infty} w_i x_i$, where $x_i = 0$ for $i > N$ and where the w_i are constants determined by the disclosure rule. *Disclosure* occurs whenever $S(X) > 0$. S is *subadditive* if $S(X \cup Y) \leq S(X) + S(Y)$, so indeed the union of two non-disclosure cells remains a non-disclosure cell. The (subadditive upper) linear sensitivity measure corresponding to the (n, k)-rule is

$$S_D(X) = \sum_{i=1}^{n} ((100\text{-}k)/100)x_i + \sum_{i=n+1}^{\infty} (-k/100)x_i. \tag{2.1}$$

The linear sensitivity measure corresponding to a (p, q)-prior-posterior ambiguity rule is

$$S_{p,q}(X) = (p/100)x_1 - \sum_{i=n+1}^{\infty} (q/100)x_i. \tag{2.2}$$

That corresponding to the n-threshold rule is

$$S_n(X) = n - \sum_{i=1}^{\infty} x_i. \tag{2.3}$$

The two main results of Cox (1981b) are to characterize subadditive linear sensitivity measures completely and to study the problem of determining when a cell union Z containing a sensitive cell X will be non-sensitive. The subadditivity result is that $S(X) = w_o + \sum_{i=1}^{\infty} w_i x_i$ is subadditive if and only if the sequence of weights $\{w_i\}_{i=1}^{\infty}$ is non-increasing (Cox (1981b, Theorem 1)). This result is both powerful and easy to use. Immediately one sees that $S_D(X)$ and $S_{p,q}(X)$ are subadditive, results which otherwise would have to be developed directly. For more complicated rules such as rules which combine assumptions about coalitions and prior knowledge, a direct verification of subadditivity might be difficult or impossible, whereas the non-increasing condition on the w_i is easily checked.

The other main result of Cox (1981b) has two parts. The first part establishes a necessary condition for a cell union Z containing a sensitive cell X to be non-sensitive. This condition is stated in terms of $V(Z)$ (Cox (1981b, Proposition 3)). Specifically, Z must satisfy

$$V(Z) \geq S(X)/|w| + V(X), \tag{2.4}$$

where $w = \min \{w_i\}_{i=1}^{\infty}$. This result is of crucial importance because it links disclosure definition to disclosure protection. The second part of the second main result establishes that necessary conditions such as (2.4) which are stated only in terms of $V(Z)$ and ignore individual respondent contributions may fail the test of sufficiency (Cox (1981b, Proposition 4)). By combining these results, a mathematical framework is formed within which methods of disclosure definition, measurement and control can be linked.

3. Controlling Statistical Disclosure: Cell Suppression

The methods of the preceding section, although applicable generally, are particularly well-suited to the requirements of the disclosure control technique *cell suppression*. For that reason, we discuss cell suppression first. Cell suppression can be characterized as withholding properly chosen cell values from publication in order to ensure that only non-sensitive data can be inferred from linear relationships between the published values. These relationships define a class of cell values which are *effectively published* in the sense

that they may be derived through linear manipulation of published cell values. We continue to restrict attention to *disclosure from above* in two-way tables, i.e., disclosure which occurs when the value $V(Z)$ of a cell union provides a close upper estimate of the value of sensitive data. In the language of linear sensitivity measures, a *sufficient* cell suppression methodology is one which suppresses enough cell values from publication to ensure that, for any cell or cell union $\underline{Z}$ whose value $V(Z)$ is (effectively) published, the condition $S(Z) \leq 0$ holds. This means, of course, that all sensitive cells $\underline{X}$ are suppressed (i.e., all cells $\underline{X}$ in the table for which $S(X) > 0$), together with a number of appropriately chosen non-sensitive cells $\underline{Y}$, called *complementary suppressions*, which are sacrificed in order to ensure that $S(Z) \leq 0$ whenever $Z = X \cup Y$ and $V(Z)$ is (effectively) published.

Techniques of cell suppression are aimed at making good choices of the $\underline{Y}$'s in the sense that such choices minimize the consequent information loss to the data user. Cell suppression is a complex undertaking, even in the simple case of a single two-way table, primarily because of the many cell unions which are possible and, in the case of aggregate magnitude data, because of the myriad ways that respondent data can be distributed over the table cells and combined in cell unions. To focus our discussion, we restrict attention to tables in which the marginal totals entries are non-sensitive and further assume that no marginal cells are to be sacrificed as complementary suppressions. This assumption allows us to define information loss due to cell suppression unambiguously, as follows. This is two-tier definition.

Assuming that each of two sets $\underline{A}$ and $\underline{B}$ of complementary suppressions within the table are *sufficient* (i.e., $S(Z) \leq 0$ for all (effectively) published values $V(Z)$), then the *information loss* due to $\underline{A}$ is less than that due to $\underline{B}$ if $\underline{A}$ involves fewer complementary suppressions than $\underline{B}$. If $\underline{A}$ and $\underline{B}$ involve the same number of complementary suppressions, then the information loss due to $\underline{A}$ is less than that due to $\underline{B}$ if the total value of the complementary suppressions in $\underline{A}$ is less than that in $\underline{B}$. Cox (1980) investigates cell suppression methodology within this framework. We summarize these and recent results.

Denote the cell in the (i,j)-position of our two-way table $\underline{T}$ by $\underline{X_{ij}}$. Whenever an individual cell $\underline{X_{ij}}$ is suppressed either as a disclosure cell or as a complementary suppression, we replace $V(X_{ij})$ by a variable $\underline{v_{ij}}$ in the table. Let $\underline{W}$ denote the set of (i,j)-values for which variables $\underline{v_{ij}}$ already have been defined, so that, initially, $W = \{(i,j)\colon \underline{X_{ij}} \text{ is sensitive}\}$. At any stage in the suppression process, each row and column of $\underline{T}$ gives rise to a linear equation in the $\underline{v_{ij}}$ of the form

$$\sum_{i \text{ (or } j)} v_{ij} = j\text{th column (or } i\text{th row) marginal total.} \tag{3.1}$$

Within the system of linear equations generated by the equations of (3.1), a one-to-one correspondence can be established between the set of those linear combinations of the equations of (3.1) involving only the coefficient value $\underline{1}$ and unions of table cells $Z = \bigcup_{(i,j)\in Q} X_{ij}$ whose values are (effectively) published. Explicitly, if

$$\sum_{(i,j)\in Q} v_{ij} = C(Q) = \text{constant} \tag{3.2}$$

is one such linear combination, defined by some subset Q of $\underline{W}$, then $V(Z)=C(Q)$ is effectively published, where

$$Z = \bigcup_{(i,j)\in Q} X_{ij}. \tag{3.3}$$

If any of the constituent $\underline{X_{ij}}$ *of* $\underline{Z}$ are sensitive, then $\underline{Z}$ must be examined for sensitivity as well. If $\underline{Z}$ is determined to be sensitive, then additional cells must be suppressed, i.e., the set $\underline{W}$, and hence the set of v-variables, must be expanded to ensure that (3.2) cannot be derived.

Within this framework, we observe a hierarchy of conditions which characterize a sufficient cell suppression strategy. (1) Each of the equations (3.1) must involve at least two variables. (2) Given a sensitive table cell $\underline{X}_{ij}$, any derived equation (3.2) involving $\underline{v}_{ij}$ must satisfy

$$C(Q) \geq S(X_{ij})/|w| + V(X_{ij}). \tag{3.4}$$

In particular, (3.2) must involve at least two variables. (3) If $Z = \bigcup_{(i,j)\in Q} X_{ij}$ is effectively published, then $S(Z) \leq 0$ must hold, i.e., respondent values must not combine to create disclosure in any cell union.

This hierarchy serves two important purposes. First, it allows us to characterize the problem of cell suppression in frequency count tables as a special, simpler case of cell suppression for aggregate magnitude data, as follows. Because each respondent contributes the value $\underline{0}$ to all but one cell in a frequency count table (to which it contributes the value $\underline{1}$), conditions (1) and (2) suffice to characterize disclosure in frequency count tables, viz., in a frequency count table, disclosure control is sufficient if for any effectively published cell or cell union $\underline{Z}$, the condition $V(Z) \geq n$ holds. A second simplifying feature of frequency count data is that lower estimates of the value of a sensitive cell are unimportant whereas, for aggregate magnitude data, disclosure protection must be provided both from above and from below.

The second and more important advantage of the hierarchy (1)–(3) is that it points out three levels at which methodological research must be aimed and facilitates concise formulation of the corresponding problems mathematically. This three-level characterization of the disclosure problem also distinguishes between solved and unsolved problems. We discuss this hierarchy of problems in the implied order of complexity.

The problem of determining the minimum number of additional suppressions necessary to produce a suppression pattern which is sufficient in terms of condition (1) has been solved (Cox (1980, Theorem 4)). Excepting trivial cases, this minimum number equals either the number of rows or the number of columns in the initial suppression pattern in which additional suppression is necessary, whichever is larger. The Theorem 4 also provides a constructive algorithm for generating such patterns. *Branch and bound* methods may be incorporated into this algorithm to produce a final suppression pattern which is both sufficient in terms of condition (1) and optimal with respect to the previously stated definition of information loss. These methods have been implemented successfully at the U.S. Bureau of the Census in a computer program called INTRA (for "intra-table complementary cell suppression"). Complementary suppression for the U.S. 1977 Economic Censuses was performed using INTRA, as well as for U.S. 1982 Economic Censuses.

Moving to condition (2), we observe two theoretical limitations of the Theorem 4. The first is that even when attention is restricted to a single row or column of the table which contains suppressions, the requirement that this row or column contain at least two suppressions is not sufficient. Specifically, if one of the suppressed cells $\underline{X}$ is sensitive and if $\underline{Y}$ denotes the other suppressed cell in the row or column, then it is possible that for $Z = X \cup Y$, $V(Z) = C(Q)$ fails (3.4). Fortunately, this limitation does not often present practical problems because of the typically many minimal patterns possible. The solution is to instruct the algorithm to ignore patterns which would fail (3.4) in any particular row or column. In the event that no such patterns exist, the algorithm of Theorem 4 is repeated to find cells Y_1, Y_2, etc. in such rows and columns until all corresponding cell unions $Z = X \cup Y_1 \cup Y_2 \ldots$ satisfy (3.4). These features are incorporated within the INTRA program.

The second limitation of the Theorem 4 is more fundamental. It amounts to the specification of a (computationally efficient) procedure for the following: identify all equations of the form

$$v_{ij}=\text{constant}, \qquad (3.5)$$

which are expressible as linear combinations of the row and column equations (3.1) of $\underline{T}$. This problem is unsolved. However, methods under investigation avoid this problem by imbedding "rectangles" of suppressions within suppression patterns to ensure that *no equations of the form (3.5) are derivable.*

Condition (2) gives rise to a linear programming problem which is more general than the problem cited above: within the system of linear constraints generated by the equations (3.1), for each sensitive tabulation cell $\underline{X}_{ij}$, determine the maximum value $\underline{M}_{ij}$ of the objective function $f=v_{ij}$.

From the general theory of mathematical programming, $M_{ij} = V(Z)$ for some cell union $\underline{Z}$ containing $\underline{X}_{ij}$. *If* $C(Q) = M_{ij}$ fails (3.4), then sufficient disclosure protection has not been provided to the respondents of $\underline{X}_{ij}$ and more suppression is necessary. Formulations for the efficient implementation of *disclosure checking procedures* such as (3.4) are discussed in Cox (1978) under the heading "validation and linear estimation". Although the present discussion is limited to checking for disclosure from above, analogous procedures based on lower linear sensitivity measures can be developed to check for disclosure from below.

Summarizing the sufficiency requirements of conditions (1) and (2), we obtain the following set of (difficult) mathematical problems: given a two-way table containing suppressed cells, determine best-possible (*exact*) upper estimates $\underline{M}_{ij}$ of $\underline{v}_{ij}$ corresponding to sensitive cells $\underline{X}_{ij}$; next, determine the minimum number of additional complementary suppressions necessary to ensure that each of these estimates satisfies the corresponding condition (3.4); and construct one (or more or all) such minimal patterns.

Solutions to the problems just stated would solve completely the cell suppression problem for frequency count data but, in the case of aggregate magnitude data, may still fail condition (3). The key difference between conditions (2) and (3) is that condition (3) formulates the disclosure problem at the respondent level rather than at the cell level. The sufficiency of condition (2) is a necessary condition for condition (3) to hold. I am not aware of any general formulations for solving (3). A brute force specification of a mathematical programming model which represents the constraint system at the respondent level offers a framework within which to address this problem, albeit computationally inefficient. Condition (3) amounts to the explicit construction and examination of all cells $\underline{Z}$ with the following properties: (a) $V(Z)$ is effectively published; (b) $\underline{Z}$ contains a sensitive tabulation cell $\underline{X}$; and (c) $\underline{Z}$ is minimal with respect to the operation "cell union" among all cells satisfying (a) and (b). This is very demanding computationally and calls for specially designed mathematical programming formulations which remain undiscovered. Practical implementations such as INTRA treat conditions (2) and (3) partially by heuristic methods.

4. Controlling Disclosure: Rounding

Cell suppression, including such derivatives as combining definitional categories, is the only disclosure control method to have been applied practically to tables of aggregate magnitude data. A wider range of techniques exists for the simpler case of frequency count data. These techniques involve *perturbing* the values of all (or most) tabulation cells rather than suppressing a select few.

The most thoroughly explored set of perturbation techniques are those which *round* the frequency count values to control disclosure.

Under a threshold definition of disclosure, all cells in the table (including marginal totals cells) are rounded to a multiple of the integer $\underline{n}$, where $\underline{n}$ is the threshold value. Thereby, users of the data are prevented from making inferences of the form $V(X) < n$ with certainty, so that disclosure control is achieved. To minimize *data distortion* within the table, rounding criteria require that the rounded values $R(X)$ of cells $\underline{X}$ within the table satisfy

$$|R(X)-V(X)| \leq n, \tag{4.1}$$

i.e., $V(X)$ is rounded up or down to an *adjacent integer multiple* of $\underline{n}$. Explicitly, $V(X)$ is rounded to $n[V(X)/n]$ or to $n([V(X)/n] + 1)$ where [ ] is the *greatest integer function*. The criteria on which rounding methods are judged are (1) whether the rounded marginal values also satisfy (4.1), (2) whether the array of marginal values is also a table (i.e., additivity along rows and columns is maintained), and (3) with respect to an appropriate measure of data-distortion, whether data distortion due to rounding has been minimized.

