

2004:05

**Designutredning för KPI:
Effektiv allokering av urvalet för
prismätningarna i butiker och tjänsteställen**

**Examensarbete inom Matematisk statistik utfört på
Statistiska centralbyrån i Stockholm**

I serien Bakgrundsfakta presenteras bakgrundsmaterial till den statistik som avdelningen för ekonomisk statistik vid SCB producerar. Det kan röra sig om produktbeskrivningar, metodredovisningar samt olika sammanställningar av statistik som kan ge en överblick och underlätta användandet av statistiken.

Utgivna publikationer från 2001 i serien Bakgrundsfakta till Ekonomisk statistik

- 2001:1 Offentlig och privat verksamhet – statistik om anordnare av välfärdstjänster 1995, 1997 och 1999
- 2002:1 Forskar kvinnor mer än män? Resultat från en arbetstidsundersökning riktad till forskande och undervisande personal vid universitet och högskolor år 2000
- 2002:2 Forskning och utveckling (FoU) i företag med färre än 50 anställda år 2000
- 2002:3 Företagsenheten i den ekonomiska statistiken
- 2002:4 Statistik om privatiseringen av välfärdstjänster 1995–2001. En sammanställning från SCB:s statistikkällor
- 2003:1 Effekter av minskad detaljeringsgrad i varunomenklaturen i Intrastat – från KN8 till KN6
- 2003:2 Consequences of reduced grade in detail in the nomenclature in Intrastat – from CN8 to CN6
- 2003:3 SAMU. The system for co-ordination of frame populations and samples from the Business Register at Statistics Sweden
- 2003:4 Projekt med anknytning till projektet “Statistik om den nya ekonomin”. En kartläggning av utvecklingsprojekt och uppdrag
- 2003:5 Development of Alternative Methods to Produce Early Estimates of the Swedish Foreign Trade Statistics
- 2003:6 Övergång från SNI 92 till SNI 2002: Underlag för att bedöma effekter av tidsseriebrott
- 2003:7 Sveriges industriproduktionsindex 1913–2002 – Tidsserieanalys
The Swedish Industrial Production Index 1913–2002 – Time Series Analysis
- 2003:8 Cross-country comparison of prices for durable consumer goods: Pilot study – washing machines
- 2003:9 Monthly leading indicators using the leading information in the monthly Business Tendency Survey
- 2003:10 Privat drift av offentligt finansierade välfärdstjänster. En sammanställning av statistik
- 2003:11 Säsongrensning av Nationalräkenskaperna – Översikt
- 2003:12 En tillämpning av TRAMO/SEATS: Den svenska utrikeshandeln 1914–2003
- 2003:13 A note on improving imputations using time series forecasts
- 2003:14 Definitions of goods and services in external trade statistics

Fortsättning på omslagets tredje sida!

Ovannämnda rapporter, liksom övriga SCB-publikationer, kan beställas från:
Statistiska centralbyrån, SCB, Publikationstjänsten, 701 89 ÖREBRO,
telefon 019-17 68 00 eller fax 019-17 64 44.

Du kan också köpa SCB:s publikationer i **Statistikbutiken**:
Karlavägen 100, Stockholm

2004:05

**Designutredning för KPI:
Effektiv allokering av urvalet för
prismätningarna i butiker och tjänsteställen**

**Examensarbete inom Matematisk statistik utfört på
Statistiska centralbyrån i Stockholm**

Johanna Arvidson

LITH-MAT-Ex-2003

Producent Statistiska centralbyrån
Avdelningen för Ekonomisk statistik

Förfrågningar Anders Norberg, tel 08-506 946 79 (handledare)
Statistiska centralbyrån
e-post: anders.norberg@scb.se

Johanna Arvidson
Examinator: Per Gösta Andersson, Matematiska institutionen LiTH,
Linköping i november 2003

© 2004 Statistiska centralbyrån

ISSN 1650-9447

Printed in Sweden

Innehållsförteckning

Sammanfattning	5
Förord	7
1 Inledning	8
1.1 Bakgrund.....	8
1.2 Problemformulering.....	8
1.3 Syfte.....	9
1.4 Avgränsningar	9
1.5 Målgrupp	9
1.6 Läsanvisningar / Disposition	9
2 Om SCB	10
2.1 SCB:s organisation.....	10
2.2 SCB:s uppgift.....	10
3 KPI – Konsumentprisindex	11
3.1 Beskrivning av KPI.....	11
3.2 Hur data till KPI samlas in.....	14
3.3 Vad KPI används till.....	14
3.4 Historik om KPI.....	15
4 Metoder för statistiska undersökningar	16
4.1 Hur en statistisk undersökning går till på SCB	16
4.2 Undersökning.....	18
4.3 Urvalsundersökning.....	19
4.3.1 Argument för slumpurval	19
4.3.2 OSU.....	19
4.3.3 Stratifierat urval.....	19
4.3.4 PPS-urval.....	19
4.3.5 Urvalssäkerhet.....	19
4.4 Granskning	19
4.4.1 Interaktiv grafisk granskning.....	20
4.5 Fel i undersökningen	21
4.6 SCB:s IT-verktyg.....	21
5 Prisinsamling i KPI	22
5.1 Central- och lokal prisinsamling till KPI	22
5.2 Studie av kundbesök.....	22
5.3 SCB:s urval vid mätning av KPI.....	23
6 Kostnadsstudien	24
6.1 Min del i kostnadsstudien	24
6.2 Syfte med tidsstudien.....	24
6.3 Population, objekt, variabler, redovisningsgrupp.....	24
6.4 Mätinstrument.....	24
6.5 Urval och utsändning av tidsstudien.....	25
6.6 Mätning.....	25
6.7 Insamling, avprickning, manuell förgranskning.....	25
6.8 Granskning av undersökningen	26
6.9 Bearbetning.....	27
6.10 Fel i tidsstudien.....	28
6.11 KPI:s totala budget.....	28
7 Varians i KPI	32
7.1 Diskussion om de båda metoderna	32
7.2 Design- och modellbaserad variansskattningsmodell	32
7.3 Tillämpning av den designbaserade variansskattningsmodellen.....	36

8	Allokering	41
8.1	<i>Allokeringsproblemet</i>	41
8.2	<i>Optimal allokering</i>	41
8.3	<i>Systematiskt och slumpmässigt sökarbete</i>	41
9	Resultat	43
9.1	<i>Indata</i>	43
9.2	<i>Vad är viktigast att KPI mäter samt alternativa beräkningar av varianser</i>	43
9.3	<i>Resultat från allokeringen</i>	44
9.3.1	Resultat från systematisk addera.....	44
9.3.2	Resultat från systematisk subtrahera.....	54
9.3.3	Resultat från slumpmässig.....	55
10	Slutsats	58
	För den systematiska modellen med gräns:.....	58
	För den slumpmässiga modellen:.....	59
	Referenser	60
	Appendix A	61
	Appendix B	62
	Appendix C	63
	Appendix D	64
	Appendix E	66
	Appendix F	67
	Appendix G	68
	Appendix H	69
	Appendix I	70
	Appendix J	71
	Appendix K	72
	Appendix L	73
	Appendix M	74

Sammanfattning

Konsumentprisindex är intressant bland annat för att transfereringar inom socialförsäkrings-systemet grundar sig på det. Exempel på dessa transfereringar är barnbidrag, pensioner och socialbidrag. För att mäta KPI görs en statistisk undersökning. Vid statistiska undersökningar finns det alltid en viss osäkerhet. Då man vill förbättra en undersökning behövs ett större urval – men det kostar pengar. Det innebär att man alltid måste göra ett avvägande mellan storlek på osäkerhet och kostnaden att utföra undersökningen. För att förbättra den statistiska kvaliteten på KPI-undersökningen kan man göra en bättre fördelning av det totala butiksurvalet, urvalet av representantvaror samt antal observationer per butik och representantvara.

Syftet med rapporten är att utföra en optimal allokering av butiks-, varu- och produkturvalen i KPI som dras årligen. Ett annat sätt att uttrycka sig på är att vi vill utnyttja KPI-anslaget bättre genom att minimera osäkerheten. För att uppnå syftet har jag delat upp det i flera frågor som vi ska besvara. För det första måste vi undersöka hur stor kostnaden är för prismätning i ytterligare en butik, hur stor kostnaden är för ännu en representantvara samt hur stor kostnaden är att utföra ännu en prisnotering i en butik. Sedan ska vi undersöka hur många butiker som vi ska prismäta i, hur många representantvaror vi ska ha och hur många produkter som ska prismätas. De avgränsningar som görs för denna rapport är att lösa allokeringproblemet för tre branscher nämligen för klädeshandeln, möbelhandeln och restauranger.

För att finna det bästa urvalet har jag låtit en dator finna optimum med stort sökarbete. Programmen (två systematiska och ett slumpmässigt) gjordes i SAS och hade till uppgift att minimera variansen för given kostnad.

Resultat systematisk:

Antalet klädesbutiker, möbelbutiker och restauranger ska alla minskas. Det går inte att säga något generellt för alla produkter eller produkter inom en grupp (kläder, möbler och restaurangtjänster).

För möbelbutiker och restauranger bör antalet butiker vara fler i slutet av året än i början av året. För klädesbutiker ska urvalet vara stort redan i början av året. Det skulle kunna bero på att omsättningshastigheten på kläder är extremt hög. I januari finns januarirean med dramatiska prisändringar.

Vid arbete med KPI måste man veta vad som är viktigast att KPI mäter. Vill man optimera KPI:s urval för att skatta årlig förändring (inflation) ska man titta på variansen för butiker och varor under andra halvåret, det vill säga varianser beräknas som medelvärde för korttidslänkar månad 7–12 år 2001–2002. Om man prioriterar förändringen av inflation ska man beräkna varianser för månadsförändringar. Vi har kunnat beräkna varianser för varje månad 2001–maj 2003.

Resultat slumpmässig:

Det finns ett negativt samband mellan antal klädesbutiker och antal varor i klädbutiken. Man kan alltså till viss del godtyckligt välja fler butiker och färre varor eller vice versa. För möbelbutiker och restauranger har man större valmöjlighet då man ska välja en kombination av antal butiker och antal produkter för prismätning.

En gräns har valts av praktiska skäl på att man max ska välja 30 produkter i en butik. Anledningen är dels att inte arbetet ska bli för betungande för intervjuaren samt att det inte ska störa butiksägaren alltför mycket. En sak som är värt att notera är att för restauranger och kläder är det många värden som är otillåtna, det vill säga hamnar utanför gränsen på 30 produkter. Däremot ger inte begränsningen på möbler något utslag.

Idag används 1 050 000 kronor för klädesundersökningen, den undersökningen borde kosta 1 250 000 kronor medan möbel- och restaurangundersökningarna ska göras 200 000 kronor billigare.

Förord

Det här examensarbetet är ett arbete om 20 poäng inom matematisk statistik. Det är utfört på Statistiska centralbyrån i Stockholm i samförstånd med Linköpings tekniska högskola. Att arbeta med detta projekt har varit mycket givande ur flera aspekter såsom att jag fått arbeta med ett mycket intressant problem, fått använda och utveckla mina kunskaper inom matematisk statistik samt fått praktisk erfarenhet hur en undersökning på en statistik byrå görs.

Jag vill framförallt tacka min handledare, Anders Norberg på metodfunktionen under ekonomisk statistik på SCB, som med stort stöd och engagemang hjälpt mig och gjort denna tid så lärorik. Bättre handledare kan man inte ha! Därefter vill jag tacka designgruppen, Magnus Andersson, Kamala Krishnan, Martin Ribe samt Anders Norberg som låtit mig vara delaktig i deras projekt och visat mig hur arbetet på SCB går till. Ett tack vill jag även rikta till metodfunktionen med Lennart Nordberg som chef för att jag fått ingå i deras arbetslag. Dessutom vill jag tacka Jörgen Dalén för nyttiga synpunkter då det gäller variansskattning.

Jag vill även tacka min examinator Per Gösta Andersson vid Linköpings tekniska högskola för synpunkter om min rapport.

Slutligen vill jag tacka min familj och Henrik Åkerström för stöd och uppmuntran innan och under tiden för examensarbetet.

Johanna Arvidson
Stockholm, September 2003

1 Inledning

Det första kapitlet är den inledande delen av uppsatsen, här kommer bakgrunden till varför uppsatsen har ett värde att diskuteras. I denna del presenteras problemformulering, syftet med uppsatsen och de avgränsningar som gjorts. Kapitlet avslutas med vilken målgrupp uppsatsen riktar sig mot samt läsanvisning för uppsatsen.

1.1 Bakgrund

Konsumentprisindex, KPI, är intressant av en rad olika skäl såsom att transfereringar inom socialförsäkringssystemet grundar sig på det. Vid statistiska undersökningar finns det alltid en viss osäkerhet - det gäller även vid mätning av KPI. Antag att inflationen, vilken mäts som årstakten i förändringen av KPI, är $2\% \pm 0,3\%$ och att transfereringarna är i storleksordningen 100 000 000 000 under ett år. Nästa år blir överföringarna $102\,000\,000\,000 \pm 300\,000\,000$. Då osäkerheten är 0,3% innebär det att slumpen avgör om det blir en feltransferering på 300 000 000. För att kunna få lägre osäkerhet behövs ett större urval, men att göra ett större urval kostar pengar. Hur omfattande storleken på urvalen ska vara är ett svårt avvägande, möjligtvis kan urvalsstorlekarna väljas effektivare.

Inom SCB:s avdelning för ekonomisk statistik bedriver Prisprogrammet, med stöd av Metodfunktionen, ett utvecklingsprojekt med syfte att förbättra den statistiska kvaliteten i KPI. Projektet jag arbetar med är en del av ett större projekt, kallat designprojektet. Avsikten med projektet är att:

- Skaffa nödvändig information om kostnader och variationer¹ och göra ett bättre urval.
- Utvärdera nuvarande kvalitetsvärderingsmetod och alternativa metoder.
- Utvärdera instruktionspärmen om KPI som intervjuarna använder.

Prisprogrammet har haft osäker information om vad det i genomsnitt kostar att skicka en intervjuare till en slumpmässigt vald butik samt vad det kostar per produkt att prismäta den när intervjuaren väl etablerat sig i butiken. Med mer tillförlitlig information om kostnader och variation i prisutveckling (varians) kan det troligtvis göras en bättre fördelning av det totala butiksurvalet, urvalet av representantvaror och antal observationer per butik och representantvara. (Norberg (2003b))

1.2 Problemformulering

Idag går knappt hälften av kostnaderna för intervjuarna till prismätning. Resten används till andra kostnader såsom resa till och från butik. Frågorna som ska utredas är:

- Hur stor är kostnaden för prismätning i ytterligare en butik och i hur många butiker ska prismätning ske?
- Hur stor är kostnaden för ännu en representantvara² och hur många representantvaror ska det finnas?
- Hur stor är kostnaden för ytterligare en produkt (prismätning) och hur många produkter ska prismätas?
- Hur stor är kostnaden för varje intervjuare och hur många intervjuare ska utföra prismätningen?

Jag har inte möjlighet att mäta variationen mellan intervjuarnas arbete (det kommer Anders Norberg att arbeta med i ett annat projekt) vilket innebär att jag inte kan föreslå något annat antal av intervjuare än det vi har idag. Det betyder att den sista frågan inte kommer besvaras utan endast är med för att se alla dimensioner av problemet.

¹ Spridningen av värdena.

² En representantvara är en typ av vara som SCB har definierat, till exempel: damkänga – i skinn, ofodrad eller tunt foder, sula av gummi, och/eller syntet. Det finns flera produkter i varje butik som passar in under representantvara.

1.3 Syfte

Syftet är att utföra en optimal allokering av butiks-, varu- och produkturvalen i KPI som dras årligen. Ett annat sätt att uttrycka det på är att vi vill utnyttja KPI-anslaget bättre genom att minimera osäkerheten.

1.4 Avgränsningar

De avgränsningar som görs för denna rapport är att lösa allokeringsproblemet för tre branscher nämligen för klädeshandeln, möbelhandeln och restauranger. Avgränsningen görs för att inte den stora mängden av data skall bli en fördröjningsorsak. Dessa varugrupper är gynnsamma att analysera då SCB mäter priset på minst två produkter inom varje kombination av försäljningsställe och representantvara, vilket är en nödvändig förutsättning för variansskattningar. Varugrupperna är intressanta tillsammans eftersom mätningen av kläder är dyr och har stor varians medan restaurangtjänsterna mäts relativt billigt via telefonsamtal och dess varians är liten. Kostnaden för mätningen av möbler är mindre än kostnaden för mätning av kläder men större än kostnaden för mätning av restaurangtjänster. Då programmen för att lösa allokeringsproblemet för dessa branscher är gjorda kommer Norberg att utföra uträkningarna för övriga branscher.

1.5 Målgrupp

Målgruppen för denna rapport är studerande på matematikprogram främst inom matematisk statistik, andra intresserade med motsvarande förkunskapskrav samt anställda på ekonomisk statistik på SCB främst inom metodfunktionen och prisprogrammet.

1.6 Läsanvisningar / Disposition

I kapitel 2 beskrivs SCB, i kapitel 3 redogörs för konsumentprisindex, kapitel 4 behandlar metoder för statistiska undersökningar. Kapitel 5 innehåller information om hur prisinsamling i KPI görs, kostnadsstudien presenteras i kapitel 6 och varians för KPI beskrivs i kapitel 7. Kapitel 8 innehåller allokering av urvalet för KPI och 9 innehåller resultat. Uppsatsen avslutas med kapitel 10 som är slutsatser.

2 Om SCB

Kapitlet om SCB är en fristående del, avsikten är att ge läsaren djupare kunskap om SCB. Kapitlet innehåller en skildring över företagets organisation, jag redogör noggrannare för hur de avdelningar som jag varit i kontakt med är organiserade. Kapitlet avslutas med en beskrivning över vilken uppgift SCB har.

2.1 SCB:s organisation

SCB har fyra områden som de arbetar inom, dessa är arbetsmarknad och utbildningsstatistik, befolkning och välfärdsstatistik, ekonomisk statistik samt miljö- och regionalstatistik. Varje område motsvaras av en avdelning med många underenheter, till exempel under ekonomisk statistik finns prisprogrammet och metodfunktionen. Gemensamt för hela SCB är ett antal funktioner såsom stab och personal. (SCB:s hemsida (030212))

Under avdelningen för ekonomisk statistik finns metodenheten, som har tio anställda, fyra i Stockholm och sex i Örebro. Chefen på metodenheten sitter i Örebro, han åker till Stockholm minst en dag i veckan. Övriga anställda åker ibland till Stockholm respektive Örebro för att träffas. Ett annat sätt de löser möten på är att ha videokonferensmöten då alla tio kan vara närvarande. De som arbetar på metodenheten sysslar med metodfrågor, de arbetar ut mot andra enheter inom ekonomisk statistik. Oftast arbetar personerna på metodenheten med olika projekt; vilket innebär att de egentligen har ganska lite arbete tillsammans. På övriga enheter, exempelvis priser, finns det också personer som arbetar med metodfrågor, man kan säga att metodenheten är ett stöd då det gäller metodfrågor i olika projekt.

På prisprogrammet arbetar 37 anställda, de flesta som är anställda på prisprogrammet är ekonomer, statistiker eller matematiker. Prisprogrammets har ansvar för produktion av statistik om priser och prisutveckling. (SCB:s intranät (030303)) Intervjuarprogrammet består av 147 fältintervjuare som är bosatta över hela landet. Några av dessa 147 är pensionerade som arbetar per timme vid behov. Dessutom har intervjuarprogrammet en telefongrupp. Landet delas in i 7 regioner, varje region har en samordnare. Bland de cirka 150 intervjuarna arbetar för närvarande cirka 90 med insamling av priser till KPI.

Figur 1
Organisationen för ekonomisk statistik på SCB

Ekonomisk statistik				
Ledning	Analys-funktionen	IT-funktionen	Finansmarknad	Forskning och informatik
Industri-indikatorer	Industri	Metod-funktionen	National-räkenskaper	Offentlig ekonomi
Priser	Företagsregister	Tjänsteäring	Utrikeshandel	

De som är skrivna med fet stil är de program som jag arbetat med.
SCB:s intranät (030303)

2.2 SCB:s uppgift

SCB har till uppgift att ta fram aktuell, tillförlitlig och objektiv officiell statistik på olika samhällsområden och göra statistiken lättillgänglig. Exempel på de samhällsområden där de tar fram statistik är arbetsmarknad, ekonomi, näringsliv och priser, befolkning och välfärd samt bostäder och byggande. All officiell statistik ska vara lättåtkomlig och kostnadsfri. (SCB:s hemsida (030212))

3 KPI – Konsumentprisindex

Kapitlet inleds med en beskrivning av KPI, följt av en redogörelse av hur KPI mäts, dess användning och historia.

3.1 Beskrivning av KPI

Det statistiska måttet konsumentprisindex visar hur konsumentpriserna utvecklas. Ett enkelt sätt att förklara KPI är att tänka sig en korg med de varor och tjänster ett hushåll konsumerar; KPI är då måttet för prisutvecklingen för den korgen. KPI-resultatet kommer en gång i månaden. År 1954 började konsumentprisindex beräknas månadsvis, då till och med i december 1980 var basåret 1949. Idag är basåret 1980 och det är detta års resultat som motsvarar KPI-talet 100. För att åskådliggöra hur KPI fungerar följer ett exempel. I mars 2000 var KPI-talet 259,9, det innebär att 100 kronor 1980 svarar mot 259.90 kronor i mars 2000. (Ribe (030217a))

KPI har två uppgifter, det ena är att visa prisläget månaden innan och det andra är att visa prisutvecklingen från decennium till decennium, det vill säga visa prisutvecklingen på kort- och långsikt. Jag börjar med att förklara på kortsikt; varje månad följer SCB prisutvecklingen för en korg jämfört med priset på korgen i december året innan, till exempel i februari 2003 får vi uppgifter om prisförändringen på korgen mellan december 2002 och januari 2003. Den korg man följer svarar mot konsumtionsmönstret för föregående år, exempelvis så svarar den korg som man följer 2003 mot de konsumtionsmönster som fanns år 2002. Då det gäller långsikt; en gång per år får man ett långsiktigt KPI-tal. Det långsiktiga KPI-talet är prisutvecklingen för hela året, mellan december året innan till december det år korgen gäller. Vill vi exempelvis ha en prisutveckling på korgen för 2002 räknas prisutvecklingen från december 2001 till december 2002 för konsumtionen under 2002 ut. Alltså finns det i december varje år två KPI-tal, ett kort- och ett långtidsindex. (Ribe (030217c))

Skillnaden mellan långsiktig (L) och kortsiktig (K) länkning är att den långsiktiga använder kvantiteten Q_y från år y och den kortsiktiga använder kvantiteten Q_{y-1} från år $y-1$, det vill säga kvantiteten från detta år respektive året innan. Formeln nedan är den teoretiska formeln, senare beskrivs hur denna formel är konstruerad för att praktiskt kunna användas.

$$L_{y-1,12}^{y,12} = \frac{\sum_k P_k^{y,12} Q_k^y}{\sum_k P_k^{y-1,12} Q_k^y}$$

$$K_{y-1,12}^{y,m} = \frac{\sum_k P_k^{y,m} Q_k^{y-1}}{\sum_k P_k^{y-1,12} Q_k^{y-1}}$$

Summeringen sker över N produkter där $k=1, \dots, N$. Q står för kvantitet och P för prisutvecklingen. (SCB, Ekonomisk statistik, Priser, (2001)) Kvantiteten är volymen, exempelvis kan det vara kilo, styck eller kilowattimmar. Det beror naturligtvis på vilken enhet priset uttrycks i, till exempel pris/kilo, pris/styck eller pris/kilowattimme. Kvantiteten är ganska osäker med mycket bortfall i mätningarna, i prisutvecklingen är det däremot inget bortfall för det är de priser som intervjuarna samlar in, dock finns en viss osäkerhet där med eftersom inte alla butiker undersöks.

Nedan beskrivs formlerna. Den långsiktiga länkningen avser december varje år. Summering i täljaren sker med avseende på prisutvecklingen och kvantiteten. Priset är det som gäller under december innevarande år och kvantiteten är hur mycket folk köper under hela detta år. I nämnaren görs summering med avseende på priset i december föregående år och kvantiteten är samma som i täljaren. Efter divisionen mellan dessa båda summeringar fås den långsiktiga indexlänken. Innebörden av divisionen är att förhållandet mellan de båda årens kostnader för samma varukorg erhålls. Den kortsiktiga länkningen avser en given månad under året. Sum-

mering i täljaren sker med avseende på pris och kvantitet. Priset är det pris som finns under den månad det här året som man ämnar undersöka och kvantiteten är hur mycket folk köpte under hela förra året. I nämnaren görs summering med avseende på det pris som fanns i december föregående år och kvantiteten är densamma som för täljaren. Efter divisionen mellan dessa båda summeringar fås den kortsiktiga länkningen. Innebörden av divisionen är att förhållandet mellan de båda årens kostnader för samma varukorg fås för varje månad.

Vitsen med långtidskorgen är att den, till skillnad mot korttidskorgen, tar hänsyn till att olika prisutveckling för olika varor påverkar konsumenternas konsumtionsmönster. Korttidskorgen baseras på förra årets konsumtionsmönster som inte överrensstämmer med årets. Med andra ord kan korttidskorgen innehålla ”för mycket” av det som blivit dyrare och ”för lite” av det som blivit billigare, vilket kan innebära att korttidskorgens prisökning blir större än den faktiska prisökningen. (Ribe (030217c)) Detta åskådliggörs med ett exempel: om apelsiner blir dyrare och äpplen billigare kommer konsumenterna att köpa lite mindre apelsiner och lite mer äpplen än de gjort tidigare och ändå vara lika nöjda. Långtidskorgen tar hänsyn till denna förändring i konsumtion, men det gör inte korttidskorgen.

