



Statistiska centralbyrån
Statistics Sweden

Fråga rätt!

Utveckla, testa, utvärdera
och förbättra blanketter

Juni 2001

INNEHÅLL		Sida
Förord		5
1 Varför satsa på mätteknik?		7
1.1 Primära mål		7
1.2 Positiva sideeffekter		7
1.3 Vad händer om blankettutvecklingen försummas?		8
1.4 Utvecklingstendenser		9
2 Databildning		11
2.1 Centrala begrepp		11
2.2 Hur går databildningen till?		12
2.3 Kognitiv modell för svarsprocessen		15
2.4 Måtfelsmodell för effektivare resursallokering		16
3 Faser för mättekniskt arbete - en sammanfattning		17
4 Fas 1 – Definiera undersökningen		19
4.1 Inledning		19
4.2 Gå från sakproblem till statistikproblem!		19
4.3 Skaffa information om undersökningsenheterna!		20
4.4 Välj datainsamlingsmetod!		20
4.5 Vilka hjälpmedel finns?		25
4.6 Checklista		27
5 Fas 2 Blankettkonstruktion		29
5.1 Inledning		29
5.2 Allmänna tips till frågekonstruktören		29
5.3 Tänk på uppgiftslämnaren		31
5.4 Vanliga feltyper i blanketter till individer		32
5.5 Speciellt för företagsundersökningar		39
5.6 Frågor för jämförelser över tiden		40
5.7 Vilka hjälpmedel finns?		40
5.8 Layout		42
5.9 Checklista		48
6 Fas 3 - Kognitiva test		49
6.1 Innehåll		49
6.2 Kognitivt test i sammandrag		49
6.3 Rekrytering av testpersoner		52
6.4 Verktyg		53
6.5 Test av blanketter till företag		56
6.6 Värdering av testresultat		59
6.7 Översikt, checklista och resultat		60
7 Fas 4 - Experiment		63
7.1 Innehåll		63
7.2 Metodik		63
7.3 Exempel på beslut efter experiment vid SCB		64
7.4 Översikt, checklista och resultat		65

8	Fas 5 - Anpassning till produktionen	67
8.1	Innehåll	67
8.2	Produktionsanpassning av blanketten	67
8.3	Verktyg för anpassning	68
8.4	Felsignaler och åtgärder	69
8.5	Provintervjuer - exempel	69
8.6	Checklista och resultat	70
9	Fas 6 - Utvärdering	71
9.1	Innehåll	71
9.2	Kvalitetssäkring	71
9.3	Revidera blanketter i löpande undersökningar	76
9.4	Exempel	77
9.5	Checklistor och resultat	80
10	Fas 7 - Kvalitetsredovisning	83
10.1	Innehåll	83
10.2	Anvisningar och förutsättningar	83
10.3	Mätteknisk kvalitet – exempel på redovisning	85
	Bilaga	89
	Mätfelsmodell för effektivare resursallokering	89

Generaldirektörens förord

Fråga rätt! Utveckla, testa, utvärdera och förbättra blanketter har tagits fram inom ramen för kvalitetsarbetet på Statistiska centralbyrån. Den redovisar i handboksform tillgängliga och rekommenderade metoder för mättekniskt arbete med inriktning på blankett- och frågekonstruktion. Genom att lägga ner arbete på att utveckla en bra blankett läggs grunden för att ta fram statistik med god kvalitet. Handboken kan användas av personer som ansvarar för att samla in data både på SCB och på andra myndigheter, organisationer och företag.

Svante Öberg

Författarnas förord

Denna handbok visar på metoder att utveckla blanketter och informationsmaterial i statistiska undersökningar. Handboken beskriver en systematisk arbetsprocess från att man definierar variablerna till att man deklarerar undersökningsresultatens kvalitet. Genom att man följer en standardiserad arbetsgång blir blankettutvecklingen mer rationell och effektiv.

Centrala avsnitt handlar om metoder för att testa preliminära versioner av en blankett under realistiska villkor, så att svagheter i dem kan identifieras och åtgärdas. Detta är ett område som många undersökningar försummar och får betala ett högt pris för i form av sämre kvalitet och dyrare produktion. Framställningen är inriktad på de vanligaste undersökningsmetoderna som används för att samla data från företag, kommuner, organisationer, personer och hushåll. Många av dessa undersökningar är periodiska, medan andra görs bara en gång eller intermittent.

Handboken är i första hand avsedd för personer som utvecklar egna blanketter och arbetar med insamling av data. Men även statistik användare kan dra nytta av den för att bedöma om de material de analyserar har samlats in med betryggande mätteknisk kvalitet. De kan kontrollera om arbetet i de olika faserna av blankettutveckling har utförts väl eller försumrats.

Handboken är inte tänkt att läsas från ”pärm till pärm” utan kan användas som uppslagsbok. Man läser om den fas man tänker arbeta med för att få idéer och tips. Den avspeglar SCB:s uppfattning om vad som för närvarande är bra metoder för det mättekniska arbetet enligt den förebild och serie av handböcker som vi brukar benämna Current Best Methods (CBM). En schematisk bild över det mättekniska arbetet finns på omslagets tredje sida.

Detta är en första utgåva angående mätteknik. Vi tar tacksamt emot synpunkter och räknar med att uppdatera och komma med en ny upplaga inom några år.

Huvudansvarig för arbetet med handboken har varit **Håkan L Lindström**. Följande personer har bidragit till olika avsnitt: **Gunilla Davidsson, Birgit Henningson, Anette Björnram** och **Helén Marklund**. Dessutom har en läsgrupp hjälpt till med att lämna synpunkter. Hans Berglund har språkgranskat.

1 Varför satsa på mätteknik?

1.1 Primära mål

Mättekniskt arbete innebär att utveckla en undersöknings innehåll, insamlingsmetod och blankett så, att uppgiftslämnaren med minsta ansträngning och tillräcklig noggrannhet kan lämna de uppgifter som statistiken behöver. Dessutom är det till hjälp när undersökningens resultat ska kvalitetsdeklarerars.

För att uppnå detta krävs att man:

- har omfattande praktiska erfarenheter och teoretiska kunskaper när det gäller att identifiera och definiera undersökningsvariabler
- kan välja datainsamlingsmetod anpassad till undersökningssituationen
- kan utveckla frågor, svarsalternativ och anvisningar bl.a. genom att använda kvalitativa och kvantitativa testmetoder
- kan mäta och deklarerat tillförlitligheten och relevansen i de uppgifter som senare samlas in, dvs. bedöma hur blanketten fungerat i produktionen.

De **omedelbara målen** för mättekniskt arbete är att

- ge undersökningen ett **väldefinierat och mätbart innehåll** genom att sätta sig in i uppgiftslämnarnas verklighet och kapacitet att svara
- genom expertkunskap samt kognitiva och kvalitativa metoder utveckla frågor, svarsalternativ och blanketten i sin helhet samt anpassa den till datainsamlingsmetod och uppgiftslämnare så att de ger **minsta möjliga svarsbias och svarsvarians**
- genom observationer, kvalitativa studier, experiment och evalveringar **ta fram mått** eller åtminstone indikatorer **på den mättekniska osäkerheten hos resultaten.**

1.2 Positiva sidoeffekter

En systematisk utveckling och testning av blankett och insamlingsmetod effektiviserar och förenklar insamling och bearbetning. Viktiga effekter både för uppgiftslämnaren och för undersökningens kostnader och produktionstid är följande:

Uppgiftslämnarbördan minskar. Med en bra blankett går det åt mindre tid för den som svarar att hitta rätt i blanketten, att läsa anvisningar och mindre tid och eftertanke att komma fram till svaren. Minskad uppgiftslämnarbörda kan få både objektsbortfall och partiellt **bortfall** att minska.

Dillman ger exempel på att en kompakt och rörig blankett minskat svarsfrekvensen med 3-4 procent. Akkerboom (1997) redovisar ett extremfall då svarsfrekvensen ökade från ca 50 till 90 procent efter test och revidering av en blankett.

Behovet av granskning minskar när uppgiftslämnaren får möjlighet att svara på ett sätt som han/hon har möjligheter till.

Volymen på granskningen indikerar hur stora problemen varit att lämna uppgifter och pekar på i vilken mån det kan löna sig att revidera blankett och

anvisningar. Generellt anses granskningen ta 30-40 % av företagsundersökningarnas totala resurser i anspråk och 15-20 % av individundersökningarnas enligt Granquist (1999). I en studie av Industristatistiken 1990 beskrivs att ca 13 % av alla uppgifter rättades och att 65 % av de svarande företagen berördes av åtminstone en rättning (Hedlin (1992). I flera fall påvisades direkta samband mellan brister i blankett och felaktiga uppgifter. För en variabel (antal manliga respektive antal kvinnliga ägare) måste 70 % av uppgifterna ändras p.g.a. att anvisningarna stod i en fotnot, som inte uppmärksammades. För variabeln "sysselsatta inom övriga verksamheter" ledde en missvisande layout till att 50 % av uppgifterna måste korrigeras.

Behovet av **återkontakter** minskar om man kan avlägsna missledande frågor, som orsakar inkonsistenta svar och svar av fel storleksordning.

1.3 Vad händer om blankettutvecklingen försummas?

Mätfel uppstår när frågor och blanketter är utformade så att uppgiftslämnaren inte förstår dem korrekt, när det inte går att få fram de begärda svaren med god tillförlitlighet och då uppgiftslämnarbrödan är för hög. Oftast beror detta på att man inte insett svårigheterna och därför inte satsat tillräckligt på att testa blanketten systematiskt. Men det förekommer också att man valt att studera svårsmått storheter trots att man är medveten om vilken osäkerhet och begränsning i användningen detta för med sig.

Anvisningarna för kvalitetsdeklarationer i Sveriges Officiella Statistik (SOS) sammanfattar orsakerna till mätfel på följande sätt.

Mätning/observation

Det händer att en lämnad uppgift inte överensstämmer med det "sanna" värdet enligt variabelns definition. Det finns en mängd skäl till att så kan vara fallet, t.ex. att blankettfrågan inte passar ihop med uppgiftskällans bokföring, frågan är tvetydigt formulerad, personer har bristande minne, uppgiftslämnare kan vara slarviga (eller ännu värre, medvetet vilseledande), fysikaliska mätmetoder kan vara behäftade med brister, m.m. Allmänt talar man om att mätfel föreligger. Mätfel bidrar naturligtvis till statistiks osäkerhet, och kan göra det såväl på ett systematiskt sätt (medförande skevhet) som på ett tillfälligt sätt (som inte leder till skevhet, men ökar osäkerheten).

En empirisk studie av Christianson och Polfeldt (1996) identifierar ett antal exempel på orsaker till mätfel. Utan anspråk på att vara fullständig ger den en översikt av felrisker som man måste försöka parera, när man gör blanketter. Studien klassificerade mätfel efter orsak hos sammanlagt 206 utvalda variabler, fördelade på 135 produkter vid SCB. Rapporteringen byggde betydligt oftare på bedömningar och indikationer än på evalveringsstudier. Nedan följer de vanligaste feltyperna enligt den genomgång som utredarna gjorde:

Feltyper rangordnade efter hur vanliga de var.

A	Definitionsproblem. Statistikens definitioner var obekanta eller användes inte hos uppgiftslämnaren.
B	Minnesfel. Var vanligast i individ- och hushållsundersökningar men förekom också i företagsundersökningar.
C	Uppgiftslämnaren måste göra bedömningar. Exakta uppgifter fanns inte alls eller inte för rätt period.
D	Periodiseringsproblem. Uppgifterna avsåg en annan referensperiod eller -tidpunkt än den efterfrågade, t.ex. då ett företag har brutet räkenskapsår.
E	Krav på beräkning. Variabeln fanns inte, utan svaret måste kalkyleras från andra uppgifter.
F	Korrelerad variabel användes. T.ex. svarade man med uppgift om leveranser, då produktionen efterfrågades men inte var tillgänglig.
G	Redovisningsproblem. Företaget hade inte sina uppgifter uppdelade på de kategorier eller objekttyper som statistiken efterfrågade.
H	Klassificeringsfel i bakgrundsvariabler p.g.a. att de var svåra att mäta.
I	Sant värde saknades.
J	Felaktig inkludering av komponenter i en summavariabel. Uppgiftslämnaren räknar felaktigt in komponenter som inte efterfrågas.
K	Fel vid registreringen i blanketten som inte upptäcks.
L	Felaktig exkludering av komponenter i en summavariabel. Uppgiftslämnaren glömmar att räkna in en eller flera komponenter som efterfrågades.

Speciellt i ekonomiska undersökningar förekommer också följande fel.

- Sortfel då uppgiftslämnaren inte observerar i vilken enhet svaret ska ges. T.ex. rapporteras i tusental kronor i stället för i miljoner kronor.
- Uppgiftslämnaren kan inte få fram den efterfrågade kvantitativa uppgiften och väljer att lämna svarsrutan blank, markera med ett streck eller kanske skriva 0.
- Felaktig exkludering av en komponent i en variabel medför ibland felaktig inkludering av komponenten i en annan - ofta en övrig- eller restpost.

Vilka feltyper som kan uppträda i en enskild undersökning beror på faktorer som ämne, svårighetsgrad, insamlingsmetod och uppgiftslämnarens motivation. Dessa kan man inte förutse enbart med hjälp av ämneskunskap och mätteknisk erfarenhet. Man måste också testa blanketten på verkliga uppgiftslämnare innan den sätts in i produktion.

1.4 Utvecklingstendenser

In på 1980-talet var mättekniska metodstudier oftare inriktade på att evalvera effekter hos använda mätmetoder än på att identifiera orsakerna till mätfel. Numera används en lång rad metoder och verktyg för att studera och analysera hur uppgiftslämnaren svarar på frågorna i en blankett. Med ett litet test av en preliminär version av blanketten kan man undanröja många orsaker till fel innan huvudundersökningen startar. Längst har man kommit med metoder för att testa blanketter till individer. Men allt fler erfarenheter samlas också om test av blanketter i företagsundersökningar.

Det mättekniska metodområdet utvecklas snabbt. Säkrare kriterier utvecklas nu för val av bästa metod och principer för samtidig användning av flera metoder. För att man successivt ska kunna förbättra löpande undersökningar behöver man en systematik för att från produktionsprocessen ta fram indikatorer på förekomsten av mätfel. Utnyttjande av ADB-stöd för att ta fram statistik över insamlings- och mätprocessen kommer att öka i takt med att allt flera insamlingar blir datorstödda.

De flesta statistiska centralbyråer (inklusive alla de nordiska) har numera en specialistgrupp för utveckling och test av blanketter. De brukar kallas "Questionnaire Design Resource Center" (QDRC) eller något liknande. Gruppen vid SCB kallas Mättekniska laboratoriet (ML).

Referenser

1. Akkerboom, H. and Dehue, F. (1997) *Examples of whole packages tests*. Proceedings of the Workshop on Minimum Standards in Questionnaire Testing held i Örebro 19 - 21 October 1997.
2. Christianson, A. and Polfeldt, T.(1996). *Response quality improvement initiatives at Statistics Sweden*. Proceedings of the 2nd International Conference on Methodological Issues in Official Statistics. Stockholm, September 23-24 1996.
3. Dillman, D.A. *Progress in the design of respondent-friendly selfadministered questionnaires*.
4. Granquist, L. m.fl. (1999) *On improving quality by modern editing*. SCB-rapport.
5. Hedlin, D. (1992). *Jämförelse av granskade och ogranskade data i industristatistiken*. Gransknings-PM nr 29. SCB-rapport.

2 Databildning

2.1 Centrala begrepp

Mättekniska begrepp används med skiftande innebörd och ofta utan att definieras. För att undvika oklarhet anger vi här hur ett antal centrala begrepp används i denna handbok. I några fall finns inte någon etablerad term på svenska. Vi anger då inom parantes den motsvarande benämningen på engelska för att förenkla vidareläsning i facklitteraturen. I enstaka fall är den engelska termen så etablerad och enkel att vi har föredragit den framför en svensk översättning.

Undersökningsobjekt - person, kommun, skola, jordbruksföretag osv. som man begär uppgifter om. Den som svarar på frågorna i en intervju eller fyller i en blankett kallas **uppgiftslämnare**. (Andra vanliga benämningar är svarande, svars-person och respondent.)

I en **individundersökning** lämnar en person uppgifter om sig själv. Undersökningar som är riktade till företag, myndigheter, organisationer, föreningar osv. får här den samlande benämningen **företagsundersökningar**. Detta svarar mot hur man på engelska ofta använder begreppet '*Establishment surveys*'. Dessa skiljer sig från individundersökningar främst genom att det inte finns en unik uppgiftslämnare, utan man måste först försöka identifiera "den mest insiktsfulle uppgiftslämnaren" bland flera möjliga. I **hushållsundersökningar** ska en person svara för t.ex. hela sitt hushålls inkomster eller utgifter. I dessa undersökningar har man samma typ av svårighet att bestämma vem som är den bästa uppgiftslämnaren.

Vi använder begreppet **fråga** för den text som specificerar vilken uppgift som ska lämnas och de svarsalternativ som erbjuds. Begreppet används både då blanketten innehåller egentliga frågor och då en begäran om information uttrycks i annan språklig form, t.ex. "Antal anställda vid årsskiftet", "Omsättning inklusive moms under tredje kvartalet 1999".

För **faktavariabler**, som mäter t.ex. åkerareal, levererad kvantitet, tidsåtgång, ålder och utbildningsnivå, finns ett **sant värde**. Man måste definiera variabeln precis genom att förklara begrepp, ange vilken referensperiod frågan gäller och ange vilken sort storheten ska mätas i. Med tillräckligt stora ansträngningar skulle en annan person än uppgiftslämnaren komma fram till samma svar. För en **attitydvariabel** finns det inte något sant värde på samma sätt - inte ens med motsvarande preciseringar. Däremot kan man föreställa sig förekomsten av ett **operativt sant värde**. För detta krävs dock att fråga, anvisningar och svarsalternativ är så tydligt utformade att uppgiftslämnaren åtminstone i genomsnitt kommer fram till detta svar i en tänkt serie av upprepningar. Ju mer komplex och hypotetisk en fråga är, desto svårare är det att tänka sig att den har ens ett operativt sant värde.

För variabler där man varken kan föreställa sig faktiskt eller operativt sanna värden, kan man heller inte avgöra när en frågeformulering är bättre än en annan eller när ett svar är mer eller mindre riktigt. För skattningar som bygger på sådana variabler saknar begrepp som konfidensnivå och bias egentlig mening. Ett påstående om tillförlitlighet kan då inte längre bygga på samplingteorin utan måste hämta sitt stöd på en annan typ av modellresonemang från en annan utgångspunkt.

Vi använder **blankett** som generell benämning och undviker andra vanliga benämningar som (fråge-) formulär, enkät och mätinstrument. **Att utveckla en blankett** innebär att utforma anvisningar, frågor och svarsalternativ, att göra en lämplig layout på blanketten och att ta fram informationsmaterial.

SIS handbok STG 126, som också räknar in elektroniska blanketter, definierar: *Blanketter används för att underlätta överförandet av information i form av skriven text från en uppgiftslämnare till en uppgiftsmottagare.* (sid. 104). Traditionellt betyder formulär en förebild för utformning av blanketten, t.ex. i form av tryckanvisningar. Blanketten i sin tur är det tryckta papper eller elektroniska blankett som uppgifterna skrivs in i/registreras på. Den ifyllda blanketten i sin tur kallas dokument. Med hänsyn till att blankettkonstruktion numera nästan alltid görs i datormiljö och svaren ofta lämnas med datorstöd, har begreppet formulär spelat ut sin roll.

Med **uppgiftslämnarbörda** menar vi i första hand den tidsåtgång och arbetsinsats som krävs av uppgiftslämnaren för att sätta sig in i frågorna, skaffa fram information, fylla i blanketten och skicka in den. Vid felaktigt ifyllda blanketter ökar uppgiftslämnarbördan då statistikproducenten gör återkontakter för rättning och komplettering. När en produkt använder elektroniska blanketter läggs ofta granskningsvillkor in i dessa, vilket betyder att rättning läggs på. Till uppgiftslämnarbördan kan också räknas att behöva svara på känsliga frågor.

Denna handbok handlar främst om att utveckla blanketter med standardiserade frågor och fasta (strukturerade) svarsalternativ - i första hand för statistikproduktion. Samma principer gäller dock för blanketter med administrativ användning. **Standardiserade frågor** är frågor som ställs med samma formulering till alla uppgiftslämnare. Olika frågor kan ställas till olika uppgiftslämnare beroende på vilken delgrupp de tillhör. Svaren på selektionsfrågor med åtföljande hopp-anvisningar leder till att alla uppgiftslämnare inte svarar på alla frågor. I en fråga med **fasta svarsalternativ** ska uppgiftslämnaren fylla i en kvantitativ uppgift av angiven sort på anvisad plats, kryssa för/ringa in ett eller flera givna svarsalternativ eller välja ett eller flera av de svarsalternativ intervjuaren läser upp. Med **öppna svarsalternativ** menas att uppgiftslämnaren själv får formulera sina svar. Kostnaden och tidsåtgången för att koda och registrera svaren blir avskräckande. Öppna svarsalternativ brukar därför bara användas för enstaka frågor och för att ge uppgiftslämnaren tillfälle att ge avslutande översiktliga kommentarer om undersökningen.

Mättekniska testmetoder använder sig i stor utsträckning av **kvalitativa metoder** (metoder som detaljerat belyser egenskaperna hos datainsamlingsmetod och blankett). Man använder både standardiserade och icke standardiserade fördjupningsfrågor med öppna svar för att förstå hur uppgiftslämnaren uppfattar frågorna och kommer fram till sina svar.

2.2 Hur går databildningen till?

Vid val av mätmetod, vid konstruktion av en blankett och när man avgör hur noggrann en uppgift behöver vara, tar man hänsyn till hur uppgiftslämnaren kan få fram, bearbeta och lämna de efterfrågade uppgifterna - dvs. hur de enskilda variabelvärdena/grunddata kan komma att bildas. Vilka krav frågan ställer på uppgiftslämnaren beror på

- vilken **skaltp** variabeln ska mätas på
- hur den **grundläggande databildningen** ska ske
- vilka **bearbetningar** man kräver att uppgiftslämnaren ska utföra

Fyra skaltyper

Man skiljer på fyra **skaltp** eller **mättnivåer** efter hur en variabel kan mätas.

Skaltp/ (Mättnivå)	Egenskaper hos skalan Mätvärdena kan:
Kvotskala	Särskiljas, rangordnas, mätas med konstant måttenhet och med absolut nollpunkt
Intervallskala	Särskiljas, rangordnas och mätas med konstant måttenhet
Ordinalskala	Särskiljas och rangordnas
Nominalskala	Särskiljas

[Valet av skala hänger också samman med vilken bearbetning och vilken typ av analys man kommer att kunna göra. Den diskussionen ligger dock utanför denna handbok.]

Att lämna uppgifter på **kvot- och intervallskalenivå** ställer minst krav på uppgiftslämnaren när det gäller att resonera, eftersom måttenhet och skalsteg är väl definierade. Man får lämna uppgift om t.ex. drivmedelsförbrukning, temperatur, försäkringskostnader. Däremot kan det krävas en hel del förarbete, t.ex. med att mäta och beräkna, innan man fått fram de begärda uppgifterna.