The most familiar form of rounding is *conventional rounding* (i.e., each $V(X)$ is rounded to $n[(V(X)/n) + 0.5])$. Although conventional rounding satisfies criterion (1) and is optimal with respect to criterion (3), it is undesirable as a disclosure control method because of its deterministic nature, viz., in many examples it is possible to work back to the original values. Nargundkar and Saveland (1972) introduced *random rounding*: for each cell $\underline{X}$, a random number $\underline{r}$ satisfying $0 \leq r \leq 1$ is generated; if, for

$$p = V(X)/n - [V(X)/n], \tag{4.2}$$

we have $0 \leq r \leq p$, then $V(X)$ is *rounded up* to n $([V(X)/n] + 1)$; otherwise, $V(X)$ is *rounded down* to n $[V(X)/n]$. In the language of probability, $V(X)$ is rounded up with probability $\underline{p}$ and down with probability $(1-\underline{p})$. Random rounding satisfies criterion (1) but, like conventional rounding, often fails criterion (2). However, it does guarantee that the *expected value* of the row and column sums of rounded values equals the (unrounded) row and column totals values. This feature is something of a substitute for (2) and is desirable relative to (3). Random rounding has been used by Statistics Canada to provide disclosure control in frequency count tables.

Fellegi (1975) improved the efficiency of random rounding relative to (2) for the case of one-way tables only. This technique, called *controlled random rounding*, is a random rounding algorithm in which careful attention is paid to the values of subtotals (*partial sums*) as each value on the row is rounded. The result is that the totals value for the one-way table (and, indeed, specified intermediate subtotals), is forced to satisfy (2). Unfortunately, a corresponding result for two-way tables does not exist.

Controlled rounding was developed by Cox and Ernst (1982). This technique is based upon a specialized mathematical programming model known as the *transportation model*. Controlled rounding answers affirmatively the question of whether an arbitrary two-way table of frequency counts always admits a rounding to base $\underline{n}$ which satisfies criteria (1) and (2). It then employs the transportation model to produce tables of rounded values which are optimal (minimizing) with respect to any of a large class of measures of data distortion, viz., the sum of absolute (or squared or m^{th} power) differences between corresponding rounded and unrounded values, or the maximum absolute value of such differences. If desired, zero values may be fixed before the rounding is performed.

Rounding methods vary in complexity, conventional rounding being the simplest and controlled rounding the most complex, but all can be implemented in a computationally

efficient manner. Rounding problems are interesting mathematically and enjoy a broad range of applications. For example, controlled rounding has several important statistical applications such as optimally *raking counts* in tables, viz., optimally estimating values of internal cells in a table from fixed values of the marginal totals, and to perform *controlled selection* of sampling units for statistical surveys (Causey et al. (1983)).

5. Concluding Comments

Statistical disclosure is an important statistical topic which presents interesting mathematical problems. This article attempts to illustrate these problems and their interrelationships. Solutions to this class of problems would form the basis of a unified disclosure methodology. Analogous sets of mathematical problems exist in other statistical areas such as editing questionnaire data for completeness and consistency. These areas are also in need of a unifying methodology.

Theoretical formulations which solve these problems are appealing to the theoretician and would provide the unifying framework needed to study statistical disclosure. However, statistical disclosure and similar problems are inherently practical problems – if disclosure is likely to occur then something (practical) must be done. By searching in the realm of mathematics for appropriate methods, the researcher seeks to avoid methods which are faulty or clumsy in favor of those which are efficient and broadly useful. A cell suppression or rounding algorithm will be applied thousands of times by a statistical agency. The underlying method must be both theoretically valid and computationally efficient.

For these reasons, I believe that researchers should look to mathematical programming theory for solutions to many problems in statistical disclosure. Controlled rounding of frequency counts is a mathematical programming formulation. Gordon Sande of Statistics Canada has performed some unreported work on applying mathematical programming models and linear sensitivity measures to the cell suppression problem. To ensure successful implementation, such new methods must exploit specialized mathematical structure inherent in the problems. Problems of theoretical interest arise in devising mathematical programming models which reduce to linear programs or to even more structured models such as transportation models which can be solved with great computational efficiency.

6. References

American Statistical Association (1977): Report of Ad Hoc Committee on Privacy and Confidentiality. The American Statistician, 31,2,pp.59–78.

Causey, B., Cox, L. and Ernst, L. (1983): Applications of Transportation Theory to Statistical Problems. Manuscript.

Cox, L. (1978): Automated Statistical Disclosure Control. American Statistical Association, Proceedings of the Survey Research Methods Section, Washington, DC, pp.177–182.

Cox, L. (1980): Suppression Methodology and Statistical Disclosure Control. Journal of the American Statistical Association, 75, 370, pp.377–385.

Cox, L. (1981a): Statistical Confidentiality in Microdata Release. Journal of Clinical Computing, 9, 6, pp.192–214.

Cox, L. (1981b): Linear Sensitivity Measures in Statistical Disclosure Control. Journal of Statistical Planning and Inference, 5, pp.153–164.

Cox, L. and Ernst, L. (1982): Controlled Rounding. INFOR, 20, 4, pp.423–432.

Cox, L. and Sande, G. (1979): Techniques for Preserving Statistical Confidentiality. International Statistical Institute,

Proceedings of the 42nd Session, Manila (to appear).

Dalenius, T. (1974): The Invasion of Privacy Problem and Statistics Production – An Overview. Statistisk tidskrift, Statistics Sweden, Stockholm, pp.213–225.

Dalenius, T. (1977): Towards a Methodology for Statistical Disclosure Control. Report o19, Confidentiality in Surveys Report Series, University of Stockholm, Stockholm.

Dalenius, T. and Reiss, S. (1978): Data Swapping – A Technique for Disclosure Control. American Statistical Association, Proceedings of the Survey Research Methods Section, Washington, DC, pp.191–194.

Denning, D. (1982): Cryptography and Data Security. Addison–Wesley, Reading, MA.

Fellegi, I. (1975): Controlled Random Rounding. Survey Methodology, 1, Statistics Canada, Ottawa, pp.123–133.

Nargundkar, M.S. and Saveland, W. (1972): Random Rounding: A Means of Preventing Disclosure About Individual Respondents in Aggregate Data. American Statistical Association, Proceedings of the Social Statistics Section, Washington, DC, pp.382–385.

U.S. Bureau of the Census (1978): Information Privacy and Statistics; A Topical Bibliography, by Tore Dalenius (Bureau of the Census Working Paper No. 41). U.S. Government Printing Office, Washington, DC.

U.S. Department of Commerce (1978): Statistical Policy Working Paper 2: Report on Statistical Disclosure and Disclosure Avoidance Techniques. U.S. Government Printing Office, Washington, DC.

Confidentiality Issues in Health Statistics:

An American Perspective

by Leo K. Lichtig[1]

0. Abstract

Conduct of health research and generation of health statistics increasingly involves use of large computerized data sets containing sensitive medical information on individual patients. These data sets are often collected by public agencies for a variety of uses, and are also made available to others for research use. *Protecting the confidentiality of these data is* both an ethical requirement of the health professions and a right that patients providing the information should expect.

Data release policies are needed that protect confidentiality while maximizing the utility of a data set. Developing and implementing such policies requires careful evaluation of a data set as well as an understanding of likely uses for the data. Finally, a researcher receiving data must take responsibility for protecting the confidentiality of that data and upholding the credibility of the agency which collected the data.

1. Introduction

The requirement that health practitioners strictly maintain the confidentiality of patient information dates back at least to the time of Hippocrates (460–377? BC), who included in his famous physician's oath:

> "Whatever, in connection with my professional practice or not in connection with it, I see or hear in the life of men, which ought not to be spoken of abroad, I will not divulge, as reckoning that all such should be kept secret." (Stedman and Garver (1939)).

This oath was easy to maintain when most care for a patient was provided by a single physician who kept his own patient records. Today, accurate patient information must be shared among physician specialists, insurance companies, law enforcement and other government agencies, epidemiologists, health service researchers, and so on. While the vast *majority of this information sharing has a distinct* patient or social benefit, it demonstrates that an individual physician taking the Hippocratic oath no longer has direct control over patient information.

While health data are only a special case of data on individuals, disclosure of health information, including erroneous information, can have a greater personal impact than disclosure of most other types of information. Disclosure of health information can potentially affect a patient's ability to obtain employment, insurance, credit, education and other equally important elements of life. In addition, disclosure could inhibit a patient's future openness with medical practitioners to the detriment of the patient's own health. For these reasons, there has been much public discussion in the United States on confidentiality of health data (e.g. Beggs-Baker et al. (1974); Curran et al. (1969); Martin (1974); Privacy Protection Study Commission (1977); Simmons (1980); Westin (1976)), as well as proposal and adoption of numerous laws and public policies (e.g. U.S. Public Law 93–579 (The Privacy Act of 1974); National Center for Health Statistics (NCHS) (NCHS (1978a, 1978b, 1980)); U.S. House of Representatives (1980). In addition, various professionals who handle health data

[1] Senior Policy Advisor, Blue Cross of Northeastern New York, Albany, New York, USA.

have adopted codes of ethics which require them to protect the data as stringently as physicians are pledged to. Notably, the American Medical Record Association (AMRA) has adopted a detailed policy statement on the confidentiality of patient health information in addition to their code of ethics (AMRA (1982)). This is appropriate since medical record professionals are the keepers of much of the raw, detailed information from which health related data sets are developed.

Adherence to such codes of ethics is essential if any system to prevent disclosure is to function, and if the public is to be confident that their information is properly protected. This latter point is extremely important since a *perceived* threat to confidentiality of privacy can be just as damaging as an actual violation (e.g. Meyers and Oliver (1978); Tanfer (1982)).

Access to health information is generally one of two types; individual access or statistical access. Individual access involves seeking information on a specific person or persons for such reasons as judicial or legal proceedings, audits by health insurance companies or agencies, public health and safety emergencies, or simply continuity of care for a patient. Statistical access is for research or statistical reporting purposes and involves seeking information on a large number of patients with certain characteristics. The focus of this article is on the latter type, and deals mostly with the application of data release policies specific to health data.

As a final word of introduction, this article only deals with policies regarding patient information. To what extent information about providers of health care should be considered confidential is a legal and ethical issue beyond the scope of this article.

2. Typical data sets

In order to develop data protection policies, one needs to know the content of data sets being protected and the purposes for developing the data set(s). The following data sets are typical of those used in the United States.

2.1. *Medical records*

A medical record is an organized collection of information about an individual patient describing an episode or episodes of care. Components of a medical record generally include medical history, diagnoses and treatments, responses to treatments, social influences on the patient's health such as job, family or financial problems, results of physical examinations and diagnostic tests, and any other information a health care provider may feel important to current or future care for a patient.

Raw medical records are generally not computerized, so protecting the confidentiality of their contents requires physical measures, such as locked files, proper identification of those seeking patient information, and only releasing information relevant to the needs of a requestor. With all of the concern over protection of computerized data files, many lose sight of the likelihood that more unethical people can break into a file cabinet and read records, or can misrepresent themselves as eligible to receive confidential information, than can break into a computer system and retrieve information stored there (see also Dalenius (1979)).

2.2. *Medical discharge abstracts*

Medical discharge abstracts are computer summaries of information in patients' hospital medical records including age, sex, diagnoses, procedures and clinical service. In the United States, virtually all hospital discharge abstracts contain a minimum data set called the Uniform Hospital Discharge Data Set (UHDDS) (NCHS (1980a)). Table 1 lists the UHDDS which will be used in examples later in this article.

Discharge abstracts provide an efficient and effective means for satisfying numerous information needs without having to reopen a patient's detailed medical record; this limits po-

Table 1. Functional Classification of UHDDS Data Items

Data Item	Function: Identification	Classification: Key[1]	Classification: Non-Key	Information: Sensitive	Information: Non-Sensitive
1. Hospital No.		X(3,4,+)			
2. Medical Record No.	X				
3. Admission Date		X(1,4,+)			
4. Discharge Date		X(1,3,+)			
5. Sex			X		
6. Race			X		
7. Birth Date		X(1,3,or4)			
8. Diagnoses				X	
9. Procedures				X	
10. Discharge Status					X
11. Payment Status				X	
12. Surgery Dates		X(1,+)			
13. Attending Physician Code				X	
14. Operating Physician Code				X	
15. Zip Code		X(1,3,or4)			

1 The numbers in parentheses are other data items generally needed to be able to identify an individual patient. Only likely combinations are listed; a "+" indicates that at least one other key item would be needed to identify a patient. An implicit key item is any value of a classification data item containing a single patient (e.g., if only one person was admitted from a specific zip code, then that zip code would identify that patient).

tential disclosure to the information on an abstract. Uses for medical discharge abstracts include hospital management information systems, reports required for accreditation, medical quality assurance systems, regional health planning, hospital payment and a variety of research projects inside and outside of a hospital.

2.3. *Third party payor claims*

Private or public payment for individual patient care is almost universal in the United States with about 90% of the U.S. population having public or private health insurance coverage (Health Insurance Association of America) (HIAA) (1982). These third party payors have a need for detailed medical information when the individuals they insure receive covered health care services. While claims payment falls into the category of individual access to health data, payors also use the claims data for statistical reporting, actuarial and research purposes.

Third party payor claims contain information to identify the provider and patient, a patient's clinical condition(s), and the services rendered. In addition, information is often included to identify place of employment, and other available insurance coverage.

In order for providers to release information to payors, patients sign data release forms permitting such release. This allows payors to obtain any and all information necessary to pay a claim. However, it does not relieve any provider from responsibility to release only necessary information, and it does not permit a payor to release information it obtains except to other payors covering a patient's care.

2.4. *Registries*

Disease registries collect information on all individuals with specific diseases. For example, cancer registries collect information on every person diagnosed as having cancer, including: name, address and other identification data;

type, site and stage of the cancer; data on the work and home environment; etc. Such registries are critical to long term epidemiological research, and they must contain detailed, identifiable, personal data. Their legal protection must, therefore, be a prime concern when creating a registry.

Similarly, vital statistics registries are maintained for birth, death and marriage records. Except for the fact that vital statistics registries can contain information on virtually every member of a population, they are similar to disease registries in their needs for protection.