KPI är uppbyggt som en kedja med en länk för varje år. För de år som varit används respektive års långtidslänkar som kommer av de långtidskorgar som finns för åren och för pågående år (inklusive december) brukas den korttidslänk som kommer av korttidskorgen för året. (Ribe (030217c))

Kedjeindexet från år 0 till månad m i år Y blir

$$KPI_0^{Y,m} = I_0^{0,12} * L_{0,12}^{1,12} * \dots * L_{y-1,12}^{y,12} * \dots * L_{Y-2,12}^{Y-1,12} * K_{Y-1,12}^{Y,m}$$

där $I_0^{0,12} = L_{-1,12}^{0,12} / \frac{1}{12} \sum_m K_{-1,12}^{0,m}$

$I_0^{0,12}$ kommer från år 1980, det kan beskrivas som prisutvecklingen från ”1980 i genomsnitt” till december 1980. Senare år hoppar man från december till december. Korttidslänken K sträcker sig från december föregående år till vilken månad som helst innevarande år. Långtidslänken L betecknar index för de hela år som gått. Detta innebär att i långa loppet beror KPI bara av långsiktig länkning som har vägning enligt konsumtionen under samma år som prisutvecklingen avser. I den kortsiktiga länkningen används viktning från föregående år. Den kortsiktiga termen ersätts med den långsiktiga vid årets slut. (SCB, Ekonomisk statistik, Priser, (2001)) Det vill säga varje månad läggs det till en ny faktor som är kvoten mellan den månaden och december året innan. Nästa månad görs en ny beräkning och KPI uppdateras; den faktor som innehöll föregående månad tas bort och den nya faktorn läggs till. I december görs en ny uppdatering och skillnaden mellan december detta år och december året innan beräknas. Denna faktor ligger kvar för det år som varit. Ett kort exempel följer som en illustration: KPI:s årsmedel under 2000 var 260.7, i mars 2001 var det 264.6 och i juni 2001 var det 268.3, för hela året 2001 var det 267.1. Den procentuella ökningen av KPI presenteras i tabell 1. Ur tabellen visas att under år 2001 var KPI-ökningen lägre än den var i juni 2001 men högre än den var i mars 2001, alla beräknas med avseende på år 2000. Detta exempel visar vikten av att vara tydlig med att ange för vilken månad eller år KPI gäller. (SCB:s hemsida (030212))

Tabell 1
Procentuell ökning av KPI

Tidpunkt	KPI	Procentuell ökning mellan 2000 och mars-, juni- respektive hela 2001
År 2000	260.7	–
Mars 2001	264.6	$(264.6-260.7)/260.7=1.50\%$
Juni 2001	268.3	$(268.3-260.7)/260.7=2.92\%$
År 2001	267.1	$(267.1-260.7)/260.7=2.45\%$

Sverige är det enda landet som använder två index, vanligt i andra länder är att de endast har en motsvarighet till vår korttidslänk som de kedjar årligen eller mer sällan.

I praktiken används formeln nedan vid beräkning av KPI:s länkar, där $V = PQ$. Den är densamma som den tidigare men uttryckt på ett lite annorlunda sätt.

$$\text{Långsiktig vikt för } y-1 \text{ är } W^{y-1,L} = \frac{V^{y-1} / \frac{P^{y-1}}{P^{y-2,12}}}{\sum (V^{y-1} / \frac{P^{y-1}}{P^{y-2,12}})}$$

och stoppas in i det långsiktiga indexet

$$L_{y-2,12}^{y-1,12} = \sum W^{y-1,L} \frac{P^{y-1,12}}{P^{y-2,12}}$$

$$\text{Kortsiktig vikt för år } y \text{ är } W^{y,K} = \frac{V^{y-1} \frac{P^{y-1,12}}{P^{y-2,12}} / \frac{P^{y-1}}{P^{y-2,12}}}{\sum (V^{y-1} \frac{P^{y-1,12}}{P^{y-2,12}} / \frac{P^{y-1}}{P^{y-2,12}})}$$

och stoppas in i det kortsiktiga indexet

$$K_{y-1,12}^{y,m} = \sum W^{y,K} \frac{P^{y,m}}{P^{y-1,12}}$$

V^{y-1} är konsumtionen i kronor, $\frac{P^{y-1}}{P^{y-2,12}}$ är priskorrigeringen av värdet. y avser det aktuella året, $y-1$ föregående år, $y-1,12$ december föregående år och $y-2,12$ december för två år sedan.

$$P^{y-1} = \frac{1}{12} \sum_{m=1}^{12} P^{y-1,m} \quad \text{det vill säga medelpris per år.}$$

Nedan följer fyra formler för index aggregat och uppgifter om hur bra modellerna är. Genom formeln går man från priser till index. Detta är beräkningar på en "lägre nivå" än det som står tidigare i 3.1, man beräknar indexet för priser och det kan sedan användas i beräkningen av KPI.

$$A = \frac{\frac{1}{n} \sum_{ij}^n P_{ij}^t}{\frac{1}{n} \sum_{ij}^n P_{ij}^0} \quad \text{Ganska bra}$$

$$R = \frac{1}{n} \sum_{ij}^n \frac{P_{ij}^t}{P_{ij}^0} \quad \text{Förbjuden}$$

$$G = \left(\prod_{ij}^n \frac{P_{ij}^t}{P_{ij}^0} \right)^{\frac{1}{n}} \quad \text{Bra}$$

$$RA = \frac{\frac{1}{n} \sum_{ij}^n P_{ij}^t / \frac{1}{2}(P_{ij}^0 + P_{ij}^t)}{\frac{1}{n} \sum_{ij}^n P_{ij}^0 / \frac{1}{2}(P_{ij}^0 + P_{ij}^t)} \quad \text{Används!}$$

p är pris, 0 står för december förra året och t för aktuell månad. Formeln RA är den som används, i specialfall används formeln A. (Norberg (2003c))

3.2 Hur data till KPI samlas in

Prismätning för KPI utförs på så vis att antingen går anställda intervjuare från SCB ut och besöker butiker för att prismäta varor eller så ringer de till butiker och frågar om priserna eller så görs prismätningen centralt av SCB:s personal. Centralt bestäms om intervjuaren ska utföra prismätningen via besök eller telefon. Dock kan intervjuarna ibland behöva resa till en butik trots att den är definierad som telefonbutik, ett exempel är blomsterbutiker. Anledningen till att intervjuaren behöver resa till butiken kan vara att personalen inte svarar i telefonen eller säger sig inte ha tid att svara.

Eftersom det varken går att samla in priser på alla produkter, eller besöka alla butiker som finns i Sverige måste urval göras. (Ribe (030217a)) Vid val av butiker görs ett slumpmässigt urval med hjälp av det företagsregister, vilket innehåller butiker i olika branscher, som SCB har. Urvalet görs med hänsyn till det antal anställda butiken har, vilket i detta fall räknas som en indikator på hur stor butiken är, se 5.3 för mer information. (Ribe (030217b)). De representantvaror som ska finnas med i urvalet bestäms centralt av SCB, det är inget slumpmässigt urval utan är ett noga övervägt val. För mer information se 5.3. Slutligen väljer intervjuarna själva i respektive butiker en eller flera produkter per representantvara, produkterna måste dock falla inom ramen för beskrivningen av representantvaran och den/de är per definition den/de mest sålda inom aktuell vara. Det kan dock vara svårt att veta vilken som är den/de mest sålda till exempel inom klädbranschen då plaggen inte finns i sortimentet särskilt länge, då får intervjuarna välja den/de produkt som de tror kommer bli den/de mest sålda. Undantag är i livsmedelsbutiker där representantvaran är en definierad produkt, bestämd centralt av SCB. Ett exempel på en bestämd produkt i en livsmedelsaffär är: två kilo vetemjöl, Kungsörnen. Olika produkter och tjänster har olika prisutveckling. Det KPI gör är att visa den genomsnittliga prisutvecklingen för alla typer av produkter och tjänster. Inom samma år är det samma produkter som prismäts, undantag görs om produkten slutar säljas. Att utbudet ideligen ändras är vanligt framförallt inom klädbranschen. Då det inträffar måste intervjuaren välja en ny produkt, detta leder naturligtvis till osäkerhet. Vid val av ny produkt kvalitetsvärderar intervjuaren efter egen bedömning, det innebär att intervjuaren väger in sina egna värderingar. Till exempel kan en intervjuare mena att en prisökning på en TV rättfärdigas av att bildskärmen är större, medan en annan intervjuare inte anser det. Kvalitetsvärdering utförs inte på produkter i livsmedelsbutiker, eftersom ingen ny produkt väljs där om en produkt inte längre finns. Anledningen till att jämförelsen görs mellan samma produkter är att det då går att jämföra resultaten mellan olika månader. (Ribe (030217a))

Vid uträkning av KPI krävs mycket data, både priser på produkter och uppgifter om hur mycket konsumenterna konsumerar. Dessa uppgifter fås av månatliga insamlingar av priser respektive försäljningsstatistik samt undersökningar där konsumenterna tillfrågas om sina utgifter. (Ribe (030217a))

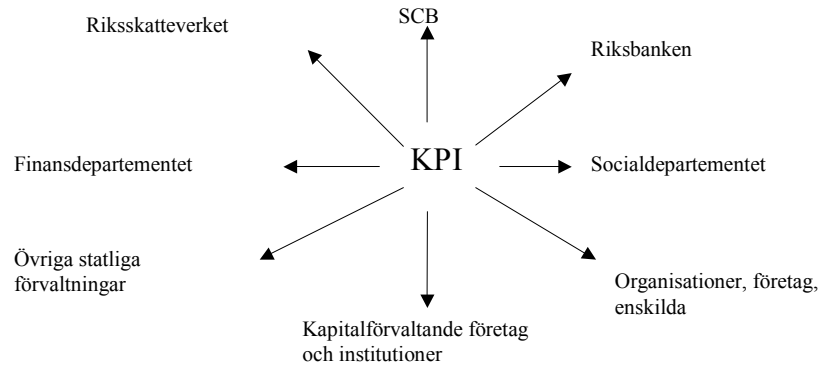
3.3 Vad KPI används till

KPI är av betydelse för hela befolkningen, för att kunna kompensera exempelvis löntagare och pensionärer då priserna på varor och tjänster blir högre och pengarna inte räcker till på samma sätt. (Ribe (030217a)) KPI används dessutom som mätare för inflationstrycket i landet. Det är orsaken till att exempelvis Riksbanken bevakar KPI-talet. (Ribe (030217b))

KPI används flitigt av olika departement såsom finansdepartementet och socialdepartementet. Finansdepartementet använder det som utgångspunkt vid beslut inom konjunktur- och stabiliseringspolitiken och socialdepartementet vid bestämmande av prisbasbeloppet, vilket styr exempelvis pensioner och studielån. Även Riksskatteverket och Riksbanken använder KPI. Det första för uträkning av de omräkningstal som behövs vid realisationsvinstbeskattning av fastighet och uträkning av brytpunkten i inkomstskatteskalen. Riksbanken använder förändringen av KPI som uttrycklig målvariabel samt som grund för avgörande i frågor beträffande penningpolitiken. Andra som också använder KPI är SCB själva för omsättningsstatistiken och deflatering i nationalräkenskaperna, andra statliga förvaltningar (till exempel

Statens Jordbruksverk och KI³), organisationer, företag och privatpersoner. De tre sistnämnda vid indexreglering i avtal samt vid omräkning av värdebelopp till fast penningvärde. Slutligen så använder kapitalförvaltande företag och institutioner KPI som stöd för bedömning av kommande räntenivåer och realavkastning. (Källa PR0101)

Figur 2
Använder KPI



Egen figur. (SCB:s hemsida (030212))

3.4 Historik om KPI

Från juli 1954 har KPI beräknats varje månad, (sedan 1949 finns det kvartalsvisa uppgifter). Att beräkna KPI utförs inte på samma sätt i alla länder. Det finns däremot andra ”mått”, HIKP (Harmoniserade konsumentprisindexet), som beräknas med hjälp av KPI som har samordnade mätmetoder för alla länder inom EU. HIKP har framställts för att göra internationella jämförelser möjliga. (Källa PR0101)

³ Konjunkturinstitutet

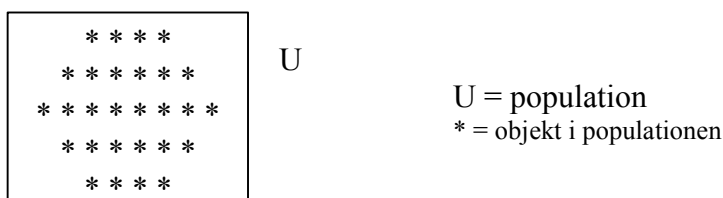
4 Metoder för statistiska undersökningar

Kapitlet inleds med ett avsnitt om hur en statistisk undersökning går till. Sedan följer beskrivningar av undersökning, urvalsundersökning, granskning samt fel i undersökningen.

4.1 Hur en statistisk undersökning går till på SCB

Sakproblem är de problem en statistisk undersökning kan hjälpa till att reda ut, på så vis att en undersökning genererar information som behövs för att lösa problem. Undersökningens värde står i proportion till hur informationen kan nyttjas för att belysa sakproblemet. Dessa problem är ofta av invecklad natur, det kan vara politiska-, sociala- eller ekonomiska problem. Ett *statistiskt problem* är i sin tur det problem som kan lösas med hjälp av statistisk metodik. (SCB:s hemsida (030226)) *Objekt* definieras som företag, personer, föremål, händelser eller liknande där deras egenskaper eller attityder är det som man vill undersöka. (SCB:s hemsida (030226)) Alla objekt måste vara namngivna så att de går att skilja på dem. Objekten namnges ofta som $1, \dots, N$ där N är antalet objekt i populationen. Figuren nedan illustrerar detta. Mängden av objekten (till exempel personer) utgör *populationen*. Förteckningen eller listan över undersökningens populationsobjekt kallas *ram*. (Rosén (2002))

Figur 3
Objekt i populationen U

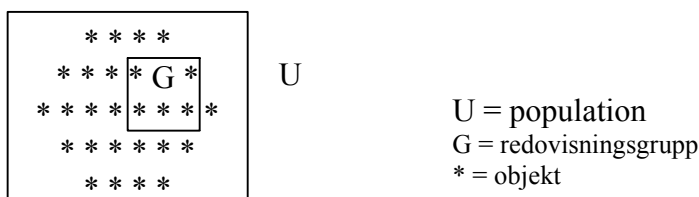


(Rosén (2002))

Urvalet är en del av ramen, det väljs med hjälp av statistiska metoder från ramen. Vid ett urval är det de utvalda objekten som undersöks och slutsatser kan dras om hela populationen. Urval beskrivs närmare i 4.3. (SCB:s hemsida (030226)) En storhet som antar olika värden för olika objekt kallas *variabel*, vilken kan vara kvantitativ (till exempel inkomst) eller kvalitativ (till exempel kön). En kvantitativ variabel kan vara kontinuerlig eller diskret. Exempel på den förra är ålder och längd, exempel på den senare är barnantal och rum i en lägenhet. (SCB:s hemsida (030226)) En kvantitativ (eller kodad kvalitativ) variabel y redovisas av sina variabelvärden $y = (y_1, y_2, \dots, y_N)$ där y_i = variabelvärdet för objekt i , $i=1, 2, \dots, N$. (Rosén (2002))

Den indelning man har vid presentation av statistik kallas *redovisningsgrupper*. Figuren nedan ger en förklaring på vad en redovisningsgrupp är.

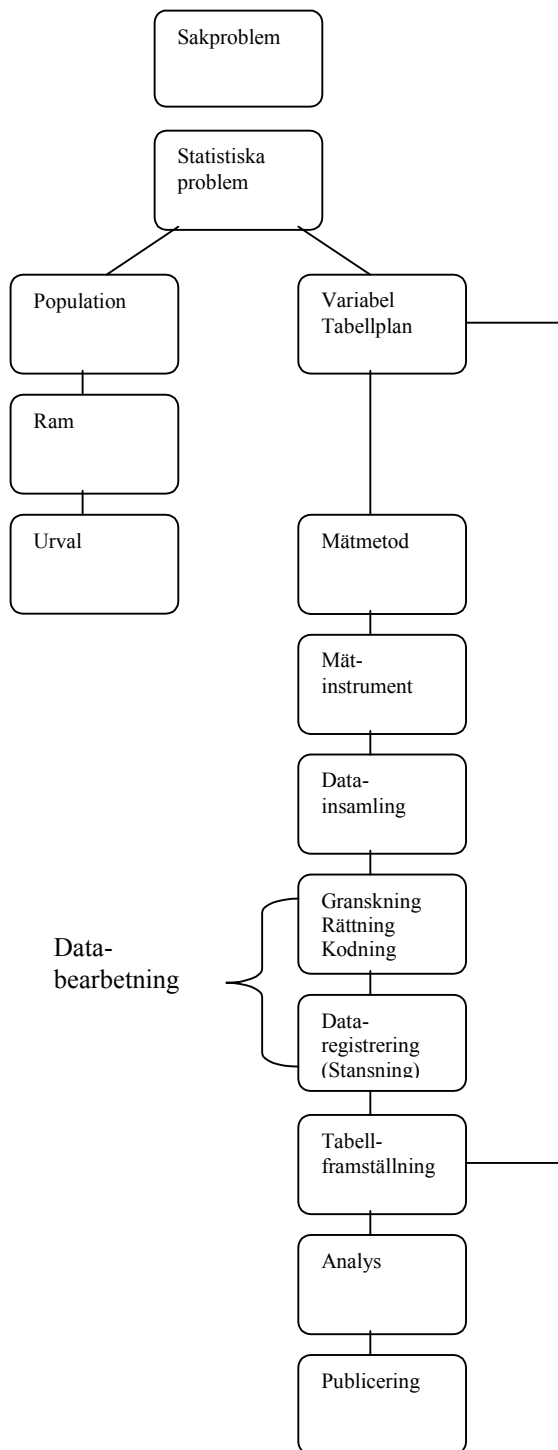
Figur 4
Redovisningsgrupp G i populationen U



(Rosén (2002))

Undersökningens resultat ska presenteras i tabeller. Då dessa tabeller är tomma kallas de *tabellplan*. På det sätt man ordnar fram upplysning om undersökningsvariablerna kallas *mätmetod*, de metoder som kan nyttjas är registerdata, brevenkät, telefonintervju samt besöksintervju. Det formulär med frågor som ska ställas kallas *mätinstrument*. *Datainsamlingen* resulterar i svar via enkät och intervju. *Databearbetning* delas upp i granskning, kodning samt dataregistrering. Granskning innebär att rätta fel i materialet som samlas in. Kodning avser att översätta svaren till ett homogent språk, med andra ord översätta text till siffror (exempelvis så kan kön kodas så att 1 står för man och 2 för kvinna). Granskning beskrivs mer i 4.4. *Dataregistrering* innebär att data läggs in i filer. Resultaten av undersökningen produceras i *tabellframställningen*, vilken görs enligt tabellplanen. Exempel är diagram, tabeller och medelvärden. Vid *analys* undersöks vad som fåtts fram av undersökningen och vid *publikation* sammanställs resultaten för att publiceras. (SCB:s hemsida (030226))

Figur 5
Hur en statistisk undersökning på SCB går till



(SCB:s hemsida (030226))

4.2 Undersökning

Det är av största vikt att utföra en god planering innan en undersökning genomförs. Det är möjligt att felaktigt tro att om det är så att man har en stor undersökning (många respondenter) så är den automatiskt god. Så är icke fallet då det finns exempel på flera stora undersökningar som gjorts där det inte gått att få fram något vettigt då de haft ett felaktigt upplägg. (Blom, Holmquist (1998))

4.3 Urvalsundersökning

Då en undersökning ska ge exakta värden på statistiska storheter måste en totalundersökning göras, med andra ord måste variabelvärden för samtliga objekt i populationen samlas in. Totalundersökningar är både kostsamma och tar mycket tid i anspråk. Då man undersöker ett urval eller stickprov av populationens objekt är det troligt att erhålla en bra uppfattning om förhållande i hela populationen. Om en jämförelse mellan totalundersökning och urvalsundersökning görs kan man finna att fördel med urvalsundersökning är att det tar kortare tid att få fram statistiken, den är billigare samt att det leder till mindre uppgiftslämnarbörda. (Rosén (2002)) Man kan då använda resurserna till säkrare mätmetoder vilket kan leda till att den statistiska total kvaliteten blir bättre vid en urvalsundersökning än vid en totalundersökning.

4.3.1 Argument för slumpurval

Om man gör ett urval som har någon form av systematik och sedan använder skattningar är det möjligt att resultatet av undersökningen betvivlas då man kan mena att det kan bli ett skevt urval på grund av att systematiken medförde förbindelse mellan ”komma med i urvalet” och ”undersökningsvariabelns värde”. För att minimera risken för ett skevt urval ska ett urval göras där det inte finns någon form av systematik. Det som är mest osystematiskt är slumpen och att göra ett slumpurval kan därför anses vara det bästa rent teoretiskt sätt. (Rosén (2002))

4.3.2 OSU

Ett urval från $U=(1,2,\dots,N)$ är ett obundet slumpurval (OSU) med urvalsstorlek n om detta gäller:

- Urvalet har fix storlek n .
 - Varje U -delmängd med n (olika) objekt har samma sannolikhet att bli urvalet.
- (Rosén (2002))

4.3.3 Stratifierat urval

Med ramen/populationen U uppdelad i stratan A_1, A_2, \dots, A_N skapas ett stratifierat urval genom att:

- Från vart och ett av stratana dras ett sannolikhetsurval.
- Urvalen från de olika stratana dras oberoende av varandra.

4.3.4 PPS-urval

Probability Proportional to Size (PPS) innebär att urvalen görs med sannolikhet proportionell mot något storleksmått. (Dahmström (1991)) Här följer ett exempel som åskådliggör PPS-urval: Man ska välja fem kulor ur 15 grupper som alla innehåller olika antal kulor. Sannolikheten är proportionell mot antalet kulor i gruppen; det vill säga att den grupp med flest kulor får högst sannolikhet att väljas, den med minst lägst sannolikhet att väljas och så vidare.

4.3.5 Urvalsosäkerhet

Det finns alltid en osäkerhet då det gäller statistiken från en urvalsundersökning. Anledningen beror på att den korrekta informationen för alla inte kan fås vid undersökning av några. Även andra faktorer, såsom störningar och misstag, bidrar till skattningarnas totala osäkerhet. (Rosén (2002))

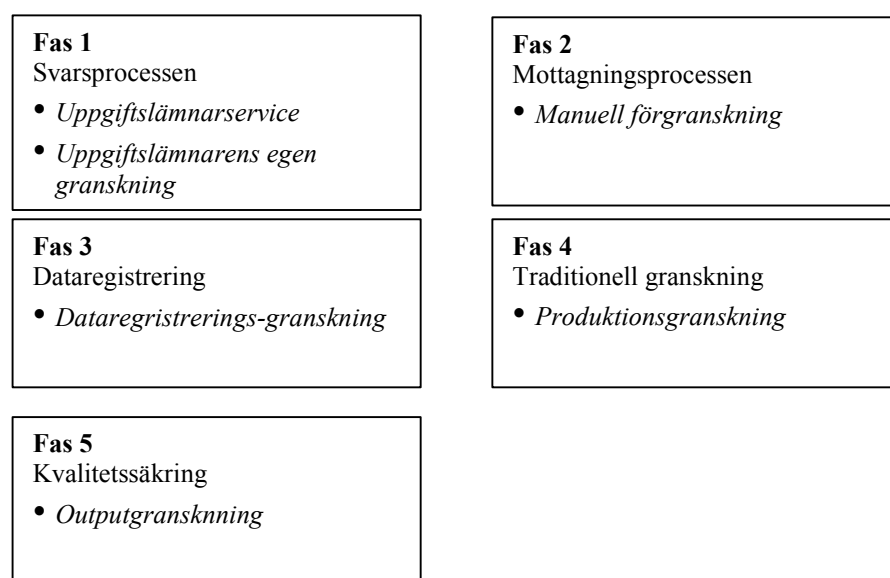
4.4 Granskning

Målet med granskning är att lära av misstag och olämpliga metoder för att undersökningen ska göras bättre nästa gång, det vill säga att avstyra att fel uppstår, att öka kvaliteten på data i aktuell undersökning samt att medverka till kvalitetsbedömning av statistiken. Då en undersökning görs är det viktigt att ställa sig frågor som vilka granskningsprocesser som krävs och vid val av flera delprocesser hur de ska koordineras så att dubbelarbete och onödigt arbete ska kunna uteslutas. (Granquist, Arvidson, Elffors, Norberg, Lundell (2003))

Granskning av data kan göras vid ett eller flera tillfällen i insamlings- och produktionsprocessen; uppgiftslämnargranskning, manuell förgranskning, dataregistreringsgranskning, pro-

duktionsgranskning samt outputgranskning. Uppgiftslämnargranskning innebär att uppgiftslämnaren själv eller i samråd med intervjuaren granskar sina svar i undersökningen. Manuell förgranskning innebär att personal som utför undersökningen snabbt läser igenom den ifyllda blanketten för att se om den är så mycket ifylld att det är meningsfullt att låta den fortsätta i granskningsprocessen. Vid dataregistreringsgranskning, som är nästa steg, kontrolleras variabelvärden samtidigt som de registreras. Produktionsgranskning görs för att undersöka om svaren är rimliga. Slutligen görs outputgranskning för att se om resultaten av undersökningen är rimliga. (Granquist, Arvidson, Elffors, Norberg, Lundell, (2003))

Figur 6
Faser vid granskning av postenkäter



Egen bild, fakta från (Granquist, Arvidson, Elffors, Norberg, Lundell, (2003)).

4.4.1 Interaktiv grafisk granskning

Interaktiv grafisk granskning innebär att alla objekt visas som punkter eller staplar i diagram, misstänkta objekt markeras och dess variabelvärde visas på skärmen för verifiering, effekterna av ändringarna kontrolleras och då effekten inte påverkar skattningen ska man sluta med granskningen. Det kan vara en effektiv metod för att finna såväl extrema värden som oväntade mönster. Grafisk granskning kan ge underlag för val av acceptansgräns till den traditionella granskningen. För att illustrera hur viktigt grafisk granskning är följer ett ordspråk: en grafisk bild säger mer än hundra datasidor. (Granquist, Arvidson, Elffors, Norberg, Lundell, (2003))

Det finns flera typer av diagram som kan nyttjas vid grafisk granskning. Valet av diagram beror på vad man vill undersöka. I boken guiden till granskning ger författarna tips om vilken typ av diagram som är lämplig att använda för att ta reda på olika saker. Nedan presenteras dessa tips.