Frågor om uppgifter på **ordinalskalenivå** förutsätter att uppgiftslämnaren ska kunna tolka verbala beskrivningar av skalans ändpunkter och ta ställning till hur långt varje skalsteg (svarsalternativ) är. T.ex. måste uppgiftslämnaren bestämma sig för betydelsen av "Mycket optimistisk", "Ganska optimistisk", "Osäker", "Ganska pessimistisk", "Mycket pessimistisk" som svar på en fråga hur man bedömer framtidsutsikterna för sitt företag, sin privatekonomi eller något annat.

Frågor om uppgifter på **nominalskalenivå** kräver att uppgiftslämnaren ska kunna skilja mellan olika alternativ. Ibland är valen relativt direkta och enkla, t.ex. då man ska ange sin födelseort, utbildningsnivå eller civilstånd. I andra fall krävs beräkningar och sammanvägningar, t.ex. för att bestämma vilken socioekonomisk grupp ett hushåll tillhör, vilken arbetskraftsstatus en person har eller vilken näringsgren ett företag tillhör.

Om man räknar med att flertalet av uppgiftslämnarna inte kan, inte hinner eller inte vill lämna uppgifter på högsta (teoretiskt möjliga) nivå, får man överväga hur variabeln bäst ska mätas på "lägre skalnivå". T.ex. kan "antalet resor med lokaltrafiken" eller "antalet transporter av ett varuslag" under en period mätas alternativt genom att uppgiftslämnaren:

1. Anger exakt antal (kvotskala)
2. Markerar en storleksklass, t.ex. någon av 0, 1-5, 6-10, fler än 10 (ordinalskala)

3. Väljer bland verbalt beskrivna alternativ t.ex. (ordinalskala)
 dagligen
 några gånger varje vecka
 några gånger varje månad
 mera sällan
 aldrig
4. Anger om aktiviteten (nominalskala)
 förekommer
 ej förekommer

Tre kategorier av databildning

Man kan särskilja tre huvudtyper av direkt databildning, med hänsyn till hur uppgiftslämnaren tar fram svaren på blankettens frågor:

- A. Uppgifterna finns redan registrerade hos uppgiftslämnaren och behöver "bara" hämtas fram. Uppgiftslämnaren förväntas göra en korrekt **avskrift eller kopiering** - manuellt eller elektroniskt. Det kan t.ex. vara en uppgift ur ett företags bokslut eller personaladministrativa system, kostnaden för en personlig försäkring, en utbildningsuppgift för en person.
- B. Undersökningen begär att uppgiftslämnaren ska göra **mätningar eller observationer och registrera uppgifter som annars inte skulle ha registrerats**. T.ex. läsa av bensin-, el- eller vattenmätare den sista dagen varje månad eller bokföra transporter, resvanor, tidsanvändning eller hushållsutgifter under en period.
- C. Blanketten frågar efter **personliga uppgifter, som enbart finns i uppgiftslämnarens medvetande**, t.ex. kunskaper, planer, minnesbilder, attityder och bedömningar. I fall C till skillnad från A och B ger uppgiftslämnaren **svår direkt**.

Frågor som förutsätter databildning av kategorierna A och B gäller dels kvantitativa uppgifter i intervall- eller kvotskala (t.ex. priser, kostnader, volymer, antal anställda), dels klassificeringar (t.ex. utbildningsnivå, verksamhetsområde, fastighetstyp). De kan i praktiken bara ställas i postenkäter eller motsvarande elektroniska insamlingar, eftersom svaren inte kan ges direkt utan måste eftersökas och hämtas fram av uppgiftslämnaren. Större delen av databildningen i löpande företagsundersökningar och även i hushållsundersökningar är av kategori A eller B. Uppgifterna kan då lämnas av varje person som har befogenhet och kunskaper att göra det och dessutom fått tillräckliga och begripliga anvisningar.

Frågor som kräver databildning av kategori C gäller till stor del kvalitativa uppgifter på nominal- eller ordinalskalennivå. Dessa kan oftast bara besvaras tillförlitligt om urvalspersonen själv är uppgiftslämnare. Om man vill vara helt säker på att svaren inte påverkas av andra personer, är man begränsad till att göra intervjuundersökningar. Även engångsundersökningar till företag kräver ibland databildning av kategori C, t.ex. om rekryteringsplaner, bedömningar av konjunkturen och förväntningar. I sådana fall torde postenkät vara tillförlitligare, eftersom uppgiftslämnaren då kan förankra sina svar hos "rätt" instans.

Uppgiftslämnarnas bearbetningar av egna data

Statistikproducenten begär ofta att uppgiftslämnarna inte bara ska lämna avskrifter eller direkta observationer, utan dessutom att de ska göra sammanvägningar eller bearbetningar av sådana. För att bearbetningarna ska bli enhetligt

gjorda, behövs precisa anvisningar i blanketten om hur det ska utföras. Bearbetningarna kan vara av olika typer.

1. I det enklaste fallet begär man att uppgiftslämnaren ska göra summeringar, bilda kvoter eller utföra andra matematiska eller logiska operationer på ”enkla data”. Ibland ska både enkla data och resultatet av bearbetningen föras in i blanketten, ibland bara resultatet. [När statistikproducenten tar över bearbetningen kallas sådana beräknade storheter för ’härledda variabler’]
2. I mer krävande fall måste uppgiftslämnaren dessutom själv samla in uppgifter på sin arbetsplats och göra beräkningar innan han kan fylla i de uppgifter, som blanketten frågar efter.
3. Också vid ”svar direkt” ställs ibland stora krav på att uppgiftslämnaren snabbt ska kunna väga samman och bearbeta erinringar, värderingar, kunskaper osv. ”i huvudet”. För att få alla uppgiftslämnare att tänka på samma sätt, får man anstränga sig att skriva tydliga frågor och anvisningar så att uppgiftslämnarna kommer fram till samma referensram.
4. I undersökningar med intelligenta elektroniska blanketter uppmanar man uppgiftslämnaren att själv rätta de svar som felsignaleras. Detta kräver omprövning av redan givna svar.

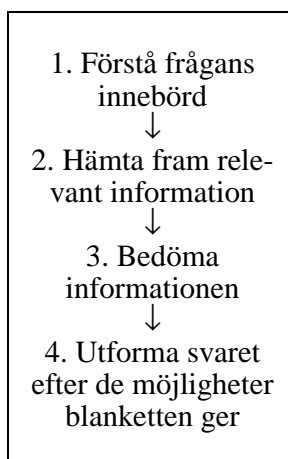
Föreställning och verklighet

Att data bildas på ett visst sätt betyder inte nödvändigtvis att uppgiftslämnaren följer intentionerna. Även när det finns registrerade uppgifter att kopiera händer det att man i stället lämnar approximativa ”svar direkt”. Det kan hända när uppgiftslämnare är jäktade, tycker att de inte behöver läsa anvisningar eller inte är tillräckligt insatta i området för att kunna svara korrekt. Hur databildningen faktiskt går till undersöker man med hjälp av särskilda testmetoder. En tidig studie inom prisstatistiken visade att över 50 % av företagen lämnade priser på sin försäljning på den svenska marknaden i stället för, som efterfrågades, fakturerade priser.

2.3 Kognitiv modell för svarsprocessen

En kognitiv modell är en föreställning om hur uppgiftslämnaren tar till sig och förstår anvisningar och frågor i blanketten, hämtar fram information, värderar denna, anpassar sitt svar till de alternativ som ges och avger det. En ofta refererad kognitiv modell omfattar följande fyra steg från det att uppgiftslämnaren får frågan tills uppgiften lämnas:

1. **Att förstå ord och begrepp i blanketten och innebörden av uppgiften.** Den som gör en blankett tror gärna att uppgiftslämnarna spontant ska förstå och använda samma terminologi och därmed ge ord och begrepp samma innebörd som han/hon själv gör. Erfarenheter från blanketttest har visat att detta långt ifrån alltid är fallet ens då uppgiftslämnaren är ”professionell”, dvs. lämnar uppgifter i sin yrkesroll.
2. **Att hämta fram information ur minnet respektive från noteringar/ bokföring och liknande.** Informationen är inte alltid direkt tillgänglig eller strukturerad på det sätt som frågan förutsätter. I sitt sökandet efter information kommer uppgiftslämnaren ofta att använda sig av sina egna referensramar, som



inte nödvändigtvis stämmer med undersökningens.

3. **Att bearbeta sin information och värdera vad av den som verkligen efterfrågas.** Den information man tar fram spontant stämmer inte alltid exakt med den efterfrågade utan måste avgränsas, kompletteras och ordnas med hänsyn till frågans krav på detaljering och noggrannhet.
4. **Att utforma sitt svar med hänsyn till blankettens utformning och passa in det i rätt ram eller svarsalternativ.** I enkla fall handlar det om att rapportera i rätt sort och för rätt period. I mer krävande fall kan det gälla att bestämma vad ändpunkterna och skalstegen i en ordinalskala står för och avgöra vilket alternativ som stämmer på en själv.

Den kognitiva modellen ger en uppfattning om på vad sätt en blankett kan innehålla svårigheter. Det kan vara att uppgiftslämnaren inte förstår frågan, att den kräver för mycket av uppgiftslämnarens erinringsförmåga eller att svaret kräver alltför mycket av beräkning eller bedömning. Därmed ger den också uppslag om hur ett kognitivt blanketttest ska utformas för att belysa graden av svårighet och orsaken till den. I kapitlet "FAS 3 Kognitiva test" beskrivs hur dessa genomförs och hur olika testverktyg används till att peka ut var risker finns för att systematiska och slumpmässiga fel uppstår.

2.4 Mätfelsmodell för effektivare resursallokering

Storleken och betydelsen av de mätfel, som följer av att använda dåliga blanketter underskattas av många användare och köpare av statistiska undersökningar. De inser inte nyttan av systematisk blankettutveckling utan tror att undersökningen blir billigare att genomföra om de låter bli att testa sina blanketter. Med ett modellresonemang kan man enkelt visa att det med små omfördelningar av givna resurser går att genomföra ett blanketttest. Undersökningens tillförlitlighet kan avsevärt förbättras under mycket allmänna förutsättningar. Modellresonemanget är med nödvändighet tekniskt. Det är däremot inte nödvändigt att läsa det för att förstå och tillämpa principerna för blankettutveckling. Sådana resonemang är viktiga för att visa hur skattningarnas tillförlitlighet i högsta grad påverkas av att man satsar på att reducera systematiska och icke systematiska mätfel.

Modellen presenteras i bilaga.

3 Faser för mättekniskt arbete - en sammanfattning

Utvecklingen av datainsamlingsmetod och blankett samt kvalitetsbevakning beskrivs i denna handbok som en huvudprocess indelad i sju på varandra följande delprocesser eller faser. Den bygger på uppfattningen att samverkan mellan ämnes-, produktions- och mättekniska kompetenser är nödvändig. Indelningen i faser gör det tydligare hur man undviker merarbete och onödiga kostnader genom att upptäcka och hantera problem i rätt ordning. Arbetsgången är tillämplig på blanketter för

- alla typer av uppgiftslämnare
- både engångsundersökningar och löpande undersökningar
- alla insamlings- och mätmetoder

Fas 1 Definiera undersökningen

Arbetsuppgift: Att definiera undersökningens innehåll, variabler, tidsperiod, population och de statistiska parametrar som ska uppskattas. Samtidigt bestäms datainsamlingsmetod. Kunden är den ansvariga personen och måste fatta de viktigaste besluten.

Resultat: En variabellista, en tabellplan och/eller en analysplan.

Fas 2 Blankettkonstruktion

Arbetsuppgift: Att översätta variabellistan till en blankett, fråga för fråga. Blanketten ska anpassas till datainsamlingsmetod. Definitioner bestäms och instruktioner skrivs.

Resultat: En blankett, som är språkligt, logiskt och tekniskt riktig. Layouten ska vara tydlig och se trevlig ut. Blanketten har inte testats på ”riktiga uppgiftslämnare”.

Fas 3 Kognitiva test

Arbetsuppgift: Även om frågekonstruktören är nöjd med sin produkt kanske inte uppgiftslämnaren är det. Man måste konstatera om uppgiftslämnaren kommer att förstå de ord, fraser och begrepp som används. Kan uppgiftslämnaren skaffa fram informationen, hur kommer hon/han fram till ett svar? Tycker uppgiftslämnaren att vissa frågor är känsliga eller alltför privata? Kognitiva, vanligen kvalitativa test kommer att visa detta.

Resultat: En reviderad version av blanketten, som är anpassad till uppgiftslämnarnas förmåga att förstå frågor och anvisningar och till deras kapacitet och vilja att svara. Blanketten revideras och förbättras men måste kanske testas en gång till.

Fas 4 Experiment

Arbetsuppgift: Även en bra blankett kanske inte fungerar i en huvudundersökning. T.ex. kanske uppgiftslämnarbördan är för stor. För att man ska kunna se om det går att göra en undersökning med acceptabel reliabilitet, kan ett experiment som ger kvantitativa resultat vara nödvändigt.

Resultat: Beslut om en metod är tillräckligt bra eller vilken metod som är bäst enligt några kvantitativa kriterier. Man kanske avbryter undersökningen om det bästa alternativet inte är tillräckligt bra.

Fas 5 Anpassning till produktionen

Arbetsuppgift: Beroende på tidsplaner och resurser kanske testblanketten inte är färdig för produktion. Ytterligare anpassningar måste göras. T.ex. kan blanketten ha testats med en pappersversion och undersökningen ska göras med CATI. Förberedelser för skanning kanske inte var färdiga. Ett vanligt problem är att blanketten utformas för en datainsamlingsmetod men flera metoder används i produktionen.

Resultat: En blankett som fungerar både för uppgiftslämnaren och för producenten. Indikatorer på reliabiliteten är möjliga.

Fas 6 Utvärdering

Arbetsuppgift: Beräkna och uppskatta felindikatorer. Använda dessa mått i nästa undersökning för att identifiera felkällor och ta bort dem vid nästa tillfälle undersökningen upprepas. Arbetet görs dels under själva datainsamlingen, dels efter att arbetet avslutats. Producenten är den som har översikten av arbetet i denna fas.

Resultat: Ett antal kvalitetsmått – på olika informationsnivåer. I löpande undersökningar också ett antal användbara indikatorer på olika typer av felkällor. Försök att identifiera dem och utveckla verktyg för att ta bort eller reducera felen.

Fas 7 Kvalitetsredovisning

Arbetsuppgift: Under fas 1 – 6 har ett antal olika observationer och indikatorer på datakvalitet producerats. Många är lokaliserade till producent och process. De måste presenteras på ett sådant sätt att informationen om kvaliteten blir så användbar som möjligt för kunden och andra användare.

Resultat: Information om mätkvalitet ska inkluderas i en systematiskt kvalitetsrapport som täcker alla relevanta aspekter på undersökningskvalitet. Vi följer kvalitetsdeklaration enligt SOS.

4 Fas 1 – Definiera undersökningen

4.1 Inledning

Arbetsuppgifter: Fas 1 omfattar det som behöver göras innan man börjar formulera konkreta frågor och utforma blanketten i övrigt.

I denna ingår att

- lösa innehålls- och relevansfrågor
- göra en variabelförteckning och tabell- eller analysplan.

Samtidigt måste man bestämma

- datainsamlingsmetod
- urvalsram, urvalsförfarande och urvalsstorlek.

Innan man går vidare till fas 2 är det viktigt att noga tänka igenom om undersökningen överhuvudtaget är genomförbar med hänsyn till datatillgång, uppgiftslämnarbräda, kostnadsramar och tidshållning.

Arbetsfördelning: För att en god grund ska läggas för undersökningens kvalitet, behövs från början ett nära samarbete mellan beställare, producent och frågekonstruktör. Producenten representeras av en undersökningsansvarig person. Vi använder uttrycket frågekonstruktör för att beskriva en person som arbetar med utarbetande av frågor och framställning av frågeblanketten. Ibland är detta densamma som undersökningsansvarig person, ibland anlitar eller samarbetar denne med en mätteknisk specialist för dessa moment. På SCB finns för detta ett speciellt mättekniskt laboratorium och ett nätverk av personer som jobbar med fråge- och blankettutformning.

Beställaren är drivande i denna fas genom att specificera sitt behov av information, medan frågekonstruktörens och producentens uppgifter främst är att visa på vilka hjälpmedel som finns och vilka tekniska lösningar som kan fungera. Det är beställaren som måste avgöra om planeringsarbetet ska gå vidare eller inte.

Det räcker inte att beställaren och/eller producenten har stor ämneskunskap och kännedom om vilken information som redan finns om undersökningsområdet. För att senare kunna skriva väl fungerande frågor behöver frågekonstruktören vara med redan när sakproblemet översätts till ett statistikproblem. Detsamma gäller när man väljer urvalsram och reder ut hur väl denna beskriver undersökningsobjekten (dvs. dem man ska samla uppgifter om, oavsett om dessa är personer, skolor, företag, mjölkkor eller något annat). Likaså när man bestämmer vilken datainsamlingsmetod som ska användas. Hur detta påverkar när man konstruerar blanketten framgår av det följande.

4.2 Gå från sakproblem till statistikproblem!

För att en undersökning ska ge användbara resultat, måste beställaren föra en grundlig diskussion med den som ska utföra den och den som ska skriva frågorna. Man måste precisera undersökningsområdet och klarlägga hur man ska kunna mäta det med tillgängliga resurser. Beställare, producent och frågekonstruktör måste komma fram till en gemensam syn på hur uppgiften ska kunna lösas. Likaså måste man planera hur analysen efter datainsamlingen ska göras.

Det finns mycket att vinna på ett väl genomtänkt förarbete, som ger tydliga definitioner och avgränsningar av vad som ska mätas med en statistisk undersökning. Arbetet med att översätta allmänt formulerade undersökningsmål som att mäta - t.ex. ”svenskarnas kostvanor”, ”svenskarnas skattemoral” eller ”företagens innovationsverksamhet” till preciserade frågor kräver tid och eftertanke. När beställaren har en vag bild av sina frågeställningar, tvingas frågekonstruktören att arbeta med flera tänkbara tolkningar av hur sakproblemet ska översättas till statistiskt mätbara storheter. Att frågekonstruktören deltar i arbetet redan i fas 1, medför att avsevärt mindre arbete behöver läggas ner i de följande faserna.

4.3 Skaffa information om undersökningsenheterna!

Utöver de egentliga frågorna om undersökningens ämne krävs ibland att man ställer undersökningstekniska frågor för att skattningarna ska kunna beräknas korrekt. Frågekonstruktören behöver få besked från den som är ansvarig för estimationen i undersökningen vilken information som behövs. De följande fyra typerna av frågor behöver övervägas i varje undersökning.

- En urvalsram innehåller nästan alltid några objekt som utgör **övertäckning**. Frågor måste ställas som identifierar de objekt som inte tillhör intressepopulationen. Svaren används för att avbryta en intervju och/eller för att i efterhand sortera bort svar från dem som inte skulle vara med.
- Ibland krävs särskilda frågor för att **bestämma urvalssannolikheten** för undersökningsobjektet. Detta kan krävas om samma undersökningsobjekt finns i mer än en urvalsram eller är sammansatt av flera urvalsobjekt.
- Andra frågor är nödvändiga för att **avgränsa undersökningsenheten**, t.ex. vilka personer som ingår i ett kosthushåll eller vilken organisatorisk enhet som svaren i en företagsundersökning verkligen avser.
- Särskilt i företagsundersökningar behöver man - för eventuella återkontakter eller följande undersökningar - uppgift om **vilken person som lämnat uppgifterna** och hur denne kan kontaktas.

4.4 Välj datainsamlingsmetod!

Hur frågor, svarsalternativ och blankett ska konstrueras beror i hög grad på valet av datainsamlingsmetod. Vilken metod som ger bäst mätkvalitet i en speciell undersökning beror på dess omfattning, på typen av frågor och på vem som ska vara uppgiftslämnare. Då man designar undersökningen väger man också in krav på statistisk precision samt på kostnader och tidsplan. Detta kan innebära att man kompromissar med ambitionen att använda den tillförlitligaste mätmetoden.

4.4.1 Postenkäten och mer moderna efterföljare

Pappersblanketter

Med **postenkät** menar vi här att blankett som ska sändas med post till uppgiftslämnaren för att fyllas i och återsändas. Vid frivilligt uppgiftslämnande rekommenderas att den inte omfattar mer än 12-16 sidor och inte tar mer än 30 till 45 minuter att fylla i.

Postenkätens mättekniska fördelar är att man kan

- ha relativt långa förklarande ingresser och/eller frågelydelser
- ha flera, även långa, svarsalternativ
- använda **enkla** tablåfrågor (T.ex. ställs samma frågor om alla hushållsmedlemmar. Man använder en svarskolumn för varje person och en rad per fråga.)
- ställa frågor om mer känsliga ämnen än i en intervju
- ge uppgiftslämnaren den tid som behövs för att lämna genomtänkta svar eller leta fram korrekta uppgifter om belopp, antal årsverken m.m.
- använda hjälpmedel som kartor, bilder, symboler eller annat som uppgiftslämnaren behöver se för att kunna besvara frågorna.

Postenkätens mättekniska begränsningar innebär att man ska:

- undvika hoppanvisningar, eftersom de ökar risken för felaktigt ifyllda svar. Har uppgiftslämnaren besvarat både frågan med hoppanvisning och frågan eller frågorna som ska hoppas över, framgår det sällan var felet är. Man måste då besluta om svaren ska betraktas som partiellt bortfall eller om en annan strategi ska användas. Hur man än väljer, påverkas resultatens kvalitet.
- undvika omfattande eller komplicerade tablåer där de svarande förväntas markera ett svar som ger information från både rad och kolumn, vilket ofta medför alltför många felaktiga markeringar
- begränsa antalet frågor till 12-16 sidor med ”tilltalande” layout. Risken för både objektsbortfall och partiellt bortfall ökar om postenkäten uppfattas som alltför lång. Det är också stor risk att uppgiftslämnarens vilja och/eller förmåga att ge genomtänkta och korrekta svar minskar om blanketten är alltför omfattande.
- ta hänsyn till att det finns risk för att personer i uppgiftslämnarens omgivning påverkar svaren, dvs. frågor som förutsätter att en bestämd person besvarar dem bör undvikas
- räkna med att uppgiftslämnaren kan bläddra både framåt och bakåt i blanketten innan han/hon besvarar en fråga. Svaret kan alltså påverkas inte bara av de frågor som föregår utan även av de som följer efter den aktuella frågan till skillnad från i intervjuundersökningar.

Anvisningar i postenkäter till företag måste tydligt ange vilken befattning eller kompetens den som svarar på frågorna ska ha. Man kan inte utgå från att mottagaren av blanketten lämnar över till den som har störst kännedom om ämnesområdet. När inte den ”mest lämpade personen” själv svarar, kan tillförlitligheten försämrats avsevärt.

Elektroniska blanketter

Elektroniska blanketter, dvs. blanketter som är tillgängliga på en webbplats eller distribueras över nätet eller på diskett, är i dag mycket vanliga i administrativ användning, t.ex. vid skadeanmälningar till försäkringsbolag, vid bankaffärer, inom sjukvården och inom enskilda företag. De utnyttjas i växande omfattning vid insamling av data för statistikframställning. Blanketter kan läggas ut på en webbsida eller distribueras via telenätet. Elektronisk distribution över nätet kommer sannolikt att öka snabbt när problem med säkerhet och sekretess blivit lösta på ett juridiskt hållbart sätt.

Vid byte från pappersblankett till elektronisk översätter man inte blanketten till den nya tekniken utan använder också dess möjligheter. Elektroniska blanketter kan vara mer eller mindre ”intelligenta” och i olika utsträckning ha inbyggda

- granskningsvillkor och felsignaler
- möjligheter att rätta felsignalerade svar och att registrera andra ändringar
- rutiner för summeringar och andra beräkningar
- hoppinstruktioner, som utförs automatiskt
- informationsfält, som visas på skärmbilden när man klickar på den aktuella frågan.