2.5. *Clinical research*

In clinical research, experimental therapies are generally tested on human subjects. Tracking the course of recovery and post-recovery health status are required to determine effectiveness of a therapy, any side effects, and the impact on quality of life. Often, this means that numerous people will have access to a patient and a patient's records. In these circumstances, a patient must be fully apprised of these potential intrusions and must consent to a limited amount of loss of privacy. However, intrusions can be kept to a minimum, and only those professionally related to the research should be allowed any personal information.

2.6. *Surveys*

Health surveys, such as the National Health Interview Survey and the National Ambulatory Medical Care Survey (NCHS (1981)), provide population-based information that is not available from samples of patients admitted to health care institutions. These surveys can involve gathering detailed data on persons' health status or perceived health status, personal habits, diet, types of illness, etc. Patient identifiers are necessary for audit purposes or for repeat surveys over time. However, identifiers can generally be removed from the actual data files used for analysis, thus reducing the risk of disclosure without affecting the quality of any results. Identifiers could be maintained on a master file to which there is limited access.

A major distinction between health survey data and any of the above data sets is the direct involvement of an individual in the data release. All of the previous data sets, with the possible exception of the vital statistics registries, are created from medical records by persons other than a patient. In surveys, a person knows that information is being released, exactly what information is being released, and to whom it is being released. This does not diminish the need to protect survey data once collected, but it does eliminate the need for an intermediary between a patient and data collector to decide on what information to release to whom. (See Merrian et al. (1980) for a general discussion of confidentiality in survey research.)

3. Evaluating a data set

"Disclosure" of confidential patient information can be defined as the "revelation of sensitive information about an individual patient, that would not have been known from other sources" (Lichtig (1982a)). Thus, disclosure of confidential information requires that:

- the identity of a patient is known,
- the information revealed is of a sensitive nature, and
- the information revealed is not already known from other sources.

This definition, together with an awareness of the potential uses for a data set, allows for development of data protection policies that minimize the risk of disclosure of confidential information while maximizing utility of a data set to users.

3.1. *Evaluation of data elements*

Data elements on a micro-data file, as opposed to a file of summary statistics, can serve three functions (after U.S. Department of Commerce (1978)):

(1) *Identifying function*. These data elements definitively single out an individual person. Name, address and identification numbers, such as the U.S. Social Security Number, all serve identification functions.

(2) *Classifying function*. These data items are used to categorize a person for statistical reporting or analytical purposes. Sex, age, smoker/non-smoker, and birth weight are all examples of classification variables. Almost any data item can be used to categorize a patient, but certain categorizations could provide a key to the identity of a patient. "Key" classification items could be date of birth, country of birth, date of admission or discharge, etc. Generally, the fewer the number of people in a category, the greater the potential for disclosure. As an example, if a person knows only a patient's date of admission to a specific hospital and the patient's date of birth, then this person might have a good chance of uniquely identifying that patient in a hospital data base containing hospital identifier, admission date and birth date.

(3) *Information function*. These data items provide information about a patient, and are generally the main reason for collecting the data in the first place.

Information can be considered *sensitive* or *non-sensitive*. Sensitive data are items that a patient may not want publically known, would cause embarrassment to a patient, or would compromise the ability of the agency collecting tha data to continue to accomplish its function. A patient's medical history or source of payment (e.g. public assistance) could be considered sensitive. Non-sensitive data items are already generally known about a person, such as their height or length of stay.

The border between sensitive and non-sensitive information is not always clear, and can be very different from patient to patient. Therefore, the ideal goal in protecting confidentiality is to prevent identification of a patient and, thus, prevent the revelation of *any* information about that patient, sensitive or not. This ideal is impossible to guarantee without totally denying access to data. Therefore, a balance must be struck where the risk of disclosure is minimized, and is outweighed by the likely benefits resulting from release of the data.

As an example of evaluating a data set, Table 1 shows the UHDDS data items by function (adapted from Lichtig (1982a)). Developing such a table is by no means an exact science. Since many data items often serve more than one of the above three functions within the same set of data, an attempt must be made to classify each item as *primarily* serving an identification function, a classification function (separated into key or non-key), or an information function (separated into sensitive or non-sensitive). When more than one function is involved, assignment is made in the following order of priority: identification, key classification, sensitive information, non-key classification, and non-sensitive information.

The borders between key vs. non-key data, and sensitive vs. non-sensitive data are not always clear, so some subjectivity is involved when making these distinctions. In separating key vs. non-key data, some considerations are: the likely number of cases in a resulting cell (the smaller the number, the greater the threat), the amount of time and effort required to use a classification data item as a key, the amount of information that must already be known about a person before the data could be used as a key and the number of other data items needed simultaneously to use an item as a key.

3.2. *Outliers*

If a record is extremely unusual, then it is potentially identifiable. Very old patients (e.g. over 100 years), and cases with very long

lengths of stay (e.g. over 1 year) are rare in acute care hospitals. Such atypical values for data items can serve as keys to a patient's identity regardless of the function that the data item serves. This is especially problematic if unusual cases have been discussed in the press.

4. Data release policies

4.1. Preparing a data set for release

The primary public policy concern is for developing a rational balance between the public's right to know and an individual's right to privacy. Much socially valuable research can be performed using health care data files, and the public has a right to expect such research to be performed; being too protective can hurt valuable public research (e.g. Dalenius (1979)). The issue basically involves structuring data sets and data release policies that provide reliable and appropriately detailed data to researchers while minimizing the risk of disclosure.

There are numerous disclosure-avoidance options available for minimizing the risk of disclosure of confidential information. Selection of the most appropriate option, or combination of options for a data set depends on a knowledge of the data involved and an awareness of the likely uses for the data. The following options are not exhaustive of the possibilities, and creative alternatives can often be developed that are tailored to the special needs of a particular data set.

4.1.1. Remove identification and key data items

Under this option, all data items that serve an identification function or a key classification function are removed from each record. Since simple removal of some items may hinder certain types of research, a data item can often be replaced with non-confidential data based on the values removed.

For example, Table 1 shows that under this option the admission and discharge dates should be removed from UHDDS files prior to release. To make the resulting file more useful, the following information could be calculated prior to removal of the dates:

- length of stay;
- day of the week of admission and discharge;
- quarter of the year of admission and discharge;
- preoperative lengths of stay.

Including these calculated data would satisfy most researchers' needs for the removed dates.

4.1.2. Removal of sensitive data from unique cases

As already discussed, very unusual cases are potentially identifiable. In order to prevent disclosure of any confidential information in the event that such unusual cases are identified, the sensitive information on these records can be removed.

Again using the UHDDS as an example, all of the items identified as sensitive in Table 1 could be removed from the records of patients over 90 years old or with length of stay over 180 days. This would prevent disclosure of any sensitive information for these cases, and should not detrimentally effect any research effort since no empirical conclusions can generally be drawn from such low volume categories of patients.

4.1.3. Release of a sample of records

Many micro-data files released by federal agencies in the United States contain only a sample of the data actually possessed by the agency. This greatly reduces the risk of disclosure since anyone using the data would only have a limited chance, such as 1 in 100 for a 1% sample, of finding any single person in the sample. In addition, any unique record in the file would not likely be unique in the entire data set.

The appropriateness of this option is often a function of the size of the data set. For large national data sets, a small sample may provide sufficient information for researchers. How-

ever, if data sets are of small to moderate size and the effects being examined are small to moderate, then a large sample may be needed for productive research. The larger the required sample, the lower the level of protection afforded by this option. Since health research often falls into the category of small to moderate effects in small to moderate populations, sample data sets are not often appropriate.

4.1.4. Randomly switch sensitive data items

Under this option, all sensitive data items for a randomly selected number of similar patients would be switched. For example, all diagnoses, procedures, etc. on one record would be switched with those on another record. By doing this, even if a patient is identified using non-sensitive data items, one cannot know whether or not the diagnoses, procedures, etc. truly reflect that patient's hospitalization.

A risk accompanying this option is the likelihood that any violation of confidentiality would be a result of unauthorized access to the data rather than misuse by a principal investigator. An unauthorized user may not know about the switch and false information may be released about a patient. False disclosure can often be worse than accurate disclosure (e.g. an insurance company believing that a patient has a cardiac condition when this is not the case).

4.1.5. Add noise to the data

By adding random noise with very clear characteristics (e.g. normal distribution with specified mean and variance), the exact values of the data would not be known, and many statistical analyses would be unaffected.

This option is difficult to properly design and implement when multivariate analyses are anticipated as they often are in health research. In addition, this option is only applicable to data such as length of stay and age (at least ordinal level data). It is not applicable to data such as diagnosis and procedure codes (nominal data) which often represent the most sensitive data items in a health related data set.

5. Some examples

5.1. Confidentiality and data release in the New York Statewide Planning and Research Cooperative System

The New York Statewide Planning and Research Cooperative System (SPARCS) serves as a comprehensive hospital care information network established to accumulate uniform data on hospitals and the care provided in them. The State legislation establishing SPARCS (Title 10, Chapter 5, Section 405.30) defines its purpose as follows:

> "SPARCS is designed to contribute to the goal of providing high quality medical care at reasonable cost to the inhabitants of the State by serving as an information source for organizations and agencies seeking to promote the efficient delivery of health care services." (405.30.b.1)

The law goes on to specify some of the purposes for which SPARCS data may be used, including: financial studies, rate setting, surveillance of hospital utilization, health planning and resource allocation studies, epidemiology, and other health care research studies.

Among the data sets collected by SPARCS, two include data on all discharges from acute care hospitals in New York State, about three million records per year. One data set is an expansion of the national UHDDS called the Discharge Data Abstract (DDA). The other data set includes billing information from the hospital Uniform Billing Form (UBF–1) for each patient discharged. Tables 2 and 3 show these data sets.

All data items in the SPARCS patient data sets are considered either deniable or non-deniable, as shown in Tables 2 and 3. Deniable data items include identification items and key classification items as listed in the SPARCS legislation. Data requests involving only non-

Table 2. Sample of New York State Discharge Data Abstract Data Set

Data Item	Deniable	Data Item	Deniable
Medical Record Number	X	Other Diagnosis Codes[1]	
Admitting Number	X	Other Procedure Codes[1]	
Admission Date (Month and year only are non-deniable)	X	Dates of other Procedures[1]	X
Admission Hour		Transfer Status	
Admission Status		Primary Source of Reimbursement	
Age (in years)		Secondary Source of Reimbursement	
Sex		Total Acute Certified Days Used	
Race		Total Certified Alternate Care Days	
Zip Code		Provider Identification Number	X
Attending Physician License Number	X	Discharge Date (Month and year only are non-deniable)	X
Operating Physician License Number	X	Discharge Hour	
Other Physician License Number	X	Disposition	
Admitting Diagnosis Code		Diagnosis Related Group	
Principal Diagnosis Code		Length of Stay	
Date of Principal Procedure	X		

1 These items repeat 4 times allowing for 4 secondary diagnoses or 4 secondary procedures.

deniable data are honored without hearings as long as they are reasonable and technically feasible.

Since many valuable studies would benefit substantially from use of deniable data, the SPARCS legislation established a Data Protection Review Board (DPRB) to review and approve requests for deniable data. The DPRB is made up of 15 members from the state and federal governments, third party payors, hospitals, consumers, physicians, and other interested parties. The DPRB cannot approve any request unless a majority of the Board is satisfied that:

"a. the request is technically feasible in terms of the ability of SPARCS to produce the requested data;
b. the purpose of the request is consistent with the purposes for which data from SPARCS may be used;
c. the applicant is qualified to undertake the intended study; and
d. the applicant requires such data and is able to insure that patient privacy is protected." (405.30.d.iii)

Table 3. Sample of New York State Uniform Billing Data Set

Data Item	Deniable	Data Item	Deniable
Medical Record Number	X	Accomodation Days[1]	
Admitting Number	X	Accomodation Rate[1]	
Admission Date (Month and year only are non-deniable)	X	Accomodation Code[1]	
Patient's Date of Birth	X	Accomodation Total Charge[1]	
Sex		Accomodation Non-Covered Charges[1]	
Federal Tax Number (Hospital Identifier)	X	Ancillary Service Codes[2]	
No Fault/Worker's Compensation Code		Ancillary Total Charge[2]	
Commercial Policy Number	X	Ancillary Non-Covered Charges[2]	
Insurance Company Code	X	Total of Covered Charges	
Commercial Insurance Company Name		Total of Non-Covered Charges	
Amount Billed to Commercial Insurance		Discharge Date (Month and year only are non-deniable)	X
Patient Responsibility		Patient Discharge Status	
Blood Pints Furnished		Statement Covers Period (From/Through)	
Total Charge for Blood			

1 This set of items repeats 5 times allowing for 5 different accommodations during a stay (e.g. patient admitted to Intensive Care, then transferred to acute care on day 3 would have two sets of accommodation data).
2 This set of items repeats 33 times allowing 33 different ancillary services (laboratory, pharmacy, radiology, etc.).

With these policies, the degree of risk of disclosure from release of deniable data items can be weighed against the merits of proposed studies on a case-by-case basis. The primary drawback is the need to assemble a knowledgeable board for at least monthly hearings to review new requests.

5.2. Confidentiality and data release in the New Jersey Statewide Case Mix Payment System

As part of efforts in the State of New Jersey to develop a hospital payment system based on patient case mix, hospital costs for treating different types of patients were estimated (New Jersey State Department of Health (NJSDOH) (1979)). One output of this process is a cost-per-case file in which individual patient records contain estimates of the treatment costs for the case. The availability of cost estimates on a patient-specific basis is very valuable to many research studies in health economics. Since the New Jersey cost-per-case file was one of the first in the United States, and is still likely the largest of its kind, there are many requests for the file. This necessitated development of a data release policy to make the file available while protecting confidentiality.