- Vid upptäckande av uppenbara fel → stapeldiagram
- Vid fastställande av acceptansgräns → lådagram
- Vid upptäckande av avvikelserfel → punktdiagram eller lådagram
- Vid påträffande av inliers → punktdiagram
- Vid outputgranskning → lådagram

(Granquist, Arvidson, Elffors, Norberg, Lundell, (2003))

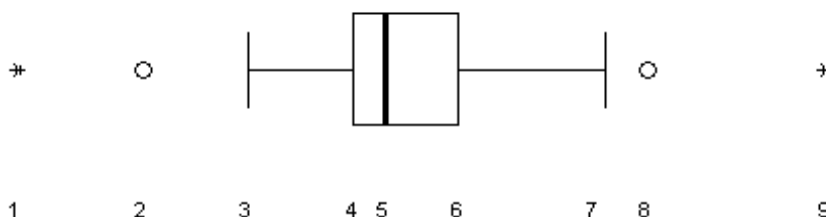
Ett lådagram har ett antal beståndsdelar, såsom låda, morrhår samt extremvärden. Den övre och undre kvartilen begränsar lådan, minst 50 procent av fördelningen ligger i lådan. Medianen är det streck som finns i lådan. Längden på morrhåren bestäms av en konstant k som multipliceras med kvartilavvikelsen. Storleken på k är valfri, standard är dock 1,5. De punk-

ter som finns utanför morrhåren är extremvärden. (Granquist, Arvidson, Elffors, Norberg, Lundell, (2003))

Figur 7 visar lådagramets delar, numreringen avser följande förklaringar:

1 och 2 står för extremvärden som ligger lägre än 1,5 kvartilavstånd från boxen. 3 är morrhårets ända vilket står för lägsta värdet inom 1,5 kvartilavstånd från lådan. 4 är undre kvartilen som är den nedre delen av lådan, medianvärdet ligger vid 5 och 6 står för den övre kvartilen. Morrhårets andra ända finns i 7, det är det största värdet inom 1,5 kvartilavstånd från lådan. Både 8 och 9 är extremvärden.⁴ (Handelshögskolan vid Göteborgs universitets hemsida (030505))

Figur 7
Lådagram



Handelshögskolan vid Göteborgs universitets hemsida (030505).

4.5 Fel i undersökningen

Fel i undersökningar kan vara av flera olika slag såsom täckningsfel, urvalsfel, bortfallsfel, mätfel samt bearbetningsfel. Vid täckningsfel så är det fel i ramen. Det finns två typer av täckningsfel, det ena är övertäckning och det andra undertäckning. Det första är då objekt som ej ska vara med i populationen ingår i ramen, och det andra är då objekt som ska vara med i populationen inte ingår i ramen. Undertäckning är problematiskt medan övertäckning kan rensas i mät- och/eller granskningsprocessen. Om undersökningen genomförs för ett urval av objekt ur populationen får man en stokastisk osäkerhet, urvalsfel. Då det inte finns mätvärden för alla objekt uppkommer bortfallsfel. Det finns två typer av bortfall, objektbortfall vilket inträffar då objektet vägrar svara eller är oanträffbar samt partiellt bortfall som uppkommer då svar inte fyllts i på alla frågor. Mätfel kan komma av fel i mätmetoden (exempelvis då intervjuaren påverkar respondenten) eller fel i mätinstrumentet; det vill säga att det som ska mätas inte mäts eller att det mäter annat än det som ska mätas. Slutligen finns bearbetningsfel som kommer av kodning-, dataregistrering- och tabuleringsfel. (SCB:s hemsida (030226))

4.6 SCB:s IT-verktyg

Det analysverktyg som SCB använder sig av är statistikprogrammet SAS, vilket står för Statistical Analysis System. SAS innehåller ett stort antal moduler som behövs vid hantering av datamaterial; statistiska analyser, presentation och grafik. (Åbo akademi datacentrals hemsida (030226)) En styrka som SAS besitter är det kan sköta stora datamängder. För att få en uppfattning vad en "stor datamängd" är så kan en tabell som innehålla upp till 13 000 variabler och tabeller med många, mer än 10 000 000, observationer gå fort att behandla i SAS. (SCB:s intranät (030303))

⁴Det är olika tecken på 1 och 2 respektive 8 och 9 men ingen skillnad mellan dem.

5 Prisinsamling i KPI

Kapitlet handlar om hur prisinsamlingen för KPI utförs, först beskrivs insamlingen av data till KPI. För att få mer insikt hur prisinsamlingen går till i praktiken gjorde jag en studie av ett kundbesök som beskrivs här. Därefter kommer ett avsnitt om central och lokal prismätning och slutligen beskrivs SCB:s urval vid mätning av KPI.

5.1 Central- och lokal prisinsamling till KPI

Det finns två tillvägagångssätt att samla in priser på, det ena är central prisinsamling och det andra är lokal prisinsamling. Vid central insamling är det den centrala personalen på prisprogrammet som samlar in priser via telefon eller postenkät. Då det gäller lokal prisinsamling har produktspecifikationerna angetts centralt, ett urval av butiker dragits och sedan uppdragits åt SCB:s intervjuare att samla in data. Då specifikationerna inte är så strama får prisinsamlaren ta den mest sålda produkten som stämmer överens med specifikationerna. Om specifikationerna är väldigt strama måste prisinsamlaren söka efter den angivna produkten, finns den inte blir de tvungna att ange att den saknas. (SCB, Ekonomisk statistik, Priser, (2001)) Uppgifter om priser samlas in varje månad för ett urval av representantvaror. (SCB:s hemsida (030226), Konsumentprisindex 1830–2001, 030221)

I det så kallade dagligvarusystemet ingår alla dagligvaror förutom färsk frukt, grönsaker, mjukt bröd och färsk fisk. Prisprogrammet på ekonomisk statistik på SCB har valt många varuvarianter och specificerat dem grundligt i blanketterna. Med andra ord ska intervjuaren inte själv välja varianter i dagligvarusystemet med undantag för kött och ost. Dagligvaruhandeln delas in i tre grupper; ICA, KF och övriga butiker. Urvalen av varuvarianter ser olika ut för dessa. Det ingår ungefär 60 butiker i undersökningen och varje urval innehåller ungefär 400 varuvarianter. Vid prismätning av kläder och PC finns separata system med skyttelblanketter, dessa rymmer ett stort antal uppgifter per variant. I prismätningen för kläder ingår omkring 90 butiker och för PC ett 20-tal. Lokalprissystemet inkluderar de varor och tjänster som inte nämnts ovan. Här bestämmer intervjuaren själv de lokala varuvarianter som skall prismätas. En kvalitetsvärdering ska göras vid variantbyte för många av dessa varor. Drygt 200 representantvaror och -tjänster hör till denna grupp. Det är omkring ca 600 försäljningsställen som omfattas, priserna samlas in per telefon eller via besök. Telefongruppen som finns i Örebro står för insamlingen av priser för bensin och diesel. För de varor där man vet att priserna är desamma i alla butiker, till exempel för alkohol och läkemedel, samlas priserna in centralt. Andra varor och tjänster där det behövs särskilda metoder mäts också centralt, gäller exempelvis för resor. Även boendekostnader, som mäts genom bland annat hyresundersökningar en gång i kvartalet, och insamling av prisuppgifter för värme och hushållsel, sköts centralt. (Norberg (2002))

5.2 Studie av kundbesök

Den 12 februari 2003 följde jag med en intervjuare då hon prismätte varor i en klädesbutik i Skärholmen. Prismätningen omfattade bland annat tröjor, kjolar och byxor. En problematisk sak som jag fick erfara var den höga omsättningshastigheten för kläder, det vill säga kläderna finns bara i butiken under ett par månader. Då intervjuaren söker en produkt som inte finns får hon välja ett annat likvärdigt plagg, till exempel har hon tittat på vad priset på ett par byxor kostar och de inte finns längre får hon välja ett par andra byxor av samma märke, det går alltså inte i detta fall att välja en kjol eller skjorta. Problemet är att välja ”rätt” produkt så att prismätningen blir så bra som möjligt (men det faller utom ramen för mitt arbete). Ett problem som dock ligger inom ramen för min undersökning är det delproblem som kommer av att byta varor som prismätts är den tid det tar att byta. Det som är mest tidkrävande är att leta efter produkter samt bestämma vilken ny produkt som ska väljas då produkten inte finns i sortimentet längre, men längst tid tar det nog att söka produkterna då de i klädbutiker ofta möblerar om bland produkterna.

Intervjuaren jag följde menar att prismätning i möbelaffärer tar kortare tid än prismätningen i klädbutiker eftersom sortimentet inte ändras så mycket i möbelaffärer. Det tar dock lång tid om en produkt utgått ur sortimentet för det är då inte bara att välja en ny eftersom det finns restriktioner att följa. Måste intervjuaren välja många nya varor tar det mycket lång tid, men oftast är så inte fallet. Att vara noggrann och skriva ner tydligt vilken produkt man prismäter är av största vikt eftersom det kan vara svårt att komma ihåg exakt hur exempelvis kjolen ser ut när man ska söka den efter en månad. Om en intervjuare inte antecknat grundligt kan det ta henne lång tid att söka igenom butiken och försäkra sig om att det var just den kjolen hon prismätte förra gången. Man kan alltså vinna mycket tid genom att skriva små ”kom ihåg” notiser. Ett annat skäl till att skriva bra beskrivningar är om intervjuaren blir sjuk och någon annan intervjuare tar över arbetet är det enklare för den nya då tydliga beskrivningar finns. En butik kan ingå i prismätningen ett eller flera år, dock högst fem år i följd. De kan naturligtvis återkomma i undersökningen. Somliga företag vill inte vara med, men de har inget val eftersom företagen har uppgiftslämnarplikt.

5.3 SCB:s urval vid mätning av KPI

Eftersom, som tidigare nämnts, det av praktiska skäl är omöjligt att undersöka alla butiker, varor och tjänster har SCB gjort ett urval av noteringsställen (butiker och betjäningsställen), representantvaror, varuvarianter och tjänster. SCB har utformat undersökningen som en panelundersökning, vilket innebär att de priser som samlas in är priserna på samma representantvara i samma butik. Då urval av butiker väljs dras de slumpmässigt branschvis från Företagsdatabasen. Butiksurvalet är ett stratifierat pps-urval, det går till på så sätt att alla butiker får en viss sannolikhet att komma med i undersökningen. Vid urvalet viktas butikerna med avseende på antal anställda, vilket innebär att de butiker som har många anställda har en större sannolikhet att bli utvald. Det är ett roterande butiksurval, vilket innebär att varje år byts 20% av butikerna ut mot nya.

Urvalet av representantvara och varuvarianter/tjänster går till på så vis att en vald vara, så kallad representantvara, får representera en grupp av produkter. Anledningen till detta är att det är omöjligt att i praktiken regelbundet samla in prisuppgifter på allt. Kriterier för val av representantvara är att de ska vara lätta att definiera, de ska finnas i butikerna under hyfsat lång tid framöver samt att de ska säljas över större delen av landet. De produkter i butikerna som prismäts väljs inte ut slumpmässigt, men SCB räknar på det som om det valdes slumpmässigt. Det finns inget sätt att räkna på det om man räknar att det inte väljs slumpmässigt. Regeln är som sagt att det ska vara den mest sålda produkten inom en viss representantvara. Men ska man prismäta i en klädesbutik så har personalen svårt att svara på det eftersom de hela tiden byter sortiment. Där får istället intervjuaren välja ett plagg som det finns många av och som man kan anta att de räknar med ska bli en storsäljare. Dagligvarorna (livsmedel) väljs ut med ett stratifierat pps-urval, varje produkt får ett tal som står i proportion till hur mycket som säljs av produkten. Försäljningsstatistiken för dagligvaror fås av tre stora livsmedelskoncerner, nämligen ICA, Coop Konsum och Axfood. Produkterna byts ut då de inte säljs mer. (SCB, Ekonomisk statistik, Priser, KPI)

6 Kostnadsstudien

Undersökningar på SCB utförs på ett likartat sätt. Nedan följer kort hur arbetet utfördes vid den undersökning som jag var verksam i. Detta kapitel skiljer sig från övriga uppsatsen på så vis att den är som en "rapport i rapporten" då den innehåller såväl syfte som resultat av kostnadsstudien.

6.1 Min del i kostnadsstudien

Då jag kom in i bilden var enkäten redan utsänd. När enkäterna samlats in har jag prickat av, dataregistrerat, granskat och analyserat dem. Vid sammanställningen använde jag mig av excel och vid en del av granskningen av villkorsstyrd formattering i excel. För att analysera data använde jag mig av SAS.

6.2 Syfte med tidsstudien

SCB:s syfte med tidsstudien är att undersöka hur mycket tid (och därmed pengar) prisprogrammet betalar för att intervjuarna skall komma i kontakt med butik (resa, telefonsamtal och förberedelse) och vad det kostar att göra en prisnotering när kontakten är skapad. Resultat erfordras per bransch och gärna per varugrupp

6.3 Population, objekt, variabler, redovisningsgrupp

Objektet i denna studie är dels butiker och dels produkter. Populationen torde vara alla butiker i alla branscher som ingår i prismätningen. Det ideala skulle vara att göra denna tidsstudie under flera månader, men det är en ekonomisk omöjlighet eftersom det tar mycket tid i anspråk framförallt av personal som ska analysera och registrera data. Februari är ett medvetet urval bland årets 12 månader. Att februari är representativt för alla månader kan inte påstås, men det är inte heller vår ambition att påstå något liknande.

De redovisningsgrupper som finns i tidsstudien är stratumgrupper. Indelningen i grupper görs för att kunna presentera resultaten på ett överskådligt sett.

Medelvärde och trimmat medelvärde beräknas för såväl övrig tid som för tid per notering för de olika stratumgrupperna. Ett trimmat medelvärde är då man utesluter de 10% största och de 10% minsta värdena och sedan beräknar medelvärdet av övriga. Lådagrammen, se appendix F, visar variationen för resa i kilometer för stratumgrupp vid besök i butik, övrig tid i kilometer för stratumgrupp vid besök i butik samt övrig tid för stratumgrupp vid telefonintervju.

6.4 Mätinstrument

Det mätinstrument som används vid tidsstudien är en postenkät till de intervjuare som gjorde KPI i februari 2003. *Enkäten* var utformad på så vis att den innehöll tre avsnitt; ett om hur instruktionspärmen fungerar, ett om hur tidsåtgången vid prisinsamlingen i februari såg ut och slutligen hur intervjuarna uppfattar att kvalitetsvärderingen fungerar. För mitt arbete är det den andra delen som är av intresse.

Delen som behandlade tidsåtgången vid prisinsamling i februari innebar att intervjuarna skulle specificera tiden för antal kilometer till butik, transporttid, hemresa i kilometer och minuter, prisinsamling samt övrigt tid vid telefonintervju. De skulle även fylla i den sammanlagda tid för KPI-arbete som är gemensam för alla butiker såsom postöppning, planering, paketering samt återsändning. Enkäten är utformad på så vis att det innan frågorna finns en *instruktion* som berättar hur blanketten ska fyllas i. Den bedömdes vara noga utformad för att minimera att den blir felaktigt ifyllt. Dock verkar det tämligen omöjligt att göra en blankett som inte misstolkas av någon, till exempel så varierar den sammanlagda tiden som är

gemensam för alla butiker väldigt mycket. Det finns också en del partiellt bortfall. En *provundersökning* med 3 intervjuare genomfördes. För att förebygga missförstånd hade det varit bra om en större provundersökning gjorts, förslagsvis med 10–20 intervjuare.

6.5 Urval och utsändning av tidsstudien

Denna undersökning är en totalundersökning, alla intervjuare som arbetar med KPI tillfrågas. Undersökningen gjordes endast vid ett tillfälle, för februari 2003. Enkäten skickades per brev till samtliga intervjuare den 5 februari 2003. Då enkäten var ifylld skickades den tillbaka till SCB i Stockholm. Ett e-brev skickades ut och påminde att det snart var dags att lämna in blanketterna och de som inte fått någon (ett finare sätt att säga ”tappat bort den”) kunde höra av sig för att få en ny. Efter sista inlämningsdatumet skickades ännu ett e-brev till intervjuarna. I brevet tackades intervjuarna för att de skickat in enkäten, det stod också att det fortfarande gick bra att skicka in för dem som inte gjort det. Man kan säga att det var ett kombinerat tack- och påminnelsebrev.

6.6 Mätning

De problem som upptäcktes vid manuell förgranskning och dataregistrering av svaren på enkäterna och vad vi gjort för att ”fixa det” sammanfattas nedan.

- Svårt att veta vilken enhet intervjuarna använt; timme och minuter eller enbart minuter. Ett svar kunde till exempel vara ”120”, betyder det 1 timme och 20 minuter eller 120 minuter (dvs 2 timmar). En granskning måste göra så förhållandet mellan kilometer och minuter verkar rimligt.
- För sammanlagd tid för alla butiker (till exempel postöppning) verkar det som om några har missuppfattat och angett något annat, troligtvis all KPI-tid. Det mest realistiska tas bort.
- För fem blanketter fanns det två intervjuarnummer på en och samma blankett, vilket innebär att intervjuarna inte samlat in priser i de butiker de fått tilldelade. Vi beslöt oss för att den intervjuare som genomfört undersökningen skulle få ”överta” butiken i fråga.
- Det finns de som endast fyllt i tid på övrig telefontid. Kan vara så om personalen i den butik som ska prismätas inte svarar, men det verkar inte så troligt att det ska vara så många butiker som är oanträffbara. Dessa bortser vi helt från.
- Någon har angett minuter det tagit att komma till en butik men inte hur många kilometer det var och ingen hemresa.
- Somliga har adderat tiden det tar att prismäta eller antal kilometer till butiker för flera butiker.
- Då intervjuarna glömt sätta klammer⁵ då de har besökt flera butiker på en och samma resa har jag lagt in en klammer.

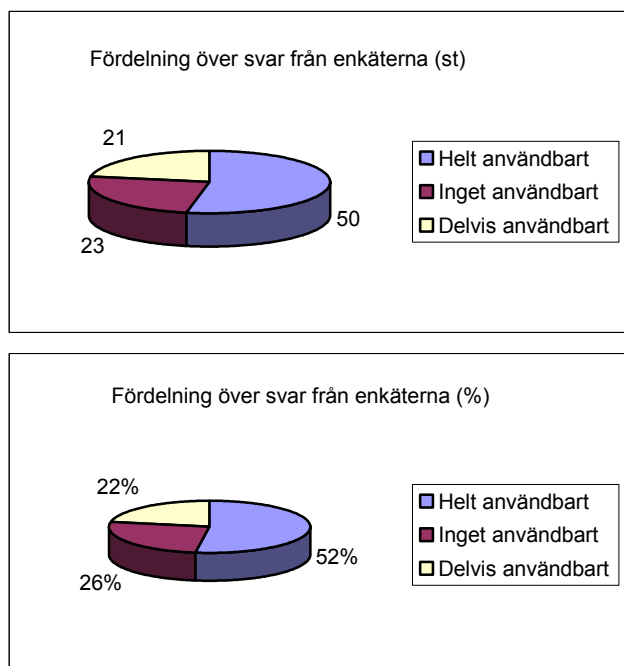
6.7 Insamling, avprickning, manuell förgranskning

För att få reda på hur många som inte skickar in blanketten görs avprickning. En förteckning över alla intervjuare med intervjuarnummer, namn och butiksnummer gjordes. Därefter prickades de som skickade in enkäten av. Av de 94 utskickade enkäterna var det var 75 som svarade. Detta innebär att det var 19 som inte svarade på enkäten, det vill säga det var ett objektbortfall på 19. Ytterligare 4 enkäter var inte alls dugliga; inget material användes från dem. Detta innebär att objektbortfallet i själva verket var 23 (19+4). Det var 21 enkäter som hade oanvändbart material på vissa frågor; det kallas partiellt bortfall. För 50 enkäter kunde allt material användas. Orsaker till bortfallet går bara att spekulera i. Det är inte speciellt troligt att intervjuarna glömde bort enkäten eftersom de har fått flera påminnelser. Man kan tänka sig att de medvetet struntat i att besvara enkäten. Man kan tycka att det är ytterst dumt att inte alla intervjuare svarat på enkäten eftersom det borde ligga i deras intresse att deras

⁵ En symbol (en klammer) görs för att visa att dessa butiker besöktes på en och samma resa.

arbete ska bli effektivare. En annan aspekt är att de som arbetar med enkäter borde veta hur viktigt det är att svara på enkäter för att kunna dra tillförlitliga slutsatser.

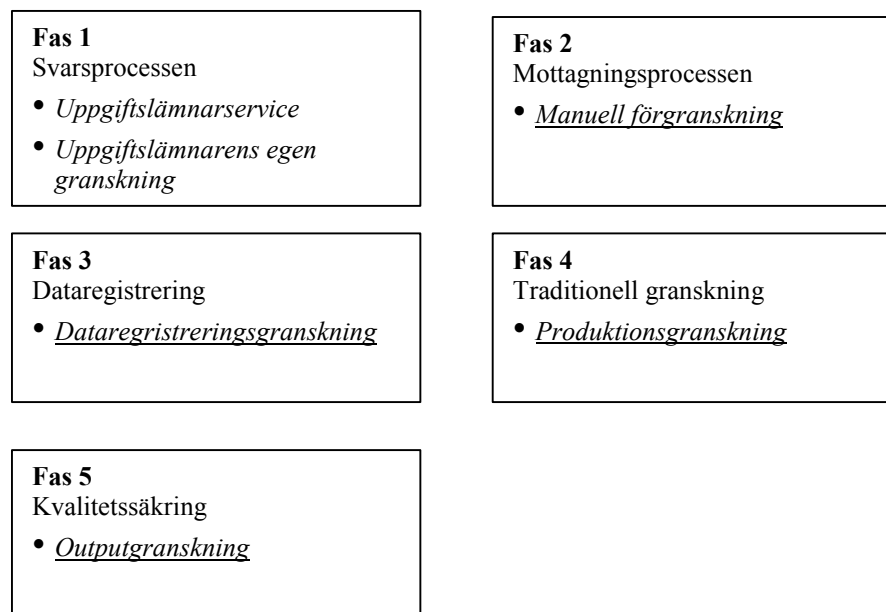
Figur 8
Fördelning över svaren i antal respektive procent



6.8 Granskning av undersökningen

Vid undersökningen gjordes manuell förgranskning, dataregistreringsgranskning, produktionsgranskning och outputgranskning. Som nämnts i 6.7 togs en lista fram med alla respondenter så att avprickning kunde ske för att se antalet som inte svarat, detta kan betraktas som manuell förgranskning. Dataregistreringsgranskning är namnet på den villkorsstyrda formateringar jag utförde. Villkorsstyrda formateringar innebär i detta fall de kontroller som programmerades i excel. Kontrollerna reagerade då något verkade orimligt, kunde till exempel vara att tiden det tog att resa till butiken inte var rimligt med avseende på antalet kilometer det var till butiken. För att undersöka om svaren var rimliga, produktionsgranska, sammanställdes data i excel. En sammanlänkning gjordes mellan data från tidsstudien, intervjuarregistret samt data som innehöll uppgifter om hur många prismätningar som gjorts per butik. Att länka samman dessa tabeller underlättar för vidare arbete vid analysering av data. Anledningen till att en samkörning mellan data från tidsstudien och intervjuarregistret gjordes var för att kunna hitta felaktigt angivna butiksnummer och korrigera felen. Då felen rättades gjordes en ny kolumn där de korrigerade värdena skrevs in. På detta sätt fanns ursprungsdata kvar om något skulle gå snett. Fel som uppstod med avseende på butiksnummer redovisas i appendix A. Att intervjuarna skrivit av butiksnumren fel och kastat om två siffror är exempel på fel i avskrivningen. En annan typ av fel är då jag stansat av fel vid dataregistreringen. De flesta fel kunde rättas, de fel som dock återstod finns i appendix B. En annan sak som kan ordnas med hjälp av registret var de butiker där inte intervjuaren angett något butiksnummer utan skrivit namnet på affären. Efter lite ”detektivarbete” kan butiksnumren anges. Anledningen till att ingen rättning kan utföras för butikerna som anges i appendix B beror på att intervjuaren inte har skickat in uppgifter om tidsåtgången vid prismätning i butiken. För den ena butiken fattas dessutom fakta om antal noteringar per butik. Slutligen tittade jag på resultaten för att se om det verkade rimligt, detta innebär outputgranskning.

Figur 9
Faser vid granskning av postenkäter



Egen bild, fakta från (Granquist, Arvidson, Elffors, Norberg, Lundell, (2003)). De som är understrukna är de som används i denna undersökning.

6.9 Bearbetning

Vid bearbetning av data har flera program i SAS gjorts, nedan redogör jag för vad programmen utför. Ett program har gjorts för att fördela gemensam KPI-tid på butikerna. Punkterna i diagrammet i appendix C står för gemensam tid för KPI-arbete för intervjuarna. Kryssen är sådana som vi valt att bortse från då vi tycker oss kunna påstå att de inte kan vara vettiga, detta innebär att vi har trunkerat bort ungefär en tredjedel av värdena. Rutorna är de värden som vi behåller och kommer att göra våra beräkningar på. Frågan om extra KPI-tid tror vi har skapat förvirring bland intervjuarna som skulle svara på enkäten. Vi misstänker att somliga intervjuare lagt samman all tid för KPI arbete och skrivit den tiden där istället för den extra tid som var avsedd att anges. Vi har eliminerat de fall där den extra gemensamma KPI tiden är för stor. Anledningen till att detta diagram gjordes var för att undersöka om det finns något samband mellan sammanlagd KPI-tid och antal butiker en intervjuare har. En idé var att den gemensamma tiden ökande med antal butiker, men det är inget som tyder på det. Det går snarare att påstå att den gemensamma KPI-tiden är ungefär lika lång oavsett antal butiker, det finns i vilket fall ingen avsevärd skillnad.

Ett annat SAS-program som konstruerades hade till uppgift att fördela ut kostnadsuppgifterna per bransch. Lådagram gjordes för att se efter om det finns extremvärden då det gäller pris per notering. Morrhåren är 3*kvartilavståndet, vilket innebär att vi tolererar ganska stora avvikelser. Finns det värden utanför morrhåren kan vi nog påstå att det är något "skumt" med dem. Anledningen till att jag valde att göra lådagram beror på att det är ett överskådligt sett att hitta extremvärden. I boken Guide till granskning (2003) rekommenderas det att göra just lådagram.