En elektronisk blankett kan göras mer komplex än en pappersblankett utan att uppgiftslämnarbördan behöver öka. Samtidigt kan konsekvensen mellan svaren förbättras. Blankettkonstruktören måste utveckla och skriva sina frågor i block som passar till datorskärmens format. Men det är troligt att de som använder elektroniska blanketter även vill kunna göra pappersutskrifter för att få överblick eller för att kunna göra anteckningar innan de svarar. Alla uppgiftslämnare kommer inte att ha möjlighet eller inte vilja fylla i elektroniska blanketter. Även om det är problematiskt att få överensstämmelse mellan blankett på skärmbild och blankett i pappersversion, kan problemet lösas. För engångsundersökningar kan programmeringskostnaderna för att utnyttja den elektroniska blankettens möjligheter bli orimligt höga.

Vid sidan om de insamlingsmetoder som tas upp här finns många andra. För insamling av ett fåtal numeriska uppgifter från varje uppgiftslämnare använder SCB Touchtone Data Entry (TDE) i löpande företagsundersökningar. Sifferuppgifterna kan kompletteras med att kommentarer talas in. Tekniker för direkta uttag ur företagens administrativa system är under utveckling, men de försenas av att datamognaden varierar starkt och att det förekommer många olika ADB-system vid företagen.

4.4.2 Telefonintervju

I undersökningar riktade till individer där bara urvalspersonen själv kan eller bör besvara frågorna rekommenderas intervju. Postenkäten är då mindre lämplig, eftersom svaren i en sådan kan komma att påverkas eller besvaras av t.ex. familjemedlemmar eller arbetskamrater. Intervjutiden i telefonintervjuer bör helst inte överstiga 30 minuter. Om längre intervjutid krävs, måste man överväga att välja andra datainsamlingsmetoder - besöksintervjuer eller postenkät. Telefonintervjuer som tar upp ett ämne som är angeläget för uppgiftslämnarna (t.ex. frågor om barnomsorg till småbarnsföräldrar i en viss kommun) fungerar oftast mycket bra även med en längre intervjutid.

Telefonintervjuns mättekniska fördelar är att

- datorstöd finns utvecklat för metoden
- få eller inga problem uppstår med filterfrågor och hoppanvisningar. När intervjun är datorstödd, sköts hoppen automatiskt och intervjuarfelen minskar (intervjuaren kan dock fortfarande råka markera fel svarsalternativ och på så sätt orsaka ett felaktigt hopp).
- det går utmärkt att ha flera och olika följdfrågor till olika grupper av uppgiftslämnare eller till följd av tidigare lämnade svar, eftersom hoppanvisningarna inte utgör något problem.

Telefonintervjuns mättekniska begränsningar:

- Frågorna måste vara korta och inte innehålla flera informationsmoment. Korttidsminnet räcker annars inte till, och då vet man inte hur frågan har blivit uppfattad och vad svaret gäller
- Av samma skäl får inte heller svarsalternativen vara många eller långa
- Frågorna måste skrivas och intervjuarna läsa dem på ett sådant sätt att de hindrar intervjupersonen från att svara för snabbt, dvs. innan han/hon hört hela frågan eller alla svarsalternativen.
- Frågor som kräver eftertanke innan man svarar lämpar sig inte lika bra i telefonintervjuer, eftersom tempot under intervjun blir förhållandevis högt.

I vissa undersökningar accepterar man att använda sig av **indirekta intervjuer** för att höja svarsfrekvensen, dvs. låta någon annan än urvalspersonen svara. Om denne inte själv kan besvara frågorna (p.g.a. sjukdom, frånvaro, osv.) kan frågor som rör faktauppgifter om honom, ställas till en närstående som är väl förtrogen med de aktuella förhållandena. Däremot ska t.ex. attitydfrågor registreras som partiellt bortfall. Indirekta intervjuer har större förutsättningar att ge acceptabla resultat för frågor av typen ”hur det brukar vara” än för frågor om ”hur det varit under en speciell period”.

4.4.3 Besöksintervjuer

En relativt dyrbar datainsamlingsmetod är besöksintervjun, dvs. uppgiftslämnaren intervjuas vid ett personligt möte med intervjuaren. Intervjutiden bör helst inte uppgå till mer än en timme. Kvaliteten i svaren riskerar i annat fall att försämrats. Både intervjupersonens och intervjuarens koncentrationsförmåga blir sämre med längre intervjutid. Besöksundersökningar används framför allt om blanketten omfattar många frågor eller är komplicerad och/eller kräver hjälpmedel t.ex. i form av svarskort eller kartor. Andra anledningar till att man använder besöksintervju som datainsamlingsmetod kan vara själva ämnesinnehållet eller platsen för intervjun (t.ex. intervjuer på flygplatser med resande). Besöksintervjun ger frågekonstruktören möjlighet att ta fram en blankett som tar till vara de flesta fördelarna från både postenkäten och telefonintervjun. Nackdelen är risken att intervjuarna genom sitt sätt att läsa frågor och genom kroppsspråket kan ha relativt stor påverkan på hur uppgiftslämnarna svarar. Framför allt tillförlitligheten i svaren på attitydfrågor kan försämrats avsevärt. Även besöksintervjuer är som regel datorstödda.

4.4.4 Datorstödda intervjuer

Att intervjuer numera oftast är datorstödda ger frågekonstruktören nya möjligheter. De främsta fördelarna är att inbyggda kontroller hindrar orimliga och inkonsistenta värden från att bli registrerade och att intervjuaren med hjälp av automatiska hoppanvisningar alltid kommer rätt till nästa fråga oavsett hur många och komplexa hoppanvisningar blanketten innehåller.

Vid långa intervjuer kan översikten bli lidande. Tyvärr är det ofta stora problem (ännu så länge) med att skapa tablåer i datorstödda intervjuer. I pappersblanketter kan ibland flera frågor samlas i en tablå, bl.a. för att ge intervjuaren större översikt över frågorna. För en datorstödd intervju är det bra att göra ett översiktsschema på papper, som blockvis visar vilka frågor som ingår i blanketten. Det gör det enklare för intervjuaren att veta var hon/han är i intervjun, speciellt om det ingår flera s.k. loopar (frågor som bara ställs under vissa förutsättningar). Översiktsschemat är till särskilt stor hjälp om intervjuaren tvingats backa ett par steg i loopen för att rätta en tidigare uppgift.

Den sammanlagda tidsåtgången blir olika beroende på vilken typ av blankett som ska utvecklas. Om pappersblanketter används, måste mycket tid läggas ner på arbetet sedan blanketten kommit tillbaka ifylld. För den datoriserade intervjun går det istället åt mycket tid när blanketten ska programmeras och noggrant testas (alla kombinationer av tänkbara svar måste testas). Att bestämma layout på skärmbilden är också tidskrävande.

4.4.5 Flera datainsamlingsmetoder i samma undersökning

Många undersökningar är ”*mixed mode*”-undersökningar, vilket innebär att man använder mer än en metod för datainsamling. Syftet är ofta att förbättra svarsfrekvensen genom att ge uppgiftslämnarna ytterligare ett sätt att svara på. Andra gånger är det en nödvändighet att utveckla blanketter för olika grupper av urvalspersoner, t.ex. i undersökningar där det ingår personer med funktionshinder (som hörselskadade och synskadade). Allmänt anses att antalet ”modes” kommer att öka i många undersökningar i takt med att insamlingen datoriseras. Detta är ett betydande mättekniskt problem, eftersom samma frågeutformning som regel inte fungerar lika bra för olika insamlingsätt.

Ofta utvecklar man blanketten för en huvudsaklig mätmetod och använder den så gott det går i de övriga. Den mest förekommande ”mixed mode”-varianten är postenkät följt av telefonintervju bland bortfallet på postenkäten. Blanketten utvecklar man oftast som postenkät och sedan intervjuar man med samma blankett. Postenkätens stora fördelar, att kunna ha långa och många svarsalternativ, att kunna ha lite längre förklarande ingresser, blir till stora nackdelar i telefonintervjun. Dessa kan alltså inte utnyttjas fullt ut i en ”mixed-mode”-undersökning. Om man inte dessutom skriver särskilda anvisningar till telefonintervjun, överlåter man åt var och en av intervjuarna att göra improviserade lösningar allteftersom svårigheter uppträder. Risken för stor variation i deras sätt att lösa problemen är betydande. Därvid blir också den korrelerade intervjuarvariansen stor.

Hur man än väljer att genomföra en ”mixed mode”-undersökning, uppstår det alltid vissa problem med tillförlitligheten. Bland annat visar det sig inte sällan att svarens fördelningar är olika då man jämför de olika datainsamlingsmetoderna. I telefonintervjun händer det oftare att en större andel svarande väljer ett svar antingen i början eller slutet bland flera upplästa svarsalternativ, än vad som är fallet bland dem som besvarat frågan i postenkäten och där sett alla svarsalternativen samtidigt. I postenkäten är det vanligare att en större andel väljer svarsalternativ som ”ingen åsikt/känner inte till/vet inte” än vad det är i telefonintervjun. ’Vet inte’-alternativet ska inte läsas upp av intervjuaren. Dessutom vill de flesta människor inte gärna framstå som okunniga eller utan åsikter inför en intervjuare, som väntar på ett svar.

Om en insamlingsmetod i en ”mixed mode”-undersökning är personligt besök, måste frågekonstruktören noga tänka igenom vilka frågeområden som riskerar att påverkas i hög grad av intervjuarbias och ta hänsyn till det. Exempel på sådana frågor är alkoholkonsumtion, utsatthet för våld i hemmet, skattemoral. Tankbara lösningar på problemet är särskild intervjuarutbildning och/eller att intervjuaren lämnar över en blankett med dessa frågor, som uppgiftslämnaren fyller i och ger intervjuaren i ett igenklistrat kuvert eller själv postar.

4.5 Vilka hjälpmedel finns?

4.5.1 Tidigare undersökningar

Om det tidigare har genomförts undersökningar inom samma sakområde, är det till stor nytta att studera frågorna. Men det är viktigt att värdera hur väl frågorna fångade in sakområdet. På Internet finns ibland ytterligare information om många undersökningsområden.

Man bör också ta reda på hur dessa undersökningar ”fungerade på fältet”, dvs. hur de genomfördes och vilken mätkvalitet de hade. Speciellt intressant är om det finns evalveringar och återintervjustudier. Det är även värdefullt att få uppgift om hur stora objektsbortfallen och de partiella bortfallen blev. Forskningsrapporter inom nya metodområden finns att söka på webben.

4.5.2 Registeruppgifter

Uppgifter hämtade från administrativa register används ofta för att ersätta vissa frågor i en blankett. Registeruppgifter kan ibland vara tillförlitligare än uppgiftslämnarnas svar i en undersökning. Det gäller bl.a. uppgifter om händelser som ligger längre tillbaka i tiden. Uppgifter om t.ex. inkomster har visat sig ha bättre tillförlitlighet i registren än då de lämnas av urvalspersonerna. Att använda registeruppgifter betyder att uppgiftsbördan minskar, vilket ibland är en förutsättning för att undersökningen ska kunna genomföras med acceptabel svarsfrekvens. Ibland kan man föredra att avstå från ny datainsamling och nöja sig med en registerstudie. Detta kan bli aktuellt i longitudinella studier och t.ex. om uppgiftslämnarna är svåra att anträffa och obenägna eller oförmögna att lämna uppgifter.

Det är viktigt att göra en avvägning mellan vad undersökningen ska studera och vad den administrativa uppgiften i registret står för. Men det är inte givet att den administrativa registeruppgiften verkligen innehåller hela den efterfrågade informationen eller beskriver den verklighet man vill studera på ett bra sätt. Den kan behöva kompletteras med frågor i blanketten.

Det finns många studier som visar på brister i täckning och relevans i administrativa uppgifter - t.ex. om den svarta sektorn i ekonomin. Antalet brott som kommit till polisens kännedom är betydligt mindre än det antal brott som rapporteras i en brottsofferundersökning. Ett annat problem är när registeruppgifterna inte har samma referensperiod och aktualitet som svaren i en urvalsundersökning. T.ex. kan taxeringsuppgifter hos RSV vara upp till två år äldre än uppgifterna i en nyss genomförd hushållsundersökning.

4.5.3 Djupintervjuer och fokusgrupper

Att använda djupintervjuer och fokusgrupper med personer som representerar populationen är effektiva metoder. Härigenom kan man vidga sitt kunnande om sakområdet eller ”kontrollera att man tänker rätt”.

En **djupintervju eller kvalitativ intervju** inriktar sig på fördjupning och förklaring av orsakssammanhang och bakomliggande faktorer. Metoden används för att identifiera de områden eller variabler som är relevanta för uppgiftslämnaren och bör tas med i blanketten. Om djupintervjun inriktar sig på personens förståelse av begrepp, hur personen uppfattar vissa frågor och på hur personen tänker för att komma fram till sitt svar, kallas intervjun dessutom för **kognitiv**. Det innebär att man gör intervjuer med hjälp av begrepp och metoder från kognitiv psykologi.

Fokusgrupp är en informell diskussion med en liten grupp om sex till åtta personer ur den population som ska undersökas. Personerna ska inte vara bekanta med varandra i förväg. Diskussionen pågår lämpligen i 1 - 1½ timme. Den leds av en "moderator" som, efter en kort introduktion, ser till att samtalet håller sig "till ämnet" och att allas åsikter kommer fram. Moderatoren får inte själv delta i diskussionen med egna åsikter, utan skjuter bara vid behov in förtydligande frågor. Diskussionen spelas in på band och analyseras i efterhand. Man använder sig ibland av observatörer för att fånga upp icke verbala signaler och komplettera moderatorns iakttagelser. Fokusgrupper med företagare håller man med fördel på förmiddagen för att undvika återbud orsakade av oförutsedda händelser på arbetet.

Exempel: Redan på planeringsstadiet av en kvantitativ undersökning som skulle undersöka allmänhetens syn på demokratifrågor, genomfördes fokusgrupper med personer ur populationen. Man ville få svar på om de frågeområden som man räknade med skulle vara viktiga verkligen också var det för "vanligt folk". Fokusgrupperna "avslöjade" också ordval, vilket var en värdefull kunskap att ha, när det blev dags att skriva frågorna till den kvantitativa undersökningen.

Att rekrytera personer till fokusgrupper kräver mycket arbete när det gäller att spåra och kontakta lämpliga deltagare. De ska inte väljas slumpmässigt ur undersökningsskiktet, utan de ska väljas ut omsorgsfullt. Tillsammans ska de representera största möjliga spridning i fråga om bakgrund som är relevant för undersökningen. Personerna ska även vara verbala, ingå i olika sociala nätverk och ha en bra bild av hur andra personer i dessa nätverk tänker och tycker.

Fokusgrupper är ett bra sätt att få nya, "fräscha" synsätt utanför de statistiska undersökningarnas föreställningsvärld. Är gruppen väl sammansatt, upptäcks om något kan verka konstigt för "vanligt folk". Därför är det viktigt att deltagarna inte hämtas från alltför nära håll eller används i flera fokusgrupper. En fördel med fokusgrupper jämfört med djupintervjuer är att det ur samspelet mellan deltagarna uppstår associationer och nya tankespår som individuella djupintervjuer inte ger.

Djupintervjuer och fokusgrupper är mycket användbara metoder för att samla in och få grepp om relevanta frågeområden, både innehållsmässigt och beträffande språkbruk. Genom djupintervjuer och fokusgrupper kan man också på ett enkelt sätt få kunskap om arbetssätt och rutiner i företag/kommuner/organisationer m.m. Allt detta är av stort värde för frågekonstruktören att känna till. I vissa fall visar sig vara nödvändigt att ställa frågor till företag om administrativa rutiner, vilka dataprogram som används o.d. Annars går det inte att bedöma kvaliteten i avgivna svar eller att renodla redovisningsgrupper tillräckligt väl.

4.5.4 Variabellista och tabellplan

När man upprättar variabellista och tabellplan, får man en säker överblick över de frågeområden som frågorna skall avse. Variabellistan är en enkel förteckning, där varje ingående variabel ges ett namn och en kod. En variabel kan sedan ge upphov till en eller flera frågor, beroende på hur svår man bedömer den är att mäta. Tabellplanen går att ställa upp snabbt med hjälp av variabelkoderna. Den gör det möjligt att pricka av om man har fått med just de variabler som behövs för att ge svar på undersökningens sakproblem. Frågeområden i blanketten kan lätt stämmas av mot själva sakproblemet. Tabellplanerna är ett bra hjälpmedel för att se till att nödvändiga bakgrundsvariabler och redovisningsgrupper har kommit med.

VARIABELLISTA

Bakgrundsvariabler

BV1 Kön
 BV2 Ålder
 BV3 Utbildning
 BV4 Befattning

Undersökningsvariabler

UV1 Trafikregler
 UV2 Hjälpmedel
 UV3 Attityder till XX

Variabellista och tabellplan är enkla men viktiga hjälpmedel som bör användas vid varje blankettkonstruktion. På motsvarande sätt upprättar man och stämmer av planer för multivariata analyser eller för grafiska presentationer. Man har inte råd att låta det gå dithän att man först vid analys och presentation upptäcker att man förbisett variabler som i hög grad kunde ha bidragit till tolkningen och förståelsen av resultaten. Möjligheten att komplettera variabellistan under de följande faserna blir allt mer begränsad ju längre man kommit i planeringen. När undersökningen väl har satts i gång är det för sent.

TABELLPLAN Bakgrund BV1**BV2*BV3 BV3**BV2*BV5 BV4**BV1*BV5 Trafik UV1*BV1, BV2 UV2*BV2, BV3 UV3*BV3, BV4

4.6 Checklista

Innan frågekonstruktören övergår till själva blankettkonstruktionen, ska man under fas 1 ha blivit klar med att följande

CHECKLISTA Definiera undersökningens innehåll

1. Definiera informationsbehov och tänkt innehåll i blanketten. Detta innebär att definiera objekt, population och de parametrar man vill skatta.
2. Djupintervjua personer i den tänkta populationen av uppgiftslämnare (ibland även bland användarna) och/eller genomför fokusgrupper.
3. Bestämma urvalsförfarande, urvalsstorlek och datainsamlingsmetod, med hänsyn till kvalitetskrav, ekonomiska och andra resurser samt tidspress.
4. Undersöka vilken information om undersökningsområdet som redan finns i form av registerinformation och tidigare urvalsundersökningar. Försöka ta reda på vilken mätkvalitet dessa har och samla goda exempel på frågeområden, enstaka frågor, anvisningar och instruktioner.
5. Avgöra om undersökningen är genomförbar eller inte.
6. Sammanställa en variabellista och en tabellplan, dvs. kontrollera att allt finns med.

Slutprodukten av fas 1 är en variabellista och tabellplan. Det ska finnas beslut om kategorin av uppgiftslämnare, om hur urvalet dras och om vilken datainsamlingsmetod som ska användas. Man ska ha klart för sig vilken information från administrativa register som ska användas i stället för att fråga uppgiftslämnarna direkt.

För en helt **ny undersökning** måste arbetet starta i **Fas 1**. Redan här måste man avgöra om det är meningsfullt att genomföra undersökningen. Ju mer obekant det ämnesområde är som ska undersökas, desto större betydelse och omfattning får de inledande faserna.

5 Fas 2 Blankettkonstruktion

5.1 Inledning

Arbetsuppgifter: Denna fas handlar om att forma blanketten fråga för fråga. Frågeområdena ska alltså översättas till konkreta frågor, anpassade till den valda metoden för datainsamling. Definitioner och svarsalternativ ska fastställas, frågeordning och layout ska bestämmas, informationsmaterial och anvisningar ska skrivas. Erfarenheten har visat att vissa variabler inte kan mätas direkt. I stället måste man bryta ner dem och ställa flera frågor och bestämma regler för hur svaren på dessa skall kodas. Exempel på sådana variabler är arbetskrafts status i AKU och långvarigt sjuk i ULF. Den avsedda slutprodukten är en så bra blankett som möjligt.

- Arbetsfördelningen inom detta område växlar. Många beställare lämnar över en preliminär version av blanketten och arbetar tillsammans med frågekonstruktören fram den slutliga ”skrivbordsversionen”. Detta är ofta ett fördelaktigt arbetssätt, eftersom alla parter tvingas tänka igenom frågeuppsättningen och ta ett gemensamt ansvar för den. I andra fall överlämnar beställaren blankettkonstruktionen helt åt frågekonstruktören. Beställaren måste dock alltid godkänna den slutliga versionen.

5.2 Allmänna tips till frågekonstruktören

Den perfekta blanketten, där alla uppgiftslämnare entydigt förstår alla frågor och utan problem kan besvara dem på ett riktigt sätt, kommer aldrig att konstrueras. Men man bör självfallet sträva efter det. Redan när man sitter vid skrivbordet för att göra det första blankettutkastet, kan man se till att undvika de vanligaste feltyperna. Man behöver då inte använda resurser på ett eller flera test för att upptäcka misstag som man borde ha sett redan före testet. Förutsättningar för att en blankett ska fungera bra är att den har en tydlig och motiverande introduktion, är lätt att läsa, förstå och fylla i och att uppgiftslämnaren leds på ett tydligt sätt genom blanketten.

En frågas formulering och dess placering är avgörande för hur uppgiftslämnarna uppfattar den och besvarar den. Här följer en lista över några grundläggande språkliga och organisatoriska tumregler.

Ordval

- Undvik negationer
- Tänk på tempus. Presens för pågående verksamheter och imperfekt för avslutade.
- Undvik förkortningar, svåra ord och facktermer. Förklara deras innebörd om de trots allt måste användas.
- Använd ord och formuleringar som uppfattas som neutrala. Annars kan frågan uppfattas som positivt eller negativt laddad och bli ledande.

Precisera

- Specificera sort och önskad noggrannhet i svaren
- Ange referenstidpunkt/-period tydligt. Byt inte referenstidpunkt/-period mellan olika frågor om det inte är absolut nödvändigt.
- Ange vilka komponenter som ska inkluderas och vilka som inte ska det (*t.ex. 'inrikes försäljning', 'exklusive moms', 'under 3:e kvartalet 1999'*).

Syntax och begriplighet

- Använd enkelt och klart språk - även när du frågar fackfolk!
- Ställ korta och konkreta frågor. (I långa komplexa frågor kan det vara svårt att urskilja vilken del som uppgiftslämnaren ska besvara.)
- Undvik meningar med många och långa ord.
- Formulera frågan så att uppgiftslämnarna kan ge ett entydigt svar
- Se upp med abstraktioner och hypotetiska frågor.
- Fråga om en sak i taget - undvik summerande frågor.
- Se över frågor som innehåller bindeord som ”och”, ”samt”, ”eller”. Frågan riskerar då att innehålla flera frågor. En sådan fråga bör delas upp i flera.

Ordning/orientering

- Placera en anvisning så nära frågan som möjligt. Begränsa längden.
- Tänk på att få en logisk följd mellan frågorna.
- Dela upp blanketten i logiska block.
- Förse blocken med titlar och dela upp blanketten i avsnitt, även grafiskt. Detta gäller särskilt s.k. omnibussundersökningar, som ofta har tvära övergångar mellan frågeområdena och behöver förmedlande text mellan dessa.
- Använd grafiska signaler för att visa var man hittar anvisningar, viktiga begrepp och den logiska vägen genom blanketten.
- Placera om möjligt frågor som kan vara svåra att besvara i slutet av enkäten. Viktigt att uppgiftslämnaren inte ”kör fast” på de första frågorna.
- Placera frågor som riskerar att uppfattas som känsliga eller integritetskränkande men ändå måste vara med i slutet av blanketten.