Table 4. Functional Classification of New Jersey Cost–Per–Case File

Data Item	Function: Identification	Classification: Key[1]	Classification: Non-Key	Information: Sensitive	Information: Non-Sensitive
1. Hospital No.		X(7,8,+)			
2. Hospital Type			X		
3. Medical Record No.	X				
4. Patient Billing No.	X				
5. Admission Class			X		
6. Admission Hour		X(1,7)			
7. Admission Date		X(1,8,+)			
8. Discharge Date		X(1,7,+)			
9. Residence Code		X(1,3,or4)			
10. Sex			X		
11. Race			X		
12. Birth Date		X(1,7,or8)			
13. Age			X		
14. Referral Source					X
15. Diagnoses				X	
16. Procedures				X	
17. Discharge Status					X
18. Clinical Service				X	
19. Payment Status				X	
20. Transfer Destination					X
21. Special Care Units					X
22. Days in Special Care Units					X
23. Surgery Dates		X(1,+)			
24. Diagnosis Related Group Number				X	
25. Code Scheme					X
26. Consultations				X	
27. Physician Codes				X	
28. Zip Code		X(1,7,or8)			
29. Length of Stay					X
30. Costs by Cost Center					X

1 The numbers in parentheses are other data items generally needed to be able to identify an individual patient. Only likely combinations are listed; a "+" indicates that at least one other key item would be needed to identify a patient. An implicit key item is any value of a classification data item containing a single patient (e.g., if only one person was admitted from a specific zip code, then that zip code would identify that patient).

Most of the contents of the New Jersey cost-per-case file are listed in Table 4, along with categorization of each item into their primary function (Lichtig (1982a)). The data release policy adopted was to create two versions of the cost-per-case file (Lichtig (1982a, 1982b)). One version could be released with little risk of disclosure since the following modifications were made:

– Hospital number	Replaced with randomly assigned number unique to a hospital.
– Medical record number	Replaced with randomly assigned number unique to a record.
– Patient billing number	Deleted from record.
– Admission date	Replaced with day of the week of admission, and quarter of the year of admission.

– Discharge date	Replaced with day of the week of discharge, and quarter of the year of discharge. "Length of stay" was calculated from admission and discharge dates.
– Residence code	Deleted from record.
– Birth date	Replaced with age.
– Surgery dates	Replaced with preoperative length(s) of stay.
– ZIP code	Same as "residence code".
– Referral source code	Same as "residence code."

In addition, all sensitive data items were removed from the records of patients (a) over 90 years old, (b) with length of stay over 100 days, or (c) who are the only patient in a diagnosis related group (DRG) in a hospital.

There are numerous studies that would be detrimentally affected if certain of the data items, particularly hospital identification number and residence codes, were not available. Therefore, a second version of the file was developed that contained the residence code information (residence code and ZIP code) and hospital identification (in the form of a key to the randomly assigned numbers on the file). This second version could only be released after careful evaluation of a requestor's need for the additional information. The evaluation is performed by State Department of Health staff, and by the federal government under a contractual stipulation related to the hospital reimbursement project which was federally funded. This federal involvement brings to cost-per-case file under the Privacy Act of 1974 (PL 93–579) in addition to State regulation.

6. Conclusions

In the United States, as in other countries, maintaining accurate detailed medical information that can be easily retrieved by health care providers is an important component of high quality medical care. In addition to benefiting individual patients, such information systems contain valuable data that can be used in research to benefit society as a whole. Patients have a right to expect their medical information to remain confidential; society has a right to expect research to be performed that can generally improve health and the quality of life. We statisticians have an obligation to support policies that allow release of information useful for research while ensuring that the risk of disclosing any confidential information is kept appropriately low. Those of us engaging in health care research and using confidential information have an obligation to protect the data from disclosure. The latter point is significant because any indication of a violation of data in a researcher's possession may not only hurt the researcher's credibility and eligibility to receive data in the future, but may hurt the ability of the data collection agency to continue to obtain accurate, confidential information.

Data release policies, such as those described above, can be used to restrict release to data necessary for research and to minimize the risk of disclosure from the data that are released. Codes of ethics and constant sensitivity to data confidentiality are methods for data users to help ensure that necessary research data continue to flow.

7. References

American Medical Record Association (1982): Confidentiality of Patient Health Information: A Position Statement of the American Medical Record Association. Journal of the American Medical Record Association, 53, pp. 51–64.

Beggs-Baker, S., Nick, W., and Chase, R. (1974): Individual Privacy Considerations for Computerized Health Information Systems. Medical Care, 12, pp. 75–84.

Curran, W., Stearns, B., and Kaplan, H. (1969): Privacy, Confidentiality and Other Legal Considerations in the Establishment

of a Centralized Health-Data System. New England Journal of Medicine, 281, pp. 241–248.

Dalenius, T. (1979): Data Protection Legislation in Sweden: A Statistician's Perspective. Journal of the Royal Statistical Society, Series A, 142, pp. 285–298.

Health Insurance Association of America (1982): Soure Book of Health Insurance Data: 1981–1982. Washington, D.C.

Lichtig, L. (1982 a): Development of a Micro-Data Tape of New Jersey Cost per Case for Public Research. Trenton, N.J.: New Jersey State Department of Health.

Lichtig, L. (1982 b): Implementation of a Data Release Policy. Journal of the American Medical Record Association, 53, pp. 34–42.

Martin, M. (1974): Statistical Legislation and Confidentiality Issues. International Statistical Review, 42, pp. 265–281.

Merrian, I., Boruch, R., Dalenius, T., Duncan, J., and Jabine, T. (1980): Confidentiality of Survey Data: A Panel Discussion. Statistical Reporter, 80–12, pp. 423–433.

Meyers, S. and Oliver, J. (1978): Privacy and Hostility Toward Government as Reasons for Non-Response in the National Medical Care Expenditure Survey. American Statistical Association, Proceedings of the Social Statistics Section, pp. 509–513.

National Center for Health Statistics (1978 a): National Center for Health Statistics Policy Statement on Release of Data for Individual Elementary Units and Special Tabulations. Washington, D.C.: U.S. Government Printing Office (PHS 78–1212).

National Center for Health Statistics (1978 b): National Center for Health Statistics Staff Manual on Confidentiality. Washington, D.C.: U.S. Government Printing Office (PHS 78–1244).

National Center for Health Statistics (1980 a): Uniform Hospital Discharge Data: Minimum Data Set. Washington, D.C.: U.S. Government Printing Office (DHEW (PHS) 80–1157).

National Center for Health Statistics (1980 b): The Model State Health Statistics Act. Washington, D.C.: U.S. Department of Health, Education and Welfare (PHS 80–1458).

National Center for Health Statistics (1981): Data Systems of the National Center for Health Statistics. Washington, D.C.: U.S. Government Printing Office (PHS 82–1318).

New Jersey State Department of Health (1979): A Prospective Reimbursement System Based on Patient Case-Mix for New Jersey Hospitals 1976–1982. Washington, D.C.: U.S. Department of Health, Education and Welfare (HCFA 03009 9–79).

Privacy Protection Study Commission (1977): Personal Privacy in an Information Society. Washington, D.C.: U.S. Government Printing Office (052–003–00395–3).

Simmons, W. (1980): Confidentiality of Data in the Cooperative Health Statistics System. Washington, D.C.: National Center for Health Statistics (PHS 80–1459).

Stedman, T. and Garber, S. (1939): Stedman's Practical Medical Dictionary. Baltimore, Md.: The Williams & Wilkins Co.

Tanfer, K. (1982): Consent and Privacy in the National Survey of Family Growth. Washington, D.C.: National Center for Health Statistics (PHS 82–1365).

U.S. Department of Commerce. (1978): Statistical Policy Working Paper 2: Report on Statistical Disclosure-Avoidance Techniques. Washington, D.C.: U.S. Government Printing Office (003–005–00177–8).

U.S. House of Representatives. (1980): Federal Privacy of Medical Information Act (HRP 96–832). Washington, D.C.: U.S. Government Printing Office.

Westin, A. (1976): Computers, Health Records and Citizen Rights (National Bureau of Standards Monograph 157). Washington, D.C.: U.S. Department of Commerce.

Confidentiality in a Decentralized Statistical System

by Joseph W. Duncan[1]

1. Introduction

Tore Dalenius has developed a special reputation in the area of confidentiality. Therefore, this article is dedicated to recognition of his substantial contribution to the subject of statistical confidentiality.

2. The statistician's responsibility

The statistician has long insisted that the protection of data confidentiality is essential to ensuring the accuracy of statistical programs. Such protection provides an environment in which a high level of voluntary participation of respondents can be obtained.

Several years ago the importance of statistical confidentiality was well stated by Margaret Martin in an article entitled "Statistical Legislation and Confidentiality Issues", which appeared in the *International Statistical Review* in December, 1974. She stated:

> "Even when responses to requests for information are required by law, the success of a statistical program depends in large measure on the willing cooperation of respondents. Respondents who understand the purpose of the inquiry, who sympathize with the intended use of the information, and who believe that providing the government with the requested information will not harm them are more likely to answer truthfully and with a minimum of effort on the part of the data collection agency. One element in the enlisting such cooperation is the assurance of harmlessness to the respondent, and one of the most common methods for making such assurance in statistical data collection is the provision for keeping the replies confidential."[2]

It is generally recognized that the statistician's interest in the individual response is primarily related to the examination of the quality of the information provided. This interest in the individual response is designed to evaluate the reliability and utility of statistical aggregates derived from the individual records. *The statistician is not interested in information about individuals for action at the individual level.* The statistician is interested in information about individuals so aggregates can be estimated and an appropriate analysis can be formed. Unfortunately, this distinction is frequently not fully understood by those who want to restrict the access of statisticians to individual records.

The purpose of this article is to examine the procedure for permitting statistician access to individual records in a decentralized system in a manner which will not compromise the right of the individual or create undue stress on the societal system that provides the basis of information.

3. Basic principles

In recent years there has been a growing concern about the capability of linking record systems maintained in computer environments to create files of information about individuals and/or corporations which will violate their basic privacy. There are a variety of concerns. One concern relates to the tracking of individual transactions which may reveal insights in-

[1] Corporate Economist and Chief Statistician. The Dun & Bradstreet Corporation, New York, USA.

[2] Martin, M. (1974): Statistical Legislation and Confidentiality Issues. International Statistical Review, Vol. 42, No. 3, p. 275.

to business or individual, patterns behavior which are intended to be private. An equally disturbing concern, especially to individuals, is a fear that the information contained in their files may be incorrect with the result that the individual is unable to obtain credit, employment, or some other benefit.

Thus, the general fear of the public is that data provided for an explicit purpose will be misused *to affect their rights, benefits, or privileges or to provide an opening to investigation and/or regulation.* The computer is seen as the enabling mechanism where information provided in various aspects of one's social and business contacts will be collated, compared, analyzed, and used to restrict the freedom of the individual or corporation.

Since this concern is growing as computers become more pervasive, it is more and more difficult for the statistician to discuss record linkage or access to administrative records without compounding the fear of the individual that the results will be available to the non-statistician for administrative purposes.

In order to approach the issue of statistical confidentiality, it is first necessary to address the basic private and confidentiality issue. In a report issued nearly a decade ago, the Department of Health, Education, and Welfare in the United States (HEW) set forth basic principles which should be observed when developing certain data systems. These principles are:

(1) "The individual must be informed when asked to supply personal data for the system whether he is legaly required to, or may refuse, to supply the data requested, and also of any specific consequences for him of providing or not providing such data."

(2) The agency should "assure that no use of individually identifiable data made is not within the stated purposes of the system as reasonably understood by the individual, unless the informed consent of the individual has been explicitly obtained."

(3) The agency should further "assure that no data about an individual are made available from the system in response to a demand for data made by means of compulsory legal process, unless the individual to whom the data pertain a) has been notified of the demand and b) had been afforded the access to the data before they are made available in response to the demand."[3]

These principles were generally embodied in the United States in the Privacy Act of 1974. Further, with respect to statistical standards, these principles have been in place, at least since May 3, 1974, standards which were issued by the Statistical Policy Division of the Office of Management and Budget.[4]

While the basic principles just outlined are now generally accepted they have not been extended to meet the direct needs of the statistical community for linking files as a means of developing more accurate statistical aggregates. The requirement is to both protect the individual from misuse of individual records and to build confidence in the public that the bureaucracy does, in fact, honor that protection. In order to provide the environment for effective matching of records for statistical purposes, I have proposed three additional standards. These are:

i) Statistical agencies should have mandated legislative protection for the confidentiality of information collected solely for statistical purposes. This should apply to both corporate and personal data. The element of trust which is involved in voluntary submission of data should be backed up by clearly mandated pro-

[3] *Report of the Secretary's Advisory Committee on Automated Personal Data Systems,* U.S. Department of Health, Education, and Welfare, U.S. Government Printing Office, July, 1973 (Stock No. 1700–00116).

[4] The author was associate director of the Office of Management and Budget at the times these standards were issued on May 3, 1974. They were issued in OMB Circular No. A–46. (See especially Section 7A.)

tections so there is no uncertainty concerning the confidential nature of the data submission and so that voluntary data collection programs are effective. Even in mandated data collection efforts, it is essential to have cooperation of respondents if the data submission is to be accurate and comprehensive. Protection from disclosures helps assure that the quality of submission is of the highest possible order.[5]

ii) The uses of statistical data must be restricted to prevent their use in identifiable form for making determinations which affect a particular respondent. While this is partially covered in the first principle, it should be explicit that the confidentiality of the statistical data means that these data sets are not available for other regulatory, administrative, or judicial purposes within the same agency or department collecting the data. Hence, environmental data collected for statistical purposes should not be used for regulatory purposes. Statistical uses must be made clear at the outset. There must be no possibility of divergence in these uses. In effect, statistical data in statistical agencies, is thus placed in a "protected enclave."

iii) Exchange of data among the "protected enclave" should be feasible under controlled conditions. Comprehensive data systems concerning the interrelationships among various aspects of social and economic patterns require that various data sets be combined and studied jointly. Once the principle is set forth that the data will only be used for statistical purposes, there should be no concern about exchange of information among statistical agencies which have "protected enclave" status in law.

This principle for statistical data systems is by far the most controversial, especially among those individuals who wish complete knowledge and control of uses of data pertaining to them held by Federal agencies. For the long-range development of sound statistical information or social processes, however, I believe it is essential.

4. Building statistical enclaves

The first step in achieving this situation is the development of a clear legal status for "protected enclaves" for selected statistical agencies in the major departments. The statistical agency must be free of intervention in terms of unauthorized access to data. Employees should be subject to strict ethical standards established with respect to data handling. Once the individual has agreed to provide information for statistical purposes, there should be a mechanism for transferring identifiable data among agencies under controlled condition. At a minimum this requires:

a. A statement at time of data collection about the character of potential statistical uses;

b. A review agency that has power to authorize transfers;

c. A clear set of criteria that specify when transfer of identifiable data would qualify as

[5] The HEW report suggests the following features for protection against compulsory disclosure:

"The data to be protected should be limited to those used exclusively for statistical reporting or research. Thus, the protection would apply to statistical reporting and research data derived from administrative records, and kept apart from them, but not to the administrative records themselves."