Jag skapar 19 branschgrupper av branscherna, se appendix D. Anledningen till att branschgrupper skapas är att det för vissa branscher är för lite data för att analysera. Indelningen i branschgrupper har gjorts enligt följande kriterier. Prisinsamling för branscher inom samma branschgrupp måste vara av samma slag, det vill säga huvudsakligen telefonintervju eller besök, mix får ej förekomma. Svårhetsgraden på prisinsamlingen ska överrensstämma. Med det menar jag att det är svårare att mäta möbler än livsmedel därför är det ytterst olämpligt att dessa båda skulle ingå i samma branschgrupp. Däremot kan man påstå att sättet och svårigheten att mäta priser i hemtextilbutiker och bostättningsaffärer (där man kan köpa till

exempel stekpannor) är liknande vilket gör att de båda branscherna kan ingå i samma branschgrupp. Medelvärde för resa i kilometer, övrig tid samt prisinsamlingstid per notering beräknas. Övrig tid innebär tiden att resa (hem- och ditresa) samt övrig tid vid telefonintervju. Se appendix G. Priserna på produkterna i apoteken är samma vilket innebär att det kvittar i vilket apotek som prismätningen sker.

Tre diagram har gjorts se appendix F, resa i kilometer för stratumgrupp vid besök i butik, övrig tid för stratumgrupper vid besök i butik samt övrig tid för stratumgrupper vid telefonintervju. I de båda diagrammen om restid i kilometer och övrig tid vid butiker som besöks går det att utläsa från vilka län de avvikande värdena härstammar. Jag förväntade mig att de långa resavstånden skulle vara i Norrland, Värmland och Dalarna eftersom det är förhållandevis långt mellan städerna där. Delvis kan denna tanke stärkas men det finns långa resor som görs i andra län. Om man undersöker i vilka län som en resa görs där tur och retur är tio mil eller mer är det i gruppen Norrland, Värmland och Dalarna samt i länen Småland, Västergötland och Stockholm. Som kan utläsas i appendix F figur 3 beror de flesta extremvärdena i figuren "Övrig tid för stratumgrupper vid telefonintervju" av att intervjuaren besökt butiker där priserna skulle samlats in över telefon. Den övriga tiden är som bekant restid till butik samt övrig tid vid telefonintervju. De kan vara så att butiken inte svarar i telefon vilket gör att intervjuaren blir tvungen att resa till butiken.

6.10 Fel i tidsstudien

De typer av fel som finns i tidsstudien är framförallt bortfallsfel. Båda typerna, objektbortfall och partiellt bortfall, förekommer. Anledningar till att vi fick bortfall torde vara olika då det gäller objekt- och partiellt bortfall. Att påstå att det mesta av det partiellt bortfallet beror på att enkäten missuppfattades borde inte vara något uppseendeväckande uttalande. Antagligen borde provenkäten utförts på fler personer för att få fram fler funderingar så att enkäten skulle kunna göras bättre. Däremot då det gäller objektbortfallet är det knivigare att förklara orsaken, det kan exempelvis bero på att intervjuaren inte vill eller orkar fylla i enkäten. Att hon/han skulle ha missat den verkar högst osannolikt beroende på att enkäten skickades ut i samma kuvert som blanketterna intervjuarna skulle ha för arbetet i februari. Sedan skickade ett e-brev som påminde om enkäten och frågade om det var någon som inte fått enkäten. Senare skickades två e-brev till, en där det stod att hade de några frågor om enkäten kunde de kontakta mig och en som ett som tackade för hjälpen och innehöll uppmaningen att det fortfarande var möjligt att skicka in om någon missat det.

6.11 KPI:s totala budget

Indata hämtas från appendix L och resultaten presenteras sist i detta kapitel samt i appendix M. För att kunna lösa problemet i 8.2 måste man veta hur stor kostnad den totala budgeten för KPI är (C_0). Vi låter sedan kostnaden vara lika stor som den är idag när vi söker effektivaste urvalsstorlekar.

C står för kostnader, det är indexet som anger vilken kostnad det är fråga om.

$$C_{hg} = 13 * (TidNotR_h * int + a_r)$$

C_{hg} står för kostnad per notering (kr/not) under ett år, h för butiker, g för varor och 13 innebär att det avser en 13 månaders period. Anledningen till att det är en 13 månaders period är att indexlänken startar i december och slutar i december året därpå, vilket blir 13 månader. Den trimmade tiden för prisnoteringen i minuter betecknas $TidNotR_h$, int är intervjukostnaden i kronor per minut, a_r står för dataregistrering och granskning båda i kronor per notering.

$C_{hg} = C_g$ för alla h, h är butiker, g är varor. Det vill säga vi antar att det inte finns någon skillnad i kostnaden att prismäta kläder i varuhus, stormarknader, sportbutiker eller klädbuti-

ker. Skattningen att mäta kläder är tid/not i klädbutiker. Orsaken är att i till exempel sportbutiker säljs även andra produkter än kläder.

Startkostnaden (kostnaden att förbereda besök, resa och etablering i butik) är olika beroende på om butiken klassas som en besöks- eller telefonbutik.

$$C_h = \text{int} * (urv_r + \text{sortund}_r + 13 * \text{otidr}_h) \text{ för besöksbutik}$$

$$C_h = \text{int} * (urv_r + \text{sortund}_r + 12 * \text{otidr}_h + \text{engresa}) \text{ för telefonbutik}$$

C_h står för startkostnad per butik (kr/butik) för ett år, h för butiker. Den tid det tar att välja och granska en butik betecknas urv_r och tiden det tar vid sortimentsundersökningen kallas sortund_r , båda är uttryckta i minuter per butik. Den trimmade tiden för övrig tid i minuter per butik betecknas otidr_h och slutligen är int intervjukostnaden i kronor per minut. Intervjuaren ska en gång per år besöka telefonbutiken. Den resan betecknas här med engresa (en gångs resa). Resan beräknas schablonmässigt som medelvärdet av resa för kläder, möbler, hemtextil/bosättning/vitvaror/leksaker, radio/tv, sport samt skor. Ingångsdata till engresa presenteras i se appendix I.

$$\text{engresa} = \frac{43.4027 + 46.4152 + 32.5780 + 33.0185 + 38.2986 + 37.4693}{6} = 38.53 \text{ minuter}$$

För besöksbutikerna innebär 13 att det avser en 13 månaders period. Anledningen till att det är en 13 månaders period är att indexlänken startar i december och slutar i december året därpå, vilket blir 13 månader. För telefonbutiker finns en schablonpris som uppkommer då intervjuaren reser till butik ($\text{int} * \text{engresa}$) och sen ringer intervjuaren under 12 månader. Detta blir 13 månader vilket motsvarar en indexlänk.

Nedan följer uträkningar av kostnaderna för de tre branscherna den här rapporten avser. Övriga uträkningar görs på samma sätt. Resultaten för C_{hg} presenteras i appendix H och resultaten för C_h presenteras i appendix I. De värden som används vid beräkningarna finns även de uppställda i appendix H och I.

- För kläder, vilket är en besöksbutik:

$$C_{hg} = 13 * (3,19361 * (355/60) + 4,83793 + 6,28931) = 390.3 \text{ kr/not och år.}$$

$$C_h = (390/60) * (20 + 70 + 13 * 43,4027) = 4252.5 \text{ kr/butik och år.}$$

$$a_1 = \text{reg} + \text{grans}$$

$$\text{reg} = \frac{150000}{2385 * 13} = 4,83793 \text{ kr/not, } \text{grans} = \frac{500 * 390}{2385 * 13} = 6,28931 \text{ kr/not.}$$

Dataregistreringen kostar 150 000 kr/år, det görs 2385 stycken klädnoteringar i månaden. Varje år går det åt 500 timmar till granskning av kläder, de personer som granskar kostar 355 kr/h.

- För möbler, vilket är en besöksbutik:

$$C_{hg} = 13 * (1,84948 * (355/60) + 1.8043) = 165,7 \text{ kr/not och år.}$$

$$C_h = (355/60) * (20 + 70 + 13 * 46,4152) = 4102,6 \text{ kr/butik och år.}$$

$$a_2 = \text{blankett, granskning}$$

$$a_2 = \frac{355 * 653}{13 * 9883} = 1.8043 \text{ kr/not}$$

Det tar 653 timmar att granska de 9883 lokalpriserna, kostnad för intervjuarnas är 355 kr/h, hela uttrycket beskriver granskningskostnaden per notering för möbler.

- För restaurangtjänster, vilket är en telefonbutik:

$$C_{hg} = 13 * (0,6404 * (355/60) + 1.8043) = 72.7 \text{kr/not och år.}$$

$$C_h = (355/60) * (20 + 70 + 12 * 9,0132 + 38.53) = 1400.4 \text{kr/butik och år.}$$

$a_3 = \text{blankett, granskning}$

$$a_3 = \frac{355 * 653}{13 * 9883} = 1.8043 \text{ kr/not}$$

På samma sätt som för möbler.

Intervjuarna kostar 355 kr/h = 355/60 kr/min, de som granskar blanketter kostar även de 355/60 kr/min.

$C_g = \text{urvalrep} * \text{cent}$

C_g står för kostnaden för en ny representantvara. Urvalrep står för tiden det tar att välja och definiera en ny representantvara, cent står för kostnaden som kommer av att central personal ska bestämma den nya representantvaran.

$$\text{Cent} = \frac{355 + 425}{2} = 390 \text{kr} \text{ Den centrala personalen som arbetar med att bestämma nya}$$

representantvaror är assistenter och handläggare. De båda grupperna arbetar till lika stor del med urvalet men de kostar olika mycket, där handläggarna är den dyrare gruppen.

Kläder:

$$C_g = 80 * 390 / 5 = 31200 / 5 = 6240 \text{ kr/vara}$$

Dividering med fem innebär fem års avskrivning av "investering" i ny vara. Investeringen består av kostnaden för att uppdatera blankett, dataregistreringsprogram, granskningsprogram och tabellprogram. Kostnaden uppstår endast om man vill ha någon ny representantvara.

Möbler:

$$C_g = 0.36 * 390 = 140.4 \text{ kr/vara}$$

Restaurang:

$$C_g = 0.36 * 390 = 140.4 \text{ kr/vara}$$

Registrering och granskning

$$\text{Pris per styck för lokalvaror} \frac{355 * 653}{13 * 9883} = 1.8043 \text{ kr/not}$$

Kläder (se ovan)

$$\text{Livsmedel} \frac{355 * 387}{13 * 11727} = 0.9012 \text{ kr/not}$$

Tabell 2
Sammanställning av resultaten

	C_{hg} (kr/not och år) (produkt)	C_h (kr/butik och år) (butik)	C_g (kr/vara) (representantvara)
Kläder⁶	390.3	4252.5	6240
Möbler	165.7	4102.6	140.4
Restaurang⁷	72.7	1400.4	140.4

⁶ Kläder innefattar prismätning av kläder i varuhus, stormarknader, klädbutiker och sporthandlar.

⁷ Prismätning i restaurang täcker såväl vanliga restauranger som personalmatsalar.

Lägg märke till att kostnaden för representantvara (C_g) uppkommer endast då man vill ha fler representantvaror än de som finns för närvarande.

Anledningarna till att kostnaden för representantvaror för kläder är dyrare än kostnaden för representantvaror för möbler och restauranger är:

- Man måste lära sig vad det är som bestämmer priserna på representantvarorna, det vill säga ”vad är kvalitet?”. Det gör man genom att prata med branschfolk. Det är svårare att bestämma kvalitet på kläder.
- Konstruera blanketter för kläder; vilka ska ritas och tryckas på tryckeri. Blanketter för möbler och restauranger skrivs ut från datorn varje månad.
- Klädesblanketten är en så kallad skyttelblankett, vilket innebär att den skickas fram och tillbaka mellan intervjuaren och central personal på SCB under 13 månader.
- Dataregistrering för kläder sker manuellt, för möbler och restauranger kan blanketten scanas.
- Bearbetningen och granskningen är mycket större hos kläder än hos möbler och restauranger. Till exempel måste datasystemet ändras.

7 Varians i KPI

Detta kapitel omfattar en diskussion och en presentation av två metoder som skattar varianser; en metod som är designbaserad och en som är modellbaserad. Kapitlet avslutas med en tillämpningsdel.

7.1 Diskussion om de båda metoderna

Den korsklassificerade modellen är en välkänd modell som används för skattning av variansen för KPI hos SCB idag. Dock har det visats efter simuleringar, utförda av Anders Norberg på SCB, att varianserna mellan produkter och butiker överskattas för lokalpriser och kläder. När man inför produkturval som en källa till varians visar det sig att detta urvalssteg kan generera en stor del av variansen.

Den korsklassificerade modellen är uppbyggd med avseende på dagligvarusystemet (det vill säga matvaruaffärer), den tar hänsyn till variationen mellan produkter och butiker. I dagligvarusystemet förekommer inget urval av representantvara (exempel på en representantvara är damkänga i läder) utan produkterna är stratifierade efter vad vi skulle kunna kalla representantvaror. Detta innebär att formlerna inte är helt tillfredställande för övriga grupper, då variansen för sista urvalssteget ej beräknas. Det är möjligt att en modellbaserad variansskattning inom en snar framtid kommer ersätta den designbaserade modellen. Jag redogör för båda modellerna, men begränsar mig sedan till att använda den designbaserade modellen vid allokeringen eftersom Norberg i skrivandets stund inte tillräckligt utrett att den modellbaserade skulle vara att föredra.

7.2 Design- och modellbaserad variansskattningsmodell

Den designbaserade variansskattning som tas upp i denna rapport är Dalén och Ohlssons skattning för korsklassificerat urval. Denna skattning är välkänd och accepterad, det är den modellen som har använts vid skattning av varians för KPI i Sverige. Skattningen har brister, de kommer att diskuteras senare, vilket gör att man på SCB undersöker om en annan metod för variansskattning skulle vara bättre att använda. Denna modell utreds av Norberg och den är en modellbaserad variansskattning.

➤ Designbaserad skattning: Korsklassificerat urval

Variansen består av tre delar; varians mellan produkter, varians mellan butiker samt interaktionen mellan butik och produkt. Denna struktur möjliggör jämförandet mellan produkt- och butikvarianserna. Det medför i sin tur att det går att jämföra urvalsallokeringen mellan produkter och butiker.

Urvalet av butiker och produkter för KPI dras oberoende av varandra. Dagligvarusystemet i KPI har ett äkta tvådimensionellt urval med korsklassificering av butikurval och produkturval. I lokalpris- och klädersystemen kan man säga att man har korsklassificering av butiksurval och representantvaruurval, men här är urvalet av produkter (inom butik och vara) i stort sett oberoende och det är möjligen det viktigaste urvalssteget. Förutsättningarna är därmed inte uppfyllda för att använda Dalén och Ohlsson.

Variansen för skattningen \hat{I} av I är enligt Dalén och Ohlsson (1995):

$$V(\hat{I}) \approx V_{PRO} + V_{BUT} + V_{INT}$$

V_{PRO} kallas variansen mellan produkter, V_{BUT} kallas variansen mellan butiker och V_{INT} kallas interaktionen mellan produkt och butik. (Dalén, Ohlsson (1995))

$$V_{PRO} = \sum_{g=1}^G \frac{1}{m_g} \sum_{i \in U_g^R} (1 - (m_g - 1)w_i^R) \left(\sum_{h=1}^H \frac{v_{gh}}{X_{gh}} \bar{e}_{i.}^{gh} \right)^2 w_i^R$$

$$V_{BUT} = \sum_{h=1}^H \frac{1}{n_h} \sum_{j \in U_h^C} (1 - (n_h - 1)w_j^C) \left(\sum_{g=1}^G \frac{v_{gh}}{X_{gh}} \bar{e}_{.j}^{gh} \right)^2 w_j^C$$

$$V_{INT} = \sum_{g=1}^G \sum_{h=1}^H \frac{1}{m_g n_h} \frac{v_{gh}^2}{X_{gh}^2} \left[\sum_{i \in U_g^R} \sum_{j \in U_h^C} (1 - (m_g - 1)w_i^R) * (1 - (n_h - 1)w_j^C) (e_{ij}^{gh})^2 w_i^R w_j^C \right. \\ \left. - \sum_{i \in U_g^R} (1 - (m_g - 1)w_i^R) (\bar{e}_{i.}^{gh})^2 w_i^R - \sum_{j \in U_h^C} (1 - (n_h - 1)w_j^C) (\bar{e}_{.j}^{gh})^2 w_j^C \right]$$

$$\text{där } e_{ij}^{gh} = 1_{ij} (f_{ij}^1 - I_{gh} f_{ij}^0), \quad \bar{e}_{.j}^{gh} = \sum_{i \in U_g^R} w_i^R e_{ij}^{gh}, \quad \bar{e}_{i.}^{gh} = \sum_{j \in U_h^C} w_j^C e_{ij}^{gh}$$

m_g är antalet produkter i urvalet från produktgrupp g , g betecknar produktgrupp och G är antalet produktgrupper som finns.

n_h är antalet butiker i urvalet från stratum h , h betecknar butiksgrupp, H är antalet butiksstrata som finns

v_{gh} är ett vägningstal, baserad på försäljningen under ett år, för kombinationen (g,h).

$$\text{Vägningstalen är normerade så att } \sum_{g=1}^G \sum_{h=1}^H v_{gh} = 1$$

π_{gi}^R och π_{hj}^C är urvalssannolikheter för varan i inom varugrupp g respektive butik j inom butiksgrupp h .

$$\text{Indikatorn } 1_{ij} = \begin{cases} 1 & \text{om varan } i \text{ säljs i butik } j \\ 0 & \text{annars} \end{cases}$$

$$f_{ij}^0 = \frac{p_{ij}^0}{(p_{ij}^0 + p_{ij}^1)/2}, \quad f_{ij}^0 \text{ är en funktion av pris på produkt } i \text{ i butik } j \text{ vid tidpunkt } 0.$$

$$f_{ij}^1 = \frac{p_{ij}^1}{(p_{ij}^0 + p_{ij}^1)/2}, \quad f_{ij}^1 \text{ är funktion av pris på produkt } i \text{ i butik } j \text{ vid tidpunkt } 1.$$

För varje i är w_i^R ett vägningstal för produkt i och för varje j är w_j^C ett vägningstal för butik j . Dessa vägningstal beror på storlek (omsättning) av de enskilda urvalsobjekten och urvalsförfarandet. De är normerade så att $\sum_{i \in U_g^R} w_i^R = \sum_{j \in U_h^C} w_j^C = 1$ för varje g och h .

$$X_{gh} = \sum_{i \in U_g^R} \sum_{j \in U_h^C} 1_{ij} w_i^R w_j^C f_{ij}^0, \quad Y_{gh} = \sum_{i \in U_g^R} \sum_{j \in U_h^C} 1_{ij} w_i^R w_j^C f_{ij}^1$$

Skattningen:

$$\text{Låt } \hat{X}_{gh} = \sum_{i \in U_g^R} \sum_{j \in U_h^C} 1_{ij} w_i^R w_j^C f_{ij}^0, \quad \hat{Y}_{gh} = \sum_{i \in U_g^R} \sum_{j \in U_h^C} 1_{ij} w_i^R w_j^C f_{ij}^1$$

U_g^R betecknar radstratum nummer g, produktgrupp

U_h^C betecknar kolumnstratum nummer h, butiksstratum

$\hat{X}_{gh} = 0$ då produkt i inte handlas i butik j, samma sak gäller för \hat{Y}_{gh}

$$\hat{I}_{gh} = \frac{\hat{Y}_{gh}}{\hat{X}_{gh}}, \quad \hat{I}_{gh} \text{ är ett cellindex.}$$

$$\text{KPI korttidslänk blir } \hat{I} = \sum_{g=1}^G \sum_{h=1}^H v_{gh} \hat{I}_{gh}$$

$$\text{Inför } \hat{e}_{ij}^{gh} = 1_{ij} (f_{ij}^1 - \hat{I}_{gh} f_{ij}^0), \quad \hat{e}_{i.}^{gh} = \frac{1}{m_g} \sum_{j \in S_h^C} \hat{e}_{ij}^{gh}, \quad \hat{e}_{.j}^{gh} = \frac{1}{n_h} \sum_{i \in S_g^R} \hat{e}_{ij}^{gh}$$

Skattningarna för variansen för produkter, butiker samt interaktionen:

$$\hat{V}(I) = \hat{V}_{PRO} + \hat{V}_{BUT} + \hat{V}_{INT}$$

$$\hat{V}_{PRO} = \sum_g \frac{1}{m_g(m_g - 1)} \sum_{i \in S_g^R} (1 - \pi_{gi}^R) \left\{ \sum_h \frac{v_{gh}}{\hat{X}_{gh}} \hat{e}_{i.}^{gh} \right\}^2$$

$$\hat{V}_{BUT} = \sum_h \frac{1}{n_h(n_h - 1)} \sum_{j \in S_h^C} (1 - \pi_{hj}^C) \left\{ \sum_g \frac{v_{gh}}{\hat{X}_{gh}} \hat{e}_{.j}^{gh} \right\}^2$$

$$\hat{V}_{INT} = \sum_g \sum_h \frac{v_{gh}^2}{\hat{X}_{gh}^2} \frac{1}{n_h(n_h - 1)} \frac{1}{m_g(m_g - 1)} * \left\{ \sum_{i \in S_g^R} \sum_{j \in S_h^C} (1 - \pi_{hj}^C)(1 - \pi_{gi}^R) (\hat{e}_{ij}^{gh} - \hat{e}_{i.}^{gh} - \hat{e}_{.j}^{gh})^2 \right\}$$

π är urvalssannolikhet, $(1 - \pi_{hj}^C)$ är ändlighetskorrrelation, den används vid ändlig population.

(Dalén, Ohlsson (1995)) Om man har ett stort urval är ändlighetskorrrelationen nästan 0 och har man ett litet är den nästan 1. Ett exempel förklarar detta: om man vill undersöka möbelaffärer och det finns 1200 möbelaffärer i Sverige och man undersöker 1000 stycken blir urvalssannolikheten stor; närmare 1; vilket gör att ändlighetskorrrelation blir nära 0. Undersöker man bara 100 av 1200 blir ändlighetskorrrelationen nära 1 och kan nästan bortses ifrån.

➤ Modellbaserad variansskattning

Låt Y_{ijk} = logaritmen av prisförändringen (dec år y-1 till månad m år y) för en produkt k inom varan i och butik j.

Formulera en modell som innebär att stokastiska värden Y_{ij} genereras av fyra olika och oberoende stokastiska variabler enligt

$$Y_{ijk} = \mu + \beta_i + \gamma_j + \delta_{ij} + \varepsilon_{ijk}, \text{ där}$$

$$i = 1, \dots, m_g, \quad j = 1, \dots, n_h, \quad k = 1, \dots, r_{ij}$$

μ är ett generellt medelvärde utan varians, β_i är en representantvarueffekt, γ_j är en butikseffekt, δ_{ij} är en interaktionseffekt mellan representantvaror och butiker och ε_{ijk} är en produktbjudandeeffekt.

Antag att β_i är normalfördelad med väntevärde 0 och varians σ_β^2 , γ_j är normalfördelad med väntevärde 0 och varians σ_γ^2 , δ_{ij} är normalfördelad med väntevärde 0 och varians σ_δ^2 och slutligen att ε_{ijk} är normalfördelad med väntevärde 0 och varians σ_ε^2 . I så fall är Y_{ijk} normalfördelad med varians $\sigma_\beta^2 + \sigma_\gamma^2 + \sigma_\delta^2 + \sigma_\varepsilon^2$.

Antag att vi har ett urval. Låt:

$$Y_{...} = \sum_{ijk} Y_{ijk}$$

$$Y_{i..} = \sum_{jk} Y_{ijk}$$

$$Y_{.j.} = \sum_{ik} Y_{ijk}$$

$$Y_{..k} = \sum_{ij} Y_{ijk}$$

$$Y_{ij.} = \sum_k Y_{ijk}$$

Vid variansanalys delas kvadratsumman $SS_Y = \sum \sum \sum (Y_{ijk} - Y_{...})^2$ upp i följande i två delar.

Den variation som förklaras av faktorerna

$$\sum_{i=1..m_g} (Y_{i..} - Y_{...})^2 + \sum_{j=1..n_h} (Y_{.j.} - Y_{...})^2 + \sum_{i=1}^{m_g} \sum_{j=1}^{n_h} (Y_{ij.} - Y_{i..} - Y_{.j.} + Y_{...})^2$$

och den del som inte kan förklaras, den s.k. residualkvadratsumman:

$$SSE = \sum_{i=1}^{m_g} \sum_{j=1}^{n_h} \sum_{k=1}^{r_{ij}} (Y_{ijk} - Y_{...})^2$$

SAS har en procedur för variansanalys med randomiserade effekter och som tillåter varierande antal observationer per cell, r_{ij} . Resultatet ger underlag för beräkning av skattningar av $\hat{\sigma}_\beta^2$, $\hat{\sigma}_\gamma^2$, $\hat{\sigma}_\varepsilon^2$ och $\hat{\sigma}_\delta^2$

När vi har dessa skattningar kan vi också skatta variansen för medelvärdet av logaritmerade priskvoter för ett urval av representantvaror (m_g), butiker (n_h) och totalt antal produktbjudanden (p_{gh}), åtminstone om urvalet av produktbjudanden är någorlunda lika stort för alla representantvaror och butiker. Så länge vi kan generera data med en process får vi

$$\begin{aligned} V\left(\sum_{i=1}^{m_g} \sum_{j=1}^{n_h} \sum_{k=1}^{r_{ij}} Y_{ijk}\right) &= V\left(\sum_{i=1}^{m_g} \sum_{j=1}^{n_h} \sum_{k=1}^{r_{ij}} \beta_i + \gamma_j + \delta_{ij} + \varepsilon_{ijk}\right) = \\ &= V\left(\sum_{i=1}^{m_g} \sum_{j=1}^{n_h} \sum_{k=1}^{r_{ij}} \beta_i\right) + V\left(\sum_{i=1}^{m_g} \sum_{j=1}^{n_h} \sum_{k=1}^{r_{ij}} \gamma_j\right) + V\left(\sum_{i=1}^{m_g} \sum_{j=1}^{n_h} \sum_{k=1}^{r_{ij}} \delta_{ij}\right) + V\left(\sum_{i=1}^{m_g} \sum_{j=1}^{n_h} \sum_{k=1}^{r_{ij}} \varepsilon_{ijk}\right) = \\ &= \sum_{i=1}^{m_g} V\left(\beta_i \sum_{j=1}^{n_h} r_{ij}\right) + \sum_{j=1}^{n_h} V\left(\gamma_j \sum_{i=1}^{m_g} r_{ij}\right) + V\left(\sum_{i=1}^{m_g} \sum_{j=1}^{n_h} r_{ij} \delta_{ij}\right) + V\left(\sum_{i=1}^{m_g} \sum_{j=1}^{n_h} \sum_{k=1}^{r_{ij}} \varepsilon_{ijk}\right) = \\ &= \sum_{i=1}^{m_g} V\left(\beta_i r_i\right) + \sum_{j=1}^{n_h} V\left(\gamma_j r_j\right) + V\left(\sum_{i=1}^{m_g} \sum_{j=1}^{n_h} r_{ij} \delta_{ij}\right) + V\left(\sum_{i=1}^{m_g} \sum_{j=1}^{n_h} \sum_{k=1}^{r_{ij}} \varepsilon_{ijk}\right) = \\ &= \sigma_\beta^2 \sum_{i=1}^{m_g} r_i^2 + \sigma_\gamma^2 \sum_{j=1}^{n_h} r_j^2 + \sigma_\delta^2 \sum_{i=1}^{m_g} \sum_{j=1}^{n_h} r_{ij}^2 + \sigma_\varepsilon^2 \sum_{i=1}^{m_g} \sum_{j=1}^{n_h} r_{ij} \end{aligned}$$

När vi drar våra urval från ändliga populationer av butiker, representantvaror och produkt-erbjudanden bör vi kunna använda följande estimator.

$$\text{Låt det totala antalet observationer vara } r = \sum_{i=1}^{m_g} \sum_{j=1}^{n_h} r_{ij} = \sum_{i=1}^{m_g} r_{i.} = \sum_{j=1}^{n_h} r_{.j}$$

Variansskattning för KPI-data är

$$\hat{V} \left[\frac{1}{r} \sum_i \sum_j \sum_k Y_{ijk} \right] = \left(1 - \frac{m_g}{M_g} \right) \cdot \frac{\sum_{i=1}^{m_g} r_{i.}^2}{r^2} \cdot \hat{\sigma}_\beta^2 + \left(1 - \frac{n_h}{N_h} \right) \cdot \frac{\sum_{j=1}^{n_h} r_{.j}^2}{r^2} \cdot \hat{\sigma}_\gamma^2 +$$

$$+ \left(1 - \frac{m_g \cdot n_h}{M_g \cdot N_h} \right) \cdot \frac{\sum_{i=1}^{m_g} r_{.j}^2}{r^2} \cdot \hat{\sigma}_\delta^2 + \left(1 - \frac{r}{R} \right) \cdot \frac{1}{r} \cdot \hat{\sigma}_\varepsilon^2$$

(Norberg (2003a))

7.3 Tillämpning av den designbaserade variansskattningsmodellen

Låt oss titta på mellanvaruvariensskattningen

$$\hat{V}_{PRO} = \sum_g \frac{1}{m_g(m_g - 1)} \sum_{i \in S_g^R} (1 - \pi_{gi}^R) \left\{ \sum_h \frac{v_{gh}}{\hat{X}_{gh}} \hat{e}_{i.}^{gh} \right\}^2$$

Ett problem då man vill minimera \hat{V}_{PRO} med avseende på m_g är att urvalssannolikheten π_{gi}^R , som är en funktion av m_g och M_g , är olika för varje urvalsobjekt i urvalet S_g^R .