Övrigt

- Undvik oförmedlade kast mellan frågor med olika typer av krav på databildningen, dvs. mellan avskrift, bearbetning och ”svar direkt”.
- Undvik att byta riktning och läge vid placeringen av svarsrutor.
- Använd samma indelning i svarsalternativ så långt som möjligt.
- Ta aldrig med frågor bara för att ”de kan vara bra att ha”.

5.3 Tänk på uppgiftslämnaren

*Filling in a questionnaire is not a priority
(Allen Gower)*

Uppgiftslämnaren svarar ofta snabbt och många gånger med bristande engagemang. Man ska inte tro att den genomsnittlige uppgiftslämnaren anstränger sig särskilt mycket för att förstå en oklar eller svår fråga. I stället svarar man så gott man kan på det man tror frågan gäller. Klart språk och entydigt innehåll är ett måste för att få en tillförlitlig uppgift.

Tänk från början på vilka som verkligen ska besvara blanketten och inte bara vem blanketten är ställd till. Frågekonstruktören behöver inse i hur hög grad uppgiftslämnarnas utbildning, kunskaper i svenska, vana vid att uttrycka sig själva i skrift osv. varierar - och ta hänsyn till detta vid den språkliga utformningen. Behovet av definitioner eller förklaringar måste anpassas efter urvals-personernas förutsättningar.

Man måste vidare veta om de verkliga uppgiftslämnarna har kännedom och nödvändiga fackkunskaper om det som frågorna handlar om. De måste förstå vilka uppgifter de förväntas lämna och de måste kunna få fram uppgifterna ur sin erfarenhet, minne, räkenskaper, dagböcker, olika delar av verksamheten, osv. Man måste göra klart för sig vilken typ av databildning man kräver och hur stor uppgiftslämnarbördan egentligen är. I exemplen beskrivs situationer där kraven inte hade anpassats till uppgiftslämnarnas förutsättningar.

Exempel 1:

I en blankett ställd till kommuner frågade man om förskoleverksamheten. Vid en genomgång i efterhand med uppgiftslämnarna i några kommuner frågades också hur uppgifterna hade samlats in. Det visade sig att i många kommuner hade en centralt placerad person fått i uppgift att skicka kopior av blanketten till de olika barnstugorna. Den personen samlade också in dem, sammanställde svaren och skickade slutligen in till SCB. Det framkom att det varit omöjligt för denne att avgöra om svaren från de enskilda barnstugorna var rimliga eller ej, eller om något blivit missförstått. Den centralt placerade personen hade inte tillräcklig kännedom om de enskilda barnstugorna för att kunna avgöra detta.

Exempel 2:

I ett test av hälsofrågor som gjordes med barn i årskurs 4, visade det sig att de benämningar på sjukdomar eller besvär som uppdragsgivaren och frågekonstruktören befarade skulle behöva förklaras inte innebar några problem alls. Andra, skenbart enkla, ord var däremot svåra att förstå. Ett sådant exempel är "yrsel", som barnen inte alls kände till. Det fick bytas ut mot "yr i huvudet".

I undersökningar där **undersökningsobjekten är individer** är dessa som regel själva uppgiftslämnare. I sådana undersökningar behöver man väga samman informationsbehovet och en bedömning av vad uppgiftslämnarna kan klara av att svara på med tillräcklig säkerhet. Den som ska svara på frågorna förutsätts ha svaren (uppgifterna) "i huvudet", t.ex. livshistoria, konsumtion, minne av faktiskt handlande, kunskaper, åsikter, attityder. När frågorna handlar om hela hushållet, antar man att de kan besvaras av mer än en person i samma hushåll. I undersökningar av t.ex. utgifter, inkomster, resvanor och tidsanvändning uppmanar man uppgiftslämnaren att använda sig av dagbok, deklarations-

blankett, kassakvitton, försäkringsavier, taxeringsuppgifter m.m. för att få fram tillförlitliga uppgifter.

I förstagångsundersökningar där urvalsenheten är företag/kommuner/organisationer m.m. är det ofta ett problem att veta vilken person eller funktion som blanketten ska adresseras till, även då man vet vilken kompetens som krävs för att besvara den. Ett vanligt sätt att försöka nå rätt uppgiftslämnare är att adressera till en person med en viss befattning, t.ex. VD, inköpschef eller personalchef. Men oavsett vem blanketten adresseras till, är det inom företaget man avgör vem som ska besvara blanketten. Utgå inte från att uppgiftslämnaren automatiskt är en person med bra kunskaper om ämnesinnehållet.

I löpande undersökningar strävar man efter att etablera kontakt med de personer som har bra förutsättningar att lämna uppgiften. Särskilt viktigt är detta för stora företag, vars svar har stor genomslagskraft i statistiken.

I undersökningar med uppgiftslämnarplikt kan uppgiftslämnarbördan vara betydligt större än i undersökningar med frivilligt deltagande. Kraven på att använda mättekniskt väl utformade blanketter är därför om möjligt större i undersökningar med uppgiftslämnarplikt.

Svarens kvalitet är beroende av den tid som uppgiftslämnaren måste lägga ner på att få fram korrekta uppgifter. I företagsundersökningar är det uppgiftslämnarens arbetstid. Arbetsgivaren avgör hur mycket tid som kan avsättas. I undersökningar där många av uppgiftslämnarna är småföretagare finns risk för att en hög uppgiftslämnarbörda medför låg tillförlitlighet i svaren.

Samma person kan få besvara flera olika blanketter som skickats ut från såväl SCB som andra myndigheter, branschorganisationer och frågeinstitut. Det är en fördel att uppgiftslämnaren med tiden i en mening blir ”professionell” och får stor vana att förstå begrepp och fylla i blanketter. Å andra sidan ökar risken att uppgiftslämnaren lämnar samma uppgifter om flera undersökningar efterfrågar liknande, *men inte identiska*, uppgifter. Samordning och användning av standardiserade frågor minskar uppgiftslämnarbördan och ökar relevansen i svaren.

5.4 Vanliga feltyper i blanketter till individer

Det är svårare att beskriva hur man gör en bra blankett än att peka på frågekonstruktioner och blankettutformningar som fungerat dåligt. På detta område gäller i hög grad att det är av andras och egna misstag man lär sig vad som fungerar. Därför redovisas här några vanliga typer av fel som kan undvikas redan vid konstruktionen vid skrivbordet.

Definiera frågorna i tid och rum

Exempel

Läser du någon kvällstidning?

Frågan saknar förankring i tiden och kan tolkas på flera sätt. Vissa uppgiftslämnare tror att frågan gäller om man läser kvällstidningar så gott som varje dag och svarar ’Nej’ om de bara brukar läsa dem på lördagar och söndagar. Andra uppgiftslämnare i samma situation svarar ’Ja’, eftersom de ju läser kvällstidningar på helgerna. Hur frågan bör se ut avgörs av syftet med den. Om man t.ex.

vill identifiera dem som läser kvällstidningar mest varje dag, kan frågan ha lydelsen:

”Läser du någon kvällstidning minst fem dagar i veckan?”

Exempel

”Hur länge har du bott här?”

Här saknas rumslig förankring och därför är många tolkningar möjliga. 'Här' kan tolkas som bostaden, villaområdet, kommunen, Sverige. I frågetexten behöver alltså betydelsen av 'här' finnas med.

Uppgiftslämnaren har en annan referensram än undersökaren

Exempel

”Har du någon långvarig sjukdom, besvär efter olycksfall, något handikapp eller annan svaghet?”

Frågeformuleringen kan ibland, som i exemplet, förleda en del uppgiftslämnare att utelämna långvariga sjukdomar/besvär som ska med. I frågan ingår flera ord som kan ge intryck av att endast grava besvär efterfrågas. Det gör att personer med lättare former av åldersdiabetes, högt blodtryck eller psoriasis ibland svarar nej på frågan. Att intervjupersonen har en annan referensram än undersökaren är vanligast om personen sällan eller aldrig har/har haft besvär av sin långvariga sjukdom, eftersom den antingen inte har några svåra symtom eller hålls tillbaka av en väl fungerande medicinering.

Flera frågor blandas ihop i en

Exempel

I en undersökning om arbetsmiljö uppmanas uppgiftslämnarna att beskriva hur de upplever sina arbeten. De får följande instruktion:

Här nedan har vi ställt upp några rader med rutor, som går från den ena ytterligheten till den andra. Beskriv hur det vanligen är för dig genom att sätta ett kryss på varje rad. Ju längre till vänster du sätter ditt kryss, desto mer stämmer det som står till vänster. Ju längre till höger du sätter ditt kryss, desto mer stämmer det som står till höger. Näst längst ut innebär alltså att du bara delvis håller med.

Alldeles för mycket att göra	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Varken eller	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Alldeles för lite att göra
Alldeles för litet ansvar	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Alldeles för stort ansvar
Enformigt arbete	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Omväxlande arbete
Psykiskt påfrestande arbete	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>		<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Lugnt och behagligt arbete

fler frågor följer

De första två frågorna är båda egentligen två frågor i en. I den ena handlar det om i vilken grad man har för mycket att göra (resp. för litet ansvar) och i den andra i vilken grad man har för litet att göra (resp. för stort ansvar). Den som

varken har för lite eller för mycket att göra, sätter ett kryss i mittrutan. Det har visat sig att även den som ibland har för lite och ibland för mycket att göra sätter sitt kryss i mitten, dvs. väger samman en arbetssituation som pendlar mellan ytterligheterna. Det går alltså inte att uppskatta hur stor andel som menar sig ha lagom mycket att göra.

Uppgiftslämnaren kan dessutom förvirras av att de upplever att båda ytterligheterna beskriver negativa förhållanden. Andra rader (fråga 3 och 4) går i stället från en negativ ytterlighet till en positiv.

Det här är dessutom ett exempel på en instruktion som behöver skrivas enklare och klarare om den ska underlätta för uppgiftslämnaren att besvara frågan.

Exempel

Ett av USA:s största opinionsmätninginstitut, The Harris Poll, ställde faktiskt frågan: *”Har du ofta, ibland, nästan aldrig eller aldrig haft skuld känslor därför att du varit otrogen mot din fru?”*

1 % svarade ofta, 14 % ibland eller nästan aldrig och 85 % att de aldrig hade skuld känslor på grund av sin otrohet. Här fanns en underförstådd (filter-)fråga om man varit otrogen eller inte. För dem som skulle ha svarat 'Nej' på denna borde det funnits en hoppanvisning. Frågekonstruktionsfel av denna typ är ofta förrädiska och inte alls lika uppenbara som i exemplet.

Filterfrågor fungerar betydligt bättre i intervjuundersökningar än i postenkäter. I de senare bör man undvika filterfrågor och hoppanvisningar, eftersom de ofta missförstås och i stället kompletteras med ytterligare ett svarsalternativ (i det här fallet att man aldrig varit otrogen mot sin fru.)

Exempel

Följande fråga ställdes till barnavårdscentraler:

*”Genomför ni i dag test/undersökning/observation av **samtliga** barn i andra åldersgrupper än fyraåringar, för att fånga upp barn med MBD-problematik?”*

Svarsalternativen var 'Nej' och 'Ja'.

Även om den här frågan egentligen bara frågar efter **en** sak, är den uppbyggd av flera delfrågor, som också går att besvara med 'Ja' eller 'Nej'. Hur noggrant olika uppgiftslämnare läser frågorna varierar. Om frågan innehåller flera delar, som går att besvara, är det därför svårt att veta vad man fått svar på. Bästa lösningen är att dela upp frågan i flera delfrågor.

Ledande frågor och laddade ord

Exempel

”Fler människor har sett filmen 'Borta med vinden' än någon annan film som producerats det här århundradet. Har du sett den?”

Formuleringar som denna antyder att uppgiftslämnarna på något sätt avviker från ”normen” om de svarar nej, och är därigenom ledande. Frågor ska vara neutrala eller balanserade, dvs. om ett svarsalternativ motiveras i frågan, så bör övriga alternativ också motiveras.

En ledande fråga gör det lättare att välja ett svarsalternativ framför ett annat på grund av själva frågeformuleringen, t.ex. *”Tycker du själv att du är positivt*

inställd till ... ?” Här stämmer ett ’ja’-svar med ordvalet i frågan, medan ett ’nej’-svar tycks motsäga själva meningen med frågan. Frågan blir betydligt mer balanserad om den skrivs: *”Är du positivt eller negativt inställd till ... ?*” Frågor blir också ledande om man utnyttjar människors vilja att föredra status quo, spelar på prestige eller använder sig av välkända personers eller organisationers namn. I stället för att skriva t.ex. *’Framstående forskare som XX menar att ... Vilken är din åsikt?’* är det bättre att skriva *’Vissa forskare menar ..., medan andra forskare Vilken är din åsikt?’*.

Exempel

I ett metodexperiment ställde ett surveyinstitut i Sverige följande två närliggande frågor.

Den första versionen var:

”Inom EU arbetar man för att bilda en valutaunion EMU, med gemensam valuta för de länder som är med i EU. Är du för eller emot att Sverige går med i EMU?”

Svarsfördelningen blev: ”För” 38 %, ”Emot” 48 %, ”Vet ej” 13 %.

Den andra versionen löd:

”Sannolikt kommer Belgien, Holland, Luxemburg, Italien, Portugal, Spanien, Irland, Frankrike, Österrike, Tyskland och Finland gå med i EMU från start. Om det blir så, tycker du då att Sverige också bör gå med i EMU eller tycker du inte det?”

Svarsfördelningen blev: ”Bör gå med” 50 %, ”Bör inte gå med” 42 %, ”Vet ej” 8 %.

Frågorna är inte exakt lika i innehåll. Det är inte heller svarsförutsättningarna. Version 2 förutsätter att man vet vad EMU är. Frågorna kan alltså inte förväntas ge exakt samma resultat. Men skillnaden i svarsfördelning visar tydligt hur man kan ge olika bilder av opinionen genom att ge olika förutsättningar för valet av alternativ. Särskilt kan osäkra uppgiftslämnare påverkas och deras svar styras i önskad riktning.

Vaga frågor och vaga svarsalternativ

Exempel

”Har du någon bestämd läkare som du brukar vända dig till?”

Begreppet ’bestämd läkare’ är alltför vagt. Det framgår inte om man menar just doktor Nilsson, eller att man har en husläkare eller att man regelbundet besöker en specialistläkare som t.ex. ögonläkare eller gynekolog. Ordet ’brukar’ ställer till bekymmer för dem som är listade hos en husläkare men aldrig har behövt gå dit.

Exempel

”Hur ofta gick du i kyrkan förra året?”

Svarsalternativen är: ’Aldrig’, ’Sällan’, ’Då och då’, ’Regelbundet’.

I det här exemplet är svarsalternativen alltför vaga. Om olika personer blir tillfrågade vad ord som ’sällan, då och då, ibland eller ofta’ innebär för dem, får man mycket varierande svar på vad orden betyder i praktiken. Innebörden i orden varierar inte bara mellan olika personer, utan också beroende på vad det är

man frågar om. Det är stor skillnad på hur många gånger 'ofta' förväntas vara om man frågar om någon ofta äter glass på sommaren eller om någon ofta har öroninflammation.

I exemplet ovan kan dessutom ordet 'regelbundet' tolkas på flera sätt. Det kan tolkas både som att man gått i kyrkan med jämna mellanrum under året, men också att man har följt sitt över året regelbundna beteende och gått i kyrkan t.ex. till påsk, advent och julottan.

I stället för dessa vaga svarsalternativ är det bättre att konstruera svarsalternativ med tidsreferenser, dvs. antal händelser per dag, vecka, månad eller år. Man får dock inte glömma att ta hänsyn till att de flesta personer uppfattar mitten på skalan som ett "normalvärde", t.ex. så många timmar som en genomsnittstittare ser på TV under en vecka. Man placerar in sig själv efter detta om man inte kan svara med ett mer precist svar. Med olika skalor kommer svaren att fördela sig på olika sätt.

Frågan begär för mycket av uppgiftslämnaren

Frågor ska skrivas så att det är möjligt för uppgiftslämnaren att komma fram till ett svar och kunna lämna detta. Beställaren kan ha stort intresse av att få detaljerad kunskap om speciella fenomen. Frågekonstruktören måste då skriva frågor på ett sådant sätt att man inte begär för mycket av uppgiftslämnaren, men ändå får så mycket information som möjligt.

Exempel 1

53.	Vilken typ av ost brukade Du äta för cirka 10 år sedan? Ange hur mycket per dag, per vecka eller per månad?					
	<input type="checkbox"/> Jag åt sällan eller aldrig ost → <i>Gå vidare till fråga 54</i>					
	skivor/dag		skivor/vecka		skivor/mån	
Ost, 24 % fett eller mera	<input type="text"/>	eller	<input type="text"/>	eller	<input type="text"/>	
Mager ost, 17% fett eller mindre	<input type="text"/>	eller	<input type="text"/>	eller	<input type="text"/>	
Dessertost	<input type="text"/>	eller	<input type="text"/>	eller	<input type="text"/>	
	matskedar/dag		matskedar/vecka		matskedar/mån	
Mjukost (1 dl = 7 matskedar)	<input type="text"/>	eller	<input type="text"/>	eller	<input type="text"/>	
Keso, kvarg (1dl = 7 matskedar)	<input type="text"/>	eller	<input type="text"/>	eller	<input type="text"/>	

Här är det inga problem att förstå frågan, men kan alla finna ett svar? Även om man finner ett svar - kan alla avge det? Hur svarar t.ex. den som minns att han, för 10 år sedan, ibland själv gjorde sin frukost, (men hur ofta var det?) och då för det mesta åt 4-6 smörgåsar med ost på, men när hans eller hennes partner gjorde frukosten, brukade pålägget variera?

Frågan är hämtad från en undersökning som vände sig till personer som hade en viss form av cancer. Det fanns också en kontrollgrupp som saknade sjukdomen. Eftersom man tror att det finns ett samband mellan sjukdomen och de matvanor

man hade för 10 år, 20 år eller ännu längre tillbaka i tiden, är det väsentligt att få så exakta svar som möjligt. För varje fråga som frågekonstruktören skriver, måste han/hon ställa sig frågan om just den valda frågeutformningen är den bästa för att uppnå syftet med frågan. I det här exemplet fick man en mycket stor andel 'vet ej'-svar. Med en annan utformning hade förmodligen en större andel av uppgiftslämnarna kunnat avge ett exaktare svar. Ett sätt vore att först ställa några frågor, vars uppgift är att få uppgiftslämnaren att så väl som möjligt minnas sitt liv vid den här tiden – 'var det under studietiden? när barnen var små? när jag jobbade som ... ?' Först därefter närmar man sig vilka matvanor personen hade. Till sist kan man ställa mer detaljerade frågor. Ytterst sällan kan dock frågor besvaras så exakt som önskas i exemplet, efter så många år. Minnet färgas av vanor som ligger närmare i tiden.

Var försiktig med Ja-Nej-frågor när det gäller attityder, åsikter och värderingar

Exempel

Två olika frågor är tänkta att mäta populariteten för statsministern, men resulterar i olika stora positiva andelar.

"Tycker du att statsminister XX gör ett bra jobb?"

Svarsalternativ: Ja, Nej.

"Hur tycker du att XX är som statsminister?"

Svarsalternativ: Bra, Ganska bra, Ganska dålig, Dålig.

Att det är lättare att hålla med och säga 'Ja' än att säga emot är en mänsklig egenskap som ofta utnyttjas i attityd- och opinionsundersökningar. Det betyder i det här exemplet att den första frågan får en större andel positiva svar än den andra frågan. Det är en bättre fråga rent mättekniskt genom att den tillåter den svarande att redovisa en mer nyanserad uppfattning.

Exempel

"Anser du att det vore riktigt att höja skatten för att de långtidssjuka som vill ska få ha ett eget rum på vården?"

Svarsalternativen är 'Ja' och 'Nej'.

Förutom 'Ja-Nej'-problematiken, försvåras den här frågan ytterligare av att den är hypotetisk. Det är oerhört svårt att mäta åsikter som skulle hålla om det hypotetiska förslaget blev verklighet. Eller att mäta sannolikheten för ett framtida beteende som efterfrågats med en hypotetisk fråga. Eftersom de hypotetiska frågorna inte förpliktar till någonting, är det mycket lättare att hålla med om något än att säga emot. Detta gäller särskilt om det dessutom är socialt mer acceptabelt att hålla med.

En sådan attitydfråga bör hellre delas upp i flera frågor som tillsammans kan visa vilken attityd den svarande har. Uppgiftslämnarna bör bl.a. få möjlighet att kunna uttrycka att de håller med i sakfrågan, men anser att det kan ske utan skattehöjning.

Tablåer och ”instämmer i”-frågor

Erfarenheten har visat att tablåer med flera dimensioner sällan blir fullständigt ifyllda, hur praktiska och logiska de än kan förefalla för frågekonstruktören. Många uppgiftslämnare gör bara markeringar i ’ja’-kolumnen i tablå. De kryssar för t.ex. de besvär, sjukdomar eller symtom de har haft, men bryr sig inte om att kryssa för ’nej’ eller ’vet inte/minns inte’ där något av de alternativen är det rätta. Resultatet blir ett stort partiellt bortfall i stället för vetskap om vilka som inte har haft besväret, sjukdomen eller symtomet. I regel är det bäst att ställa varje delfråga som en självständig fråga.

I en speciell typ av tablåer utgörs ”frågorna” av påståenden som uppgiftslämnaren i olika hög grad ska instämma i eller ta avstånd från.

Exempel:

	Instämmer helt	Instämmer delvis	Vet inte	Tar delvis avstånd	Tar helt avstånd
Jag riskerar hälsan på jobbet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jag får tillräckligt stöd i arbetet av min chef	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

med flera påståenden

Att göra en sådan här fråga mättekniskt bra är ingen lätt uppgift. Tablåen visar på flera svårigheter. Den första är det olämpliga i att använda två skalor, en för ’instämmande’ och en för ’avståndstagande’. Frågekonstruktören måste noggrant tänka igenom vad det innebär att bara delvis instämma eller delvis ta avstånd. Kanske är den verklighet som ryms i att delvis instämma precis samma som ryms i att delvis ta avstånd? I just det här exemplet finns dessutom ett helt felaktigt mittalternativ, eftersom ’vet inte’-alternativet inte ingår i skalan utan ligger helt utanför som ännu ett möjligt svarsalternativ.

De påståenden som ingår i tablåen måste skrivas med stor eftertanke om det ska vara möjligt att kunna lämna entydiga svar i alla svarsalternativ. Framför allt ska man försöka undvika adjektiv som ’tillräckligt’, ’bra’ och liknande som beskriver det man frågar efter. Det är svårt för uppgiftslämnarna att veta vad det innebär att delvis instämma i att man får ’tillräckligt stöd’, eftersom detta egentligen är en dikotom variabel - antingen är stödet tillräckligt eller också inte. Om ordet ’tillräcklig’ tas bort i exemplet kan påståendet lättare graderas av uppgiftslämnaren.

Flera tolkningar av frågans innebörd

Följande exempel visar att det inte är lätt att i förväg upptäcka fel - trots granskning av blankett och provintervjuer med debriefing i det här fallet. Ett kognitivt blanketttest hade däremot kunnat upptäcka denna typ av svårighet.