"The protection should be limited to data identifiable with, or traceable to, specific individuals. When data are released in statistical form, reasonable precautions to protect against 'statistical disclosure' should be considered to fulfill the obligation to disclose data that can be traced to specific individuals."

"The protection should be specific enough to qualify for non-disclosure under the Freedom of Information Act exemption for matter "specifically exemted from disclosure by statute.' 5 U.S.C. 552 (b) (3)."

"The protection should be available for data in the custody of all statistical-reporting and research systems, whether supported by Federal funds or not."

"Either the data custodian or the individual about whom data are sought by legal process should be able to invoke the protection, but only the individual should be able to waive it."

"The Federal law should be controlling; no State statute should be taken to interfere with the protection it provides."

being of sufficient public interest to justify the transfer; and

d. A set of procedures to provide for removal of identifiers or destruction of the basic data file after the basic purposes of the transfer have been achieved.

Administrative data sets should be accessible to statisticians for some statistical uses unrelated to the original data collection. In certain cases statistical agencies need to use administrative records for establishing sample frames for verifying the total universe characteristics. Identifiable data extracted from administrative records for statistical purposes should be held confidential by the statistical agency which receives them in the same manner that data collected directly from the respondent are held confidential.

In essence, this suggests the creation of a "protected data set" composed of those items derived from administrative sources for use in the "protected enclave." Thus, subpoena and other access to the original identifiable data would be through the original administrative submission, not through the statistical agency. At the same time, the controlled exchange of data extracted from administrative records among statistical agencies would not be restricted further than the process defined in iii) above (Section 3), would imply.

The above principles place statistical data in a special class of information. To summarize, it must be made clear from the outset in the laws which would implement these principles that: (a) these data may not be used for determining the benefits, rights, and privileges of individuals or of corporations, and (b) the sole use of these data is for use in determining statistical aggregates.

Such protection of these data should be uniquely strong. Therefore, a controlled exchange of statistical data could appropriately be encouraged to improve the accuracy and comprehensiveness of the various measures employed, as well as to assure reduced costs of data collection and minimum reporting burden.

5. Conclusion

Once the data are in the statistical system, I have proposed a set of protected enclaves (statistical units) which are immune from outside access. High ethical standards should be enforced upon the members of these units. Given this protection and high standards of performance, it should then be established that controlled flow of data from one enclave to another is appropriate when necessary for improving the accuracy timeliness, or quality of important statistical information.

Probability – Knowledge – Action
Comments on Some Recent Controversies on Applications of Probabilistic Methods

by Leif Björk[1] and Jan Hagberg[2]

1. Introduction

The terms "probability" and "risk" are used in several different meanings in theoretical literature, in applied studies ("classical" and "Bayesian" statistics, management science, operations analysis, etc.) and in public debate. The discussion of practical risk problems is frequently characterized by lack of precision and by vacillation between different conceptions of probability, a fact that often makes it difficult for a reader to judge the purport and practical importance of presented calculations and statements.

Examples of confusion caused by lack of clarity in the interpretation of probability statements can be found in the discussion of nuclear risks in connection with the "Rasmussen report" (U.S. Department of Commerce (1975)) and similar studies, e.g. in West Germany (Deutsche Risikostudie Kernkraftwerke (1979)) and in Sweden.

In an earlier article (Björk and Hagberg (1982)), we have at some length discussed several of the topics treated below.

2. Different concepts of probability

In scientific literature, mainly four types of probability concepts are used. Different authors may present different definitions of one and the same general type of concept.

The *classical concept of probability* (probability = ratio between the number of favorable cases and the number of possible cases) is based on assumptions of symmetry ("equal possibilities") and represents an early stage in the development of probability theory.

The *statistical (frequentist) concept of probability* (probability = (limit of) relative frequency) is well known to those who apply the "classical" methods of mathematical statistics. Propositions about statistical probability refer to repetitive mass phenomena, not to single events.

By *inductive or logical probability* (Keynes, Carnap) is meant a logical relation between a hypothesis and a body of evidence. J.M. Keynes states:

When once the facts are given which determine our knowledge, what is probable or improbable in these circumstances has been fixed objectively, and is independent of our opinion. (Keynes (1921, p. 47)).

No proposition is in itself either probable or improbable, just as no place can be intrinsically distant; and the probability of the same statement varies with the evidence presented, which is, as it were, its origin of reference. (Keynes (1921, p. 7)).

A man may rationally believe a proposition to be probable, when it is in fact false. (Keynes (1921, p. 10)).

The *subjective concept of probability* plays a central role in modern decision theory and operations research. The idea is also to be found in works by early economic theorists, such as F.H. Knight's pioneering work "Risk, Uncertainty and Profit" (Knight (1921)). An orientation in more recent lines of thought is given by Fellner (1965).

[1] Consultant, Economics and Operations Analysis, Statistics Sweden.
[2] Research Director, Department for Enterprise Statistics, Statistics Sweden.

Fellner regards the "frequentist-objectivist" position as inadequate. He states (Fellner (1965, p. 45)):

A person may watch drawings of odd and even, or of red and black, to the end of his days, and yet he will not observe the theoretical value of any frequency in the sense in which he observes that a number is odd or a color is red. Nor will he ever observe the theoretical value in the sense in which he observes the actual values of frequencies in finite samples. Insofar as a person makes a probability judgment about a physical process *of any kind*, he is making the subjective judgment that it is appealing to apply a mathematically defined measure to that process. This is *always* a subjective judgment. It is a subjective judgment not only if the decision maker is applying it to an event with distinctive individual characteristics, but also if by that judgment he merely approves of the proposition that in Mises' matrix the frequencies appear to converge.

Knight (1921) makes a distinction between *measurable* uncertainty, or "risk" proper-type insurable risk and *unmeasurable* uncertainty. A corresponding distinction is made by Georgescu-Roegen (1967) in his discussion of quantification problems in economic analysis. Under the heading "Concepts, numbers, and quality" he declares (Georgescu-Roegen (1967, p. 45)):

Undoubtedly, for the sciences concerned with phenomena almost devoid of form and quality, measure usually means expanded knowledge. In physics, which has quite appropriately been defined as the quantitative knowledge of nature, there is no great harm if measurement is regarded as an end in itself. But in other fields the same attitude may lead to empty theorizing, at the very least. The advice "look for number" is wise only if it is not interpreted as meaning "you must find a number in everything". *We do not have* to represent beliefs by numbers just because our mind feels similarly embarrassed if it has to predict the outcome of a coin-tossing or the political conditions in France ten years from now. The two events are not instances of the same phenomenon. A measure for all uncertainty situations, even though a number, has absolutely no scientific value, for it can be obtained only by an intentionally mutilated representation of reality. We hear people almost every day speaking of "calculated risk", but no one yet can tell us how he calculated it so that we could check his calculations. "Calculated risk" if taken literally is a mere parade of mathematical terms.

In the models of decision-making under uncertainty associated with "statistical decision theory" or "operations analysis", it is usually assumed that the decision-maker can assign numerical weights ("probabilities") to imagined alternative events or results of actions. In certain cases it may be a question of estimations concerning relative frequencies, but it may also be a question of assessments that cannot be given a frequentist interpretation – e.g. an assessment of alternatives in the face of a unique investment decision. No distinction is made between "measurable" and "unmeasurable" uncertainty in the sense intended by Knight and Georgescu-Roegen.

Assessments of "probabilities" or "risks" based on "statistical decision theory" or "Bayesian analysis" do not presuppose any possibility of empiric verification. de Finetti (1974) states (p. x):

My thesis, paradoxically, and a little provocatively, but nonetheless genuinely, is simply this:

PROBABILITY DOES NOT EXIST.

The abandonment of superstitious beliefs about the existence of Phlogiston, the Cosmic Ether, Absolute Space and Time . . ., or Fairies and Witches, was an essential step along the road to scientific thinking. Probability, too, if regarded as something endowed with some kind of objective existence, is no less a misleading misconception, an illusory attempt to exteriorize or materialize our true probabilistic beliefs.

3. Methodology of the Rasmussen report

The Rasmussen study was sponsored by the U.S. Atomic Energy Commission to "estimate the public risks that could be involved in commercial nuclear plants of the type now in use . . . The objective of the study was to make a realistic estimate of these risks and, to provide perspective, to compare them with

non-nuclear risks to which our society and its individuals are already exposed." (U.S. Department of Commerce (1975, p. -A-1)).

The Rasmussen group performed calculations concerning the risks of radioactive releases from nuclear power plants based on the event-tree/fault-tree method. The results of these calculations were used as inputs to a consequence model for calculation of different risks for the population.

The probabilistic model used in the Rasmussen report gives a simplified picture of conceivable sequences of events in a nuclear power plant. It is assumed that a specific accident, such as a core melt or a rupture of containment, may occur as a consequence of a series of failures. The model describes alternative courses of events after different initiating events.

In each step of the description of sequences Boolean (binary) logic is applied. This implies, e.g., that a valve is assumed to be either open or closed, but not to leak.

Another simplification is the deliberate disregard of the effects of changes during the lifetime of a reactor (e.g. aging), unplanned human intervention, sabotage, war etc., as well as of several conceivable courses of events that have been considered to be of small importance.

Each type of initiating events and failures (including faults committed by operators) is assigned a numerical "probability". By application of the addition and multiplication rules of the probability calculus, "overall probabilities" are computed for various kinds of accidents, such as core melt, rupture of containment, radioactive releases and their consequences.

The numerical "probabilities" assigned to different events in the model are based on sources of various kinds. In the report it is pointed out that direct experience is relatively limited with regard to events in a reactor leading to radioactive releases. Some statistical data from nuclear plants have been used, but to a great extent the assessed probabilities are based on statistics concerning similar plants or similar components, or on expert judgment (e.g. concerning human errors).

The results are presented in the form of point estimates and error bounds for the "probability" or "frequency" of different events. One typical assessment that has been frequently quoted in the controversies around the Rasmussen report is formulated thus: ". . . the probability of a core melt accident is on the average one in 20 000 per reactor per year."

The event trees and fault trees used in the Rasmussen study constitute a modified form of the "decision trees" used in decision analysis. Probabilistic methods of this type have been developed for application to problems of decision-making in business, public administration, etc., where the final result of present action is uncertain because of the influence of future events. According to the Rasmussen report (Appendix I, p. I-2), the term "event trees" is more adequate than "decision trees" in studies of reactor accidents, since reactor safety systems are largely automated, so that the sequence of initiating events plus safety system responses involves few decisions.

The event-tree/fault-tree technique had been used by the National Aeronautics and Space Administration (NASA) as an aid in the development and construction of components for space vehicles. Its ambitions were more limited than those of the Rasmussen study. In a letter quoted in Attachment 3 to the Rasmussen report the Comptroller General of the U.S. stated (Main report, p. 196):

NASA experts believe that "absolute" reliability numbers are misleading and that the time required to develop them is better spent on critical-component reliability analyses. It [NASA] does make predictions during development to compare design alternatives and to evaluate components.

The Rasmussen report is a committee product – not a logical deduction based on a decision-maker's probability judgment as described in text books on decision analysis. The nu-

merical results presented have their origin in a series of decisions based on group discussions, expert memos, etc. The decisions refer to the choice of models, causal factors included and excluded, parameter values, mathematical techniques, mode of presentation, etc. All these types of decisions have been treated in public discussion.

4. Nuclear risk studies in West Germany and Sweden

Several recent studies of nuclear risks have applied methods more or less analogous to those used by the Rasmussen group. It may especially be noted that in studies performed in West Germany and Sweden the directives given have explicitly prescribed acceptance of basic premisses of the Rasmussen study, thus substantially restricting the scope of independent judgment.

The German Risk Study for Nuclear Power Plants (Deutsche Risikostudie Kernkraftwerke (1979)) was performed on behalf of the Minister of Research and Technology of the Federal Republic of Germany. According to the directives, the study was to apply as far as possible the methods of the Rasmussen study to German plant and site conditions. A direct transfer of the American results was not deemed justified, mainly in view of differences between the design of the reference plants of the Rasmussen study and German nuclear power plants and differences in population density in the two countries.

In August 1977, The Swedish Energy Commission and the company AB ASEA–ATOM signed a contract concerning a safety study for a Swedish nuclear power plant under construction (Forsmark 3). We quote from the contract:

The study is to be performed with respect to Forsmark Power Station, Aggregate 3, and aim at an investigation of the probability of different sequences of events leading to core melt, based on the systematics accounted for in the American report WASH–1400 (the so-called "Rasmussen report").

– – –

The study of Forsmark 3 to the extent proposed is to be done on a comparative basis, the BWR plant in WASH–1400 (Peach Bottom BWR 3200) serving as reference plant.

– – –

Deviations from assumptions and methodology used in WASH–1400 shall be clearly accounted for and justified.

5. Probability statements – interpretation and use

One of the basic issues in the discussion of risk studies of the Rasmussen type concerns the interpretation of probability statements. The difference between "classical" and "Bayesian" statistics is frequently referred to.

The only theoretical work referred to in the main report of the Rasmussen study is Raiffa (1970): "Decision Analysis". Raiffa stresses the difference between classical statistics and "the so-called Bayesian viewpoint, which uses both utilities and subjective probabilities" and tells the user of his recommended method (pp. 109–110):

. . . all you have agreed to do is apply a suggestive label to a dubious concept; you have in no way committed yourself to work with this judgmental probability as if it were a bona fide, objective, tangible, real-world, frequency-based probability. The remaining part of this chapter aims at convincing you that *so far as action is concerned* – and this is all that is of interest here – you should feel free to use these judgmental probabilites just as if they were the "real" thing, and you should not worry about mixing some objective and some subjective or judgmental probabilities together in the same formula.