Vi förlänger både täljaren och nämnaren med $(1 - \frac{m_g}{M_g})$. Anledningen till denna förlängning är att i genomsnitt är $\pi_{gi}^R = \frac{m_g}{M_g}$. Det medför att $(1 - \pi_{gi}^R) \approx (1 - \frac{m_g}{M_g})$, utom för riktigt stora objekt som har urvalssannolikheten nära 1,0.

ning är att i genomsnitt är $\pi_{gi}^R = \frac{m_g}{M_g}$. Det medför att $(1 - \pi_{gi}^R) \approx (1 - \frac{m_g}{M_g})$, utom för riktigt stora objekt som har urvalssannolikheten nära 1,0.

$$\Rightarrow \hat{V}_{PRO} = \sum_g \frac{1}{m_g(m_g - 1)} \sum_{i \in S_g^R} \frac{1 - \frac{m_g}{M_g}}{1 - \frac{m_g}{M_g}} (1 - \pi_{gi}^R) \left\{ \sum_h \frac{v_{gh}}{\hat{X}_{gh}} \hat{e}_{i.}^{gh} \right\}^2$$

Förkortar:

$$\Rightarrow \hat{V}_{PRO} \approx \sum_g \frac{1}{m_g(m_g - 1)} \sum_{i \in S_g^R} \left(1 - \frac{m_g}{M_g} \right) \left\{ \sum_h \frac{v_{gh}}{\hat{X}_{gh}} \hat{e}_{i.}^{gh} \right\}^2$$

$1 - \frac{m_g}{M_g}$ beror inte av i vilket gör att det kan flyttas utanför summeringen.

$$\Rightarrow \hat{V}_{PRO} \approx \sum_g \frac{1}{m_g} \left(1 - \frac{m_g}{M_g} \right) \underbrace{\frac{1}{(m_g - 1)} \sum_i \left\{ \sum_h \frac{v_{gh}}{\hat{X}_{gh}} \hat{e}_{i.}^{gh} \right\}^2}_{S_g^2} = \sum_g \left(\frac{1}{m_g} - \frac{1}{M_g} \right) S_g^2$$

Eftersom om vi inte har (sparat) information om M_g ersätts det med skattningen

$$\hat{M}_g = \sum_{i \in g} \frac{1}{\pi_{gi}^R}$$

\hat{M}_g är skattning av hur många varor som finns i urvalsramen, m_g är antalet varor som tagits med i urvalet och π_{gi}^R är sannolikheter. S_g^2 är ett mått på variationen, inklusive effekten av vägningstal, för produktgrupp g .

Ur programmet som skattar varianserna fås en storhet vbg . Sambandet mellan vbg och S_g^2 är följande:

$$vbg \approx \frac{1}{m_g} \left(1 - \frac{m_g}{\hat{M}_g}\right) S_g^2 \text{ och vi kan plocka ut}$$

$$\Rightarrow S_g^2 = m_g \frac{1}{1 - \frac{m_g}{\hat{M}_g}} vbg$$

Vid minimering av variansen är det intressanta:

$$\left(\frac{1}{m_g} - \frac{1}{\hat{M}_g}\right) S_g^2$$

↑ ↑

av "konstant"

intresse

det vill säga, vi vill minimera följande: $\sum_g \frac{1}{m_g} S_g^2$.

Låt oss titta på mellanbutiksvariansskattningen

$$\hat{V}_{BUT} = \sum_h \frac{1}{n_h(n_h - 1)} \sum_{j \in S_h^C} (1 - \pi_{hj}^C) \left\{ \sum_g \frac{v_{gh}}{\hat{X}_{gh}} \hat{e}_{.j}^{gh} \right\}^2$$

Ett problem då man vill minimera \hat{V}_{BUT} med avseende på n_h är att urvalssannolikheten π_{hj}^C , som är en funktion av n_h och \hat{N}_h , är olika för varje urvalsobjekt i urvalet S_h^C .

Vi förlänger både täljaren och nämnaren med $(1 - \frac{n_h}{N_h})$. Anledningen till denna förlängning

är att i genomsnitt är $\pi_{hj}^C = \frac{n_h}{N_h}$. Det medför att $(1 - \pi_{hj}^C) \approx (1 - \frac{n_h}{N_h})$, utom för riktigt stora objekt som har urvalssannolikheten nära 1,0.

$$\Rightarrow \hat{V}_{BUT} = \sum_h \frac{1}{n_h(n_h - 1)} \sum_{j \in S_h^C} \frac{1 - \frac{n_h}{N_h}}{1 - \frac{n_h}{N_h}} (1 - \pi_{hj}^C) \left\{ \sum_g \frac{v_{gh}}{\hat{X}_{gh}} \hat{e}_{.j}^{gh} \right\}^2$$

Förkortar:

$$\Rightarrow \hat{V}_{BUT} \approx \sum_h \frac{1}{n_h(n_h - 1)} \sum_{j \in S_h^C} \left(1 - \frac{n_h}{N_h}\right) \left\{ \sum_g \frac{v_{gh}}{\hat{X}_{gh}} \hat{e}_{.j}^{gh} \right\}^2$$

Vi har ett PPS-urval, det vill säga de n_h , butikerna, väljs med olika sannolikheter c. Förkortningen fungerar bra om alla π_{hj} är ganska små, till exempel $\leq \frac{1}{10}$. Då varierar $(1-\pi_{hj})$ inom

intervallet 0.90–1.0, det vill säga alla $(1-\pi_{hj})$ är relativt nära $1 - \frac{n_h}{N_h}$. I branscher där det

finns företag som väljs med säkerhet, π nära 1, kan denna förkortning leda till fel.

$1 - \frac{n_h}{N_h}$ beror inte av j vilket gör att det kan flyttas utanför summeringen.

$$\Rightarrow \hat{V}_{BUT} \approx \sum_h \frac{1}{n_h} \left(1 - \frac{n_h}{N_h}\right) \underbrace{\frac{1}{(n_h-1)} \sum_j \left(\sum_g \frac{v_{gh}}{\hat{X}_{gh}} \hat{e}_{.j}^{gh}\right)^2}_{S_h^2} = \sum_h \left(\frac{1}{n_h} - \frac{1}{N_h}\right) S_h^2$$

Analogt med produktsidan ersätter vi N_h med skattningen

$$\hat{N}_h = \sum_{j \in h} \frac{1}{\pi_{hj}^C}$$

\hat{N}_h är skattning av hur många butiker som finns i urvalsramen, n_h är antalet butiker som tagits med och π_{hj}^C är sannolikheter. S_h^2 är ett mått på variationen, inklusive effekten av vägningstal, för butikgrupp h.

Ur programmet som skattar variansen fås storheten vbh. Sambandet mellan vbg och S_h^2 är följande:

$$vbh \approx \frac{1}{n_h} \left(1 - \frac{n_h}{\hat{N}_h}\right) S_h^2 \text{ och vi kan plocka ut}$$

$$\Rightarrow S_h^2 = n_h \frac{1}{1 - \frac{n_h}{\hat{N}_h}} vbh$$

Vid minimering av variansen är det intressanta:

$$\left(\frac{1}{n_h} - \frac{1}{\hat{N}_h}\right) S_h^2$$

↑ ↑

av "konstant"

intresse

det vill säga, vi vill minimera följande: $\sum_h \frac{1}{n_h} S_h^2$.

Låt oss titta på interaktionsdelen

$$\hat{V}_{INT} = \sum_g \sum_h \frac{v_{gh}^2}{\hat{X}_{gh}^2} \frac{1}{m_g(m_g-1)} \frac{1}{n_h(n_h-1)} * \left\{ \sum_{i \in S_g^R} \sum_{j \in S_h^C} (1 - \pi_{gi}^R)(1 - \pi_{hj}^C) (\hat{e}_{ij}^{gh} - \hat{e}_{i.}^{gh} - \hat{e}_{.j}^{gh})^2 \right\}$$

Ett problem då man vill minimera \hat{V}_{INT} med avseende på m_g är att urvalssannolikheten

π_{gi}^R , som är en funktion av m_g och M_g , är olika för varje urvalsobjekt i urvalet S_g^R och

med avseende på n_h är att urvalssannolikheten π_{hj}^C , som är en funktion av n_h och N_h , är

olika för varje urvalsobjekt i urvalet S_h^C . Vi förlänger både täljaren och nämnaren med

$(1 - \frac{m_g}{M_g})$ och $(1 - \frac{n_h}{N_h})$. Anledningen till dessa förlängningar är att i genomsnitt är

$\pi_{gi}^R = \frac{m_g}{M_g}$ och $\pi_{hj}^C = \frac{n_h}{N_h}$. Det medför att för $(1 - \pi_{gi}^R) \approx (1 - \frac{m_g}{M_g})$ och att

$(1 - \pi_{hj}^C) \approx (1 - \frac{n_h}{N_h})$, utom för riktigt stora objekt som har urvalssannolikheter nära 1,0.

$$\Rightarrow \hat{V}_{INT} = \sum_g \sum_h \frac{v_{gh}^2}{\hat{X}_{gh}^2} \frac{1}{m_g(m_g-1)} \frac{1}{n_h(n_h-1)} * \left\{ \sum_{i \in S_g^R} \sum_{j \in S_h^C} \frac{1 - \frac{m_g}{M_g}}{1 - \frac{m_g}{M_g}} (1 - \pi_{gi}^R) \frac{1 - \frac{n_h}{N_h}}{1 - \frac{n_h}{N_h}} (1 - \pi_{hj}^C) (\hat{e}_{ij}^{gh} - \hat{e}_{i.}^{gh} - \hat{e}_{.j}^{gh})^2 \right\}$$

Förkortar:

$$\Rightarrow \hat{V}_{INT} \approx \sum_g \sum_h \frac{v_{gh}^2}{\hat{X}_{gh}^2} \frac{1}{m_g(m_g-1)} \frac{1}{n_h(n_h-1)} * \left\{ \sum_{i \in S_g^R} \sum_{j \in S_h^C} \left(1 - \frac{m_g}{M_g}\right) \left(1 - \frac{n_h}{N_h}\right) (\hat{e}_{ij}^{gh} - \hat{e}_{i.}^{gh} - \hat{e}_{.j}^{gh})^2 \right\}$$

$1 - \frac{m_g}{M_g}$ beror inte av i och $1 - \frac{n_h}{N_h}$ beror inte av j vilket gör att de kan flyttas utanför sum-

meringen. v_{gh}^2 och \hat{X}_{gh}^2 beror inte heller av i eller j vilket innebär att de kan flyttas in i sum-

$$\Rightarrow \hat{V}_{INT} \approx \sum_g \sum_h \frac{1}{m_g} \left(1 - \frac{m_g}{M_g}\right) \frac{1}{n_h} \left(1 - \frac{n_h}{N_h}\right) \frac{1}{(m_g-1)(n_h-1)} * \underbrace{\left\{ \sum_{i \in S_g^R} \sum_{j \in S_h^C} \frac{v_{gh}^2}{\hat{X}_{gh}^2} (\hat{e}_{ij}^{gh} - \hat{e}_{i.}^{gh} - \hat{e}_{.j}^{gh})^2 \right\}}_{S_{gh}^2} =$$

$$= \sum_g \sum_h \left(\frac{1}{m_g} - \frac{1}{M_g}\right) \left(\frac{1}{n_h} - \frac{1}{N_h}\right) S_{gh}^2$$

Eftersom om vi inte har (sparat) information om M_g och N_h ersätts de med skattningarna

$$\hat{M}_g = \sum_{i \in g} \frac{1}{\pi_{gi}^R}, \hat{N}_h = \sum_{j \in h} \frac{1}{\pi_{hj}^C}$$

\hat{M}_g är skattning av hur många varor och \hat{N}_h är skattning av hur många butiker och som finns i urvalsramen, m_g är antalet varor och n_h är antalet butiker som tagits med i urvalet,

π_{gi}^R och π_{hj}^C är urvalssannolikheter. S_{gh}^2 är ett mått på interaktionen (inklusive effekten av vägningstal) för produktgrupp g och butikgrupp h .

Ur programmet som skattar variansen fås $vbgh$. Sambandet mellan $vbgh$ och S_{gh}^2 är följande:

$$vbgh \approx \frac{1}{m_g} \frac{1}{n_h} \left(1 - \frac{m_g}{\hat{M}_g}\right) \left(1 - \frac{n_h}{\hat{N}_h}\right) S_{gh}^2 \text{ och vi kan plocka ut}$$

$$\Rightarrow S_{gh}^2 = m_g \frac{1}{1 - \frac{m_g}{\hat{M}_g}} n_h \frac{1}{1 - \frac{n_h}{\hat{N}_h}} vbgh$$

Vid minimering av variansen är det intressanta:

$$\left(\frac{1}{m_g} - \frac{1}{\hat{M}_g}\right)\left(\frac{1}{n_h} - \frac{1}{\hat{N}_h}\right) S_{gh}^2 = \left(\frac{1}{m_g} \cdot \frac{1}{n_h} - \frac{1}{\hat{N}_h} \cdot \frac{1}{m_g} - \frac{1}{\hat{M}_g} \cdot \frac{1}{n_h} + \frac{1}{\hat{M}_g} \cdot \frac{1}{\hat{N}_h}\right) S_{gh}^2$$

+ + - + - + - -

+ är av intressen, - är ”konstant”.

Empiriskt vet vi att skattningarna av interaktionstermerna (\hat{V}_{INT}) är små jämfört med skattningarna av mellanvaru- och mellanbutikstermerna (\hat{V}_{PRO} och \hat{V}_{BUT}). Urvalsfraktionerna är små i de flesta strata i såväl varu- som butiksurvalen ($m_g \ll M_g$ och $n_h \ll N_h$).

Med denna motivering vill vi minimera följande: $\sum_g \sum_h \frac{1}{m_g} \frac{1}{n_h} S_{gh}^2$

Det är en del av variansskattningen som fås fram som skall minimeras, här finns n_h och m_g i termerna:

$$V \approx \sum_{g \in G} \frac{1}{m_g} S_g^2 + \sum_{h \in H} \frac{1}{n_h} S_h^2 + \sum_{g \in G} \sum_{h \in H} \frac{1}{m_g} \frac{1}{n_h} S_{gh}^2$$

8 Allokering

Kapitlet består av en redogörelse för allokeringsproblemet, en del om optimal allokering och en del om Lagranges multiplikatorsats. Kapitlet avslutas med en presentation om systematiskt och slumpmässigt sökarbete.

8.1 Allokeringsproblemet

För att ge ett urval en optimal urvalsallokering behövs, dels en variansfunktion och dels en kostnadsfunktion. Då vi ska finna det optimala urvalet måste vi ha värden på kostnaden och variansen i dag. Kostnadsfunktionens komponenter beräknas i kapitel 6, Kostnadsstudien, variansen skattas i kapitel 7, Variansen i KPI.

8.2 Optimal allokering

Minimera variansen med bivillkor att kostnad skall vara som den hittills har varit. Vi känner ej den ”sanna” variansen så vi minimerar en skattning av den.

$$\begin{aligned} \text{Minimera } V &= \sum_{g \in G} \frac{1}{m_g} S_g^2 + \sum_{h \in H} \frac{1}{n_h} S_h^2 + \sum_{g \in G} \sum_{h \in H} \frac{1}{m_g} \frac{1}{n_h} S_{gh}^2 \\ \text{då } C_0 &= \sum_{h \in H} n_h C_h + \sum_{h \in H} \sum_{g \in G} n_h m_g C_{hg} + \sum_{g \in G} m_g C_g \end{aligned}$$

S_h^2 är ett uttryck för skattad varians för butikstratum h , n_h är urvalet av butiker i stratum h , S_g^2 är ett uttryck för skattad varians för produktgrupp g , m_g är urvalet av varor i produktgrupp g , S_{gh}^2 är ett mått på interaktionen. C_0 är den totala kostnaden för KPI, C_h är kostnaden som finns för varje butik och C_{hg} är kostnaden för en vara i en butik.

$C_{hg} = 0$ för omöjliga kombinationer. (Ett exempel på en omöjlig kombination är kläder i en möbelaffär).

Idag mäts det priser på 23 stycken produkter i klädeshandeln, 21 stycken produkter i möbelbutiker och 28 stycken restaurangtjänster. Av praktiska skäl för att inte butiksarbetet skall bli för omfattande för intervjuare och butiksägare har vi infört maxgränser för antal produkter att prismäta. Det är inte möjligt för en intervjuare att prismäta alltför många priser i en butik. Intervjuarna menar att personalen i mindre butiker (inte stora kedjor) inte vill att intervjuarna ska vara inne i butiken så länge då det av kunderna kan uppfattas som obekvämt att en person går runt i affären och tittar och gör anteckningar om produkterna. Vi kommer välja maxgränserna till 30 kläder (produkter), 30 möbler (produkter) samt 30 restauranger (produkter). Anledningen till gränserna är dels att inte arbetet ska bli för betungande för intervjuaren samt att det inte ska störa butiksägaren alltför mycket. Vi har också en lägsta gräns på tre butiker och två produkter. Dessa väljs för att man ska kunna räkna varianser.

Det vi är intresserade att finna är alltså de värden på n_h och m_g som ger min V .

8.3 Systematiskt och slumpmässigt sökarbete

Med Lagrange multiplikatorsats klarar vi inte av att hitta lösningar på grund av ytterliggare restriktioner som ska läggas på urvalet, såsom minst två varor på samma ställe. Då det inte går att lösa problemet via Lagrange multiplikatorsats kan det vara lämpligt att låta en dator finna optimum med stort sökarbete. Tre modeller av simuleringar har utförts, två ”systema-

tiska” och en ”slumpmässig”. Programmen gjordes i SAS och hade till uppgift att minimera variansen.

I det systematiska programmet ”Systematisk addera” adderas tillfälligt 0.1 enheter i tur och ordning till varje n_h och varje m_g , en i taget. Sedan undersöks vilken ökning av butik- och varuurlösen som ger minst varians. För det butiksstratat eller produktgruppen adderas 0.1 och övriga behåller sitt ursprungsvärde. Med hjälp av kostnadsfunktionerna justeras alla n_h och m_g litet nedåt så att kostnaden på nytt stämmer med budget. Dessa värden på n_h och m_g blir de nya startvärdena. Proceduren upprepas tills stabilitet uppnåtts, det behövdes 2000 gånger. Det systematiska programmet ”Systematisk subtrahera” utförs på samma sätt med undantag för att det istället dras bort 0,1.

I det slumpmässiga programmet ”Slumpmässig” adderas ett slumpantal som halverar, fördubblar eller något där emellan urvalet för varje n_h och m_g . Kostnadsfunktionerna används även här för att proportionellt justera n_h och m_g så att budgeten går ihop.

$$n_h = \max(n_{hmin}, (n_h * (0,5 + slump)))$$

$$n_g = \max(n_{gmin}, (n_g * (0,5 + slump)))$$

slump är ett slumpantal mellan 0 och 1 (även 0 och 1 ingår), $n_h * (0,5 + slump)$ innebär att antalet butiker kan minska med 50% (det vill säga halveras) eller öka med 50% eller något däremellan, på samma sätt för produkter.

Jag kommer att använda mig av alla tre metoderna. Programmet ”Systematisk subtrahera” skall dock enbart vara ett stöd för att påvisa att man genom ”Systematisk addera” kommer fram till en trovärdig lösning, man kommer alltså hitta samma lösningar från två olika håll.

9 Resultat

I detta kapitel kommer resultaten från rapporten presenteras.

För att ge överskådligare figurer i detta kapitel kommer produkterna ingå i produktgrupper som vi uttrycker enligt följande:

- Möbel 1 – Matbord/Stol (Stratumgrupp 310)
- Möbel 2 – Säng/Hyllor och skåp/Madrass (Stratumgrupp 320)
- Möbel 3 – Fåtölj/Soffa (Stratumgrupp 330)
- Mat 1 – Rödvin/Starköl/Snaps/Mellanöl (Stratumgrupp 710)
- Mat 2 – Öl klass två/Läsk (Stratumgrupp 712)
- Mat 3 – Pizza/Snabbmat (Stratumgrupp 720)
- Mat 4 – Middag/Lunch à la carte (Stratumgrupp 731)
- Mat 5 – Dagens lunch (Stratumgrupp 732)
- Mat 6 – Kaffe/Kaffebröd (Stratumgrupp 733)
- Kläder 1 – Strumpbyxor/Nattlinne (Stratumgrupp 930)
- Kläder 2 – Damkläder (Stratumgrupp 940)
- Kläder 3 – Herrhandskar/Herrskinnjacka (Stratumgrupp 945)
- Kläder 4 – Jeans/Herrjoggingdress (Stratumgrupp 950)
- Kläder 5 – Herrkläder (Stratumgrupp 960)
- Kläder 6 – Barnkläder (Stratumgrupp 970)

9.1 Indata

Resultaten från tidsstudien sammanfattas i tabellen nedan:

Tabell 3
Sammanställning av resultaten

	C_{ng} (kr/not och år) (produkt)	C_h (kr/butik och år) (butik)	C_g (kr/vara) (representantvara)
Kläder⁸	390.3	4252.5	6240
Möbler	165.7	4102.6	140.4
Restaurang⁹	72.7	1400.4	140.4

Notera att kostnaden för representantvara (C_g) uppkommer endast då man vill ha fler representantvaror än de som finns för närvarande.

9.2 Vad är viktigast att KPI mäter samt alternativa beräkningar av varianser

Vid arbete med KPI måste man ställa sig frågan om vad som är viktigast att KPI mäter. Varianser kan beräknas på olika vis beroende på vad som man avser att mäta. Nedan följer några påstående om vad som är viktigast att KPI mäter och några alternativ hur varianser kan skattas. Efter det skrivs de olika sakerna som KPI kan mäta var och en för sig och sedan beskrivs hur varianserna ska skattas i dessa fall.

Viktigast att KPI mäter:

- Indextal med 1980 som bastal.
- Inflationstakten¹⁰, viktig i ekonomisk analys, riksbanken använder till exempel denna form.

⁸ Kläder innefattar prismätning av kläder i varuhus, stormarknader, klädbutiker och sporthandlar.

⁹ Prismätning i restaurang täcker såväl vanliga restauranger som personalmatsalar.

¹⁰ Förändring av KPI på tolv månaders sikt.

- Förändringen i inflationstakten mellan två på varandra följande månader, exempelvis använder riksbanken även denna form.

Alternativa beräkningar för varianser:

1. För varje korttidslänk (K), 2001–maj 2003.
2. Medelvärde för korttidslänkar månad 7–12 år 2001–2002.
3. För månatliga förändringar, varje månad 2001–maj 2003, genom att förra månaden ersätter egentliga basen (december år $y-1$).

Notera att i figurerna som följer i kapitel nio kommer jag ange vilken av dessa alternativ för beräkning av varianser som används.

Indextal med 1980 som bastal.

Bästa urvalsallokering är densamma som för beräkning av inflationstakten, se inflationstakten.

Inflationstakten (tolvmånadersförändring)

Vi har gjort antagandet att det är bättre att ta ett medelvärde av varianser för till exempel alla månader juli–december, än att bara välja variansen för december, för att få ett stabilt beslutsunderlag att göra allokeringar. Den enda månad då inflationstakten är lika med en korttidslänk är i december. Om man vill ha en länk från exempelvis maj år $y-1$ till maj år y baseras den länken på två korttidslänkar; en maj–december år $y-1$ och en december–maj år y .

Förändring av inflationstakten

Kan approximeras med att man tar bort en månadsförändring och lägger till en ny månadsförändring. Om man exempelvis vill se förändringen mellan april 2001 och april 2002 och har länken mellan mars 2001 och mars 2002, tar man bort månadsförändringen för mars–april 2001 och lägger till månadsförändringen mars–april 2002.