Exempel

Har du någon gång varit hemma minst 6 månader i följd för att ta hand om egna barn (inkl. partners barn, adoptivbarn och fosterbarn)?

Genom att en tränad observatör följde med intervjuaren till ett flertal besöksintervjuer upptäcktes att personer som varit hemma redan innan de fick

barn, svarade både 'ja' och 'nej' på frågan. Vissa tolkade den som att de skulle ha stannat hemma speciellt för att ta hand om barnet och svarade 'nej', eftersom de redan var hemma. Andra i samma situation svarade 'ja', därför att de ju var hemma och tog hand om sitt barn. Resultatet av en sådan tvetydighet i svaren blir att man inte kan veta vad statistiken står för.

Exemplet är lärorikt genom att det visar på begränsningen i skrivbordsmetoder och tumregler. Det kan finnas saker man missar, även om inget verkar misstänkt på papperet. Det går med andra ord inte att ersätta eller utesluta grundliga kvalitativa studier om man vill belysa hur frågorna fungerar i en undersökning.

5.5 Speciellt för företagsundersökningar

Många undersökningar riktade till företag innehåller frågor om planer, ambitioner och attityder till företeelser i samhället. Här gäller samma förutsättningar vid frågekonstruktion som för motsvarande undersökningar till individer.

För undersökningar som huvudsakligen frågar om kvantitativa storheter och riktas till företag, kommuner, organisationer m.m., konstrueras frågorna med förutsättningar som är annorlunda än för flertalet blanketter till individer. De uppgifter som efterfrågas finns som regel att hämta i den bokföring och verksamhetsredovisning som företagen/kommunerna/organisationerna m.fl. har skyldighet att föra. Variablerna är oftast kvantitativa - volymer, mängder, antal, krontal, osv. Uppgiftslämnaren ombeds att fylla i blanketten med hjälp av benämningar på de efterfrågade uppgifterna ('Omsättning', 'Levererad kvantitet', 'Antal anställda', osv.).

När frågorna ska besvaras med tal av varierande storlek, men i en bestämd sort, är en vägledande utformning av svarsrutorna en viktig del av blankettens layout, om man ska få tillförlitliga svar. Risken för bristande noggrannhet, avrundningar, sortfel och positionsfel ökar med dålig utformning av blanketten, t.ex. "för små svarsrutor". Ofta måste layouten anpassas till att blanketten ska skannas.

När en blankett ska göras till en företagsundersökning, är en förutsättning ofta att blanketten ska vara komprimerad. Skälen till detta är att man vill reducera mängden papper som hanteras vid bl.a. förprintning av uppgifter, vid utskick och vid registrering av svaren, speciellt då de ska skannas. Utrymmet för anvisningar och förklaringar blir därför mycket begränsat i själva blanketten. Då måste ett mer eller mindre omfattande informationsmaterial skickas ut tillsammans med blanketten. Detta går tvärtemot den mättekniska ambitionen, som är att utveckla luftiga blanketter med anvisningarna tryckta nära frågorna. Erfarenheten är att information och anvisningar till en fråga läses oftare då de är tryckta nära intill frågan än då de finns i en separat broschyr. När man inför elektroniska blanketter, kan man lägga in anvisningarna så att de visas då man klickar på frågan. (Se mer om skanning i avsnitt 5.8.)

Ibland förprintas företagets uppgifter från föregående undersökningstillfälle. Det ska underlätta för uppgiftslämnaren att svara på rätt frågor och att ge svaren i rätt storleksordning. Man kan inte i dag säga i vilken utsträckning förprintning fungerar som tänkt. Det kan finnas risk att inte upptäckta fel konserveras och att uppgiftslämnaren inte lägger ner tid på att bedöma om status quo gäller eller om en förändring har skett, t.ex. om företagets huvudsakliga verksamhet är densamma som föregående år eller inte.

5.6 Frågor för jämförelser över tiden

Många undersökningar görs i avsikt att jämföra resultaten med tidigare undersökningar av samma ämne. I sådana fall är det givet att man överväger att använda samma frågor. Man hoppas kunna mäta en eventuell förändring på ett så tillförlitligt sätt som möjligt. Men det finns tre problem som frågekonstruktören måste ta hänsyn till för att avgöra om återanvändning och jämförelse är möjlig.

- Är frågorna fortfarande lika relevanta? Duger samma svarsalternativ och anvisningar?
- Gav frågorna tillförlitliga resultat då de användes?
- Har språket förändrats så, att man måste välja andra ord och formulera sig på ett nytt sätt för att frågan ska förstås på samma sätt som förra gången?

Både den verklighet (t.ex. anställningsformer och sparande) man vill beskriva och attityderna till den kan ha förändrats. Relativt få äldre blanketter är systematiskt utvecklade och den svars kvalitet de ger behöver bedömas. Man måste fråga sig t.ex. Hur fungerade den ”gamla” frågan? Hur stort var det partiella bortfallet? Uppstod det problem för vissa grupper att besvara frågan? Fick intervjuarna upprepa frågan eller förklara den?

Om frågeformuleringen fortfarande är relevant men medförde problem i ursprungsundersökningen, bör frågekonstruktören och uppdragsgivaren överväga för- och nackdelar med att använda frågan på nytt. Till slut måste man ta ställning till om en jämförelse mellan otillförlitliga och möjligen även irrelevanta mätningar vid två tillfällen ska föredras framför en jämförelse mellan en otillförlitlig vid det tidigare tillfället och en tillförlitlig och relevant mätning vid det senare. Vilket mått på förändringen som ska användas är av betydelse, när man tar ställning. En och samma svaghet i frågans funktion kan ha mindre betydelse för t.ex. nivåskillnader men större för skattningar av bruttoflöden.

Särskilt vid internationella jämförelser räcker det inte med att översätta frågorna ordagrant. Man måste ta hänsyn till kulturella olikheter och skillnader i ländernas sociala och ekonomiska verklighet. Annars är risken stor att man kommer att ”jämföra äpplen med päron”. Också nationella skillnader i hur datainsamlingen organiseras kan påverka.

5.7 Vilka hjälpmedel finns?

5.7.1 Standarder för klassificering

Inom flera områden finns nationella och internationella standarder för klassificering av viktiga redovisningsvariabler, t.ex. näringsgren, utbildning, arbetskraftsstatus, regional indelning. Utöver fastställda standarder finns även andra rekommenderade eller allmänt tillämpade klassifikationer och indelningar: indelning av personer i åldersklasser, gruppering av personer i kosthushåll, indelning av företag i storleksklasser osv. För att avvika från sådana standarder bör man ha mycket starka skäl. Undersökningen förlorar då jämförbarhet med andra undersökningar och resultaten kan bli svåra att tolka för användaren.

Av SCB fastställda standarder publiceras i serien Meddelanden i samordningsfrågor, MIS. Viktiga standarder är bl.a.

- Näringsgrensindelningen SNI 92 (MIS 1992:4 och MIS 1992:6)
- Standard för svensk yrkesklassificering SSYK 96 (MIS 1998:3)
- Svensk utbildningsnomenklatur SUN (MIS 1996:1, med tillägg 1998)
- Svenska utbildningar enligt ISCED 97
- Svenska riktlinjer för åldersgruppering av individer (MIS 1988:5)

5.7.2 Variabellistan

Variabellistan, som tas fram i fas 1, är frågekonstruktörens primära hjälpmedel när frågor, anvisningar och svarsalternativ ska skrivas. Frågorna i blanketten måste prickas av mot variabellistan, så man har kontroll över att man täckt hela undersökningsområdet. I variabellistan ska frågeområden och definitioner vara bestämda. Detta är en stor hjälp för frågekonstruktören när han/hon skriver frågorna. Många gånger kan man inte direkt översätta en variabel till en fråga, utan man måste arbeta både med att definiera/konkretisera vad variabeln betyder och att utveckla en grupp av frågor som tillsammans mäter företeelsen.

Exempel

I en undersökning vill man kunna identifiera personer med rörelsehinder. I variabellistan ingår alltså 'rörelsehinder'. När frågekonstruktören skriver frågor för att mäta variabeln 'rörelsehinder', behöver han/hon ha klart för sig hur begreppet 'rörelsehinder' definieras i den aktuella undersökningen. Handlar det enbart om rörelsehinder som bottnar i problem med att använda benen? Eller ska även rörelsehinder som beror på annat, t.ex. hjärtproblem eller fetma, räknas in? Ska 'rörelsehinder' även omfatta problem med att använda armar och händer? Eftersom begreppet 'rörelsehinder' har olika innebörd för olika personer, går det inte att fråga om uppgiftslämnaren 'är rörelsehindrad', utan det krävs flera frågor för att avgöra om och i så fall i vilken grad personen är rörelsehindrad. Det kan t.ex. vara nödvändigt att fråga om uppgiftslämnaren kan springa 100 meter om det behövs, promenera en halvtimme i rask takt, stiga på resp. stiga av en buss, resa sig från en köksstol, osv. beroende på vilken definition som används för 'rörelsehindrad'.

5.7.3 Djupintervjuer, fokusgrupper, expertgranskning mm

Djupintervjuer och fokusgrupper är hjälpmedel som frågekonstruktören kan använda även i fas 2. Här för att prova ut ett speciellt språkbruk, viss frågeteknik eller ett visst frågeområde, innan den slutliga blanketten är klar för ett mer fullständigt test.

En ensam blankettkonstruktör blir lätt hemmablind. Improviserade test med några kolleger eller bekanta kan hjälpa till att hitta "onödiga" fel. Innan den slutliga skrivbordsversionen fastställs, gör man en expertgranskning. Den ska bevaka att frågekonstruktionen ser ut att fungera enligt bästa erfarenhet, att språket är anpassat till gruppen av svarande och att innehållet är det rätta ur ämnessynpunkt. Dessutom ska man försöka förutse hur blankett och datainsamling ska fungera för att produktionen ska löpa smidigt. Man använder begreppet 'expertpanel' för en systematisk genomgång som två eller tre personer gör. Genomförandet varierar i praktiken, och granskningen kan vara samtidig eller oberoende. I bägge fallen måste producenten se till att synpunkterna vägs samman.

5.8 Layout

Hur blanketten ska se ut rent grafiskt, beror på **hur** blanketten ska användas och **vem** som ska fylla i svaren. Ska blanketten kunna tas upp och ses på pc-skärmen antingen via diskett eller internet, innebär det både möjligheter och begränsningar.

Intervjublanketter

Under många år har en standard för layout funnits vid SCB/I när man gör pappersblanketter för intervjuer. Att följa standard är viktigt för att underlätta och effektivisera intervjuarnas arbete. Huvudreglerna är att:

- Teckensnittet är Arial
- Frågor och svar som ska läsas upp är skrivna med gemener.
- Svartalernativ som **ej** ska läsas är skrivna med VERSALER.
- Koden för svartalernativet, vilken skrivs till vänster om frågan och ringas in av intervjuaren
- *Kursiv stil* används för att ange instruktioner.

Exempel från Undersökningar om levnadsförhållanden (ULF):

Fråga 112	<p>Brukar du umgås med någon av dina nuvarande arbetskamrater på fritiden? Med arbetskamrater menar vi då personer som du så gott som varje arbetsdag träffar på din arbetsplats.</p>
1	JA, TVÅ ELLER FLER
2	JA, EN
3	NEJ → <i>Fråga 114</i>

För blanketter i DATI finns inte så stora valmöjligheter. På skärmbilden får man plats med en fråga i taget. WIN-DATI, som nu utvecklas för SCB:s intervjuare, ger något större möjligheter. Man kommer t.ex. att kunna visa en tablåfråga på skärmen. Men när det gäller teckensnitt och liknande används samma standard som i DATI, dvs. det går inte att få fet stil, understrykningar, kursiv text osv. Det gäller att i så stor utsträckning som möjligt följa de rekommendationer som redan finns och som intervjuarna är vana vid.

Postenkäter

Kraven på blanketter är i praktiken större i en postenkät än i en intervjuundersökning. Uppgiftslämnaren ser direkt hela blanketten i stället för att höra en fråga i taget. Ingen intervjuare finns för att argumentera för medverkan eller överbrygga eventuella svagheter i blanketten.

Man vill gärna att antalet sidor i en pappersblankett ska vara en, två eller jämt delbart med fyra av hänsyn till pappers- tryck- distributions- och registreringskostnader. Uppgiftslämnaren ska från början se att blanketten är lätthanterlig och väl strukturerad. Uppdelning i logiska frågeblock, arbete med omfattningen och placeringen av informationen används för att åstadkomma detta.

Det finns en rad formella krav som måste tillgodoses, men som inte ligger inom ramen för denna handbok att redovisa i detalj. SCB:s logotyp bör finnas med.

Introduktionsbrevet till undersökningen ska tala om vem som gör undersökningen och hur den ska användas. Man ska bl.a. ange kontaktpersoner med telefonnummer, faxnummer och e-postadress, ange senaste datum för insändning, framhålla sekretesskydd och frivillighet eller i förekommande fall meddela att uppgiftsplikt föreligger.

Det finns i dag ingen standard för layout för postenkäter som följs entydigt inom SCB eller SOS, och variationen mellan program är betydande. För rent tekniska aspekter har SIS gett ut Skriva på kontor, Standarder och rekommendationer för utformning av dokument. Om hur blanketten ska bli uppgiftslämnarvänlig kan man läsa om hos Jenkins och Dillman.

Texten bör inte vara alltför kompakt, utan blanketten ska se ”luftig” ut. Detta bidrar till att underlätta **navigeringen** - dvs. hur uppgiftslämnaren följer grafiska signaler i blanketten för att svara i rätt ordning. Den svarande ska inte behöva fundera över rörelseriktningen utan ska direkt se var nästa fråga står. Man ska inte behöva gå ibland vertikalt och ibland horisontellt när man söker efter platsen där uppgiften ska skrivas in/svarsalternativet står. Med besvärlig navigering ökar risken för partiellt bortfall.

Det går ofta inte att ställa en fråga utan att definiera ett begrepp eller förklara frågans omfattning. De **anvisningar** uppgiftslämnaren behöver för att förstå frågan korrekt ska finnas i själva frågan eller i direkt anslutning till den. Till företagsblanketter krävs ibland omfattande anvisningar. Det är vanligt att dessa samlas i ett särskilt häfte. Många uppgiftslämnare sätter dock i gång att fylla i blanketten utan att titta i häftet så länge de inte ”kör fast”. En bättre utformning har blanketten när frågorna står på högra sidan av ett uppslag och anvisningarna står mitt för dem på vänstra sidan. Man kommer därigenom ifrån risken att anvisningarna har försvunnit för uppgiftslämnaren när det är dags att fylla i blanketten.

Hopp-anvisningar missförstås ofta i postenkäter. De måste vara få och enkla och med en mycket tydlig grafisk markering av vart hoppet ska gå. Frågetabläer har också visat sig krävande att fylla i. Grupper av celler kan lätt förbigås och leder därigenom till partiellt bortfall. Hopp-anvisningarna måste göras mycket enkla och överskådliga. Både hopp-anvisningar och frågetabläer behöver uppmärksammas, när blanketten testas.

Teckensnittet i blanketter för postenkät brukar vara **Arial, 10 p -fet** för frågorna och **Arial, 10 p normal** för svarsalternativen. Introduktionsbrev och omfattande anvisningar skrivs i Times New Roman, som är mera läsbar. Instruktioner skrivs med *kursiv stil* i Times New Roman

Blanketter till företag, kommuner och liknande är ofta skrivna med **Arial, 9 p** och **Arial, 8 p**. Blanketten blir kompakt och det kan för många vara svåräst med så liten text. Möjligen har sådana blanketter en gång konstruerats för att fyllas i på skrivmaskin. Eftersom nästan ingen har kvar skrivmaskiner numera, kommer de att fyllas i för hand och utrymmet blir då otillräckligt för handskrivna svar.

Vissa blanketter har både frågor och svarsalternativ i **en spalt**. Detta ger ett luftigt intryck och kan vara lättare att hantera vid manuell dataregistrering.

Exempel 1

Samma frågor med en respektive två spalter

Frågor om dig och ditt hushåll - i en spalt

1	Vilket år är du född?
	År 19
2	Är du kvinna eller man?
	1 <input type="checkbox"/> Kvinna
	2 <input type="checkbox"/> Man
3	Är du gift/sambo eller ensamstående?
	1 <input type="checkbox"/> Gift/sambo
	2 <input type="checkbox"/> Ensamstående

Andra blanketter skrivs med frågor i en vänsterspalt och svarsalternativ i en (eller ibland flera) spalter till höger. Två (eller flera) spalter sparar utrymme och ger plats för fler frågor på samma sida. Uppdelningen i frågespalt och svarsspalt (-er) underlättar överblick och orientering. Uppgiftslämnaren behöver inte flytta undan handen varje gång han/hon har skrivit ett svar för att kunna läsa texten till nästa fråga. Med bättre överblick krävs mindre tid för att svara.

Frågor om dig och ditt hushåll - i två spalter

1	Vilket år är du född?	År 19
2	Är du kvinna eller man?	1 <input type="checkbox"/> Kvinna
		2 <input type="checkbox"/> Man
3	Är du gift/sambo eller ensamstående?	1 <input type="checkbox"/> Gift/sambo
		2 <input type="checkbox"/> Ensamstående

Exempel 2

Småföretagens villkor - i två spalter

3	Bedrivs verksamheten i företaget som kooperativ?	1 <input type="checkbox"/> Ja 2 <input type="checkbox"/> Nej
4	Är det en man eller kvinna som leder företaget?	1 <input type="checkbox"/> Kvinna 2 <input type="checkbox"/> Man
5	Vilken är företagets huvudsakliga verksamhet idag?
6	Hur stor var årsomsättningen exkl. moms 1994?	Belopp i 1 000-tal kr

Exempel 3 Flera spalter

INDUSTRINS VARUPRODUKTION 1998**1. Intäkter**

	Fördelning av undersöknings- enhetens intäkter	Värde 1997, 1 000-tal kr	Värde 1998, 1 000-tal kr
01	Industriverksamhet		
02	Försäljning av färdiga varor tillverkade vid annan enhet inom företaget		
03	Övriga verksamheter (Icke industriell verksamhet)	SNI-kod	
04		SNI-kod:	
05		SNI-kod:	
06		SNI-kod:	
07		SNI-kod:	
08		SNI-kod:	
09		SNI-kod:	
10		SNI-kod:	
11	Summa nettoomsättning inklusive internleveranser		
12	Internleveranser	för vidareförädling	
13		av egenproducerade färdiga varor	
14		av varor eller tjänster från icke industriell verksamhet	
15	Summa nettoomsättning exklusive internleveranser		

De skuggade fälten i blanketten ovan förprintas för de objekt där uppgiften finns.
Teckensnitt är Arial 8 p.

Registreringsvärden (stansvärden) står ibland till vänster och ibland till höger om kryssrutan. Om siffran står till vänster, kommer texten nära kryssrutan, - vilket är att föredra.

Exempel 4 :

- 1 0-25 km
 2 26-50 km
 3 51-100 km
 4 över 100 km
 5 Vet ej

Radavstånd ofta 3 punkter.
 (WORD: Format, Stycke, Indrag och
 avstånd, Avstånd, Efter: 3 pt)

Det finns **mallar för blanketter** som skrivs med **Crystal Reports**. Av tradition skrivs många blanketter i PageMaker, men detta program kräver särskild licens. Det går också utmärkt att skriva blanketter i **Word**. Om man vill ha två spalter, använder man parallella spalter genom att *Infoga tabell* under *Tabell*. Kryssruta hittar man under *Visa - Verktögsfält - Formulär*. Rutan får inte vara för liten. Den går att markera och göra större genom att välja större teckenstorlek, t.ex. 12 p (4 x 4 mm).

När man har större material att bearbeta, är **skanning** ett bra alternativ till manuell dataregistrering. Ett + infogas i varje hörn. Det måste vara tillräckligt stora rutor (4 x 4 mm) och de får inte vara för nära varandra (se registreringsvärden). Vid skanning räcker det att registreringsvärdena står högst upp över en kolumn med svarsrutor. De behöver inte upprepas vid varje ruta. Ibland kan det vara befogat för att underlätta granskning.

1	2	3	4
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Om man väljer att **färglägga** vissa fält, ska färgen vara svag. Ibland faxas svaren tillbaka, och om blanketter har starka färger kan resultatet av överföringen bli mycket svårtolkat. Skanning klarar inte heller av starka färger. Särskilda hänsyn måste tas till skannerns kapacitet. Testa alltid hur en färglagd blankett klarar att faxas och skannas!

Då uppgiftslämnaren får ett odelat utrymme för att skriva ner sina uppgifter dvs. _____ ökar riskerna för sort och storleksfel. Misstag på sort och positionsfel undviks dels genom att sorten anges tydligt dels genom svarsboxarnas utformning. Post- och bankgiroblanketter som kräver högsta tillförlitlighet har en svarsruta för varje siffra och grupperar dessa dessutom i tresiffrergrupper, t.ex. som

,

5.9 Checklista

CHECKLISTA Blankettkonstruktion

1. Ta reda på vilka nationella och internationella standarder för klassificering, som måste följas för att tillgodose krav på jämförbarhet med annan statistik, om jämförelser ska göras.
2. Konstruera frågorna efter beprövad erfarenhet
3. Anpassa blanketten till insamlingsmetod och andra produktionsvillkor (t.ex. hur registrering ska gå till).
4. Utveckla första blankettversionen med hjälp av
 - granskning vid skrivbordet av en kollega
 - informellt blanketttest med t.ex. några kolleger eller bekanta. Detta kan dock aldrig ersätta test med verkliga uppgiftslämnare.
 - expertpanel, dvs. flera experter går tillsammans igenom blanketten
5. Språkgranska blankett och informationsmaterial

Slutprodukten av fas 2 ska vara en blankett som är komplett och tekniskt riktig. Man har tänkt på att få med bakgrundsvariablerna och de uppgifter som krävs för att kunna göra korrekta skattningar. Frågorna ska vara anpassade till datainsamlingsmetod och layouten till planerna för dataregistrering. Fortfarande vet man inte om den är anpassad till uppgiftslämnarnas kapacitet och vilja att svara. Man vet heller inte hur den kommer att fungera i produktion.

Referenser

- Råd & definitioner för Företagsstatistik 1997.* SCB
- Bergman, L.R. och Wärneryd, B. (1982). *Om datainsamling i surveyundersökningar.* Stockholm: SCB och Liber.
- Wärneryd, B. (red., 1989:) *Att fråga.* SCB.
- Salant P and Dillman D. A. (1994). *How to conduct your own survey.* John Wiley & Sons, Inc.
- Fowler F.J. (1995) *Improving Survey Questions.* SAGE Publications.
- Developing and Using Questionnaires* (1993) från General Accounting Office, Washington, USA
- Survey Measurement and Process Quality.* (1997), (redigerad av Lyberg, Biemer, Collins, de Leeuw, Dippo, Schwartz och Trewin). John Wiley and Sons, Inc.
- Jenkins, C.R., and Dillman D.A: *Towards a Theory of Self-Administered Questionnaire Design.* Kapitel 9 i SMPQ.
- JOS (1985). *Special Issue on Questionnaire Design.* Vol1. No.2 , 1985
- JOS (1992) *Measurement Errors in Surveys Part I and Part II.* Vol. 8 1 and 3, 1992.
- Skriva på kontor,* STG handbok 126.. Utgiven av SIS april 1995.
- STG = Allmänna Standardiseringsgruppen. SIS = Standardiseringen i Sverige

6 Fas 3 - Kognitiva test

*If you cannot afford to pilot your study don't do the study!
(Seymour Sudman)*

6.1 Innehåll

Arbetsuppgifter: Att skaffa kvalitativ information om hur den ”skrivbordskonstruerade” blanketten förstås och besvaras av ”verkliga” uppgiftslämnare. Blanketten revideras så att upptäckta brister elimineras. Om många svagheter upptäcks, kan den följande blankettrevisionen bli så omfattande att även innehållet måste revideras, dvs. att inte bara fas 2 och 3 utan även fas 1 måste arbetas igenom på nytt.