The relation between the methodology of the Rasmussen report and "formal Bayesian treatment" is discussed in appendices to the report. It is asserted that "the study's results can be interpreted in a general Bayesian framework, where the data distributions are interpreted as the given priors". (U.S. Nuclear Regulatory Commission (1975, p. XI 14–4)). However, the presentation of numerical results in the report differs from a consistent Bayesian (subjectivistic) approach through

the ambition to produce "realistic" assessments of "absolute" probabilities of nuclear accidents and their consequences and compare them with statistically observable non-nuclear risks. In response to comments on the draft report, according to which "fault tree analysis is incomplete and is unable to produce reliable quantitative predictions of system failure probability", the Rasmussen group declares that "fault tree methodology *can* produce meaningful results" (l.c., p. XI 3–5).

Mathematicians and statisticians, such as the Australian T.P. Speed (University of Western Australia) and J. Hoffmann-Jörgensen and M. Vaeth (University of Århus (Denmark)), deny the possibility of arriving at meaningful prognoses concerning the frequency of rare events by the methods used in the Rasmussen report.

In a paper presented at the IAEA conference in Stockholm 1980, G.W. Parry, H.J. Teague and P.W. Winter (Safety and Reliability Directorate of the U.K. Atomic Energy Authority), after discussing the limitations on the use of statistics and reviewing some of the available tools, declared (International Atomic Energy Agency (1980, Vol. II, p. 499)):

We favour the use of a formal Bayes approach as opposed to the rather ill-defined empirical approach adopted in the RSS [the Rasmussen report].

G. Apostolakis and A. Mosleh (University of California, School of Engineering and Applied Science) state (Apostolakis and Mosleh (1979, p. 135)):

A special feature of the quantification of the risks from large technological systems, particularly nuclear power plants, is that it involves the investigation of rare events, for which, of course, significant statistical records do not exist. This lack of data has rendered most of the conventional statistical techniques inapplicable and has created conceptual and operational problems concerning the meaning of probabilities and the methods for their evaluation. The frequently resulting confusion and the fact that probabilistic statements can never be proved to be correct by experience or experiments explain, in part, the controversy that followed the release of the *Reactor Safety Study* (RSS) (WASH–1400) . . .

The terms "probability" and "risk" are frequently used, both in the Rasmussen report and by participants in the general discussion of nuclear risks, in a manner that implies (or at any rate suggests) an idea of "real" or "absolute" probabilities. Assumptions used in the Rasmussen report have been criticized and alternative assessments have been presented in a frequentist terminology. As an example we quote from the paper "New trends in safety design and analysis" presented at the IAEA conference 1980 by D. Okrent (International Atomic Energy Agency (1980, vol. I, pp. 298–299)):

Since issuance of the final version of WASH–1400 in 1975 citing an estimated core melt frequency of about 1 in 20 000 per reactor year, a considerable number of things have occurred which tend to support a thesis that the WASH–1400 estimate of the frequency of core melt (or at least core damage severe enough to threaten containment and a major release of radioactive materials) was too low.
– – –

For these and similar reasons, it appears to be difficult to demonstrate with a high degree of confidence that the frequency of severe core damage or core melt for reactors in operation or under construction is less than about one in a thousand per year. It may be smaller, but it is also conceivable that it is somewhat larger. Also, there are many paths to severe core damage or core melt so that it will be difficult to make the frequency of such an accident much smaller, with a high degree of confidence.

In its report concerning the Rasmussen study, the Risk Assessment Review Group (RARG) (1979) headed by H.W. Lewis stated (p. 272):

We are unable to determine whether the absolute probabilities of accident sequences in WASH–1400 are high or low, but we believe that the error bounds on those estimates are, in general, greatly understated. This is true in part because there is in many cases an inadequate data base, in part because of an inability to quantify common cause failures, and in part

because of some questionable methodological and statistical procedures.

In a comment published in "Nuclear Engineering International" (May 1979, pp. 47–49), N. Rasmussen disagreed with this finding of the RARG and stated:

Although I now believe that the uncertainty bounds in WASH–1400 are somewhat understated, the use of the phrase "greatly understated" with no further discussion has been interpreted by many to mean that reactor accident probabilities could be much larger than stated in WASH–1400. I do not believe that this is true because the accumulated reactor experience to date puts an upper limit on this probability that is not much greater than the upper limit in WASH–1400.

Rasmussen referred to the earlier mentioned article by Apostolakis and Mosleh:

It is interesting that this [Apostolakis and Mosleh's] analysis produces just about the same result for the upper limit of core melt frequency as the previous [the Rasmussen group's] analysis.

In reality, this comparison gives no additional support to Rasmussen's belief. Apostolakis and Mosleh declared in their article that their intention was not to "produce a definitive distribution for the frequency of core melt", but rather to "demonstrate how an expert's opinion can be formally handled and what difficulties arise in doing so". They presented some calculations concerning the frequency of core melts, based on results of the Rasmussen study, certain operational experience and some points of view presented by critics of the Rasmussen study. Their general philosophy may be illustrated by the following quotation (Apostolakis and Mosleh (1979, p. 137)):

However, the subjectivistic theory does not accept the existence of true values for the parameters of distributions, that is, values that are intrinsic properties of the phenomena that the distributions describe in a particular situation. By true value, we mean a value in which we have very high confidence . . . We have used the term true for lack of better terminology. (It is instructive here to quote de Finetti, who, in a similar context, states: "The fundamental obstacle is the difficulty of escaping from the tyranny of everyday language, whose viscosity often obliges us to adopt phrases conforming to current usage instead of meditating on more apt, although more difficult, versions.")

In West Germany and in Sweden, comparisons between two or more similar calculations have been adduced as arguments for the "credibility" or "realism" of these calculations.

In the preface to the Birkhofer study (Deutsche Risikostudie Kernkraftwerke (1979)), V. Hauff, Federal Minister of Research and Technology, states:

It should be stated as an important result, that the results of the German and the American study coincide within the error bounds. To this extent the statement in the Rasmussen report is confirmed, that failure-conditioned risks through nuclear power plants are relatively small, especially where they can be compared with other risks of civilization.

In the same vein, the Swedish Government Committee on Nuclear Reactor Safety (1979) stated (p. 33):

New studies, such as the Birkhofer study in West Germany, have obtained about the same probabilities as the Rasmussen study for accidents of different severity within the relevant limits of uncertainty. This lends more credibility to the methods and the results.

The two comparisons quoted refer to results obtained from one and the same type of theoretical model in which somewhat diverging assumptions have been inserted. Such comparisons do not *per se* say anything about the "realism" of the model or the relation of the calculation results to an observable reality.

A sceptical view of the "truth value" of risk assessments of the type exemplified by the Rasmussen report is in no manner incompatible with recommendations of their use for practical purposes, e.g. in the development of technological systems. A detailed modeling of possible events and failures may be an aid to discovery of sensitive points in a system. In an appendix to the "Kemeny-report" (President's Commission on the Accident at Three Mile Island (1979)), it is stated (p. 6):

The Wash–1400 study, in using the "event tree" and "fault tree" methodologies, borrowed from the aerospace industry, actually revealed a "weak link" in the safety of the Surry reactor. This led directly to a change in inspection procedures at Surry and reduced the probability of one major risk contributing accident... by a factor of 20.

It should, however, be stressed that the assumptions made in designing a probabilistic model – such as, in the case of the Rasmussen study, the disregard of aging and other changes during the lifetime of a reactor or unplanned human intervention – will necessarily influence the practical conclusions drawn from the numerical calculations based on the model. In a discussion with N. Rasmussen, the MIT mathematician D.J. Kleitman (1979) said (p. 36):

One reason a single number has limited usefulness is that the probability of human error depends a lot on the training and the ability of the people involved. And that, by the way, is a function of the probability you assume. If a hazard seems to be very rare or unimportant, nobody pays much attention and nothing much in the way of training is done about it. On the other hand, errors considered common or serious are likely to be prevented with appropriate training.

An illustration of consequences of a low risk assessment is given in a recent article in Science (Marshall (1983)).

Diverging views regarding the role of probabilistic risk assessments in the political and administrative decision process have been voiced in the discussion following the publication of the Rasmussen report. The "Lewis report" (RARG (1979)) says (p. 305):

The [Rasmussen] report is divided into parts: a Main Report of some 150 pages (they *are* numbered), a short and separate Executive Summary, and eleven appendices which deal in greater detail with much of the material. This format is not unusual in major reports, in that the three parts cater to three different audiences. The Main Report caters to the serious reader, who wishes to know just what was accomplished, and how. The appendices are normally intended for people who are not so much readers as workers in the subject, who wish to go beyond the findings, conclusions, and comments of the main report into the detailed technical arguments and supporting data. The Executive Summary is normally reserved to the decisionmaker, who is neither a reader (for lack of time or, often, background) nor a worker, but still needs to know, succinctly, the main conclusions, findings, and recommendations of the report.

In the case of RSS, the functions of the Main Report and of the appendices (apart from Appendix XI, which was a response to comments on the draft report) were more or less as outlined above, but the Executive Summary was assigned an entirely different function. It was written, not for the decisionmaker with a shortage of time, but for the public, and is in a general soothing tone.

The editor of Technometrics, R.G. Easterling, in a paper presented at the annual meeting of the American Statistical Association in August 1980 ("Statistical problems in the assessment of nuclear risks"), presupposes a more active intellectual effort by the decisionmaker. He questions the need for quantitative risk acceptance criteria and asserts that such criteria have three major drawbacks:

(1). They encourage charlatanism. Even "noncharlatans" will be pushed toward ways to "sharpen their pencil", which is not the same thing as making their endeavor safer.

(2). They encourage laziness and self-delusion. The "decisionmaker" can go by the numbers and ignore the subtler non-quantitative aspects of the decision.

(3). They don't allow for a grey area. Suppose the acceptance limit is 10^{-6} (in some dimension, the choice of which may be a considerable problem). What if the analysis yields 1.1×10^{-6} or 2×10^{-6}? Should major decisions rest on one's rules for rounding numbers?

6. A pedagogical problem

Observation of the public discussion of nuclear risks raises a pedagogical problem which should perhaps not be entirely insoluble, notwithstanding the "viscosity of everyday language" noted by de Finetti. Specifically, it should not be impossible to popularize the distinction between an objectivist (frequentist) and a subjectivist conception of probabili-

ty. Increased consciousness with regard to the various meanings of the terms "probability" and "risk" might to some extent change the climate of discussion.

In "Statistical Theory" L. Hogben (n.d.) refers to Carnap's distinction "between different *probabilities*, one formal and ostensibly relevant to matters of judgment ("probability one"), one empirical and relevant to external events such as those which our present theme subsumes ("probability two")" (p. 314), and asserts (p. 315–316):

If it happens that the consumer has no disposition to rely on recipes of inference deriving no sanction from experience, it is his right both to know whether such recipes are or are not referable to a definition of probability consistent with his requirements and to know that recipes referable to a purely subjective definition have no title to share the spoils of hard-won victories of a calculus conceived in empirical terms. From one viewpoint, Carnap's contribution is therefore salutary. If it receives the publicity it deserves, its only effect will be to infect the investigator in the laboratory and in the field with a sceptical temper towards the claims of statistical theory, a temper more in keeping with the traditional claim that science is the last defence of intellectual freedom in its perennial conflict with arbitrary authority.

7. What do we know?

During the hearing (February 26, 1979) before the House Subcommittee on Energy and the Environment, H.W. Lewis gave some comments on the report of the Risk Assessment Review Group and said, i.a.:

Nonetheless, having concluded that the methodology is sound and therefore holds great potential for making the subject of nuclear safety more rational, we strongly urged the Commission to work to infuse this methodology into its organism. This should only be done, however, where the universe of discourse is sufficiently limited as to permit credible calculations founded upon an adequate data base, and this carries with it the conclusion that the methodology cannot yet, if ever, be used to seek a final absolute number for the probability of an accident.

The ambition of the Rasmussen group to produce "overall probabilities" brings to our mind a saying of Immanuel Kant (Gross (1909, p. 45)):

Das methodische Geschwätz der hohen Schulen ist oftmals nur ein Einverständnis, durch veränderliche Wortbedeutungen einer schwer zu lösenden Frage auszuweichen, weil das bequeme und mehrenteils vernünftige "Ich weiss nicht" auf Akademien nicht leichtlich gehört wird.

[The methodical chat of the high schools frequently is just a tacit understanding to evade, by shifting meanings of words, a question that is difficult to solve, because the convenient and mostly sensible "I do not know" is not easily heard in academies.]

8. References

Apostolakis, G. and Mosleh, A. (1979): Expert Opinion and Statistical Evidence: An Application to Reactor Core Melt Frequency. Nuclear Science and Engineering, 70, pp. 135–149.

Björk, L. and Hagberg, J. (1982): Några synpunkter på riskstudier inom energiområdet. Statistisk tidskrift, 5. (In Swedish).

Deutsche Risikostudie Kernkraftwerke (1979). Verlag TÜV Rheinland, Köln. (The Birkhofer Study).

Fellner, W. (1965): Probability and Profit. A Study of Economic Behavior Along Bayesian Lines. Richard D. Irwin, Homewood, Illinois.

de Finetti, B. (1974): Theory of Probability. Vol. I., John Wiley & Sons, Inc., New York.

Georgescu-Roegen, N. (1967): Analytical Economics. Harvard University Press.

Gross, F. (1909): Laienbrevier. Eine Darstellung der Kantischen Welt- und Lebensanschauung für den ungelehrten Gebildeten aus Kants Schriften, Briefen und mündlichen Äusserungen. Berlin MCMIX. Verlag von Reichl & Co.

Hogben, L. (n.d.): Statistical Theory. The Relationship of Probability, Credibility and

Error. W.W. Norton & Company, New York.

International Atomic Energy Agency (1980): Current Nuclear Power Plant Safety Issues. Proceedings of an International Conference on Current Nuclear Power Plant Safety Issues held in Stockholm, Sweden, October 20–24, 1980. (In three volumes).

Keynes, J.M. (1921): A Treatise on Probability. London.

Kleitman, D.J. (1979): Discussion. Technology Review, June/July.

Knight, F.H. (1921): Risk, Uncertainty and Profit. Houghton Mifflin, Boston & New York.

Marshall, E. (1983): The Salem Case: A Failure of Nuclear Logic. The "impossible" happened twice in three days when a fail-safe device failed at a New Jersey plant. Science, Vol. 220. 15 April, 1983.

President's Commission on the Accident at Three Mile Island (1979): Technical Staff Analysis Report on WASH–1400 – Reactor Safety Study, October 1979.

Raiffa, H. (1970): Decision Analysis. Introductory Lectures on Choices Under Uncertainty. Addison–Wesley.