Den enda ”äkta” månadsförändringen är december–januari. För att få ”nästan äkta” månads-länkar för övriga månader byter jag ut baspriset ut och förra månadens pris stoppas in.

Variansen beräknas för varje korttidslänk (K), 2001–maj 2003 men används inte i slutliga allokeringen.

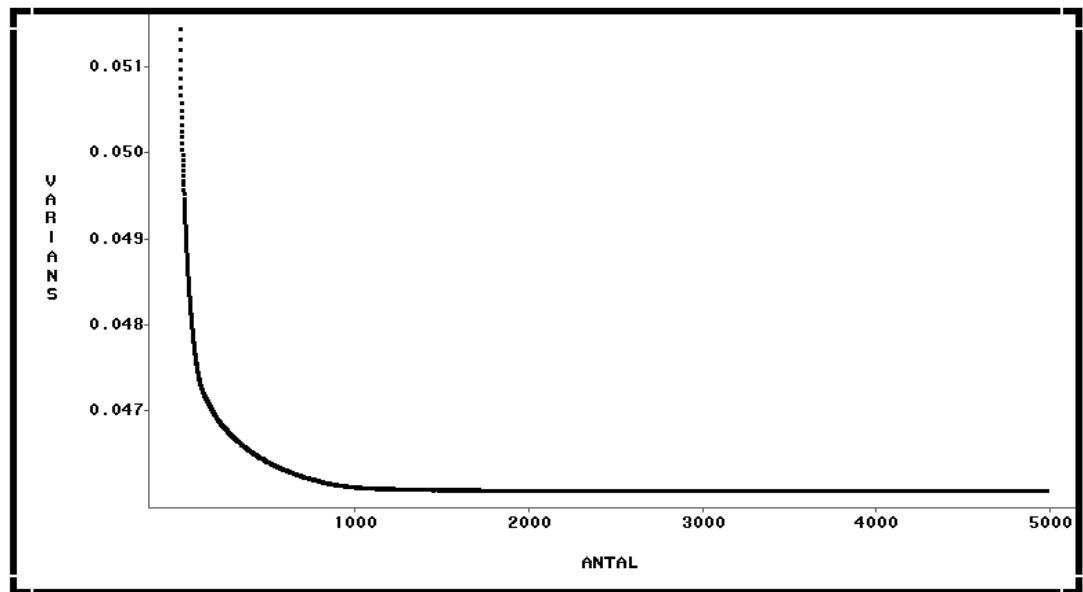
9.3 Resultat från allokeringen

I delavsnitten presenteras resultaten som kommer från SAS-programmen systematisk addera, systematisk subtrahera samt slumpmässig.

9.3.1 Resultat från systematisk addera

För de två följande figurerna finns det gränser (största tillåtna värdet) för produkter (m_g) av kläder, möbler och restaurangtjänster. Gränsen är sammanlagt 30 representantvaror för såväl kläder, möbler som restaurangtjänster. Anledningen till att vi satt gränser är som tidigare nämnts av praktiska skäl.

Figur 10
Fördelning av varianser för systematisk

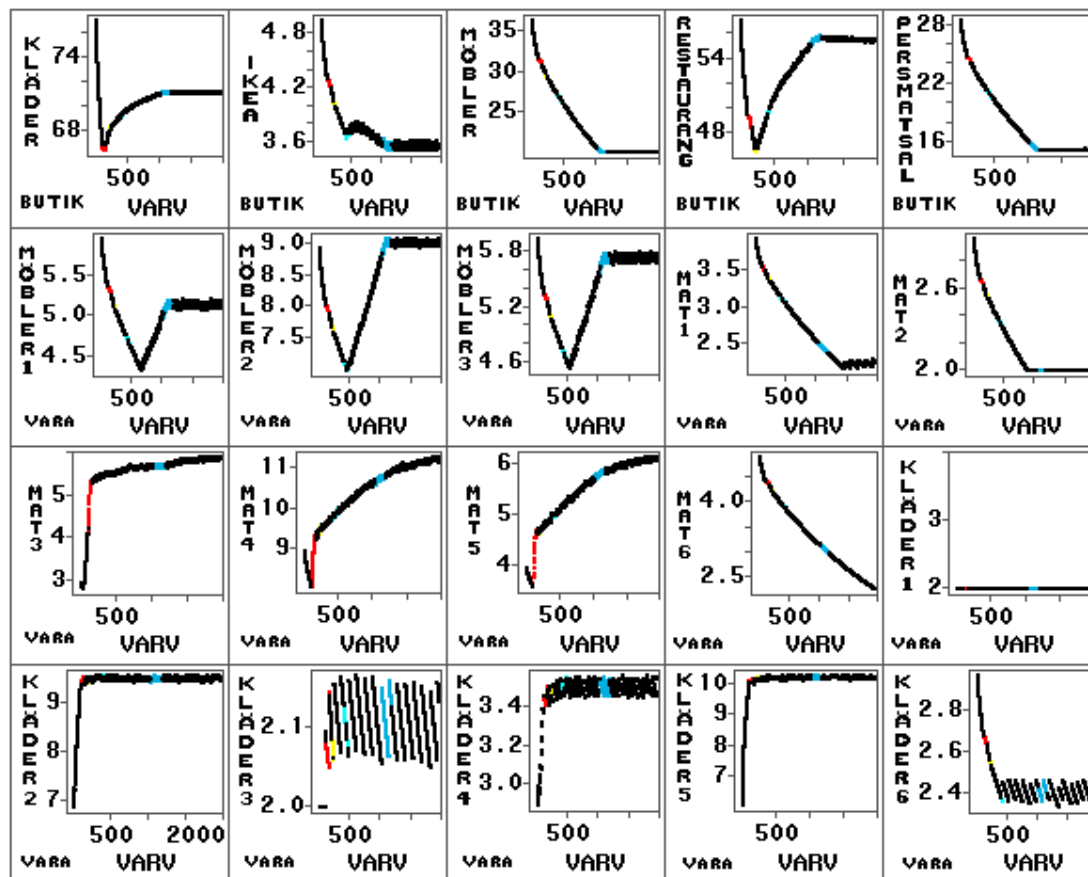


Bilden visar hur variansen sjunker med antal gånger man upprepar programmet som minimerar variansen. Vi kommer i fortsättningen välja antal¹¹ = 2000 eftersom värdet är stabilt där. Ur figuren kan man utläsa att variansen minskar från 0,0515 till 0,0461, med andra ord är det en minskning med 10%.

¹¹ Antal gånger programmet körts.

Figur 11

Antalet butiker och produkter för varv=0 till varv=2000 då det finns restriktioner på det maximala antalet produkter av kläder, möbler och restaurangtjänster. Varians enligt alternativ 2



Tabell 4

Antal butiker/produkter (restriktioner för det maximala antalet produkter i varje butik finns)

	Kläder (butik)	Ikea (butik)	Möbler (butik)	Restauranger (butik)	Permsatsal (butik)	Möbler 1 (vara)	Möbler 2 (vara)
Initialskede	77.4	5.0	36.9	57.4	29.0	6.0	9.0
Efter allokering	71.3	3.5	20.1	55.6	15.4	5.2	9.1
	Möbler 3 (vara)	Mat 1 (vara)	Mat 2 (vara)	Mat 3 (vara)	Mat 4 (vara)	Mat 5 (vara)	Mat 6 (vara)
Initialskede	6.0	4.0	3.0	3.0	9.0	4.0	5.0
Efter allokering	5.8	2.3	2.0	5.9	11.4	6.2	2.3
	Kläder 1 (vara)	Kläder 2 (vara)	Kläder 3 (vara)	Kläder 4 (vara)	Kläder 5 (vara)	Kläder 6 (vara)	
Initialskede	2.0	7.0	2.0	3.0	6.1	3.0	
Efter allokering	2.0	9.6	2.1	3.6	10.3	2.4	

Ur figuren 11 (med stöd av tabell 4) kan följande utläsas:

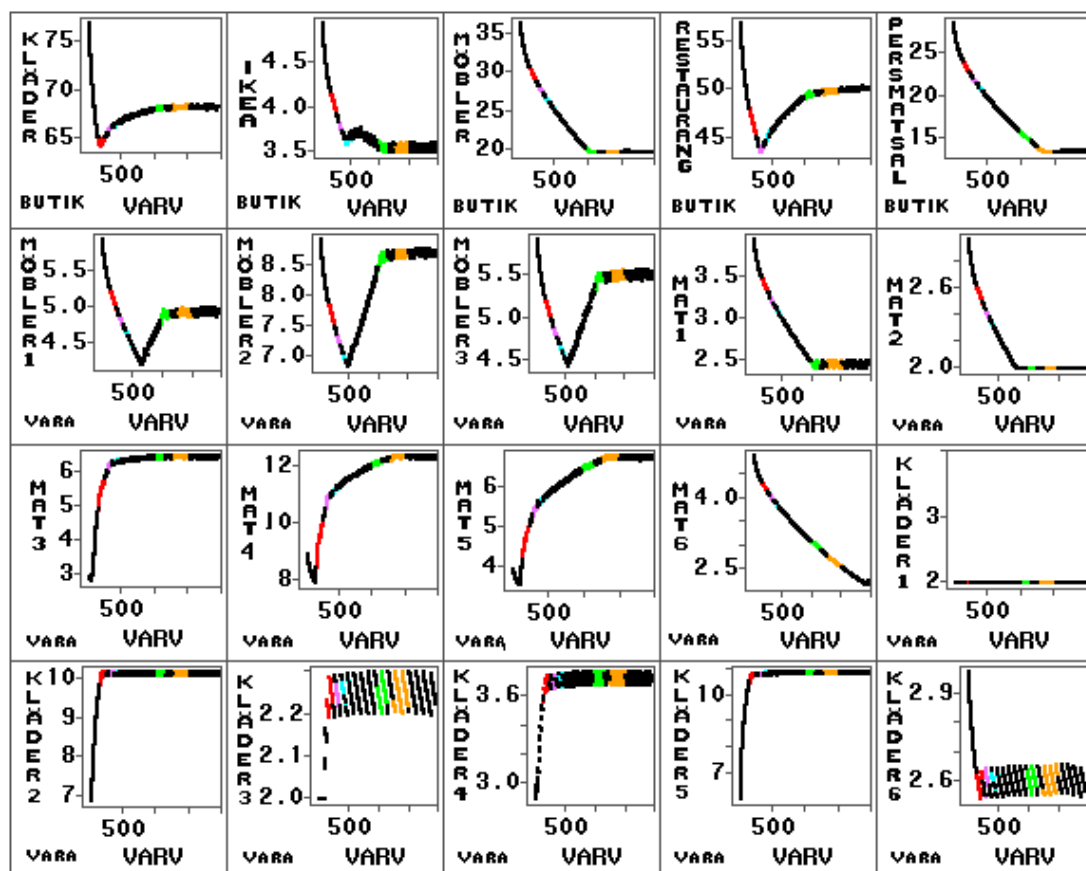
- Knät¹² i bilden över antal klädbutiker, minsta antalet klädbutikerna (66 stycken), uppkommer då summan av antalet kläder (bilderna kläder 1 – kläder 6) når övre gränsen för produkter i klädesbutiker vilken är 30 stycken. Då gränsen för produkter nåtts (det vill säga att det inte går att välja fler klädesprodukter) ökas antalet klädesbutiker igen. Det kan tolkas som att klädesundersökningen måste förstärkas ytterligare även då varuurlvalet nått maximum, vilket beror på att variansen för kläder är så stor.
- Knät i bilden över antal Ikea-butiker (3 stycken) uppkommer då summan av antalet möbler (produkter) är 17. Det är alltså inte begränsningen på 30 produkter som medför att knät i bilden uppkommer.
- Knät i bilden över antal möbelbutiker (20 stycken) (möbelbutiker förutom Ikea) uppkommer då summan av antalet möbler är 20. Inte heller här kan man påvisa att det är begränsningen på 30 produkter som medför att knät i bilden uppkommer.
- Knät i bilden över antal restauranger (46 stycken) uppkommer då summan av antalet restaurangtjänster (mat 1- mat 6) når övre gränsen som är 30 stycken. Då gränsen för produkter nåtts (det vill säga att det inte går att välja fler restaurangtjänster) ökas antalet restauranger igen.
- Knät i bilden över antal personalmatsalar¹³ (15 stycken) uppkommer då summan av antalet restaurangtjänster når övre gränsen som är 30 stycken. Då gränsen för produkter nåtts (det vill säga att det inte går att välja fler restaurangtjänster) så är antalet personalmatsalar konstant.
- Sammanfattningsvis visar figurerna att antalet butiker ska minskas. Man kan inte säga något generellt för alla produkter eller för alla produkter inom en grupp (kläder, möbler och restaurangtjänster). För sofliga produkter ska urvalet ökas, för andra ska det minskas och för några ska urvalet vara konstant.

¹² Knät uppkommer då antalet butiker går från att minska till att öka.

¹³ Personalmatsalar ingår i gruppen restauranger.

Figur 12

Antalet butiker och produkter. Det finns inte några restriktioner på det maximala antalet produkter av kläder, möbler och restaurangtjänster. Varians enligt alternativ 2



Tabell 5
Antal butiker/produkter

	Kläder (butik)	Ikea (butik)	Möbler (butik)	Restauranger (butik)	Persmatsal (butik)	Möbler 1 (vara)	Möbler 2 (vara)
Initialskede	77.4	5.0	36.9	57.4	29.0	6.0	9.0
Efter allokering	68.0	4.0	20.0	51.0	14.0	4.9	8.7
	Möbler 3 (vara)	Mat 1 (vara)	Mat 2 (vara)	Mat 3 (vara)	Mat 4 (vara)	Mat 5 (vara)	Mat 6 (vara)
Initialskede	6.0	4.0	3.0	3.0	9.0	4.0	5.0
Efter allokering	5.6	2.5	2.0	6.5	12.5	6.8	2.3
	Kläder 1 (vara)	Kläder 2 (vara)	Kläder 3 (vara)	Kläder 4 (vara)	Kläder 5 (vara)	Kläder 6 (vara)	
Initialskede	2.0	7.0	2.0	3.0	6.1	3.0	
Efter allokering	2.0	10.3	2.2	3.7	11.0	2.6	

Ur figur 12 och tabell 5 kan utläsas:

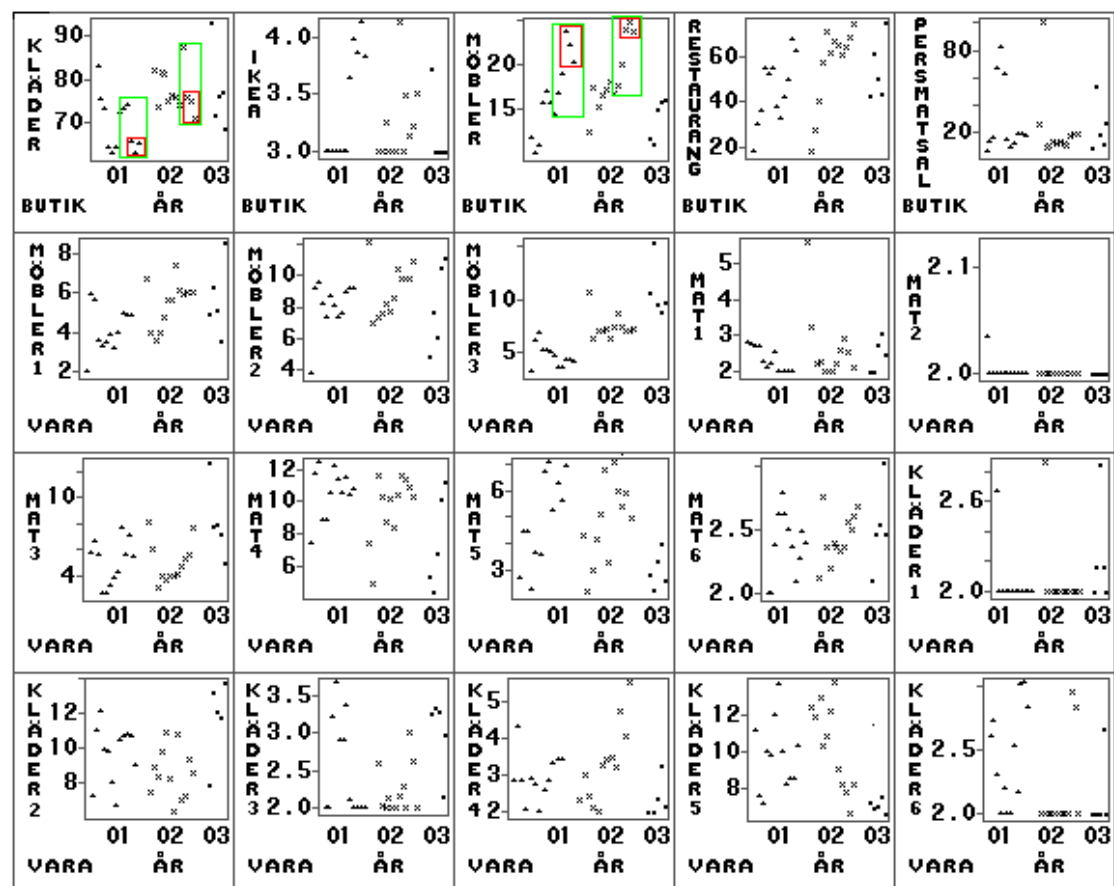
- Knät i bilden över antal klädbutiker, minsta antalet klädbutikerna (65 stycken), uppkommer då summan av antalet kläder är runt 31.
- Knät i bilden över antal Ikea-butiker (3 stycken) uppkommer då summan av antalet möbler (produkter) är 19.
- Knät i bilden över antal möbelbutiker (20 stycken) (möbelbutiker förutom Ikea) uppkommer då summan av antalet möbler är mellan 18 och 20.
- Knät i bilden över antal restauranger (44 stycken) uppkommer då summan av antalet restaurangtjänster ligger mellan 31 till 33.
- Knät i bilden över antal personalmatsalar¹⁴ (13 stycken) uppkommer då summan av antalet restaurangtjänster ligger mellan 32 och 34.
- Sammanfattningsvis visar figurerna att antalet butiker ska minskas. Man kan inte säga något generellt för alla produkter eller för alla produkter inom en grupp (kläder, möbler och restaurangtjänster). För somliga produkter ska urvalet ökas, för andra ska det minskas och för några ska urvalet vara konstant.

Vid en jämförelse mellan de båda figurerna 11 och 12 kan utläsas att det inte är någon större skillnad på vart brytpunkten är för de värde där antalet butiker börjar öka istället för minska (det vi kallat knät). Detta gäller för såväl klädesbutiker och möbelbutiker (Ikea och övriga möbelbutiker) som för restauranger (restauranger och personalmatsalar).

¹⁴ Personalmatsalar ingår i gruppen restauranger.

Figur 13

Optimala antal butiker och varor för skattning för varje korttidsindexlänk ($K_{y-1,12}^{y,m}$) under 2001, 2002 och januari till maj 2003. Varians enligt alternativ 1



De olika symbolerna (\blacktriangle , \times och \blacksquare) står för varsitt år, \blacktriangle för januari till december 2001, \times för januari till december 2002 och \blacksquare för januari till maj 2003. För att skapa denna bild jämförs prisläget för varje månad med december året innan för att få fram det antal butiker och varor inom de olika stratumgrupperna som är bäst. Notera att indexlänkarnas längd varierar mellan 1 och 12 månader, det vill säga för första månaden, januari, är indexlänken 1 och för december är månadslänken 12 månader lång.

För möbelbutiker och restauranger syns ett tydligt mönster att optimala antalet butiker ökar under ett år, det vill säga att antalet butiker bör vara fler i slutet av året än i början av året. För kläder ska urvalet vara stort redan i början av året. Det skulle kunna bero på att omsättningshastigheten på kläder är extremt hög. I januari finns januarirean med dramatiska prisändringar. Därför blir variansen för prisändringar i klädeshandeln stor redan efter någon månad. Mot slutet av året kan urvalen av butiker och varor minskas för att frigöra resurser till större urval inom möbelhandel och restaurangerna.

Slutvärdet i figur 11 är medelvärdet av månad 7–12 år 2001 och 7–12 år 2002 i figur 13. Teoretiskt borde vi hålla oss till decemberlänken eftersom den sammanfaller med en inflationstakt. Men vi trodde att osäkerheten var stor i variansskattningen så vi tog medelvärdet av varianserna för juli till december. Här kan vi ana att vårt antagande inte var helt bra. Skulle nog varit bättre att tagit medelvärdet av de tre sista månaderna (oktober–december) varje år. För kläder blir medelvärdet av juli–december (de är inringade med den större rutan) högre än medelvärdet för oktober–december (de är inringade med den mindre rutan). För möbler blir medelvärdet av juli–december (de är inringade med den större rutan) lägre än medelvärdet för oktober–december (de är inringade med den mindre rutan).

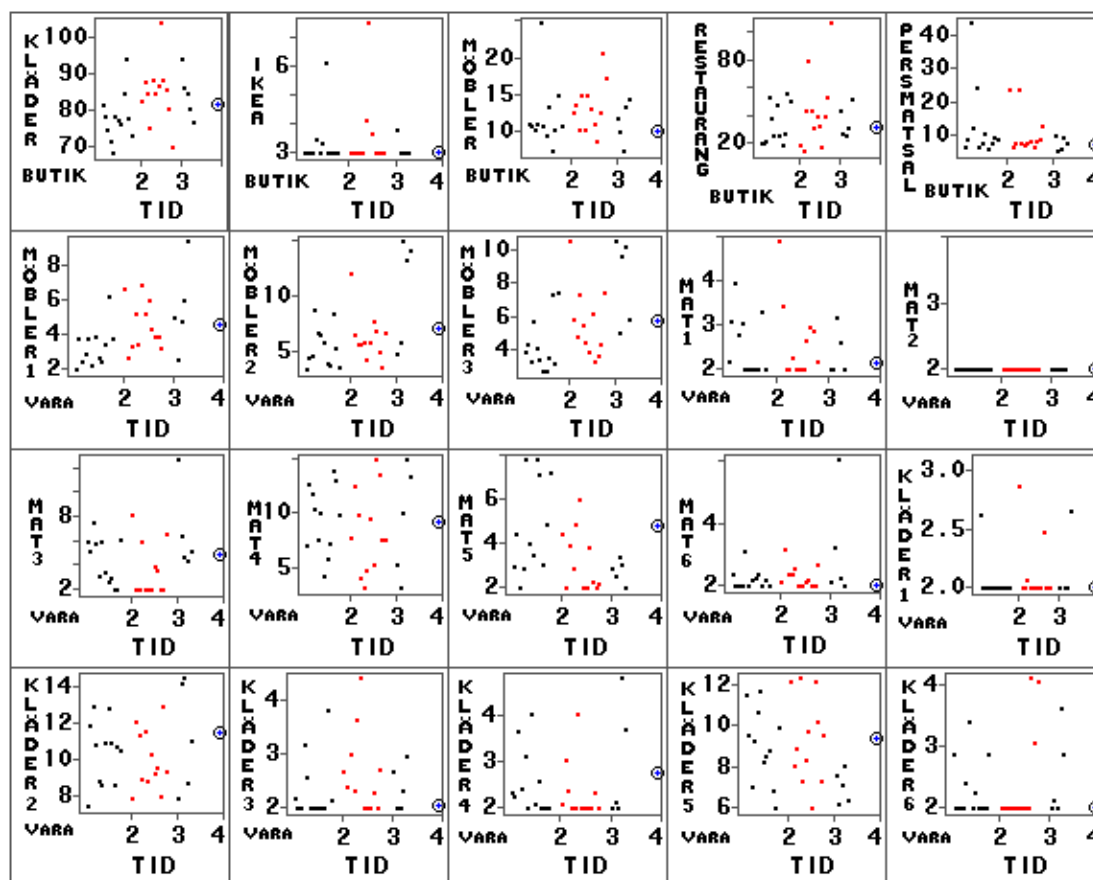
Tabell 6
Standardavvikelser för butiker och varor i genomsnitt för första respektive andra halvåret 2001–2003

År	Halvår	Månad	Kläder (butik)	Ikea (butik)	Möbler (butik)	Restauranger (butik)	Persmatsal (butik)
.	1	.	0.91769	0.000266	0.013008	0.040452	0.001742
.	2	.	1.13827	0.000882	0.043053	0.12023	0.005226
Möbler 1 (vara)	Möbler 2 (vara)	Möbler 3 (vara)	Mat 1 (vara)	Mat 2 (vara)	Mat 3 (vara)	Mat 4 (vara)	Mat 5 (vara)
0.000170	0.000562	0.000601	0.000138	0.000027	0.000566	0.001618	0.000260
0.000629	0.002186	0.000890	0.000247	0.000116	0.001786	0.004521	0.001323
Mat 6 (vara)	Kläder 1 (vara)	Kläder 2 (vara)	Kläder 3 (vara)	Kläder 4 (vara)	Kläder 5 (vara)	Kläder 6 (vara)	
0.000020	0.001276	0.061949	0.004317	0.004067	0.067496	0.001468	
0.000065	0.000483	0.037826	0.002070	0.007615	0.063777	0.002712	

Jämförelse av standardavvikelse mellan de båda halvåren, halvår 1 är januari–juni och halvår 2 är juli–december. För andra halvåret ökar standardavvikelsen för samtliga butiker och varor förutom för kläder 1, kläder 2, kläder 3 och kläder 5. Detta innebär att det är viktigt att titta på variansen för butikerna och varorna under andra halvåret om man vill optimera KPI:s urval för att skatta årlig förändring hellre än förändring mellan närliggande månader.

Figur 14

Bilder över hur det optimala antalet butiker och varor varierar över tiden när man vill mäta månadsförändring. Varians enligt alternativ 3



Jämför med månaden innan. De korsen som är inringade är medelvärdet över de 29 månaderna av antal butiker eller varor.