Kognitiva test (kvalitativa) används för att upptäcka fel i förståelsen av blanketten, relevansfel och tekniska fel men också svårigheter och brister i hanteringen av blanketten. Testen är främst inriktade på förståelse av den kognitiva processen som den beskrivs i avsnitt 2.3 under rubriken *Kognitiv modell för svarsprocessen*. De är diagnostiska i den meningen att de kan hitta orsakerna till fel och ge indikationer på hur blanketten ska revideras för att ge bästa möjliga informationsutbyte med uppgiftslämnarna. Kognitiva test använder sig främst av observationer och fördjupningsfrågor på små ”icke - sannolikhetsurval”. Vid postenkäter och då särskilt i företagsundersökningar tar man också reda på hur blanketten ska komma fram till ”rätt” uppgiftslämnare och hur denne hämtar fram uppgifterna och/eller beräknar dem.

Val av testmetod

Kognitiva test med kvalitativa resultat och experiment (Fas 4) med kvantitativa resultat ger underlag för olika typer av beslut och kompletterar varandra. Kognitiva test är oerhört mycket billigare och går snabbare att genomföra än experiment. Att göra experiment som inte föregåtts av kognitiva test är nästan alltid ett stort resursslöseri. Kognitiva test hjälper till att upptäcka de faktorer som påverkar svarens innehåll. Efter ett kognitivt test har man mycket bättre förutsättningar att kontrollera störande faktorer och att formulera precisa hypoteser om de mättekniska egenskaperna hos olika blankettversioner och datainsamlingsmetoder.

Uppgiftsplikt finns inte för provundersökningar bland företag. Om sådana alls kan göras, förutsätts att man skaffar medgivande från Näringslivets Nämnd för Regelgranskning (NNR). Risken för lågt deltagande och ointresse är stor, varför det är svårt att genomföra tillförlitliga experiment i företagsundersökningar. Små kvalitativa studier med intresserade uppgiftslämnare är i praktiken ofta det enda realistiska alternativet.

6.2 Kognitivt test i sammandrag

Problemanalys

Problemanalys består i en bedömning av vilka begrepp, frågor och uppgifter som kan vara särskilt krävande för uppgiftslämnaren att ge tillförlitliga svar på. Man bedömer också om någon grupp bland uppgiftslämnarna kan få problem med en speciell del av blanketten eller allmänt ha större problem än övriga. Problemanalysen leder till att man formulerar hypoteser om hur databildningen

och datainsamlingen i verkligheten kan komma att gå till, vilka svårigheter som kan uppträda och vilka orsakerna till detta är. Hypoteserna kan gälla

- svårigheter i att förstå frågor och anvisningar
- svårigheter att skaffa fram data
- effekter av frågeordning, orientering och layout i blanketten
- uppgiftslämnarens motivation, egna förutsättningar och inställning till undersökningen
- vid intervjuundersökningar dessutom - intervjuarens träning och agerande.

Ofta är ett antal av undersökningsvariablerna särskilt viktiga (ibland kallas dessa nyckelvariabler). För dessa är kraven på mätkvalitet särskilt höga. Det finns då skäl att ägna sådana större utrymme då man testar blanketten, även om de inte skulle vara bland de svåraste att mäta. Det är också viktigt med särskilt hög tillförlitlighet i svar som identifierar undersökningsenheten, i frågor som används för indelning av undersökningsobjekten i viktiga redovisningsgrupper och i selektionsfrågor.

Planering och organisation

En testledare ansvarar för planering, administration och analys. I detta ingår att välja testmetod, precisera verktygen i testet, ge sina intervjuare anvisningar och dokumentera testets genomförande och resultat. Till sin hjälp har testledaren ett litet antal intervjuare. Testpersonerna rekryteras i samverkan mellan testledaren och intervjuarna. Eftersom kognitiva intervjuer skiljer sig från standardiserade intervjuer i många avseenden, har dessa intervjuare fått särskild träning i att observera, ställa fördjupningsfrågor och dokumentera uppgiftslämnarens synpunkter på blanketten och agerande under testet.

Vid SCB finns sedan 1989 Mättekniska Laboratoriet. ML är organiserat som ett nätverk inom SCB. Personerna som tillhör nätverket är samhällsvetare, beteendevetare, intervjuare och statistiker med olika specialistkunskaper och lång erfarenhet från statistiska undersökningar. Samordningsfunktionen är placerad på Metodenheten, Utvecklingsavdelningen. Det finns en folder som beskriver verksamheten *SCB:s Mättekniska Laboratorium - ML hjälper dig att göra rätt från början i dina undersökningar*.

Plats för testet

Man skiljer på **fälttest** och **laborietest** av blanketter. Med fälttest menar man ett test "hemma hos" uppgiftslämnaren, på arbetsplatsen, på ett inköpsställe osv. Fälttest är oftast det bästa och ibland enda alternativet för att testa blanketter och insamling i företagsundersökningar. I dessa behöver man nämligen kartlägga samverkan mellan uppgiftslämnaren och andra på företaget för att inse hur uppgifterna hämtas fram. Laborietest genomförs i lokaler som ställts i ordning för mättekniska studier och kan ha utrustning t.ex. för audio- eller videoinspelningar av testet, och ibland ges också testledaren möjlighet att följa hur testen genomförs. Test av blanketter i individundersökningar behöver ibland göras i laboriemiljö för att testet inte ska störas av att telefonen ringer, att besökare dyker upp eller att medlemmar i testpersonens hushåll ska påverka testet.

Verktyg

Kognitiva blanketttest görs med en uppgiftslämnare i taget. En rad olika verktyg används, som olika slag av fördjupningsfrågor och observationer. De kognitiva verktygen beskrivs utförligare i avsnitt 6.4. Valet beror på vilka förhandskunskaper och hypoteser man har om frågorna och vilken typ av svårigheter man räknar med att blanketten innehåller. Men testet kan också använda verktyg som

kan upptäcka svårigheter som testledaren inte lyckades förutse. Testen designas och verktygen anpassas till valet av datainsamlingsmetod och kategorin av uppgiftslämnare i den planerade undersökningen.

Genomförande

En testintervju börjar med att testpersonen får läsa igenom informationsmaterialet. Personen får därefter fylla i blanketten om det är en postenkät, respektive svara på frågorna om det är en intervjuundersökning. Intervjuaren gör under tiden observationer men är noga med att inte störa eller påverka testpersonen. Testintervjuaren noterar viktiga iakttagelser och intervjun bandas. Efter testintervjuns slut genomför intervjuaren en systematisk utfrågning (debriefing) om hur testpersonen upplevde blanketten och frågorna. En informativ testintervju förutsätter att testpersonen är alert under hela testet och motiverad att lämna synpunkter. Den bör därför inte ta över en timme.

Vid test av företagsblanketter skulle det ta för lång tid för testpersonen att ta fram de faktiska uppgifterna. Ofta är de inte alls tillgängliga vid tiden för testet. Man kartlägger i stället hur det går till, vilka personer som blir inblandade, varifrån uppgifter hämtas, vilka kalkyler som behöver utföras och hur lång tid som krävs. (Se 6.5 Test av blanketter till företag.)

Intervjuarnas arbetsätt utvärderas längre fram

De intervjuare som medverkar i kognitiva test har dels fått en särskild utbildning, dels valts ut till denna på grund av att de tidigare visat sig skickliga i att göra standardiserade intervjuar. Även när man har resurser för att använda en observatör, skulle iakttagelserna inte säga särskilt mycket om hur väl den stora gruppen av intervjuare skulle använda blankett och anvisningar. Sådana studier kan göras först vid större experiment, fas 4, eller vid anpassning till produktionen i fas 5.

Dokumentation

Intervjuarna rapporterar var och en om sina testintervjuer efter en mall. Rapporten är mycket detaljerad. När resultaten ställs samman, ska testledaren skilja noga på testpersonernas spontana reaktioner, deras reaktion på fördjupningsfrågor, intervjuarnas iakttagelser och sina egna slutsatser. Utvärderingen av testresultaten görs i samarbete mellan testledaren och den som ska revidera blanketten. Testprotokollen innehåller inga uppgifter som identifierar testpersonerna. Protokollen förstörs, när test och revidering avslutats.

Några exempel på upptäckter genom kognitiva test

Exemplen är hämtade från Bureau of Labor Statistics (BLS) och National Centre for Health Research (NCHR), där det görs kognitiva studier med så många personer att det är meningsfullt att kvantifiera resultaten.

***Annan värdering:** BLS upptäckte att ca 40 % av de tillfrågade valde ett annat alternativ för sin huvudsakliga sysselsättning än den klass som var den rätta enligt BLS:s definition. BLS hade antalet arbetade timmar under perioden som kriterium, medan uppgiftslämnaren såg sig själv i ett mera långsiktigt perspektiv. T.ex. kunde man se sig som studerande fast man just den veckan arbetade 30 timmar på en hamburgerbar.*

***Fackterm annorlunda än vardagspråk:** NCHR upptäckte att frågan 'Har du haft besvärande smärtor i buken under senaste tre månaderna' (?) inte fungerade av två skäl. Mycket få av de svarande visste tillräckligt exakt var buken sitter. Att ordet 'besvärande' stod i frågan innebar att vissa svarande satte en mycket hög toleranströskel för smärta.*

Ett ord missförstått: BLS ställde först en fråga om uppgiftslämnaren hade läst åtminstone någon roman under föregående tidsperiod. Sedan ställdes en kontrollfråga om vilka böcker det var. Ca 20 % av de påstådda romanläsarna svarade med att uppge facklitteratur.

Saknat svarsalternativ: I ett av ML:s uppdrag testades en blankett om alkoholvanor. Det visade sig att det var viktigt för storkonsumenterna att få tala om att vissa sorter sprit drack de inte alls. De var inte nöjda med ett alternativ för låg konsumtion utan krävde ett rent 0-alternativ.

6.3 Rekrytering av testpersoner

Kognitivt test görs med ett antal **testpersoner**, som i någon mening är representativa för uppgiftslämnarna i den planerade undersökningen. Testpersonen är den som lämnar uppgifter om ett **testobjekt**, som kan vara testpersonen själv, men också ett företag, jordbruk, fastighet m.m. Antalet testobjekt är litet. I de standardiserade testen vid SCB har det varierat mellan 6 och 15. Redan ett så litet test har dock god förmåga att upptäcka brister i kommunikationen med uppgiftslämnarna. Med krav på att upptäcka mindre vanliga misstag och skillnader mellan olika grupper av uppgiftslämnare kommer det emellertid att krävas betydligt fler testpersoner.

Testobjekt och -personer **rekryteras** på olika sätt. Telefonkataloger, medlemsregister, information från uppdragsgivaren och bekantas bekanta används för att hitta och rekrytera lämpliga personer. Testpersonerna får sina resekostnader betalda och dessutom någon form av ersättning (*incentives*) för sitt deltagande. Privatpersoner brukar erbjudas lotter, medan det är bättre att erbjuda företag att ett belopp ska skickas i deras namn till en ideell organisation, som de själva får välja.

Det kräver en betydande skicklighet att rekrytera representativa testpersoner, så att testet blir informativt. De ska vara "riktiga" uppgiftslämnare i den kommande undersökningen. I individundersökningar väljs de ofta för att representera de grupper som antas ha störst svårigheter med att hantera frågorna och lämna uppgifter. Olika grupper, t.ex. efter ålder, kan ha olika kognitiva egenskaper: t.ex. tolka frågan på olika sätt, väga in olika faktorer, bearbeta resultaten eller bearbeta uppgifter på olika sätt. Ibland, när skillnaderna kan antas vara mycket stora, är det motiverat att göra olika test i sådana grupper.

I företagsundersökningar där både variabelvärden och deras vikter vid estimation varierar kraftigt tänker man på att rekrytera testobjekt vars svar ger stora tillskott till skattningarna och/eller felet i dessa.

Sannolikhetsurval används nästan aldrig. Framför allt kan det sällan ge den fördelning av urvalet som man behöver. Det skulle bli för dyrt och ta för lång tid att dra. Några vanliga rekryteringsmetoder (icke sannolikhetsurval) är då:

- Urvalet dras t.ex. med hjälp av personliga kontakter eller från någon lätt tillgänglig lista, som en telefonkatalog eller ett medlemsregister. (*Convenience sampling*)
- Urvalsdragningen baseras på ett "expertomdöme" om att urvalspersonerna är typiska eller har någon uppsättning av egenskaper som är viktiga i undersökningen. (*Purposive sampling*)
- Man definierar hur många objekt som ska ingå i varje delgrupp men lämnar åt intervjuaren att själv göra urvalet. (*Quota sampling*)

Riskabla men ibland använda metoder för att rekrytera testpersoner är att annonsera efter dem. Vanligt är också att enbart använda studenter eller att använda samma testpersoner i en rad test. De två första metoderna har svagheter genom att testpersonerna inte är representativa för uppgiftslämnarpopulationen. Den tredje genom att testpersonerna efter det första testet gärna börjar se sig själva som experter.

Rekryteringsgraden är kvoten mellan det antal testpersoner som deltog i testet och det antal personer som tillfrågades om de ville ställa upp. (Begreppen svars- respektive bortfallsfrekvens är inte särskilt meningsfulla när urvalet inte är ett sannolikhetsurval.) En rekryteringsgrad runt 50 % är många gånger möjlig för en skicklig rekryterare. Exceptionellt låg rekryteringsgrad indikerar att ämnet för studien antingen är ointressant eller för känsligt och förvarnar om risker för högt bortfall i den kommande undersökningen. Någon egentlig ”vägrarbearbetning” görs inte. Det är inte meningsfullt att övertala motvilliga testpersoner. Erfarenheten är att även om de övertalas att delta så kommer de att lämna ytterst få synpunkter på blanketten.

Av dem som deltar i ett kvalitativt test krävs inte bara mer tid än vid den vanliga postenkäten/intervjun utan även en vilja att vara öppen om bakgrunden till sina svar. Särskilt mycket avviker förutsättningarna när man testar blanketter till företagsundersökningar som har obligatoriskt deltagande i den reguljära insamlingen.

Det är rimligt att de som blir testpersoner både har större intresse för undersökningsområdet och har större språklig förmåga än de som inte vill delta, dvs. att de har mindre problem att förstå ord och formuleringar än övriga. Om antagandet är riktigt, underskattas den genomsnittlige uppgiftslämnarens kapacitet att lämna tillförlitliga svar. Knäuper m.fl. har visat hur bristande kognitiv kapacitet sammanhänger med fler vet ej, fler partiellt bortfall osv. De med låg förmåga misstolkar oftare svar, och det finns en risk att denna grupp inte är representerad i kognitiva test.

6.4 Verktyg

Att bara observera hur frågor besvaras räcker inte till för att avgöra om de besvaras rätt eller fel. Inte sällan svarar en uppgiftslämnare snabbt och övertygande utan att ha förstått frågan korrekt eller utan att vara säker på svaret. För att upptäcka när och varför sådant inträffar använder man en rad mättekniska verktyg, som kombineras på olika sätt beroende på vilka svårigheter man förutser i blanketten. Ofta använda verktyg är:

6.4.1 Fördjupningsfrågor (probes)

När blanketten fyllts i respektive intervjun genomförts ställer intervjuaren ett antal fördjupningsfrågor (**probes** eller probing questions) om hur testpersonen förstått frågorna i blanketten och kommit fram till sina svar. Probes kan ha olika syften och användas på olika sätt. Svaren på probes är ostrukturerade.

Gemensamma probes riktas till alla testpersoner. De formuleras med utgångspunkt i en hypotes om vilken den kognitiva svårigheten hos frågan är. Efter sin inriktning kan det vara olika frågor:

- frågor om hur testpersonen förstått fråga och anvisningar (*Comprehension probes*).

Ex. Current population Survey (CPS) frågade om sysselsättning föregående vecka. Enligt CPS definition sträckte sig veckan från och med söndag till och med lördag. Probes om hur uppgiftslämnarna tänkte sig veckan gav följande fördelning

17 %	Söndag t.o.m. lördag (korrekt)
54 %	Måndag t.o.m. fredag
9 %	Måndag t.o.m. lördag
6 %	Måndag t.o.m. söndag
4 %	Söndag t.o.m. Söndag
10 %	Annat sätt.

- frågor om hur testpersonen har resonerat för att komma fram till sitt svar (*Information retrieval probes*) Testpersonen kan ha svarat direkt, kontrollerat i anteckningar eller bokföring, räknat sig fram till svaren, svarat som det brukar vara osv.

Ex. CPS ställde fördjupningsfrågor om hur man kommit fram till antalet arbetade timmar senaste veckan "Visste du direkt hur du skulle svara eller fick du tänka efter?", "Hur resonerade du innan du kom fram till svaret?"

66 % angav att de alltid arbetade lika många timmar. Av återstoden sade

40 %	att de tog ett medelvärde för flera veckor
36 %	att de tänkte på en typisk vecka
24 %	att de räknade på något annat sätt

- frågor som går ut på att ta reda på om ord, fraser, situationer uppfattas av testpersonen på samma sätt som blankettkonstruktören tänkt sig. (*Frame of reference probes*).

Ex. "Vad betyder "vanligen" för Dig?". "När tycker Du att sjukdom ska kallas för långvarig?"

- frågor om hur testpersonen passat in sitt svar i de alternativ som funnits att välja på. (*Response category selection probes*)

Ex. "Varför valde du just det svaret?" "Fanns det något annat svarsalternativ som du funderade över?"

Generella (gemensamma) **probes** är allmänt formulerade frågor. De används för att söka efter svårigheter i bestämda frågor, som testledaren inte kunnat förutse.

Ex. "Kan du berätta mer om detta?" "Tyckte du att introduktionen var lätt att läsa?" "Vad det någon del av blanketten som var svår att följa?"

Random probes innebär att alla intervjuare slumpmässigt väljer ut ett antal frågor för vilka de ställer allmänt hållna (generella) probes. Random probes är inte gemensamma utan varierar mellan testpersonerna. Också de har till uppgift att fånga upp svårigheter som testledaren inte kunde förutse.

Ex. "Var det något som var otydligt med den här frågan?". "Gick det lätt att komma fram till svaret?"

Speciella (specific) **probes** är sådana som testintervjuaren själv formulerar och använder när denne under pågående test fått intrycket av att en testperson har svårigheter med en bestämd fråga. Dessa är viktiga inslag i testet och ställer

stora krav på intervjuarens observationsförmåga och förmåga att improvisera probes.

Ex. "Jag märkte att du tvekade innan du fyllde i svaret på fråga 35. Vad var skälet till att du tvekade?"

Olika "skolor" föredrar att sätta in probes under pågående test (**concurrent use**) respektive först efter avslutad intervju/ifylld postenkät (**retrospective use**). Motiveringen till att använda probes direkt när ett problem eller tveksamhet noterats, är att testpersonen inte riskerar att glömma bort vad problemet bestod i. Invändningen är att testet blir mindre likt den normala intervjusituationen, och man riskerar att testpersonen redan efter några probes börjar se sig själv som expert. Att probes ställs efteråt innebär att situationen, när blanketten fylls i respektive intervjun besvaras, är mer lik situationen i den kommande undersökningen. Vid ML:s test används probes i efterhand.

6.4.2 "Tänka högt"-metoder

"Tänka högt"-metoder (*Think aloud*) är ett effektivt verktyg men svårare att hantera än enbart probes. De tar längre tid att genomföra och kräver att intervjuaren har extra utbildning. Vid testet måste intervjuaren först själv visa hur man "tänker högt" och sedan låta testpersonen öva på en annan, men kort, blankett innan det egentliga testet startar.

"Tänka högt" innebär att testpersonen, medan han/hon fyller i en blankett, fortlöpande kommenterar den och sitt arbete med att fylla i uppgifterna. Kommentarer bandas. Eftersom intervjuaren förhåller sig avvaktande under denna fas, kommer även spontana synpunkter fram och upptäcks svårigheter, som testledaren inte lyckades förutse. Metoden passar bäst för postenkät, men kan även anpassas till en intervjusituation. Den följs upp med probes.

6.4.3 Andra verktyg i kognitiva test

Andra etablerade verktyg som hittills inte kommit till användning vid SCB och som inte har någon etablerad svensk översättning är:

- **Omformulering av frågan** (*Paraphrasing*), som innebär att testpersonen uppmanas att formulera om en fråga och uttrycka den med egna ord. Resultaten avslöjar den verkliga förståelsen av ett sammanhang och inte bara förståelsen av orden var för sig.
Ex. I blanketten står frågan. ".....?" Hur skulle du själv ha frågat om samma sak?.
- **Vinjetter** (kallas ibland "scripts") är fallbeskrivningar av olika situationer (t.ex. anställningsförhållanden), som testpersonen uppmanas att klassificera enligt blankettens svarsalternativ. De används i gränsfall där man är osäker på om den vardagliga betydelsen stämmer med den formella. Verktöget kan vara användbart redan i fas 1 tillsammans med djupintervjuer eller fokusgrupper.
Ex "En person arbetar några timmar varje vecka i sin fars företag. I gengäld får han mat och bor fritt hemma. Har han arbete enligt din uppfattning?"
- **CARD sorting** (*free sorting, dimensional sorting, vignette sorting*) går ut på att förstå hur testpersonen associerar och grupperar olika begrepp.
- Mätning av hur säker testpersonen är på sina svar. (*Confidence rating*)
Testpersonen får svara t.ex. genom att välja en skalpunkt på en i förväg given skala.

Ex "Hur säker är du på du gjorde inköpet efter årsskiftet.?", "Tror du att de resor du skrivit upp i dagboken är alla du gjorde senaste månaden?"

- **Mätning av svarstid** (*Response latency*) dvs. den tid som gått från att en fråga lästs upp till dess att testpersonen lämnat sitt svar. Svarstiden indikerar hur mycket möda testpersonen måste lägga ner för att komma fram ett sitt svar. Indikatorn lämpar sig främst för intervjuer som kan genomföras utan avbrott och särskilt vid datoriserade intervjuer då tidmätningen är enkel att genomföra. Den kan också användas vid test av blanketter för postenkät om intervjuaren kan se hur blanketten fylls i. Vid användning av intelligenta blanketter kan man få en mätning av tidsåtgången. I det fallet kan dock uppgiftslämnaren ha en rad olika skäl för att göra upphåll och avbrott.

6.5 Test av blanketter till företag

Förarbete

När man testar blanketter till företag, krävs större förarbete än när man testar blanketter till individer. När företagets storlek betyder mycket för resultaten, beaktas detta vid val av testobjekt. Situationen är mer komplex än i individundersökningar, och planeringen av testet måste därför tillåta mer flexibilitet.

Förberedande kontakter behöver tas för att klara ut relationen mellan undersökningsobjekt och uppgiftslämnare, dvs. var uppgifterna finns och vem som kan lämna dem. Uppgifterna om arbetsställe finns ibland vid företagets huvudkontor. Ibland finns uppgifterna om arbetsställena för sig och ibland för företaget i sin helhet.