Risk Assessment Review Group (RARG) (1979): Reactor Safety Study Review. Oversight hearing before the Subcommittee on Energy and the Environment of the Committee on Interior and Insular Affairs. House of Representatives. Ninetysixth Congress, first session on Reactor Safety Study Review. Hearing held in Washington, D.C., February 26, 1979, Serial No. 96–3.

Swedish Government Committee on Nuclear Reactor Safety (1979): Säker kärnkraft? Betänkande av Reaktorsäkerhetsutredningen. SOU 1979:86 (In Swedish).

U.S. Department of Commerce (1975): Reactor Safety Study: An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants. Main Report. Nuclear Regulatory Commission. (The Rasmussen Report/WASH–1400/RSS).

U.S. Nuclear Regulatory Commission (1975): Analysis of Comments on the Draft WASH–1400 Report. Appendix XI to Reactor Safety Study. October 1975.

On Applications of Random Techniques in Environmental Statistics[1]

by Thomas Polfeldt and Bo Justusson[2]

1. Introduction

1.1. In this paper, we shall first give brief descriptions of existing applications of random techniques in environmental statistics in Sweden. This review, in Section 2, will display some of the possibilities of random techniques, and also some shortcomings. There appears to be a need for extensions of the traditional ideas of population and frame.

1.2. In several of the investigations, we have encountered an apparent contrast between different types of numerical overview: administrative micro-data, monitoring results and selected typical values are three kinds of data which do not automatically convert into statistics on a national or regional level. As we shall discuss in Section 3, however, it is often possible to combine the collection of such data with a statistically sound sample for more general conclusions. The statistical device needed is simply stratification, where subjectively selected units are placed in special strata, and where some other units are also included in the sample.

1.3. Since we have found in our contacts with the data collecting institutions that it is easier, in many cases, to make ourselves understood if we refrain from using the statistical terminology, we shall follow that practice in Section 3 in order to stress the point.

1.4. Naturally, the procedures to be discussed in Section 3 are not universally applicable in environmental statistics. Some others will be briefly mentioned in the concluding Section 4.

2. Swedish applications

Basic characteristics

2.5. According to survey classification schemes (see e.g. Murthy (1982)), the following aspects have to be specified in order to describe a statistical survey with random components: type of sampling (by stages or phases), stratification or not, type of sampling frame, basic selection scheme (simple random sampling, systematic sampling, sampling with probability proportional to size, subjective selection). A discussion of the random element in surveys with subjective selection is deferred to paragraph 2.19. Table 1 below constitutes an attempt to characterize some 15 Swedish examples of the use of random survey techniques in environmental statistics.

2.6 In the table, we also include two columns which are related to the discussion in Section 3. They cover the possibility of estimating totals, directly or after model assumptions, and the interest, or lack of interest, in individual units of the sample (e.g. for administrative purposes).

2.7. Other features, like selfweighting properties, are connected with the estimation to be carried out and are not given in this table. Such aspects as sample and population size, method of data collection (enumerators, re-

[1] This is a revised version of a paper presented at the ECE Informal Meeting on General Methodological Problems in Environment Statistics, Helsinki, 2–5 May, 1983.

[2] Statistical researchers, Department of Area Statistics, Statistics Sweden.

Table 1. Swedish applications of random selection techniques in environmental statistics

Survey	Type of sampling	Frame	Stratification	Basic selection	Estimation of totals	Interest in individual units
A. National Forest Survey	s2	a	str	syst	p	0
B. Heavy metals in moss	s2	a	str	syst*	p	0
C. Objective crop yield survey	s3	l+a*	str*	pps+ srs	p	0
D. Use of pesticides, pilot survey	s1	l	str	srs	p	0
E. Use of pesticides	s1	l	str	srs	p	0
F. Soil chemicals, pilot study	s3	l+a*	str*	pps+ srs	p	0
G. Urban land use, pilot study	s1	a*	0	syst	p	0
H. Elk inventory	s1	a	str*	srs*	p	0
I. Breeding birds	s1	a	0	subj	np	0
J. Acid lakes (1979/80)	s1	o	(str)*	subj	np/pm?	i
K. Lakes with high mercury rates in fish	s1	o	(str)*	subj	np/pm?	i
L. Industrial investments for environmental protection	s1	l	str	srs	p	(i)
M. Hazardous wastes (1980)	s1	l	str	syst+ srs	p(?)	i
N. Precipitation quality	s1*	a*	(str)*	subj*	pm	0
O. Water quality - rivers	s1	a*	0	subj*	pm?	i
P. Ecological interaction	s1	a	str	subj	pm?	0
Q. Urban air quality	s1	a*	str	subj	pm	i

Type of sampling
- s1, s2, ... one, two, ... stages
- p1, p2, ... one, two, ... phases

Stratification
- str stratification used
- 0 stratification not used

Sampling frame
- a area frame
- l list
- o other

Basic selection scheme
- srs simple random sampling
- syst systematic sampling
- pps probability proportional to size sampling
- subj subjective selection

Estimation of totals
- p possible
- pm possible with model assumptions
- np not possible

Interest in individual units
- i individual data of interest
- 0 individual data not of interest

* See list of remarks, paragraph 2.8.

cords, mail, measurements, etc.), time periods, and whether the survey covers one or several subjects or variables, are also relevant for the random part of the survey but left out here.

2.8. In this paragraph we give some remarks to Table 1 and various comments.

General. Most of the surveys have been described in detail in connection with the reporting from them. No references are given here, but for most of the surveys, some pertinent facts are included in the remarks below. Only a few of the surveys are carried out by Statistics Sweden. Comments on the classifications as such follow after the remarks.

A. *The National Forest Survey* covers all wooded land. It is a two-stage survey where the first stage units are squares of land (1–3 sq.kms) and the second stage units are circular plots with 10 m. diameter, regularly placed along the sides of the square. The square grid is fixed and the squares are selected from it so as to cover the whole area during a five year cycle.

B. *Heavy metal in moss.* Moss is collected every 5 years in a subset of the squares of survey A.

C. *The objective crop yield survey* covers all arable land. The environmental surveys D and F have been included in the programme. The three stage selection consists of a pps selection of a farm (stage 1), a field of that farm (stage 2) and two circular areas within the field (stage 3). The farms are selected from the farm register after geographical stratification. The particular field is selected by pps sampling, and the choice of the circular areas is governed by random coordinates. In these plots, the crop yield is measured directly.

D. *Use of pesticides.* The pilot survey was carried out at the farms included in survey C. Two methods, personal interviews and mail enquiries, were tested. The differences were small.

E. *Use of pesticides.* In the full scale pesticide survey, the questions were included in an agricultural omnibus survey. They concerned both the area of treated fields and the amounts used of five groups of pesticides. The results were published in 1982. A special feature of the report is the inclusion of coefficients for the calculation of estimates for municipalities (small regions).

F. *Soil chemicals.* The soil content of phosphorus, potassium, mercury and cadmium (P, K, Hg, Cd) and the pH-value are measured in a small region in western Sweden. The fields selected in survey C are used. One point is randomly selected in each field, and ten small soil samples are taken from an area within about 2 m. of that point and mixed. The samples were taken during 1982. The results are about to appear.

G. *Urban land use.* A grid net was laid over the map of an urban area, and the land use was established in the grid points. The grid corresponded to 50, 75 or 100 metres. The land use was classified in 22 categories. The average number of points for an urban area and its vicinity was 1 450. In this case, *points* were thus selected from the area frame.

H. *Elk inventory.* Counties of northern Sweden were stratified according to geographical character, using maps. Each stratum was further divided into areas, all of which contained approximately equal forest acreages. The size was selected so as to give an expected number of 15–25 elks. The borders should be discernible from an aircraft. Under winter conditions, the elks were counted from the air in a random sample of these areas.

I. *Breeding birds* are counted by amateur ornithologists either in selected areas or along selected routes. Interpretation problems arise, since areas and routes are selected on subjective grounds, and since some of them are changed from year to year. Of course, no estimation of the total population is intended from these observations.

Slight acidification – most lakes have pH > 5.5 all the year

Medium acidification – most lakes have pH < 5.5 sometimes during the year

Serious acidification – at least 1/3 of the lakes have pH < 5.0 all the year

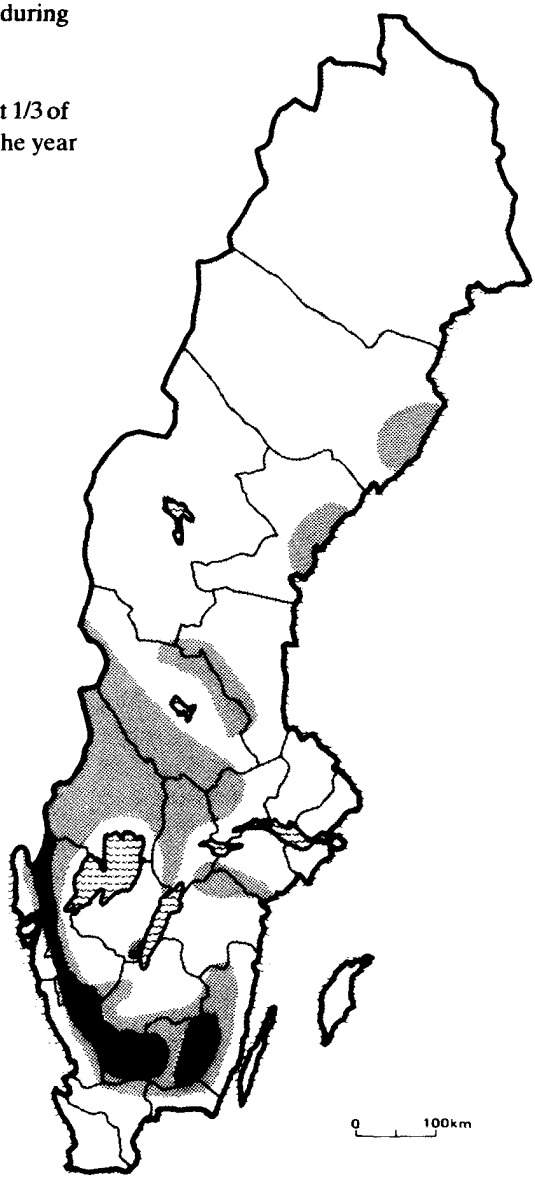


Fig. 1. pH values in the lakes of Sweden at the end of the 1970's

J. *Acid lakes.* The county authorities were asked to measure alkalinity and acidity (pH) in the lakes in 1979–1980. No specific rules were apparently given for the selection of lakes. There are about 85 000 Swedish lakes with an area of at least one hectare. Almost 10 % of these were measured, with a good geographical distribution. Techniques and times of the year varied between counties. Most of the larger lakes (more than 100 hectares) are included in the sample. Furthermore, the counties often selected lakes that were of special interest, e.g. because administrative decisions of various kinds concerning those lakes were pending.

The stratification of the lakes is by county, but it is likely that within most counties it was arranged that lakes of various types would be included – this means that a further stratification is implicit. The only frames available at the time of the survey were maps. However, the sampling, if any, has probably not been conducted strictly by areas.

Because of the vagueness of the operations, it is not possible to generalize the results using conventional statistical methods. Nevertheless, there are strong arguments supporting the lake acidity maps of Sweden which are derived from the data (see Fig. 1) – arguments based on the contiguity and slow spatial variation of the influencing factors. This means the acceptance of some not stated model. It may or may not have been possible to work out the model before the survey. In any case, the results, combined with the recently developed lake register, now provide a good basis for selecting a much smaller sample for a new survey of the same variables. The techniques of Section 3 will be useful for that purpose.

It is worth mentioning that there was a previous survey of this kind in 1972 called "A thousand lakes". The lakes selected in that survey were those that the counties considered interesting or having environmental problems. Thus, one result was a comprehensive list of such problems.

K. *Lakes with high mercury rates in fish.* There are two investigations of lakes with suspected high mercury rates in fish. One is performed continuously by the food control authorities by selecting lakes in which there is much fishing, and where mercury is likely to appear. The other was a one time survey covering a band of some 100 km width along central Sweden, in which lakes were subjectively selected considering certain characteristics: location, size, depth, and lack of industrial and sewage emissions. The latter survey is intended to measure deposition of airborne mercury and other metals.

L. *Industrial investments for environmental protection.* This survey consisted of a few additional questions in the ordinary mail survey for investments in the manufacturing industry. It covers about 1 600 enterprises stratified according to activity and size. All big enterprises are included in the survey – the sampling fraction is 1.0 in some strata. The results will be published in 1983.

M. *Hazardous wastes.* Enterprises were stratified by activity and size, and a sample was taken from 73 branches with hazardous wastes. The local authorities were allowed to add any further enterprise. This added some 150 further branches, most of them with only a few representatives in the survey. Again, this is a case where units which are interesting from an administrative point of view have been included, almost by force. The situation is discussed in the next section.

N. *Precipitation quality, O. Water quality – rivers, P. Ecological interaction.* These are all included in the Swedish monitoring programme. N and O have fixed stations for measurements, whereas in P, twenty "typical areas" have been selected. In all of them, there is that apparent vagueness about the population to be investigated and the parameters to be estimated which often hampers the sta-

tistical use of monitoring data. It is quite possible, however, that this vagueness is mainly a communication problem. If so, it is hoped that the ideas of Section 3 may help to close the gap.

The basic selection of these surveys is subjective in the geographical dimension, but systematic over time. The stratification carried out in N is not explicit but probably existant. Generally, the table entries for survey N are based on a population concept of points in time and space, cf. the description in Polfeldt (1979). From the area frame, *points* are selected for the survey.

Q. *Urban air quality* is measured by local authorities, but the instruments are standardized. There is no other co-ordination between towns – e.g. not for time periods. Site selection is subjective; often particularly exposed spots are chosen.

It is expected that towns where measuring ceases or does not exist have a fairly low degree of air pollution. The main problems consist in generalizing from a few measurements sites (often 2–6) to an entire urban area. A further definition problem is introduced when the "exposed population" is asked for.

Discussion of the classification scheme

2.9. The coarse classifications of Table 1 have several shortcomings. We shall not dwell upon *other* survey features than those included in the table. – The *type of sampling* seems to be no problem, except in survey N, where the population is not clear and all classifications are tentative. Multistage sampling is unusual. This is perhaps because a coverage of the entire population is considered essential, and it is (still) too dangerous to assume similarities between, e.g., geographically distant units. Also, stratification may be an attractive alternative. Note that in the s3 surveys (C and F), there is first a geographical stratification, and in the s2 surveys (A and B) the first stage consists of systematic sampling of areas.