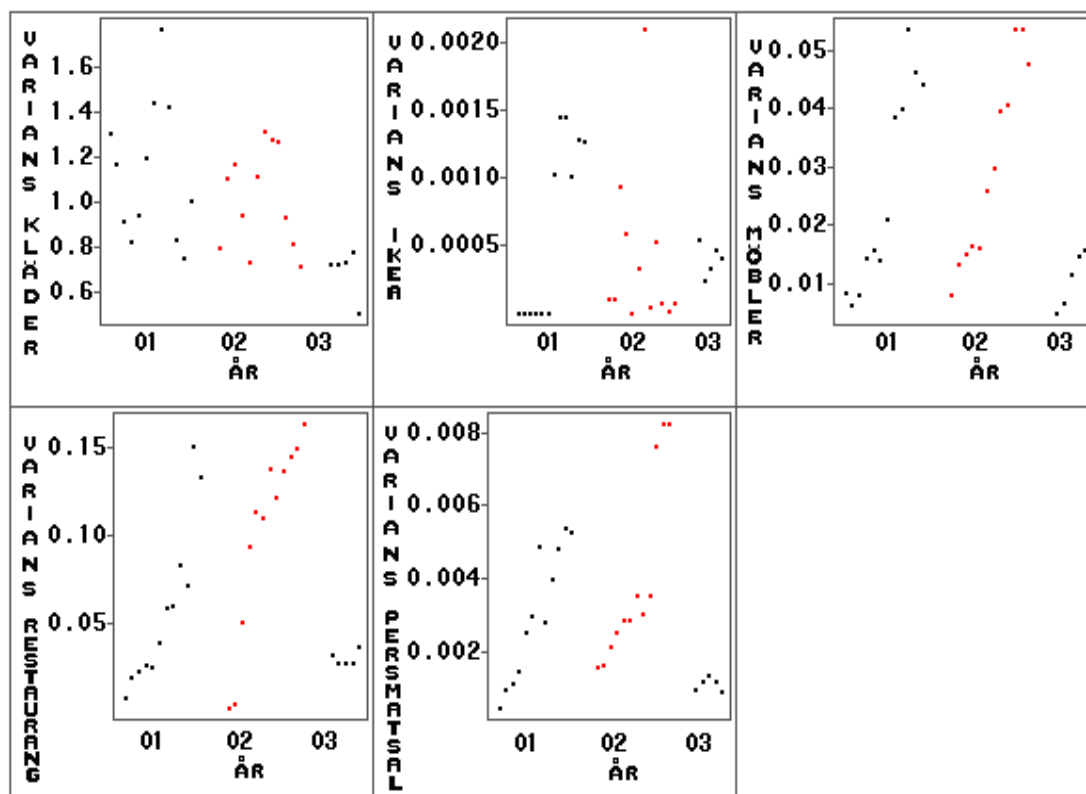
Vid beräkning av KPI jämför man den aktuella månaden med december föregående år, i denna figur har jag gjort jämförelse av den aktuella månaden mot föregående månad.

Vid en jämförelse mellan årslänkar (systematisk med gräns)¹⁵ och månadslänkar¹⁶ kan man se att man måste ta fler klädesbutiker för att få bästa skattningar av förändring av inflationen än när man vill skatta inflationens nivå. Både antalet möbelbutiker och antalet restauranger finns med i urvalet i lägre utsträckning vid månadslänkning än vad de gör i indexlänkning. I appendix K har resultaten sammanställts.

¹⁵ Jämför varje månad med december föregående år, det finns också gränser för hur många varor av kläder, möbler och mat som får mätas.

¹⁶ Varje månads prisförändring jämförs med månaden innan, även här finns gränser på varorna.

Figur 15
Varianser för butiksstratum. Varians enligt alternativ 1



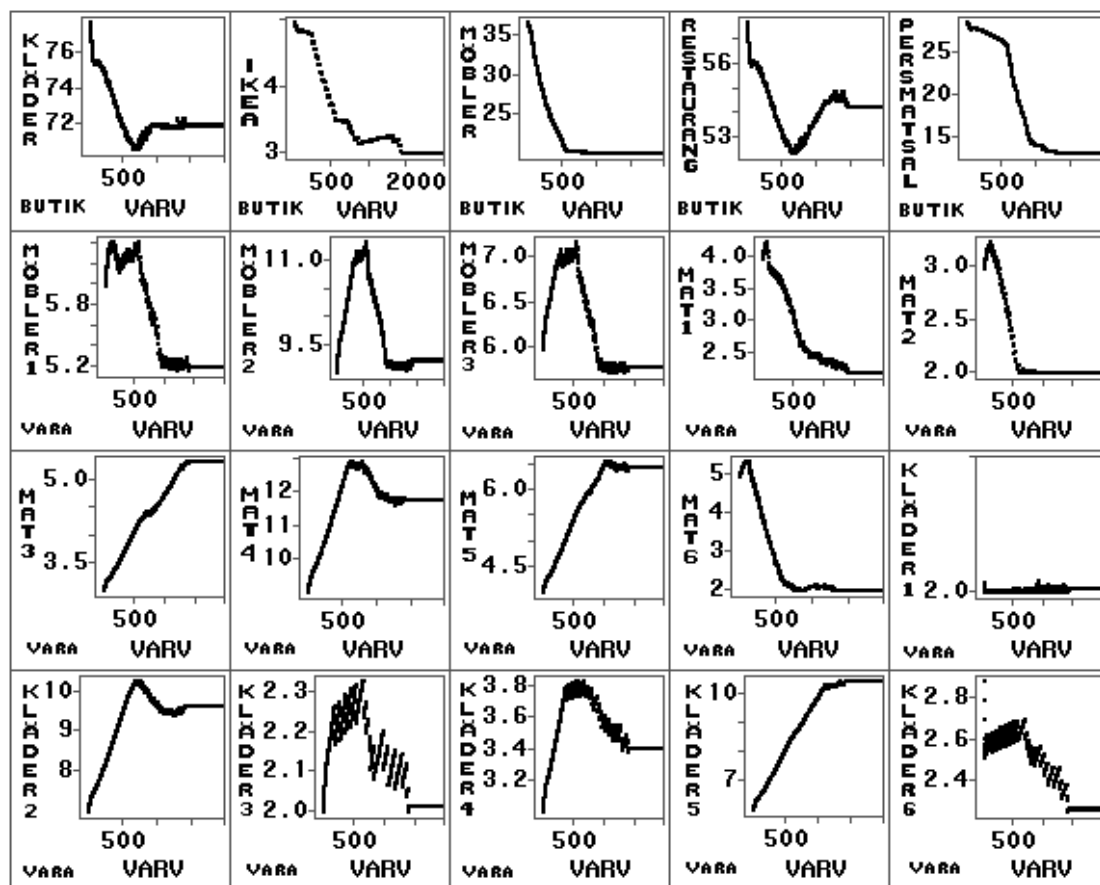
Jag har färgat variansen för 2002 för att på så vis göra så att bilden blir enklare att titta i¹⁷.

Ur figurerna kan läsas att variansen för kläder är stor redan i början av året, så är inte fallet för övriga. Speciellt kan man utläsa att för möbler, restauranger och personalmatsalar stiger variansen för varje månad. Av skalan på y-axeln ser man att variansen är betydligt större för kläder än för övriga butiker, för Ikea och personalmatsalar är den mycket liten. Detta innebär (återigen) att det är viktigt att allokera med varianser för 7–12 månaders korttidsindexlänk om man prioriterar inflationsnivån och varianser för månadsförändringar om man prioriterar förändringen av inflation.

¹⁷ Om det är en svartvit kopia på uppsatsen kan man om man tittar noga se att för år 2002 är punkterna annorlunda mot 2001 och 2003.

9.3.2 Resultat från systematisk subtrahera

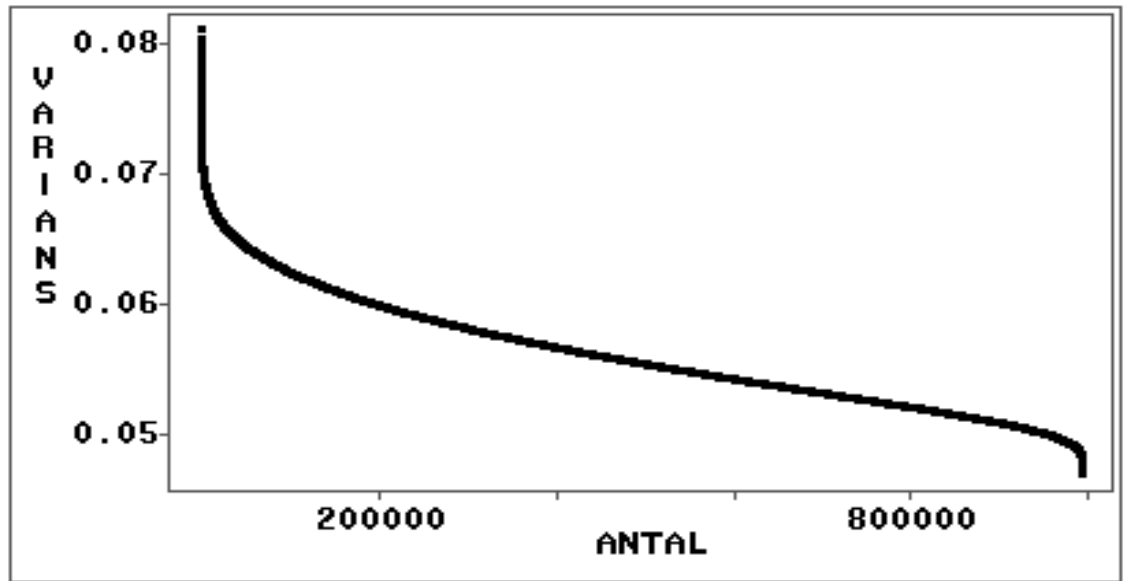
Figur 16
Antalet butiker och produkter



Programmet är detsamma som systematisk med undantag för att istället för att lägga till 0.1 drar man bort 0.1. Anledningen till att programmet görs är för att se om man kommer fram till liknande resultat om man stegar fram på ett annat sätt. Efter studerande och jämförelse av ”slutvärdet” i figurerna 11 och 16 kan man utläsa att de stämmer väldigt bra överens. Detta innebär att vi stöds i antagandet att vår metod är trovärdig.

9.3.3 Resultat från slumpmässig

Figur 17
Fördelning av varianser för slumpmässig

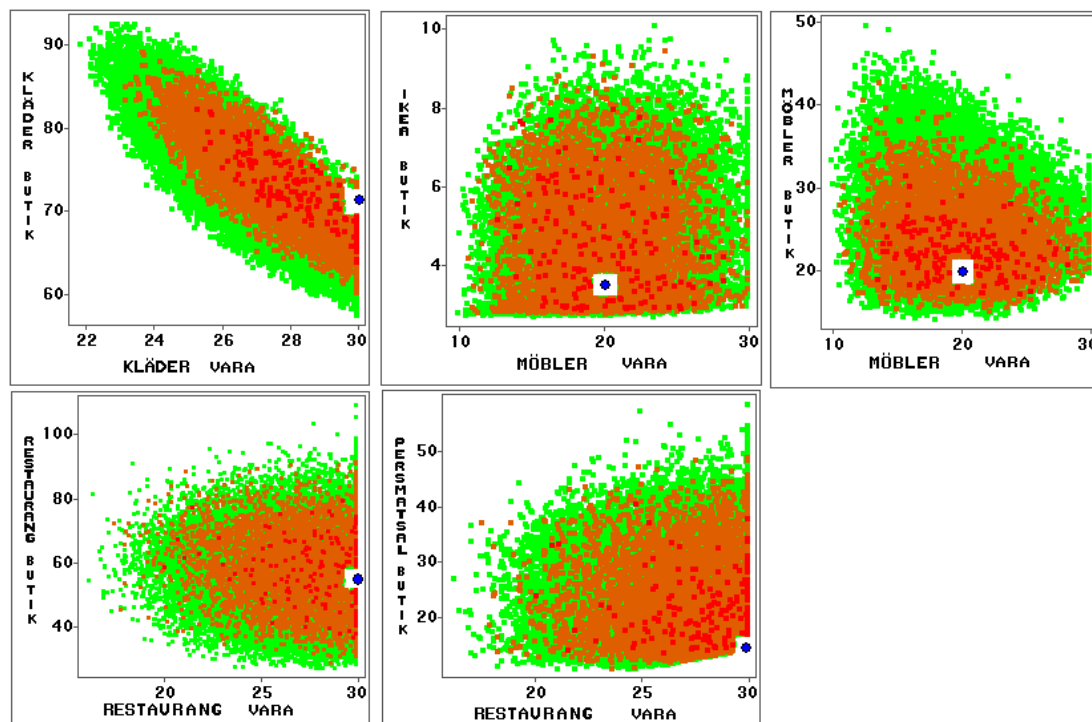


Programmet körs 1000000 gånger (antal). Gränserna på antal produkter som man kan prismäta i en och samma butik har, som tidigare nämnts, satts. De är 30 för kläder, 30 för möbler och 30 för restauranger.

Den sämsta variansen som genereras av det slumpmässiga programmet var på 0,0816 och den bästa har var 0,0470.

Med de systematiska programmen kunde man befara att i synnerhet min- och maxgränser i kombination med små steg (0.1 urvalsenheter) kunde orsaka att programmet skulle "fastna" på ett lokalt minimum. Den slumpmässiga metoden tar jättekälv, $\pm 50\%$, men det har ändå inte hittat ett bättre min-värde. Jag blir därför övertygad om att jag hittat det globala minimumet.

Figur 18
Urvalsstorlekar som ger acceptabla varianser (varians < 0.05¹⁸)



X-variabeln, kläder vara, står för summan av alla varor som prismäts i klädbutiker. X-variabeln, möbler vara, för summan av alla varor som mäts i Ikea och övriga möbelaffärer. Slutligen x-variabeln, restaurang vara, står för summan av alla all mat och dryck som prismäts i restauranger och personalmatsalar. Anledningen till att en summering av respektive varugrupper görs är för att studera sambandet mellan antalet klädbutiker och det antalet varor som prismäts i klädbutiker. Samma sak gäller för möbler och restauranger. Ringarna i de ljusare rutorna som är markerade i figurerna står för det optimala n_h (antalet butiker) och det optimala Σm_g (summan produkter). Siffrorna fås ur appendix K. De röda prickarna i diagrammen är för de värden då variansen < 0.048, de orangea är då variansen < 0.049 och slutligen gröna då variansen < 0.050.

Ur figuren kan man läsa att man kan välja butiker och Σ varor enligt följande, observera dock att det måste vara ett riktigt samband mellan butiker och Σ varor. Man kan välja 60–90 klädesbutiker och Σ varor av kläder kan vara 23–30 stycken. Då det gäller Ikea kan man välja 3–8 butiker och Σ varor kan vara 10–30 stycken. För möbler kan man välja 15–40 butiker och Σ varor kan vara 10–30 stycken. Det går att välja 30–90 restauranger (butiker) och Σ varor för restauranger ska vara 20–30 stycken. För personalmatsalar kan man välja 15–45 butiker och Σ varor ska vara 20–30 stycken.

Optimala antalet butiker och Σ varor (ringarna i de ljusa rutorna) är: För klädesbutiker ska det vara 72 stycken och Σ varor ska vara 30 stycken. Det ska vara 4 Ikea-butiker och Σ varor ska vara 20. Då det gäller möbler ska det vara 20 butiker och Σ varor ska även den vara 20. Det ska vara 56 restauranger och Σ varor är 30. För personalmatsalar ska det vara 15 butiker och Σ varor ska vara 30.

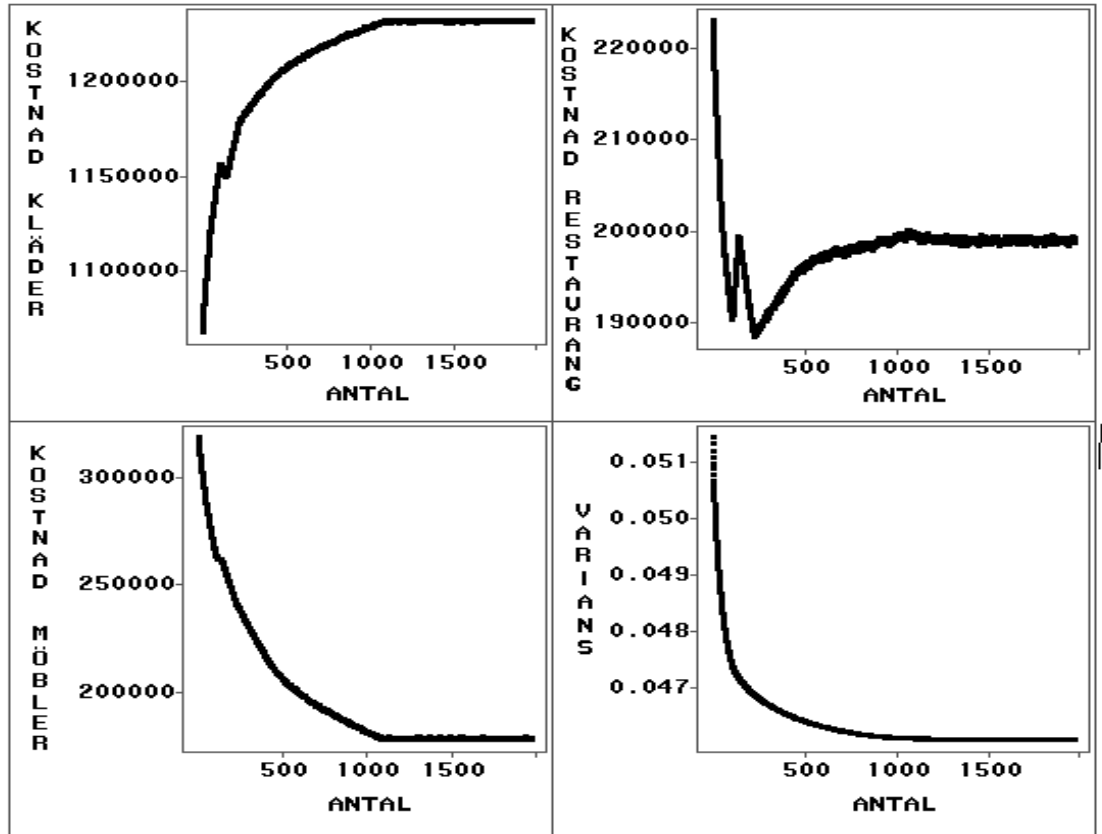
Av figuren kan utläsas att det finns ett negativt samband mellan antal klädesbutiker och antal varor i klädbutiken. Man kan alltså till viss del godtyckligt välja fler butiker och färre varor eller vice versa. För möbelbutiker och restauranger har man större valmöjlighet då man ska välja en kombination av antal butiker och antal produkter för prismätning. Ett exempel visar det genom att väljer man att mäta kläder i 80 butiker måste man hålla sig inom ramen att

¹⁸ Enligt figur 17 är den optimala variansen 0.0471.

mäta mellan 23 och 28 varor. För övriga figurer har man ett större valmöjlighet, det vill säga väljer man 4 Ikea butiker kan man välja att ta mellan 10 och 30 varor för prismätning.

En sak som är värt att notera är att begränsningarna på möbler inte ger något utslag; bilden har på inget tydligt sätt "kapats" då det är få punkter till höger i diagrammen. För restauranger och kläder har bilden "kapats", med andra ord att många värden är otillåtna, det vill säga hamnar utanför gränsen på 30 butiker.

Figur 19
Kostnad per bransch innan, under och efter körningarna



Idag används 1050000 kronor för kläderundersökningen men vi har visat att den undersökningen borde kosta 1250000 kronor medan de andra skall göras 200000 billigare.

10 Slutsats

Kapitlet består av de slutsatser som analysen mynnar ut i. Den är uppdelad på så vis att först presenteras slutsatserna från den systematiska modellen och sedan slutsatserna från den slumpmässiga modellen.

För den systematiska modellen med gräns:

- Variansen sjunker med antal gånger man upprepar programmet som minimerar variansen. Variansen minskar från 0,0515 till 0,0461, det vill säga en minskning med 10%. Efter 2000 upprepningar är värdet på variansen stabil.
- Den övre gränsen för kläder (produkter) och restaurangtjänster som båda är 30 stycken ger utslag, det vill säga då man inte får öka antalet produkter av kläder respektive restaurangtjänster måste antalet klädesbutiker respektive restauranger ökas istället. Inget sådant samband finns för möbler (produkter)
- Antalet klädesbutiker, möbelbutiker och restauranger ska alla minskas. Man kan inte säga något generellt för alla produkter eller för alla produkter inom en grupp (kläder, möbler och restaurangtjänster). För somliga produkter ska urvalet ökas, för andra ska det minskas och för några ska urvalet vara konstant. Denna punkt gäller både då modellen har gräns och då den inte har gräns.
- Vid en jämförelse mellan den systematiska modellen med gräns och systematiska modellen utan gräns (figurerna 11 och 12) kan man se att det inte är någon större skillnad på var brytpunkten för de värde där antalet butiker börjar öka istället för minska är (det vi kallat knät). Detta gäller för såväl klädesbutiker och möbelbutiker (Ikea och övriga möbelbutiker) som för restauranger (restauranger och personalmatsalar).
- För möbelbutiker och restauranger syns ett tydligt mönster att optimala antalet butiker ökar under ett år, det vill säga att antalet butiker bör vara fler i slutet av året än i början av året. För klädesbutiker ska urvalet vara stort redan i början av året. Det skulle kunna bero på att omsättningshastigheten på kläder är extremt hög. I januari finns januarirean med dramatiska prisändringar. Det ska vara 60–90 klädesbutiker, 3–4 Ikea, 10–25 möbelbutiker, 20–80 restauranger och 20–100 personalmatsalar.
- Nedan följer några påstående om vad som är viktigast att KPI mäter och sedan beskrivs hur varianserna ska mätas i dessa fall.

– Inflationstakten

Vi har gjort antagandet att det är bättre att ta ett medelvärde av sex månaders varianser, än att bara välja variansen för december, för att få ett stabilt beslutsunderlag att göra allokeringar. Kortidslänken går från december till december. Om man vill ha en länk från exempelvis maj till maj baseras den länken på flera kortidslänkar.

Vill man optimera KPI:s urval för att skatta årlig förändring är det är viktigt att titta på varianser för butiker och varor under andra halvåret¹⁹, det vill säga varianser beräknas som medelvärde för kortidslänkar månad 7–12 år 2001–2002.

Då man ska mäta inflationstakten bör man ta antal butiker och produkter enligt kolumnen ”indexlänk, juli–december 2001 och 2002 (systematisk med gräns)” i appendix K. Dessa presenteras som vita prickar i figur 18 i kapitel 9.

– Förändring av inflationstakten

Förändring av inflationstakten kan approximeras med att man tar bort en månadsförändring och lägger till en ny månadsförändring. Om man exempelvis vill se förändringen mellan

¹⁹ Juli till December.

april 2001 och april 2002 och har länken mellan mars 2001 och mars 2002, tar man bort månadsförändringen för mars–april 2001 och lägger till månadsförändringen mars–april 2002.

Om man prioriterar förändringen av inflation ska man beräkna varianser för månadsförändringar. Den enda ”äkta” månadsförändringen är december–januari, för att få ”nästan äkta” månadslänkar för övriga månader byts baspriset ut och förra månadens pris stoppas in. Vi har kunnat beräkna varianser för varje månad 2001–maj 2003.

Skulle man vara intresserad av förändringen av inflationstakten mellan två på varandra följande månader ska man välja antal butiker och produkter enligt ”månadslänk (systematisk)” i appendix K. Dessa presenteras även, som inringade plus, i figur 14 i kapitel 9.

- Vid en jämförelse mellan indexlänkar (systematisk med gräns)²⁰ och månadslänkar²¹ kan man se att man måste ta fler klädesbutiker för att få bästa skattningar av förändring av inflationen än när man vill skatta inflationens nivå. Både antalet möbelbutiker och antalet restauranger finns med i urvalet i lägre utsträckning vid månadslänkning än vad de gör i indexlänkning. I appendix K har resultaten sammanställts.
- Variansen för kläder är stor redan i början av året, så är inte fallet för övriga. Speciellt kan man utläsa att för möbler, restauranger och personalmatsalar stiger variansen för varje månad. Variansen är betydligt större för kläder än för övriga butiker, för Ikea och personalmatsalar är den mycket liten. Detta innebär (återigen) att det är viktigt att allokera med varianser för 7–12 månaders korttidsindexlänk om man prioriterar inflationsnivån och varianser för månadsförändringar om man prioriterar förändringen av inflation.

För den slumpmässiga modellen:

- Den sämsta variansen som genereras av det slumpmässiga programmet var på 0,0816 och den bästa har var 0,0470.
- Det finns ett negativt samband mellan antal klädesbutiker och antal varor i klädbutiken. Man kan alltså till viss del godtyckligt välja fler butiker och färre varor eller vice versa. För möbelbutiker och restauranger har man större valmöjlighet då man ska välja en kombination av antal butiker och antal produkter för prismätning. En sak som är värt att notera är att begränsningarna på möbler inte ger något utslag; bilden har på inget tydligt sätt ”kapats” då det är få punkter till höger i diagrammen. För restauranger och kläder har bilden ”kapats”, med andra ord att många värden är otillåtna, det vill säga hamnar utanför gränsen på 30 butiker.
- Idag används 1 050 000 kronor för klädesundersökningen men vi har visat att den undersökningen borde kosta 1 250 000 kronor medan de andra skall göras 200 000 billigare.

²⁰ Jämför varje månad med december föregående år, det finns också gränser för hur många varor av kläder, möbler och mat som får mätas.

²¹ Varje månads prisförändring jämförs med månaden innan, även här finns gränser på varorna.

Referenser

Litteratur

Blom, G. and Holmquist, H., (1998), *Statistikteori med tillämpningar*, Studentlitteratur, Lund

Rosén, B., (2002), *Teori och praktik för Urvalsundersökningar*, Studentlitteratur, Uppsala

Dahmström, K., (1991), *Från datainsamling till rapport – att göra en statistisk undersökning*, Studentlitteratur, Lund

Granquist, L. and Arvidson, G. and Elffors, C. and Norberg, A. and Lundell, L., (2003), *Guide till granskning*, SCB-tryck, Örebro

SCB, Ekonomisk statistik, Priser, (2001), *The Swedish Consumer Price Index – A handbook of methods*, SCB-tryck, Örebro

Artiklar

Dalén, J. and Ohlsson, E., (1995), Variance Estimation in the Swedish Consumer Price Index, *Journal of Business & Economic Statistics* 13, 347–356

Norberg, A., (2003a), *Egenskaper hos ett antal variansestimater för skattning av korttidsindexlänken i KPI – En simuleringsstudie*, utkast från SCB

Internet

Ribe, M., *Följa konsumentpriserna*, www.scb.se, 030217a

Ribe, M., *Enkelt om KPI*, www.scb.se, 030217b

Ribe, M., *KPI i korttid och långtid*, www.scb.se, 030217c

SCB:s hemsida, www.scb.se, 030212

SCB:s hemsida, www.scb.se, 030226

Handelshögskolan vid Göteborgs universitets hemsida, www.gu.se, 030505

Åbo akademi datacentrals hemsida, www.abo.fi, 030226

Intranät

SCB:s intranät, 030303

”Internt” material från SCB:

SCB, Ekonomisk statistik, Priser, *KPI – Intervjuarnas instruktionspärm*, SCB-tryck, Stockholm

Norberg, A., (2002), *OH-bilder Prisprogrammet – konsumentprisindex*, SCB

Norberg, A., (2003b), *Projektbeskrivning*, SCB

Norberg, A., (2003c), *Formler för indexaggregat*, SCB

Appendix A

Tabell 1
Granskning av CFARNR (=butiknummer)

Obs	CFARNR	OrdNr	UVNR	IVE	SNI92	AntNotFeb2003
1	13751513	175				.
2	15199786	226				.
3	15216294	17				.
4	22176959	446				.
5	22229491	601				.
6	27427852	176				.
7	27578830	.	27578830	14633		.
8	29538925	313				.
9	29711650	383				.
10	37704004	.				9
11	38148654	183				.