I företagsundersökningar behöver statistikproducenten ha en uppfattning om vilken befattning och kompetens uppgiftslämnaren ska ha och hur uppgifterna skaffas fram och/eller beräknas av denne. Olika kategorier av uppgiftslämnare har olika förutsättningar att ge riktiga svar. Särskilt att välja rätt är det när man frågar om bedömningar t.ex. om kommande investeringar och rekryteringsbehov och inte om avskrifter. För företaget är det inte självklart vem som ska vara uppgiftslämnare, och det kan lämna åt någon med otillräckliga insikter att svara.

CHECKLISTA: Blankettens väg till uppgiftslämnaren

1. Hur identifieras/definieras en uppgiftslämnare på rätt nivå/med rätt kunskaper?
2. Hur precist kan man adressera till "rätt" uppgiftslämnare?
3. Vilka grindvakter, dvs. postsortörer, sekreterare, växeltelefonister och andra, tar i första hand emot telefonsamtal och post?
Hur säkert och snabbt kan dessa identifiera "rätt" uppgiftslämnare och veta vem som ska ha telefonsamtalet/försändelsen?
4. Finns de uppgifter som efterfrågas verkligen på företaget och inte på en bokföringsfirma eller annat arbetsställe?
5. Hur många och vilka personer måste medverka för att ta fram uppgifterna?
6. Hur mycket arbete och tid måste uppgiftslämnaren lägga ner för att sammanställa och bereda underlaget till blanketten?

I löpande undersökningar är det en fördel att ha samma uppgiftslämnare från gång till gång. När denne lärt sig begrepp och anvisningar i blanketten, bör tillförlitligheten i svaren bli bättre. När en fast uppgiftslämnare i en löpande undersökning ersätts av vikarier, t.ex. vid tjänsteresor, semesterar och sjukdomar, ökar risken för mätfel, bortfall eller förseningar. Till och från sker ett permanent byte av uppgiftslämnare. Både kompetensnivån hos uppgiftslämnaren och byte av uppgiftslämnare i löpande undersökningar är indikatorer på tillförlitlighet och felrisker hos uppgifterna från ett företag.

Innehållet är ibland ”låst”

Vilka variabler som ska ingå i en företagsblankett är ofta givet till följd av att undersökningens resultat ska passas in i räkenskaper, index eller någon annan modell. Innehållet kan också vara bestämt av direktiv och internationella åtaganden. Innehållsmässiga ändringar kan därför vara omöjliga att göra, även då ett test visar att variablerna inte kan mätas med tillräcklig tillförlitlighet. En annan restriktion är då ändringar av blanketten insamlingen kräver medgivande från Näringslivets nämnd för regelgranskning (NNR).

Fälttest

Test av blanketter till företag görs oftast på plats, då detta är enda sättet att skapa realistiska testvillkor. Uppgiftslämnaren behöver tillgång till bokföring och annan information - eventuellt från företagets administrativa system. Medverkan av flera personer krävs ibland för att fylla i blanketten.

Sammanställningar och beräkningar innan blanketten kan fyllas i, kan ta arbetsdagar eller -veckor i anspråk och vara kostsamma för företaget. Ett blanketttest på ett företag bör inte ta mer än en timme av företagets tid i anspråk. Det går därför inte att följa och observera hela förarbetet innan blanketten fylls i, vilket kan vara mycket mer omfattande. Det är heller inte givet att uppgifterna finns tillgängliga vid testtillfället. I stället inriktas testet på hur det skulle gå till att lämna uppgifterna - eller, i löpande undersökningar, hur det brukar gå till.

Situationen är komplex och varierad

Uppgiftslämnarsituationen varierar mycket mer mellan företag än mellan individer. Bokföringssystem, redovisningsprinciper, graden av datorisering och administration inom företaget skiljer sig åt.

Även företagets verksamheter kan vara mycket olika och olika organiserade. Olika frågor besvaras av olika grupper av uppgiftslämnare. Olika blankettversioner behövs då för att undvika alltför många hoppanvisningar. Dessa åstadkommer man genom att programmera datorstyrda utskrifter.

De flesta frågor gäller kvantitativa storheter. Svaren kan variera kraftigt i storleksordning både mellan olika uppgiftslämnare och mellan olika variabler, med risk för ”sortfel” som följd. Detta uppstår när uppgiftslämnaren tänker i en sort och inte märker att blanketten frågar om en annan.

I samma mån som situationen för uppgiftslämnaren varierar, måste testet ha förmågan att upptäcka och anpassa sig till detta. Det krävs både mer förberedelser för alternativa situationer och god observationsförmåga och flexibilitet i genomförandet.

Användning av förhandsinformation i löpande undersökningar

Många undersökningar är periodiskt återkommande. För dessa kan finnas både dokumenterade och informella kunskaper hos producenten om resultatens tillförlitlighet från föregående undersökningsomgångar. Detta kan effektivisera testen och användas till att rikta in testet på:

1. frågor och avsnitt i blanketten som berett svårigheter
2. frågor och avsnitt som är särskilt viktiga
3. delgrupper eller t.o.m. enskilda företag som har haft särskilda svårigheter att lämna korrekta uppgifter
4. grupper av uppgiftslämnare vars uppgifter väger tungt i undersökningen.

När samma urval används vid flera undersökningstillfällen, kan granskade och rättade svar från föregående omgång av undersökningen förtryckas på blanketten för att vägleda uppgiftslämnaren om de nya svarens storleksordning.

Tiden för blankettrevision

En undersökning som genomförs månads- eller kvartalsvis vill beställaren kunna summera till årsnivå. Beställaren är då helt obenägen att låta producenten införa en reviderad blankett vid någon annan tidpunkt än vid ett årsskifte. En blankettrevision måste därför beslutas i så god tid före årsskiftet att man kan använda flera månads- eller kvartalsundersökningar för att utvärdera den gamla blanketten och utveckla en ny version.

CHECKLISTA: Inriktning av probes vid test av blanketter till företag

- När läses instruktioner och anvisningar och hur uppfattas och används de? Vilka instruktioner står på "rätt" plats och läses och vilka gör det inte?
- Finns det facktermer och begrepp i blanketten - definierade eller ej - som företagen inte använder eller förstår korrekt?
- Använder blanketten begrepp och benämningar konsekvent?
- Förstår testpersonen frågorna korrekt?
- Hur och från vilka källor hämtar man fram information?
- Går det alls att få fram uppgifterna?
- Finns uppgifterna så detaljerat redovisade, som statistiken begär?
 - Stämmer statistikens och företagets periodisering?
 - Hur görs omräkningar och annat förarbete i praktiken?
- Hur approximerar uppgiftslämnaren sina svar, när uppgifter inte finns i bokföringen, inte med önskad periodisering eller inte på "rätt organisatorisk nivå"? Hur osäkra är approximationerna?
- Hur orienterar uppgiftslämnaren genom blanketten? Förbises några frågor eller någon information?
- Hur lång tid går det åt för att fylla i hela blanketten och olika delar av den?

6.6 Värdering av testresultat

Ett uppenbart sätt att bestyrka tillförlitlighet hos ett test skulle vara att upprepa det, helst med oberoende testledare och intervjuare. Om upprepningen leder till samma slutsatser båda gångerna, har man åtminstone i det aktuella fallet stärkt tilltron till resultaten. Då detta är näst intill ogörligt på grund av att beställaren inte vill sätta till tillräckliga resurser, får man provisoriskt lita till processkontroll och resultatens överensstämmelse med beprövad erfarenhet. Omfattande empiriska kunskaper finns nu om kapaciteten hos olika kvalitativa metoder och hur de kan användas för att komplettera varandra.

För att ge insikter i kvaliteten hos ett kognitivt test av en blankett innan man reviderar blanketten, bör testledaren som minimum ha redovisat:

- kompetens, organisation och vana hos ”testande” organisation
- antal testpersoner och deras fördelning på ”intressanta grupper”
- kriterium för rekrytering av testpersoner, rekryteringsmetod och grad av medverkan
- testets allmänna uppläggning och använda testverktyg
- tidsåtgång för testet
- så detaljerad dokumentation av testresultat att en användare självständigt kan ta ställning till slutsatserna.

För att få en översiktlig uppfattning om hur testet fallit ut, kan man göra enkla beskrivningar av

- hur antalet kommentarer och observationer fördelar sig på testpersonerna
- hur kommentarer fördelar sig på olika delar av blanketten
- fördelningen på spontana synpunkter och synpunkter fördelar sig efter probes
- vilka typer av blankettbrister som indikerats
- hur resultaten stämmer överens med hypoteserna och mätteknisk erfarenhet.

Påtagligt ojämna fördelningar kan t.ex. peka på att några testpersoner varit överkritiska eller ointresserade, att blanketten varit tröttande eller mindre engagerande i olika avsnitt eller att testprocessen på något annat sätt inte fungerat som avsett. Fördelningen på typ av synpunkter indikerar om blanketten främst har tekniska brister eller om den varit kognitivt svår.

Varje observation måste värderas var för sig med hänsyn till mättekniska erfarenheter, eftersom fördjupningsfrågorna är öppna och svaren icke standardiserade. När testet görs med enbart ett tiotal testpersoner, har dessa rekryterats så att de ska ha olika bakgrund, och man räknar inte med att de ska uppmärksamma samma svagheter hos blanketten. Ett litet antal testpersoner räcker heller inte till för att upptäcka alla väsentliga svårigheter i blanketten. Några fel kan passera utan att uppmärksammas. Med 30 till 40 testpersoner har man betydligt bättre möjligheter att få en uppfattning om problemen är vanliga eller upptäcka mindre vanliga problem.

Att en testmetod upptäcker brister i en blankett är inte tillräckligt i sig, om man inte samtidigt lyckas göra en bra revision av den. Då man upptäckt språkliga fel, felaktigt ordval, logiska fel, att något svarsalternativ saknas, eller att hoppanvisningar har glömts bort, är det oftast uppenbart hur frågan ska förbättras. Ändringar som berör begrepp kan vara mer svårfunna. Att man ändrat en icke fungerande fråga betyder inte automatiskt att den nya frågan är invändningsfri. Ett meningsfullt kriterium på god kvalitet hos test och revidering

får man om svårigheterna är färre och mindre betydelsefulla när man testar den reviderade blanketten. Om tiden räcker till är det effektivare att göra två eller flera test med mellanliggande revision än att bara göra ett större test och sedan använda en okontrollerad revision.

De kvalitativa resultaten ger bara den första indikationen på vilken datakvalitet som kan uppnås i undersökningen. Nyttan av att förbättra en blankett mäts ytterst av hur mycket skattningarnas tillförlitlighet och produktionens effektivitet förbättrades av upptäckterna i testet och den uppföljande revideringen tillsammans. När ändringarna väsentligt påverkar uppgiftslämnarbördan och det är oklart vilka effekter de kan få på faktorer som bortfall och kostnader, är ett kvantitativt experiment motiverat.

6.7 Översikt, checklista och resultat

Blanketten kan förefalla perfekt ”på skrivbord” för statistikproducenten, men det är inte säkert att den är lika perfekt för uppgiftslämnaren. Han/hon känner kanske inte igen begrepp och definitioner, anvisningarna kan tyckas svåra att förstå och svåra att hitta, vissa uppgifter finns kanske är svåra att hitta eller kräver alltför mycket omräkningar. Layouten kan upplevas som oklar så att man inte vet vilka frågor man ska svara på.

Kognitiva test har förmågan att upptäcka hur en uppgiftslämnare förstår och uppfattar begrepp, frågor och instruktioner i blanketten, hämtar fram uppgifter ur arkiv eller minne, bearbetar och organiserar dem och passar in svaren/uppgifterna i blankettens svarsscheman/alternativ. De kan visa vilka frågor som blivit missförstådda och hur missförstånd uppstått och var alltför stora krav ställts på förmågan eller viljan att lämna och bearbeta uppgifter. Testet har också förmågan att upptäcka tekniska fel (t.ex. glömda hoppanvisningar) och innehållsfel (otillräckliga svarsalternativ) i blanketten. Det ställer ibland beställaren inför ett beslut att utelämna svärmätta variabler eller att acceptera mätningar med större osäkerhet. Om arbetet i de föregående faserna inte är tillräckligt väl genomfört för att undanröja sådana fel, dras uppmärksamheten i det kognitiva testet från de brister i blanketten som det är särskilt lämpat för att upptäcka. Ett test-retest-förfarande är nödvändigt, när det första testet avslöjat sådana brister att märkbara revideringar måste göras.

CHECKLISTA : Kognitiva test

1. Identifiera avsnitt och variabler/frågor i blanketten samt layoutaspekter, som uppgiftslämnaren kan ha svårt att hantera. Bedöm variablerna efter uppgiftslämnarbörda och kognitiv svårighetsgrad. Bedöm blankettens svårighetsgrad totalt och hur den behöver testas. Bedöm den totala uppgiftslämnarbördan.
2. Formulera hypoteser om orsaken till svårigheterna.
3. Välj testmetod och verktyg för testet, t.ex. kognitivt blanketttest med särskilda fördjupningsfrågor (probes), ”tänka-högt”-metoder eller djupintervjuer.
4. Definiera önskvärda egenskaper hos testgruppen, bestäm antalet testobjekt, plats för test och definiera principerna för rekrytering av testpersoner.
5. Genomför testet. Testa även hur man förstår och använder introduktionsbrev och informationsmaterial.
6. Dokumentera testet i detalj och redovisa vilka hypoteser som fått stöd och oförutsedda svårigheter, som upptäcktes.

7. Förbättra blanketten där brister upptäcks. Då väsentliga förändringar krävs, testas den reviderade blanketten igen (eventuellt i flera omgångar.)

Resultat: En reviderad version av blanketten, som är anpassad till uppgiftslämnarnas förmåga att förstå frågor och anvisningar och till deras kapacitet och vilja att svara. Frågorna är efter denna fas anpassade till datainsamlingsmetoden, men blanketten är normalt inte tekniskt färdig att användas i produktionen av en huvudundersökning. Alternativt kan man i detta skede anse att det är omöjligt att mäta undersökningsvariablerna med tillräcklig tillförlitlighet och därför beslutar att inte genomföra undersökningen.

Att en blankett är testad och reviderad efter bästa metod och kunskap innebär inte automatiskt att man tror att den kommer att kunna fungera felfritt för alla variabler. Det kommer att finnas signaler om att vissa variabler helt enkelt är svåra att mäta hur man än formulerar frågorna. Trots detta behåller man dem, då beställaren vill ha uppgifterna med den tillförlitlighet som kan erhållas hellre än att helt avstå från dem. Utfallet av de kognitiva testen kommer då att ge varningar om användningen av dessa frågor och uppslag om hur man ska bevaka deras tillförlitlighet i de senare faserna av processen.

Referenser

Oksenberg, m.fl. (1991) *New Strategies of Pretesting Survey Questions* JOS 91:3
Knäuper B, et al. (1997) *Question difficulty and Respondents' Cognitive Ability: The Effect on Data Quality*. JOS, Vol. 13, No. 2, 1997.

7 Fas 4 - Experiment

7.1 Innehåll

Arbetsuppgift: När en blankett är kvalitativt färdigtestad och reviderad, behöver man som regel även testa den kvantitativt för att se om den är tillräckligt bra för att användas i statistikproduktion. Man behöver kontrollera om blanketten och datainsamlingsmetoden ger tillräckligt god tillförlitlighet jämfört med beställarens krav. För den med sig oönskade effekter som högt bortfall eller höga produktionskostnader? Ibland behöver man jämföra två eller flera föreslagna lösningar för att se vilken av dem som är bäst med avseende på deras inverkan på ett antal faktorer, t.ex. mätkvalitet, bortfall, tidsåtgång och kostnader.

7.2 Metodik

Metodikerna är traditionell experimentplanering, dvs. man dimensionerar ett sannolikhetsurval så, att man kan testa hypoteser genom att skatta en metodeffekt eller en skillnad med tillräcklig noggrannhet för att kunna fatta beslut. Före engångsundersökningar gör man **fristående** experiment. I löpande undersökningar kan metodeffekter mätas genom **inbyggda** experiment, dvs. man inför den nya metoden i ett slumpmässigt delurval av hela undersökningen. Alternativt kan ett tilläggsurval användas.

Ett experiment utvärderar insamlingsförfarandet i sin helhet men kan inte upptäcka uppgiftslämnarnas svårigheter med enskilda frågor. Med hänsyn till att experiment är betydligt dyrare och tar längre tid att genomföra än kognitiva test, vore det ett anmärkningsvärt slöseri med tid och resurser om man hoppat över kognitiva test och först i denna fas upptäcker brister i blanketten. De versioner av en blankett som används i experiment och evalveringar ska därför vara kognitivt färdigtestade och reviderade. P.g.a. minskade resurser och tidspress inom statistikproduktionen har experiment blivit allt mera sällsynta som beslutsunderlag.

Fristående experiment kan göras som förberedelse till en ny undersökning eller parallellt med en löpande undersökning som behöver revideras eller evalveras. De har fördelen att kunna utformas utan låsningar till en ”gammal” insamlingsmetod. Samtidigt finns det en risk att de fristående experimenten inte tas på fullt allvar av uppgiftslämnarna, särskilt sådana som brukar delta i undersökningar med uppgiftslämnarplikt.

När experiment görs inom ramen för den ordinarie surveyproduktionen, löper man en risk att också den egna organisationen ger dem lägre prioritet än ”riktiga” undersökningar. Om experimentet t.ex. förutsätter att man anlitar intervjuare, kan detta lätt komma i konflikt med att intervjuarkapaciteten ofta behövs i de stora löpande undersökningarna. Testets komplexitet, t.ex. hur många faktorer som testas, måste vägas mot risken att försöksplanen kan hotas av oförutsedda prioriteringar.

I löpande undersökningar kan ett **experiment vara inbyggt** i en eller flera produktionsomgångar. I undersökningar med obligatoriskt deltagande är det ofta det realistiska alternativet och till fördel för svarsfrekvens, svarskvalitet och tidshållning, eftersom både producent och uppgiftslämnare kommer att ta det på allvar.

Det finns dock en konflikt i inbyggda experiment. När resultaten från experimentdelen av urvalet avviker signifikant från den del av urvalet som använder den vanliga blanketten, kan resultaten från experimentdelen inte utan vidare vägas in i skattningarna. I värsta fall måste testresultatet förkastas. Antalet urvalsobjekt som skattningarna i huvudundersökningen kan använda minskar därmed i motsvarande mån. I löpande månads- eller kvartalsvisa undersökningar kan man reducera skaderisken genom att dra små urval till experimentet under en följd av månader. Då kan experimentet avbrytas så snart det givit resultat.

Som mått på hur väl en blankett och insamlingsmetod fungerar, används ofta flera kriterievariabler i kombination. Ofta fäster man större vikt vid att mäta effekten på andra kvalitetsindikatorer än mätkvaliteten. De vanligaste är:

- nivå på objektbortfallet
- nivå på partiellt bortfall i viktiga undersökningsvariabler
- nivå på skattningarna. (Hur höga utgifter, hur många resor osv. redovisades?)
- omfattningen av granskning och rättning
- tidsåtgången för att lämna uppgifterna
- kostnaden för att samla in, registrera och granska uppgifterna.

7.3 Exempel på beslut efter experiment vid SCB

Uppgiftslämnarbördan. Undersökningen av hushållens utgifter (HUT) har under en lång följd av år haft problem att uppnå en god svarsfrekvens. Därför har producenterna inför ett antal undersökningsomgångar bestämt hur lång bokföringsperioden ska vara och vilka gåvor (*incentives*) som ska ges till uppgiftslämnarna. I experiment hade man studerat effekten av olika gåvor på bortfall, svarsfrekvens och svarskvalitet.

Incentives. HUT valde 1985 att ge uppgiftslämnarna gåvor i förväg efter ett inbyggt experiment. Där visades en positiv effekt på antalet fullständigt ifyllda kassaböcker och på andelen svarande av att ge uppgiftslämnarna en korräknare i förväg.

Datainsamlingsmetod. Undersökningarna av levnadsförhållanden (ULF) jämförde vid sin start svarsfrekvens och svarskvalitet vid hushållsintervjuer respektive vid individintervjuer i ett inbyggt experiment under hösten 1974. Relativt små skillnader förelåg. Som en följd av resultatet beslöt man att använda den mer krävande men också mer informativa med hushållsintervjuer.

Teknikbyte. Arbetskraftsundersökningarna (AKU) beslöt 1980 att gå över från telefonintervjuer med pappersblankett till datorstödda telefonintervjuer. Metodbytet föregicks av ett stort inbyggt experiment, där intervjusvaren följdes upp genom återintervjuer. Experimentet visade god överensstämmelse i svarsfördelningen i de flesta variabler men skillnader bara hos två. Den ena skillnaden åtgärdades. Den andra innebar en förbättring.

Valet av blanketttyp. Undersökningen av Hushållens utgifter (HUT) valde att använda en blankett med förtryckta spalter efter att ha jämfört fyra olika versioner med avseende på antal svar, utgiftsnivå och svarsfrekvens i ett fristående experiment.

7.4 Översikt, checklista och resultat

CHECKLISTA: Experiment

1. Beskriv frågeställningen
2. Fastställ beslutskriterier
3. Formulera testbara hypoteser
4. Designa och dimensionera experimentet
5. Genomför experimentet
6. Hypotespröva och tolka utfallet
7. Fatta beslut

Resultat

Fas 4 ska leda fram till ett antal avgöranden om undersökningen:

att genomföra en undersökning med den utvärderade metoden för mätning och insamling - alternativt med den bästa av de jämförda metoderna

att revidera undersökningens blankett och design dvs. att börja om i någon av de föregående faserna

att inte alls genomföra undersökningen.

Experiment bör utföras så, att de ger mesta möjliga information om tillförlitligheten som kan användas i den avslutande kvalitetsredovisningen om det första alternativet väljs.

Referenser

Groves Robert M. (1989) *Survey Errors and Survey Costs*. John Wiley and Sons.

8 Fas 5 - Anpassning till produktionen

8.1 Innehåll

Arbetsuppgift: För en innehållsmässigt och mättekniskt väl utprövad blankett väntar man sig att bara marginella justeringar ska behövas. Men vid produktion i full skala kan det finnas tekniska förutsättningar som inte fanns vid kognitiva test (Fas 3) och vid experiment (Fas 4). Det kan finnas moment i själva genomförandet eller problem av olika slag som inte alls var i fokus vid kognitivt test och experiment eller som inte är tillräckligt vanliga för att upptäckas hos ett litet antal objekt. Blanketten kan behöva teknikanpassas ytterligare innan den kan användas i produktion. Eftersom fler personer och funktioner medverkar i en produktion än i ett test kan utförligare anvisningar för blankettens hantering behöva skrivas ner.

8.2 Produktionsanpassning av blanketten

Det finns olika skäl till att producenten måste anpassa mättekniskt färdiga blanketter till den valda tekniken för utsändning, insamling och registrering.

Av kostnads- och tidsskäl tar man inte alltid fram en datoriserad blankett till testen - fastän man planerar att använda en sådan i undersökningen - utan testar en pappersversion. Producenten måste då anpassa blanketten till sitt system för registrering av uppgifterna och göra ett produktionsprov, t.ex. för att få veta om hoppanvisningar som programmeras in fungerar i praktiken.

Till ett kognitivt test räcker det med en blankett gjord i Word. För huvudundersökningen kan det krävas att producenten gör en omfattande programmering i ett särskilt program för blankettkonstruktion. Detta görs i fråga om massutskriften med t.ex. förtryckta adresser, kontaktpersoner, förtryckta uppgifter från tidigare undersökningsomgång, begränsning av antalet frågor i varje utskrift beroende på respektive företags verksamhetsområde.