2.10. The question of observation *times,* so important e.g. in monitoring, is explicitly included only as a possibility of giving the type of sampling as one in phases. It may also appear under stratification and basic selection. None of these allows for a completely satisfactory treatment, however.

In paragraphs 2.12 and 2.15, we introduce the possibilities to include this aspect in the population definition, or in the frame, or both.

2.11. The *populations* are of quite different types; the scheme of Table 1 does not allow for a discussion of this aspect. Traditional survey sampling assumes a population with easily identifiable discrete sampling units like persons, farms, enterprises. This is indeed the situation for some of the surveys, viz. C, D, E, F (the population consists of farmland fields), J, K (lakes), L, M (enterprises). For the surveys H and I, the population could be considered as a certain set of areas, for which the number of elks, or birds, are the measured variables. Then, this set of areas is discrete and makes up a population of the traditional type. The borders of the sampling units (areas) are to a great extent arbitrary in these cases, however. – For survey A, a reasonable population is "all forested land" (although for some variables it could be something else, e.g. "all trees"). In sampling from this population, certain subareas are selected. The locations and sizes of the subareas can be determined by the investigator – there are no "natural" sampling units in this case. The sampling techniques of forestry (e.g. line transects, 3P-sampling, etc., see various papers in Cormack et al. (1979)) are not accommodated in the scheme on which Table 1 is based. – A similar situation exists for surveys B, G, and probably P. In survey G, depending on the specific variables of interest – areas under different land use – the

sampling units are points rather than areas.

2.12. In the water and air quality surveys, O and Q, the population can be regarded as "all fresh water" or "all air", but the time aspect is more important here because of rapid changes. If the population, in the cases of surveys A and B, is a part of the (x, y)-plane (R^2), we should work here with a three dimensional space and a time component (R^3 x T). An inceptive discussion of air quality statistics from this point of view is found in Polfeldt (1983). – For survey N, it seems reasonable to use the plane and time points (R^2 x T) as the population.

2.13. The *frame* is usually an area frame or a list. In the cases of three stage sampling (C and F), the frames are of different kinds in the different stages. In contrast to the usual situation in Swedish statistics, there is a dominance of area frames. The techniques of area sampling should probably be given more attention in the field of environmental statistics. Some references are Cormack et al. (1979) and Dalenius (1965).

2.14. It was found desirable, in Table 1, to make a remark when the sample from an area frame took the form of points rather than areas. In the surveys marked "o", (J and K), it is not quite clear whether a real frame exists or not. The maps have *not* been used for a random selection of lakes in those cases. This would have been possible, but the procedure is not straightforward. It has already been mentioned that a lake register, e.g. a list, has recently been established. Thus, an easier way of obtaining random samples of lakes is now available.

2.15. **Besides list and area frames, it is possible that environmental statistics will require a conceptual extension in order to take care of the third dimension – i.e., a space frame. This** appears necessary for air quality and precipitation variables, perhaps also when waters are investigated on various depths.

The discussion in paragraph 2.12 raises the question of whether time should be explicitly included in the frame as well.

2.16. The discussion here of non-traditional populations and frames has a preliminary character. It appears important for the conceptual development of environmental statistics to explore further the population definitions and their implications.

2.17. *Stratification* is of paramount importance. It can be performed in many ways; the format of the table allows for no specifications. Have general types of stratification been described somewhere? The discussion in Section 3 will be devoted to some aspects of stratification which deserve more attention in environmental statistics.

2.18. The *basic selection* is mainly "srs" or "syst", both with substantial theoretical backing. For systematic sampling, there is a recent review in Iachan (1982). In Table 1, there is only one example of pps sampling (C and F). The reason for this scarcity is probably that sampling with probabilities proportional to size can be approximated by stratification, and this technique is better known and easier to handle in computations.

2.19. It is unsatisfactory that so many surveys have to be labelled "subj". This means that, strictly speaking, they should not have been included among the random surveys – they lack precisely that random element which allows for statistical generalization. But as was explained in the remark for survey J, there is often an implicit body of knowledge which still makes generalizations plausible. We think that it is a very important task for statisticians to find ways of reconciling the actual proce-

Table 2. How to obtain sound samples

Situation	Strategy	See paragrahps
Long-term planning	Create registers, lists or maps covering the whole country; calculate size or area of important subgroups	3.30
Specific types of units more interesting than others	Grouping; take larger samples in interesting groups	3.22–3.26
Specified units must be investigated	One group is completely investigated, one or several groups are sampled	3.22–3.26
Typical units have been selected	Estimate sizes of the groups which the typical units represent. Special measurements for this may be necessary	3.27–3.28
Estimation planning	Consider possible usefulness of including auxiliary information from sampled units	3.29, 3.32

dures, and what lies behind them, with statistical theory.

2.20. The proposal which is the subject of Section 3 consists simply in permitting subjectively selected units to be placed in special strata, and to allow for other strata as well, covering the rest of the units. This will not solve all the "subj" problems, particularly not some of the monitoring generalizations, but it *does* solve some problems.

3. The judicious grouping of investigation units

3.21. In this section, an attempt is made to keep the discussion readable also for non-statisticians. In particular, the word "grouping" or some similar expression will be used rather than "stratification". The discussion is summarized in Table 2.

3.1. Inclusion of certain units

3.22. In many situations, the environmental administrator, or the scientist, has a particular interest in some *specified* enterprises (e.g. the paper and pulp industry when looking at certain water pollutions), in specific types of lakes (e.g. those on granite bedrock in the south-western part of Sweden when acid rains are discussed), generally in some *specified units* or specific *types of units*. At the same time, there is also a requirement for a more general investigation, since the problems already identified (which are often the most pressing) may not be the only ones. From previous experience, the administrator expects a statistician to require "random samples", in order to give a wider coverage, but isn't that a foolish waste of funds?

3.23. Statisticians have been slow to point out the flexible possibilities of *grouping* the investigation units. It is entirely possible to investigate some groups very intensively, even completely, while quite small, but random samples can be taken from other groups. A scientifically sound overall picture can then be produced, *at the same time as* very detailed and accurate data are collected for the units of specific interest.

3.24. The device of grouping thus allows the possibility of *concentrating the efforts* on any group of units, and also of *including any specified units* in the survey. These two cases are slightly different from a practical point of view. This depends on the fact that sound generalizations can be made only if a sample from a particular group is selected at random, or systematically, in a way that makes the sample representative.

3.25. If, for instance, we have placed the enterprises of the paper and pulp industry in a special group, and we can afford to investigate 50 % of them, these have to be selected *at random*. This may not be in accordance with our wish to include certain big enterprises. However, we are completely free to create two (or more) subgroups: one with all the big enterprises of the paper and pulp industry, and the other subgroup with the smaller enterprises. We could then investigate 100 % of the units (enterprises) in the first group and, say, 20 % of the others.

3.26. The general procedure thus consists in placing units which *must* be included in a special group, from which a *"sample" of 100* % of the units is taken. In order to provide a complete survey, all other units are placed in one or several groups, from which *random samples* are taken. The sampling frequency may vary between the groups. The generalizations will be more efficient when units which are similar to each other, with respect to the variable(s) under investigation, are placed in the same group. (This remark is irrelevant for the special group with the 100 % sample.)

Necessary background information

3.27. For generalizations to be made, it is necessary to know the total number of units in the various groups. When the units are enterprises, divided into "big paper and pulp enterprises", "smaller paper and pulp enterprises", and "other enterprises", three totals are thus necessary. If exact numbers are not available, estimates could be used. Some information is necessary, however.

3.28. In an investigation of lakes, it may be found as the results come in, that it would be useful to distinguish between mountain lakes, forest lakes, lakes in the plains and other lakes. If the lakes have originally been selected at random (e.g. from the lake register), the sample itself can be used to provide estimates of the total number of lakes in each of the four groups.

3.29. In the study of ecological interaction, areas were selected subjectively. We advised to collect information also on such variables that were present in the nationwide surveys, mainly the National Forest Survey. This would make it possible to estimate the total areas which were similar to the investigated ones with respect to the variables of the National Forest Survey.

3.30. For a random sample to be taken, it is necessary to be able to reach, or select, all the units. Only such units which have a *possibility* of being included in the sample can be covered by statistical generalizations. If a *list* of enterprises, or lakes, is available, the problem is solved. Otherwise, it is sometimes possible to divide the country into geographical regions, each of which is small enough for a complete search of investigation units to be made. Of course, only a sample of such regions will actually be searched. Techniques of this kind are known as "area sampling". The maps used for the division into regions may be conventional, but other maps, showing e.g. vegetation zones, or climatic factors, may sometimes be more useful.

Estimation

3.31. The estimation is often straightforward

in the situation we discuss. Let us assume that we wish to estimate the total annual discharge of COD from the Swedish industry. The 36 big paper mills (with more than 500 employees) have been found to emit the amounts $x_1, x_2, \ldots, x_{36}$. Out of the 143 smaller paper mills, a sample of 30 have the discharges $x_{37}, x_{38}, \ldots, x_{66}$. The 10 009 other enterprises worthy of consideration are represented by a random sample of 102 units, with the discharges $x_{67}, \ldots, x_{168}$. The total discharge T from the enterprises under consideration, is then estimated as

$$\hat{T} = \Sigma_1^{36} x_i + \Sigma_{37}^{66} x_i + \Sigma_{67}^{168} x_i +$$

$$+ (143-30) \cdot \tfrac{1}{30} \Sigma_{37}^{66} x_i +$$

$$+ (10\,009-102) \cdot \tfrac{1}{102} \Sigma_{67}^{168} x_i .$$

This way of writing the estimate – which differs from traditional formulae – makes it clear that each of the sampled enterprises is included with its own discharge, and that average discharges are assumed for all the unsampled enterprises in groups 2 and 3. (It may of course be desirable to subdivide group 3.)

3.32. It is often possible to use some *auxiliary information* for the estimation. In the example, the total production y of a paper mill might be used (we shall assume here that one production variable is sufficient). The discharges from the unsampled paper mills, the production of which is known, could then be better estimated. We now calculate an average COD discharge per unit of production in the small paper industry[1]:

$$d = \frac{\Sigma_{37}^{66} x_i}{\Sigma_{37}^{66} y_i} ,$$

and instead of the fourth term in the expression above, we have

$$d \cdot \Sigma' y_i ,$$

where the sum Σ' now extends over all the (143–30) enterprises which were *not* sampled in the second group.

[1] It is possible that the big enterprises could be included in the calculation of this average discharge per unit. A number of conditions would then have to be met.

Presentation of results

3.33. If the estimates for the units not included in the sample are considered to be very uncertain, it is still possible to present the results by *showing the estimated figures separately*. From our example:

	Observed		Estimated for not observed enterprises		Total	
	Paper	Other	Paper	Other	Paper	Other
COD *emissions*	0000	000	000	00	0000	000

This kind of presentation has the advantage of admitting a direct grasp of the uncertain part of the total, and thus of an important aspect of the quality of that figure.

Typical values

3.34. In many instances in environmental statistics, there is a tendency to report typical values. An example is the survey P above, where twenty small areas have been selected all over Sweden. The aims are to describe ecosystems that are not too large, and which in some way are interesting and typical. Another example is afforded by some surveys of local air quality, where three or four typical cities are selected (e.g. a large residential city, a large industrial city, a smaller "average" town). A further example is the selection of eight river catchment areas in a discussion of eutrophication in the nation.

3.35. The idea behind this is clearly that the selected units represent more than them-

selves, that they give a fair picture of the prevailing conditions. But what do they represent, more specifically? It is likely that the underlying ideas could often be completed in a fruitful way. What is required is

- a definition or a description of the categories or groups which the selected units are intended to represent,
- a discussion showing that all units are, in principle, attributable to precisely one of the categories (with the possible addition of a category labelled "others"),
- the (approximate) size of the groups.

With only the extra data about size of the groups, such "typical values" would become much more valuable.

3.36. Clearly, on the basis of the indicated extra information, the procedures described in paragraphs 3.31–3.33 would also be applicable.

4. Other problems

4.37. Many monitoring systems do not fit into all the procedures described in Section 3. The section on "Typical values" may apply, since many of the fixed monitoring stations have been selected exactly in order to provide "typical data". Therefore, a statement of the extension and nature of what the data may represent would be most welcome. However, it is not easy to take samples of "all the rest", everything outside the monitoring stations.

4.38. It is not intended to give this problem a full discussion here. The superiority of fixed stations for measuring changes must be stressed. As to overviews, we mention again (cf. Polfeldt (1979, 1981)) the possibilities offered by representativeness investigations, and by Kriging and other interpolation methods. There is also often some knowledge about the scale of variation, the reasonable size of areas where the conditions are roughly similar, or a pattern for the phenomenon under scrutiny. The more such knowledge is made explicit, the better.

5. References

Cormack, R.M., Patil, G.P., and Robson, D.S. (Editors) (1979): Sampling Biological Populations. Statistical Ecology Series, Vol. 5.

Dalenius, T. (1965): Use of Sampling Methods for the Estimation of Areas. Ch. 3 in Zarkovich, S.S. (Ed.), Estimation of Areas in Agricultural Statistics. FAO, Rome. (See also other chapters in this manual.)

Iachan, R. (1982): Systematic Sampling: A Critical Review. International Statistical Review, 50, pp. 293–303.

Murthy, N.M. (1982): A Note on Symbolic Representation of Survey Designs. Survey Statistician (IASS), No. 8, pp. 3–5.

Polfeldt, T. (1979): Some Considerations on the Definition and Estimation of Environmental Quality Parameters. Proceedings of the 42nd Session of the ISI (Manila), Bulletin of the ISI, Vol. XLVIII, Book 1, pp. 341–350.

Polfeldt, T. (1981): Methodological Efforts in Environmental Statistics in Sweden. Statistics Sweden, PM NA 1981:2.

Polfeldt, T. (1983): Parameters for Air Quality Statistics. Statistics Sweden, PM MI 1983:5. Also published as SIMS Working Paper No. 25, Dept. of Statistics, Stanford University. (A revised version to appear.)

Utkom från trycket den 29 december 1983.