Appendix B

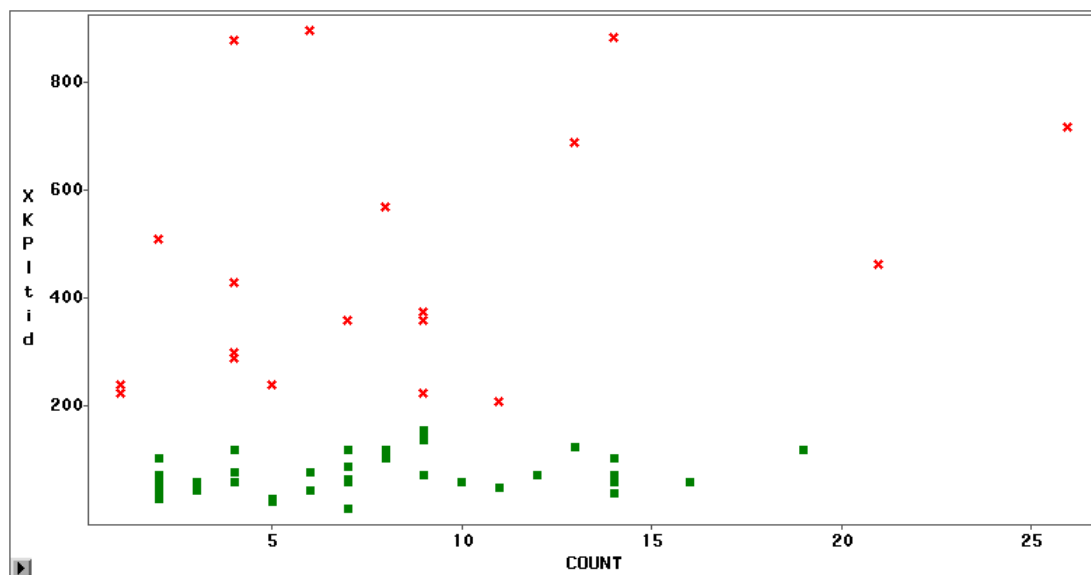
Tabell 1

Granskning av CFARNR (=butiknummer) efter genomgång

Obs	CFARNR	OrdNr	UVNR	IVE	SNI92	AntNotFeb2003
1	27578830	.	27578830	14633		.
2	37704004	.				9

Appendix C

Figur 1
Gemensam tid för KPI arbete per butik



XKPItid står för sammanlagd tid för KPI-arbete som är mer eller mindre gemensam för alla butikerna såsom till exempel postöppning, planering, paketering och återsändning. Count står för antal butiker som intervjuarna gjort prismätningar i. Varje punkt är en intervjuare. De röda kryssen är sådana som vi tycker oss kunna påstå att de inte kan vara vettiga. De gröna rutorna är de värden som vi behåller och kommer att göra våra beräkningar på.

Appendix D

Tabell 1
Stratum indelade i stratumgrupper

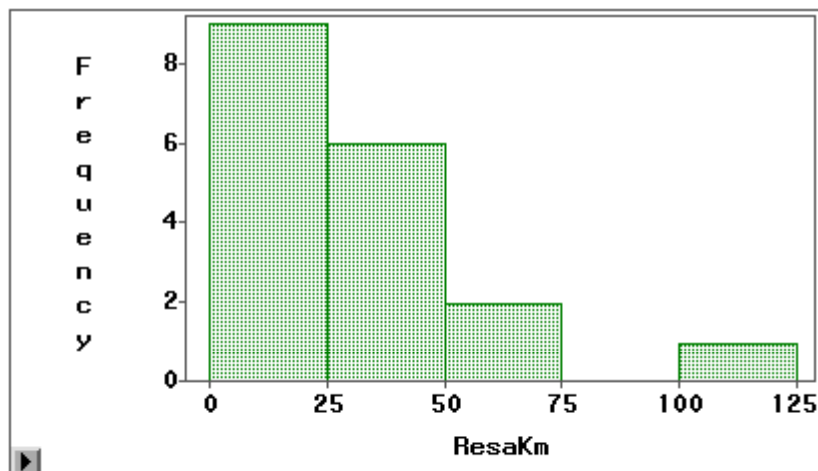
Obs	Stratum	Stratum Grupp	BT	BRANSCH
1	523100	.		APOTEK
2	501020	1	B	BILHANDEL
3	503020	1	B	BILRESERVDELAR
4	515300	3	B	BYGGVARUHUS
5	524610	3	B	BYGGVAROR
6	521111	4	B	VARUHUS
7	521112	5	B	STORMARKNADER
8	521120	5	B	LIVSMEDEL
9	522600	6	T	TOBAK
10	522710	6	T	HÅLSOKOST
11	523300	6	T	PARFYM
12	524530	6	T	SKIVOR
13	524620	6	T	FÄRG
14	524710	6	T	BOK OCH PAPPER
15	524840	6	T	GULDSMED
16	524880	6	T	ZOO
17	524950	6	T	GOLV
18	524210	8	B	KLÄDER
19	524411	9	B	MÖBLER
20	524419	9	B	MÖBLER
21	524420	10	B	HEMTEXTIL
22	524430	10	B	BOSÄTTNING
23	524510	10	B	VITVAROR
24	524860	10	B	LEKSAK
25	524520	11	B	RADIO TV
26	502010	12	T	BILVERKSTAD
27	502040	12	T	GUMMIVERKSTAD
28	524100	12	T	SYBEHÖR
29	524320	12	T	VÅSKOR
30	524440	12	T	LAMP
31	524540	12	T	MUSIKINSTR.
32	524810	12	T	OPTIKER
33	524820	12	T	FOTO
34	524830	12	T	UR
35	524990	12	T	BARNVAGN
36	524850	15	B	SPORT
37	524870	16	T	BLOMMOR
38	524930	17	B	DATOR
39	524940	18	T	TELE
40	551000	20	T	HOTELL
41	553000	21	T	REST
42	555100	21	T	PERSONALMATSAL
43	527100	22	T	SKOMAKARE
44	527200	22	T	REPARATÖR
45	714010	22	T	FILMUTHYRNING
46	714020	22	T	TV-UTHYRNING
47	852000	22	T	VET.
48	930120	22	T	TVÄTTERI

Obs	Stratum	Stratum Grupp	BT	BRANSCH
49	930210	22	T	FRISÖR
50	524310	23	B	SKOR

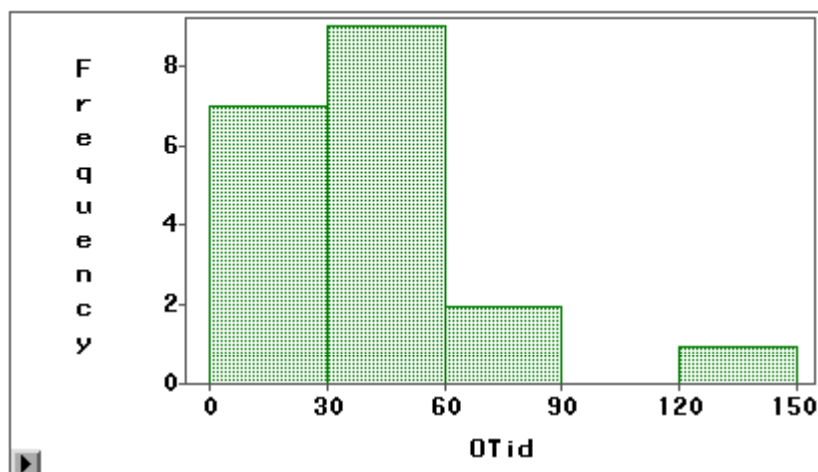
Tabellen fås ur program 3.

Appendix E

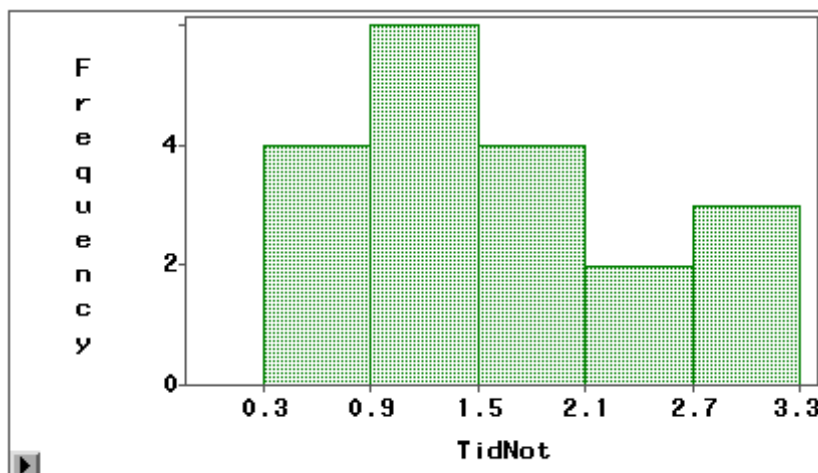
Figur 1
Stratumgruppernas fördelning över genomsnittligt antal kilometer per butik



Figur 2
Stratumgruppernas fördelning över genomsnittlig restid.

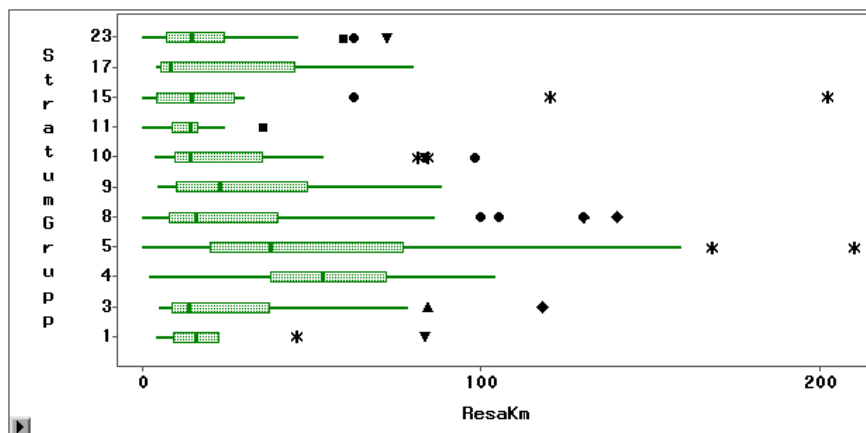


Figur 3
Stratumgruppernas fördelning över tid per notering



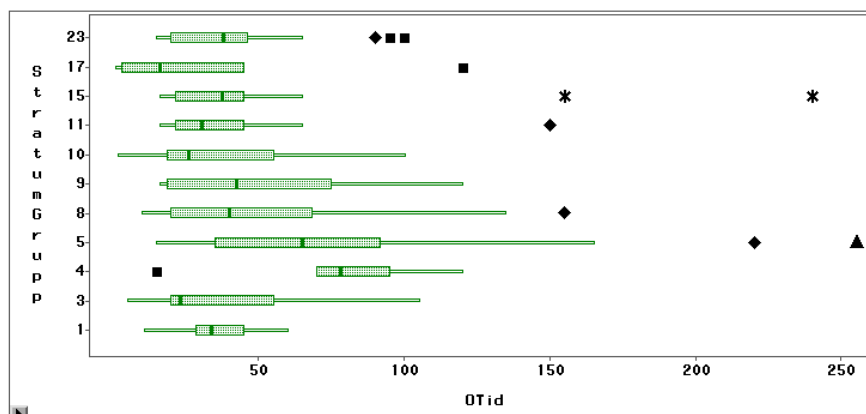
Appendix F

Figur 1
Resa i kilometer för stratumgrupp vid besök i butik



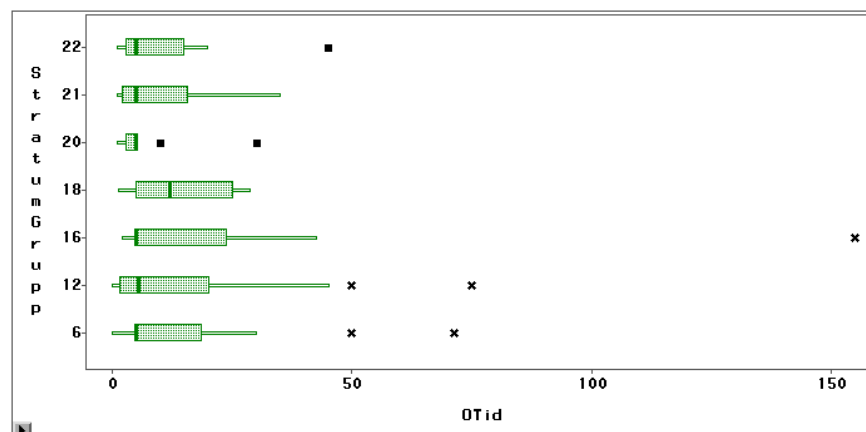
* Norrland, Dalarna & Värmland, ◆ Småland, Δ Blekinge, ● Västergötland, + Stockholm, ▽ Södermanland. Markering som är längst till vänster i stratumgrupp 10 kan vara svårt att utläsa då det är tre punkter som ligger tätt; det är två stycken * och en ▽. Markering nummer två från höger i stratumgrupp 8 är två punkter som ligger tätt, det är ett + och en ●.

Figur 2
Övrig tid för stratumgrupper vid besök i butik



* Norrland, Dalarna & Värmland, ◆ Småland, Δ Blekinge

Figur 3
Övrig tid för stratumgrupper vid telefonintervju



x rest till butik trots att det är en butik som man ska ringa till.

Appendix G

Tabell 1

Medelvärde av resa i kilometer och övrig tid samt prisinsamlingstid per notering

SG	SGNamn	FREQ	Resa Km	Resa KmN	OTid	OTidN	Pris Ins	Ant NotF	PrisTid/Not	BT
20	Hotell	14	.	0	6.5000	14	99.0	164	0.60366	T
22	Skomakare/Reparatör/ Film-uthyrning/TV-uthyrning/Veterinär/Tvätter/Frisör	26	1.2667	3	8.6857	21	65.0	47	1.38298	T
21	Restaurang/Personalmatsal	57	8.1988	10	10.2529	48	311.0	564	0.55142	T
6	Tobak/Hälsokost/Parfym/Skivor/Färg/Bok&Papper/Guldsmed/Zoo/Golv	56	10.7761	17	11.6059	50	384.0	258	1.48837	T
12	Bilverkstad/Gumverkstad/Sy behör/Väskor/Lampor/Musikinstrument/Optiker/Ur/Barnvagn	61	19.9181	17	13.5378	52	499.0	306	1.63072	T
18	Tele	11	7.4167	5	14.6970	11	105.0	88	1.19318	T
16	Blommor	32	22.0944	12	17.1532	31	306.3	361	0.84848	T
17	Dator	7	25.1667	4	30.2381	7	80.0	31	2.58065	B
1	Bilhandel/Bilreservdelar	10	23.9556	9	35.6759	9	218.0	108	2.01852	B
10	Hemtextil/Bosättning/Vitvaror/Leksaker	38	26.5012	34	35.8982	35	827.0	261	3.16858	B
3	Byggvaruhus/Byggvaror	24	29.2868	20	38.5667	21	694.0	449	1.54566	B
11	RadioTV	14	13.8590	14	40.5833	14	1173.0	493	2.37931	B
23	Skor	26	19.9144	26	41.8100	25	1017.0	335	3.03582	B
8	Kläder	69	29.9199	69	48.6036	66	5244.0	1657	3.16476	B
9	Möbler	24	31.4019	23	49.7224	22	1375.0	662	2.07704	B
15	Sport	19	31.2352	18	53.3284	17	880.0	339	2.59587	B
4	Varuhus	5	53.8667	5	75.6667	5	600.0	434	1.38249	B
5	Stormarknader/Livsmedel	43	53.8597	43	76.9915	43	8890.0	12929	0.68760	B

Appendix H

Tabell 1
Trimmade medelvärdet av tid per notering och kostnad per notering och månad

SG	SGNamn	FREQ	TidNotR (Minuter per notering för intervjuare i butik)	Kostnader för datareg och granskning (kronor per notering)	C _{hg} Kostnad per notering och år (kr/not och år)
1	Bilhandel/Bilreservdelar	10	1.88889	1.8043	168.7
3	Byggvaruhus/Byggvaror	24	1.56963	1.8043	144.2
4	Varuhus	5	1.82920	1.8043	---
5	Stormarknader/Livsmedel	43	0.66190	0.9012	---
6	Tobak/Hälsokost/Parfym/ Skivor/Färg/Bok&Papper/ Guldsmed/Zoo/Golv	56	1.83459	1.8043	164.6
8	Kläder	69	3.19361	11.127	390.3
9	Möbler	24	1.84948	1.8043	165.7
10	Hemtextil/Bosättning/Vitvaror/ Leksaker	38	2.96861	1.8043	251.8
11	RadioTV	14	1.99393	1.8043	176.8
12	Bilverkstad/Gumverkstad/ Sy- behör/Väskor/Lampor/Musik- instrument/Optiker/Ur/Barn- vagn/Foto	61	1.74373	1.8043	157.6
15	Sport	19	2.65618	1.8043	227.8
16	Blommor	32	0.74681	1.8043	80.9
17	Dator	7	4.11111	11.127	460.9
18	Tele	11	1.10714	1.8043	108.6
20	Hotell	14	0.55988	1.8043	66.5
21	Restaurang/Personalmatsal	57	0.64040	1.8043	72.7
22	Skomakare/Reparatör/Filmut- hyrning/TV-uthyrning/Veteri- när/Tvättereri/Frisör	26	2.08333	1.8043	183.7
23	Skor	26	2.67275	1.8043	229.0

Kostnaden för dataregistrering och granskning för datorer är ungefär lika stor som för kläder.

Appendix I

Tabell 1
Trimmat medelvärde av övrig tid (otid)

Stratum Grupp	Telefon- eller besöks-butik	SGNamn	FREQ	OTIDR (min/mån)	C _n Kostnad per butik och år (kr/butik och år)
1	B	Bilhandel/Bilresevdelar	10	35.7262	3280.4
3	B	Byggvaruhus/Byggvaror	24	34.1714	3160.9
4	B	Varuhus	5	81.1111	6771.3
5	B	Stormarknader/Livsmedel	43	68.0116	5763.7
6	T	Tobak/Hälsokost/Parfym/Skivor/Färg/Bok&Papper/Guldsmed/Zoo/Golv	56	8.8564	1389.3
8	B	Kläder	69	43.4027	4252.5
9	B	Möbler	24	46.4152	4102.6
10	B	Hemtextil/Bosättning/Vitvaror/Leksaker	38	32.5780	3038.3
11	B	RadioTV	14	33.0185	3072.2
12	T	Bilverkstad/Gumverkstad/Sybehör/Väskor/Lampor/Musikinstrument/Optiker/Ur/Barnvagn/Foto	61	11.0667	1546.2
15	B	Sport	19	38.2986	3478.3
16	T	Blommor	32	11.6023	1584.2
17	B	Dator	7	18.1333	1927.3
18	T	Tele	11	15.4861	1860.0
20	T	Hotell	14	4.5000	1080.0
21	T	Restaurang/Personal matsal	57	9.0132	1400.4
22	T	Skomakare/Reparatör/Filmuthyrning/TV-uthyrning/Veterinär/Tvätter/Frisör	26	7.3846	1284.8
23	B	Skor	26	37.4693	3414.5

Appendix J

Tabell 1
Kostnad per representantvara

SGNamn	Tid det tar att definiera en vara (h/vara)	Kostnad per representantvara (kr/vara)
Kläder	80	31200
Lokalpriser	0.36	140.4
Dagligvaror	0.17	66.3
PC	30	11700
Bensin	3	1170

Appendix K

Tabell 1
Månadslänk och indexlänk

	Månadslänk (systematisk)	Indexlänk, juli-december 2001 och 2002 (systematisk, med gräns)	Indexlänk, juli-december 2001 och 2002 (systematisk, utan gräns)	Indexlänk, juli-december 2001 och 2002 (systematisk, med gräns, negativ)
Kläder, butik	82,2	71,3	68,5	72,0
Ikea, butik	3,0	3,5	3,6	3,0
Möbler, butik	10,3	20,1	20,1	20,0
Restauranger, butik	31,9	55,6	50,5	54,3
Personal matsalar, butik	7,2	15,4	13,7	13,2
Möbler 1, vara	4,6	5,2	4,9	5,2
Möbler 2, vara	7,1	9,1	8,7	9,3
Möbler 3, vara	5,8	5,8	5,6	5,8
Mat 1, vara	2,1	2,3	2,5	2,2
Mat 2, vara	2,0	2,0	2,0	2,0
Mat 3, vara	5,0	5,9	6,5	5,4
Mat 4, vara	9,5	11,4	12,5	11,8
Mat 5, vara	4,9	6,2	6,8	6,5
Mat 6, vara	2,0	2,3	2,3	2,0
Kläder 1, vara	2,0	2,0	2,0	2,0
Kläder 2, vara	11,6	9,6	10,2	9,7
Kläder 3, vara	2,1	2,1	2,2	2,0
Kläder 4, vara	2,8	3,6	3,7	3,4
Kläder 5, vara	9,5	10,3	11,0	10,6
Kläder 6, vara	2,0	2,4	2,6	2,3

Appendix L

Tabell 1
Kostnader för KPI

	Intervjuare	Butiker	Representantvaror	Produkter (noteringar)
Overhead	85min/mån ²² 120h/år ²³	100h ²⁴		
Lokalpris (9883not/mån)		20min/butik ²⁵ 70min/butik ²⁶ Övrig tid per butik ²⁷	0.36h/ vara ²⁸	653h ²⁹ Prisinsamlingstid ³⁰
Dagligvar (11727not/mån)		15min/butik 70min/butik Övrig tid per butik	0.17h/vara	387h ³¹ Prisinsamlingstid
Kläder (2385not/mån)		20min/butik 70min/butik Övrig tid per butik	80h/vara	500h ³² 150tkr ³³ Prisinsamlingstid
PC (50not/mån)		20min/butik 70min/butik Övrig tid per butik	30h/vara	38h ³⁴ Prisinsamlingstid
Bensin (200not/mån)		10min/butik	3h/vara	Granskning se lok priser Prisinsamlingstid

²² Gemensam KPI tid (såsom post av brev) per intervjuare och månad, dvs årskostnaden är 85min*antal intervjuare*13 månader.

²³ Sortering och utskick av PR.

²⁴ Urvalsplanering, körning, anvisningar arbetsfördelning PR.

²⁵ Urval: Manuell granskning om butiken som slumpmässigt valts är ok att ha med i undersökningen. Totalt 500h för 1500 butiker i bruttourval.

²⁶ Sortimentundersökning: Intervjuaren åker till butik för att undersöka att sortimentet är ok och för att prata med butiksföreståndaren. 70=40+30, 40min är resa tur och retur, reduceras till igenomsnitt 40/3=13 för butiksbesök, 30min är tid inne i butik där bland annat sortimentsundersökning sker.

²⁷ Se bilaga G.

²⁸ Tid per vara det tar att definiera en ny vara.

²⁹ Blankett, granskning. 1040h/år är summan av lokalpris och dagligvaruhandel.

1040*19766/31493=652.73676≈653, 19766=2*9883, 2 står för att det är dubbelt så tidskrävande att granska lokalpriser som dagligvarupriser, det är 9883 lokalpriser och totalt 31493 priser (2*9883+11727).

³⁰ Se bilaga G.

³¹ Blankett, granskning. 1040*11727/31493=387.26324≈387, se även fotnot 7.

³² Granskning.

³³ Dataregistrering.

³⁴ Registrering och granskning.

Appendix M

Tabell 1
Resultat från prisinsamling

	Antal butiker	Prisinsamlingstid per notering, trimmad (min/not)	Övrig tid, trimmad (min/butik)
Kläder³⁵	93	3,19361	43,4027
Möbler	66	1,84948	46,4152
Restaurang³⁶	123	0,6404	9,0132
	Granskning & Dataregistrering (kr/not)	Kostnad per notering och år (kr/not och år)	Kostnad per butiker och år (kr/butik och år)
Kläder	11.1272	390.3	4252.5
Möbler	1.8043	165.7	4102.6
Restaurang	1.8043	72.7	1400.4

³⁵ Kläder innefattar prismätning av kläder i varuhus, stormarknader, klädbutiker och sporthandlar.

³⁶ Prismätning i restaurang täcker såväl vanliga restauranger som personalmatsalar.

- 2004:01 Hjälpverksamhet. Avrapportering av projektet Systematisk hantering av hjälpverksamhet
- 2004:02 Report from the Swedish Task Force on Time Series Analysis
- 2004:03 Minskad detaljeringsgrad i Sveriges officiella utrikeshandelsstatistik
- 2004:04 Finansiellt sparande i den svenska ekonomin. Utredning av skillnaderna i finansiellt sparande Nationalräkenskaper, NR – Finansräkenskaper, FiR
Bakgrund – jämförelser – analys

ISSN 1650-9447

Statistikpublikationer kan beställas från SCB, Publikationstjänsten, 701 89 ÖREBRO, e-post: publ@scb.se, telefon: 019-17 68 00, fax: 019-17 64 44. De kan också köpas genom bokhandeln eller direkt hos SCB, Karlavägen 100 i Stockholm. Aktuell publicering redovisas på vår webbplats (www.scb.se). Ytterligare hjälp ges av Bibliotek och information, e-post: information@scb.se, telefon: 08-506 948 01, fax: 08-506 948 99.

www.scb.se