Många undersökningar måste använda sig av mer än en metod för datainsamling för att hålla uppe svarsfrekvensen. När blanketten enbart har testats på den viktigaste insamlingsmetoden, måste den anpassas så att den fungerar bra också på de övriga. Detta kan bäst göras i samverkan mellan mättekniker och producent.

Beklagligtvis utvecklar man inte blanketter systematiskt i alla undersökningar utan många går in i fas 5 med felbemängda blanketter. De kommer då först här att upptäcka ett antal brister som borde ha varit åtgärdade tidigare. Konsekvenserna blir förseningar (till och med avbrott) eller/och lägre tillförlitlighet. Alla typer av brister kan heller inte upptäckas med de verktyg som används i denna fas.

Studierna i denna fas inriktas på att se hur blanketten fungerar vid utsändning, i kontakten med uppgiftslämnare, vid insamling, registrering och mäta resurs- och tidsåtgång. Man är beredd på att se över t.ex. rutiner för avregistrering och bortfallsuppföljning, informationsmaterial och intervjuarutbildning. När det gäller frågornas innehåll och utformning, räknar man dock med att det bara är marginella kvarstående brister som ska upptäckas och rättas till.

8.3 Verktyg för anpassning

För krävande engångsundersökningar och större modifieringar av löpande undersökningar behöver man göra ett ”mellanstort” test (med några hundra uppgiftslämnare), som omfattar hela produktionskedjan. Ett sådant test kallas ibland pilot eller provundersökning.. Urvalet bör vara så stort och undersökningen så designad att man får generaliserbara resultat. Man använder sig här av verktyg som också används vid kognitiva studier, speciellt observationer av uppgiftslämnandet och debriefing (utfrågning i efterhand) både av egen personal och av uppgiftslämnare.

Vid intervjuundersökningar ger debriefingen med intervjuare (och ev. uppgiftslämnare) information om vilka frågor och avsnitt som intervjuaren haft svårt att klara av. Det är viktigt att mäta hur lång tid intervjun tar, eftersom intervjutidens längd väger tungt vid beräkningen av kostnaderna för att genomföra en undersökning. När lång intervjutid visar att uppgiftslämnarbörderna blivit avskräckande höga, ger provresultaten vägledning om vilka svåra och tidskrävande frågor man kan ta bort eller förenkla.

Insamlingen i många undersökningar standardiseras så, att man i dem använder ett produktionssystem med fasta och redan utprovade produktionsprocesser. I sådana fall är det av tids- och kostnadsskäl sällan motiverat att göra ett test i hela produktionsprocessen, utan man nöjer sig att studera de processer där man avviker från sin standard och därför befärrar att oförutsedda svårigheter kan uppträda. Sådana test brukar kallas pilotundersökningar (*pilot studies*). Men denna benämning används också för andra typer av förundersökningar.

En sådan åtgärd är att göra **provintervjuer**. Provintervjuer är kompletta intervjuer, som görs precis på samma sätt som de senare ska göras i huvudundersökningen. En observatör är närvarande och intervjun spelas in på band. De utvärderas genom dels genomgång av observationer och bandinspelningar, dels debriefing av intervjuarna. Främst görs provintervjuer för att upptäcka om det finns brister i blanketten som var för ovanliga för att märkas vid ett litet kognitivt test och svårigheter att hantera information och frågor som de bättre utbildade testintervjuarna inte hade.

Provintervjuer görs av de ordinarie intervjuarna. Därför kan man inte använda sig av kognitiva verktyg som probes och ”tänka högt”-metoder. Observatören ser vad som händer på ytan av intervjun, men har inte möjlighet att upptäcka underliggande förhållanden. De ger information om problem bara i den utsträckning som uppgiftslämnarna signalerar dem och intervjuarna observerar dem spontant. Provintervjuer är därför inte ett lika kraftfullt verktyg som testintervjuer för att identifiera uppgiftslämnarens behov av förbättringar i blanketten. Jämfört med ett kognitivt test ger provintervjuer ingen information om uppgiftslämnarnas resonemang eller skälen till deras eventuella misstag. Däremot ger de tillförlitligare information om hur lång tid som går åt för att genomföra intervjun, eftersom de görs under helt ”fältmässiga” villkor. Att en fråga besvaras snabbt och med rimliga värden är ingen tillräcklig garanti vare sig för att frågan blivit rätt förstådd eller rätt besvarad. Användning av observatör gör det till en dyrare metod att använda än standardiserade test redan när man ser till styckekostnaden. Vid SCB använder sig Undersökningen av levnadsförhållanden (ULF) av provintervjuer varje gång den tar in nya frågor i blanketten.

En billigare och mindre ambitiös metod än provintervjuer kallas ”*Question rating by interviewers*” - ungefär frågevärdering av intervjuare. Intervjuare

tränas att vara observanta på feltyper och deras orsaker, men observatörer används inte. Efter att ha genomfört ett överenskommet antal intervjuer får intervjuarna klassificera frågorna efter deras svårighetsgrad enligt ett schema.

8.4 Felsignaler och åtgärder

För att försäkra sig om att insamlingen fungerar som tänkt och eventuellt kunna göra begränsade revideringar av insamlingsmetod och blankett, kan man bevaka ett antal felsignaler under insamlingens inledande skede. Åtgärder är möjliga för centralt lagrade elektroniska blanketter, t.ex. i datorstödda telefonintervjuer eller då uppgiftslämnaren hämtar en elektronisk blankett från en webbsida. Så länge bara ett litet antal har besökt denna, kan man korrigera sent upptäckta misstag. Att göra korrigeringar då man använder pappersblanketter som alla skickats ut vid ett tillfälle, blir däremot både svårt att administrera och avskräckande dyrt.

Begränsade ändringar i blanketten är lättast att göra i datorstödda intervjuundersökningar, där man kontakter uppgiftslämnarna successivt. Då kan man från ena dagen till den andra revidera blanketten. Vissa svar som givits i intervjuerna före förbättringen måste kanske kompletteras eller kodas som partiellt bortfall. I värsta fall måste hela intervjun förkastas. När alla intervjuare antingen arbetar i en central telefongrupp eller kan nås med e-post, kan producenten snabbt föra en systematisk dialog. Man kan rätta till brister i blanketten som någon upptäckt, förtydliga anvisningar och stärka undersökningens administration. En elektroniskt tillgänglig sammanställning av ”problem och lösningar” resulterar i en ständigt aktuell samling råd. Insamlingsprocessen kan förbättras successivt och större enhetlighet uppnås.

Översikt av möjliga åtgärder vid påvisade brister:

1. Punktvisa revideringar av anvisningar och variabler/frågor i centralt lagrade elektroniska blanketter.
2. Utsändning av kompletterande information och vid allvarliga fel även reviderade blanketter till uppgiftslämnare i postenkäter.
3. Fortsatt (i fas 6) och fördjupad uppföljning för att få:
 - underlag för test och revideringar av blanketten till följande omgång av undersökningen om den är löpande samt
 - till kvalitetsdeklaration.
4. Det inträffar att undersökningar med icke testade blanketter ger så starka felsignaler att undersökningen måste avbrytas. Först efter förnyad genomgång från Fas 3 eller tidigare kan undersökningen eventuellt startas på nytt.

8.5 Provintervjuer - exempel

Följande exempel är hämtat från ”Ett miljöanpassat Stockholm”. Det är en engångsstudie, som genomförts som besöksintervjuer med pappersblanketter. Beställarna hade själva tagit fram en blankett med tillhörande svars kort. De som konstruerade blanketten hade tidigare arbetat med frågekonstruktion och var väl insatta i ”vad man bör tänka på när man formulerar frågor”. Flera personer medverkade i frågekonstruktionen, och man hade ”skrivbordsgranskat” varandras frågor. Många av frågorna var hämtade från SCB-undersökningar (framför allt ULF) och därmed också utprovade i andra sammanhang. Blanketten föreföll väl genomtänkt. Man valde därför att lita till provintervjuer för att se hur lång intervjutiden blev och hur frågorna i allmänhet fungerade.

I testet deltog fem intervjuare, som tillsammans gjorde 20 provintervjuer.

- **Intervjutiden** var drygt 77 minuter i genomsnitt. Tiden var för lång. Genomsnittstiden fick inte överstiga 60 minuter.

Efter att ha hört intervjuarnas beskrivningar av vilka frågor som fungerade dåligt beslöt uppdragsgivarna vilka frågor som skulle strykas eller kortas ned.

Uppdragsgivarna förbättrade instruktionerna i enlighet med intervjuarnas önskemål. För de frågor där det inte fanns svarskort skrev man ut tydligare om det var tillåtet eller ej att läsa upp alternativen.

Vissa frågor var mycket svåra att besvara direkt. De handlade t.ex. om ”hur stor elförbrukning man hade under det föregående året” och ”hur långt man åkt med hushållets bil”. Uppdragsgivarna insåg att det var nödvändigt att på något sätt informera uppgiftslämnarna i förväg om svårbesvarade frågor, t.ex. om elförbrukning. Ett separat informationsblad togs fram som intervjuarna kunde skicka när de hade bokat in tid för intervjun.

Frågorna kom inte i logisk följd och dispositionen i blanketten visade sig vara mindre bra. Både intervjupersonerna och intervjuarna upplevde intervjun som ”lite rörig”. Uppdragsgivarna delade därför in blanketten i tio olika block, där frågorna kunde placeras in. Man fick en bättre helhetsbild.

Sammanfattning:

Provintervjuerna gav uppslag till flera förbättringar i blanketten och av intervjuarnas arbetsrutiner. Genom en lyhörd beställare, medverkan av skickliga intervjuare och **att det fanns tid**, kunde mycket av undersökningens relevans räddas. Om man följt den generella testmodellen, hade redan ett kognitivt test i Fas 3 visat att ett antal komplicerade frågetabläer måste tas bort eftersom de inte besvarades tillräckligt väl. Då hade man också kunnat utveckla blanketten för DATI, vilket hade varit ett billigare alternativ.

8.6 Checklista och resultat

Checklista: Verktyg som ger felsignaler

1. Spontana kommentarer av uppgiftslämnare - skriftligt, på telefon osv.
2. Spontana observationer som den egna personalen gör
3. Debriefing av egen personal
4. Debriefing av ett litet antal uppgiftslämnare

Dessutom i intervjuundersökningar (enklast i en CATI/CAPI - miljö):

5. Provintervjuer
6. Beteendekodning (inkl. observationer i fält) för att spåra ”svåra frågor”
7. Medlyssning, speciellt i telefonintervjuer

Resultat: En blankett som fungerar både hos uppgiftslämnaren och vid hantering, insändning, avregistrering, dataregistrering och andra produktionssteg i en fullskaleundersökning.

9 Fas 6 - Utvärdering

9.1 Innehåll

Arbetsuppgifter :

I många undersökningar är brister i mätförfarande och datainsamling den största källan till osäkerhet. Studier av dess betydelse måste därför planeras och genomföras. Informativa mått på tillförlitlighet och indikatorer på hur väl olika processer fungerar behöver beräknas. Resultatet är ett underlag till kvalitetsdeklarationen.

När fel upptäcks i en löpande undersökning, behöver man identifiera deras orsaker och eliminera dessa genom att på nytt testa och revidera blanketten. För detta räcker det inte med mått och indikatorer på felens storlek. Det krävs också kunskap om hur de uppkommit, vilket förutsätter kognitiva studier.

9.2 Kvalitetssäkring

Kvalitetssäkring innebär att man mäter om produktionsprocessen fungerar som planerat och resultaten håller den kvalitet som utlovats. Försämringar kan uppträda på grund av dåligt genomförande men också på grund av att den undersökta företeelsen eller förutsättningarna för produktionen har ändrats så att blankett och/eller undersökningens design måste ändras. Vid sidan av kvantitativa mått som man tar fram vid experiment och evalveringar, finns en rad kvantitativa och kvalitativa signaler på var det finns risk för låg svarskvalitet. Vissa kan peka på att blanketten är för omfattande och uppgiftslämnarbördan är för hög, andra på att den innehåller känsliga frågor och åter andra på att det krävs alltför mycket arbete för att ta fram svar. Mått och indikatorer tas fram med lämplig gruppindelning av de svarande så att man kan se om problemen är genomgående eller koncentrerade till delar av undersökningen. Ibland kan man upptäcka 'riskgrupper' som haft svårare än andra att svara korrekt.

När man gör datorstödda intervjuer och använder intelligenta elektroniska blanketter, kan räknare och tidtagare byggas in i blanketten så att det blir möjligt att mäta antalet felaktiga och ändrade svar samt tidsåtgången i olika avsnitt av blanketten. Det är viktigt att få ett mått på hur mycket uppgiftslämnarbördan ökar, eftersom dessa blanketter kräver att uppgiftslämnaren själv ska editera de icke godtagna svaren.

9.2.1 Debriefing

Debriefing är en utfrågning i efterhand t.ex. hur en arbetsuppgift som att intervjua i en undersökning, att granska eller att koda kunnat genomföras. De som debriefas har i detta fall aktuella erfarenheter av blanketten eller andra aspekter på datainsamlingen som uppgiftslämnare/granskare/kodare/ intervjuare. Den som debriefar skaffar sig i förväg en överblick och gör en frågelista över de inslag i arbetsprocessen som man behöver ytterligare information om. Man lämnar också tid för att lyssna på spontana synpunkter. Debriefingen görs oftast med en person i taget, eftersom man vill ha oberoende synpunkter. Debriefing i grupp förekommer dock ofta av kostnadsskäl. Jämfört med en fokusgrupp är skillnaden att man debriefar efter en väl preparerad frågelista om en process som alla deltagarna medverkat i. Gruppen är ofta homogen, och alla känner varandra i förväg.

När en intervjuare ska debriefas om hur de svarande klarat sin uppgift, vinner man i effektivitet på att träna intervjuaren på vad hon/han ska vara observant på

under intervjun. Intervjuarna missar ibland sådana problem som observeras vid t.ex. beteendekodning. De har svårt att kvantifiera sina iakttagelser, men är däremot bra på att identifiera sådana inslag i frågorna och frågeordningen som kan leda till missförstånd. Det är svårt för intervjuare att säkert skilja på vilka svårigheter som är deras egna och vilka som är uppgiftslämnarens.

Debriefing med uppgiftslämnare (*Respondent debriefing*). En betydande fördel hos metoden är att man kan välja ut uppgiftslämnare med ovanliga värden/egenskaper. Sådana är mycket svåra att hitta när man behöver dem som testpersoner i ett kognitivt blanketttest. Debriefingen bör göras omedelbart efter intervjun medan den svarande kommer ihåg sina svar och man slipper kostnaden för att ta kontakt på nytt. Jämfört med ett kognitivt blanketttest har denna typ av debriefing mer avgränsade mål och standardiserat genomförande. Urvalsmetoden gör att man måste anlita den ordinarie intervjuaren, varför kognitiva verktyg bara kan användas i mycket begränsad utsträckning. Eftersom man även kan göra suburval som ett sannolikhetsurval, är det teoretiskt möjligt att skatta effekter.

9.2.2 Beteendekodning

Beteendekodning (*behaviour coding*) innebär att man under en pågående intervju observerar den intervjuades och/eller intervjuarens beteende fråga för fråga. Observationerna kodas enligt ett i förväg fastställt schema. I ett experiment, en större "generalrepetition" (*dress rehearsal*) eller huvudundersökning ger beteendekodning kvantitativa skattningar av frekvensen svårigheter. En värdefull användning är att mäta fel som är alltför lågfrekventa eller fel i frågor som ska besvaras av för få uppgiftslämnare för att upptäckas av ett litet kvalitativt test.

Skillnader mot ett kognitivt test är att man begränsar sig till ett strikt schema för observationerna och att man inte kan använda fördjupningsfrågor. Det är en standardiserad metod, som har små möjligheter att registrera och följa upp feltyper som ligger utanför kodningsschemat. Beteendekodning ger information fråga för fråga och belyser inte direkt effekten av frågeordning och sammanhang. Den upptäcker inte om uppgiftslämnaren missförstår frågan eller "höftar till" svaren så länge svaren kommer direkt.

Undersökningar med telefonintervjuer

En standardiserad version av beteendekodning är att intervjuaren, under pågående intervju, själv kodar den svarandes reaktioner. Enklast görs detta vid datoriserade telefonintervjuer. Intervjuarna tränas att via tangentbordet registrera ett litet antal typer av reaktioner. Dessa ska vara lätta att särskilja och kunna klassificeras entydigt.

Statistics Canada har infört en standard med indelning på 5-6 koder (alla används inte alltid) för olika beteenden.

CHECKLISTA: Beteendekodning av uppgiftslämnaren

- avbryter frågan *(Interruption)*
- ber att få frågan förklarad för sig *(Clarification)*
- ber att få frågan uppläst ytterligare en gång *(Repetition)*
- besvärad av frågan *(Uncomfortable)*
- undrar hur lång tid det är kvar på intervjun *(Timeout)*
- svarar inte exakt utan gör en uppskattning *(Estimation)*

En annan indelning som använts är att uppgiftslämnaren

- tvekar om eller dröjer med svaret
- svarar innan frågan har lästs färdigt
- missförstår frågan men inser inte detta själv *(Inadequate answer)*
- förstår frågan men säger att svaret är för svårt att minnas/känna till e.d.
- vägrar explicit att svara på just den frågan.

Antalet koder får inte vara större än att intervjuaren kan hålla dem i huvudet, observera och registrera dem under pågående datainsamling. Kodningen får inte ta så lång tid att den stör den ordinarie intervjuens rytm. Endast en kod per intervjuperson och fråga tillåts. Partiellt bortfall registreras för sig. Frånvaro av registrering innebär att intervjuaren uppfattat svaret som korrekt. Statistics Canada räknar med att intervjukostnaderna ökar med 0,4 % när beteendekodning används.

En fördel med denna form av beteendekodning är att den inte kräver specialistkunskaper. Intervjuarna kan utbildas snabbt. Samma kodsystém kan användas från undersökning till undersökning. Verktöget är enkelt och billigt att använda och kan sättas in även i provintervjuer för att upptäcka vilka frågor som bereder uppgiftslämnaren svårigheter.

En starkt förenklad version av beteendekodning används vid Zentrum für Umfragungen (ZUMA) i Mannheim och kallas problemkodning. Innebörden är att intervjuaren själv anger om hon/han märkt att det finns ett problem överhuvudtaget med frågan. Hon/han behöver dock inte specificera vilken typ av problem det är. Efter avslutad intervju kan intervjuaren anteckna vad svårigheterna bestod i och senare debriefas om dem.

Undersökningar med besöksintervjuer

I undersökningar med besöksintervjuer kan man naturligtvis utföra beteendekodning enligt ovan. Man kan också låta en observatör koda både intervjuarens och den intervjuades beteenden och reaktioner. Intervjun bandas samtidigt. Intervjuarens "felsignaler" kan kodas enligt ett schema som innehåller fler kategorier och kräver svårare bedömningar än när intervjuaren kodar själv. Intervjuarens arbete kodas efter om frågan lästs upp helt korrekt, lästs med en liten avvikelse eller med en stor avvikelse.

Kodningen sker i efterhand med hjälp av protokoll och bandinspelning. Genom att en observatör krävs, är metoden jämförelsevis dyr att använda. Den kräver också mycket välutbildad personal, analysen är svårare och det tar längre tid att få fram resultaten. SCB har tidigare gjort ett antal studier av denna art i undersökningar som redan var i produktion. Trots att man gjorde många vär-

defulla observationer, var det svårt att påvisa att dessa ledde till någon förändring av sättet att fråga i undersökningarna. Det är alltid svårt att få gehör för blankettrevisioner när produktionen startat.

9.2.3 Återintervjuer

Man skiljer på två huvudtyper av återintervjustudier. Målet med den ena är att mäta svarsvariation och ibland även bias. Den kräver därför relativt stora urval. Målet med den andra är att bedöma kvaliteten i intervjuandet. Den senare görs med små urval och kan kombineras med diagnostiska frågor, som söker förklaringar till motsägelsefulla svar både i intervjuarnas arbete och i uppgiftslämnarnas situation.

Uppskatta svarsvariation och bias

Tiden mellan första intervju och återintervju är ofta en kompromiss. Den ska vara så kort att uppgiftslämnarens verklighet inte hinner förändras och så lång att uppgiftslämnaren inte längre kommer ihåg exakt hur man svarade. Två till tre veckor är en vanlig längd på intervallet. Samma datainsamlingsmetod bör användas, för att inte metodeffekter ska blandas samman med svarsvariationen. Av kostnadsskäl begränsas återintervjuerna till ett slumpmässigt draget delurval. Stickprovsstorleken behöver som regel vara några hundra, för att osäkerheten ska kunna mätas med godtagbar precision.

Om uppgiftsbördan i återintervjun hotar att bli för hög, begränsas antalet frågor. Samtidigt som man koncentrerar sig på de frågor som är viktigast att studera, får dessa inte ryckas ur sitt sammanhang så att jämförelsen störs av att sammanhanget ändras (*kontexteffekter*). Återintervjustudier för att mäta svarsvariation har t.ex. gjorts i Arbetskraftsundersökningarna samt i Undersökningarna om levnadsförhållanden.

För att uppskatta svarsbias görs återintervjuer med reconciliering. I detta fall är det inte nödvändigt att använda samma mätmetod vid det senare tillfället. Reconciliering innebär att när svaren är olika, försöker man reda ut anledningen till skillnaden och fastställa vilket som är det rätta. Ibland kan man göra detta med hjälp av logiska kontroller och hjälpinformation. I andra fall måste man också kontakta uppgiftslämnaren. Vid datorstödd insamling kan svaren i första intervjun programmeras in, så att reconcilieringen kan göras direkt efter den andra intervjun. Återintervjuer med reconciliering har främst gjorts för evalvering av Folk- och Bostadsräkningarna.

Processkontroll

Återintervjuer som processkontroll används i löpande intervjuundersökningar. I varje undersökningsomgång dras ett mindre urval (25-50). Urvalet dras så, att intervjuerna begränsas till ett mindre antal intervjuare. Återintervjuerna görs av intervjuare som har särskild utbildning för detta och som inte vet vem som gjort den ursprungliga intervjun.

Återintervjun ska kunna upptäcka brister i hela insamlingsprocessen, i blankett och instruktioner och i intervjuarens arbetssätt. Brister i blankett och instruktioner kan åtgärdas till följande undersökningsomgång. Intervjuare med högre felfrekvens kan ges en målinriktad vidareutbildning. En förenklad form är en ren återkontakt för att få bekräftat att första intervjun är gjord över huvud taget.

AM/AKU och SCB/I har utvecklat ett system för återintervjuer med reconciliering (När svaren är olika vid återintervjun försöker man reda ut anledningen till skillnaden och fastställa vilket svar som är det rätta). Eftersom AKU-intervjun görs i datoriserad miljö, kan uppgifterna i den första

intervjun programmeras in, hämtas fram och omedelbart jämföras med svaren i den andra. Återintervjuarbetet startar redan fem dagar efter den första (ursprungliga) AKU-intervjun, för att uppgiftslämnaren inte ska hinna glömma detaljerna i sin sysselsättning under referensveckan. När svaren blir olika, undersöker intervjuaren med hjälp av fördjupningsfrågor vilket som är det riktiga svaret och varför skillnaden uppstod. Variablerna 'anknytningsgrad' och 'status på arbetsmarknaden' är de mest intressanta. Återintervjuaren ställer också ett par allmänt hållna frågor om undersökningen och upplevelsen av första intervjun. För att notera förklaringar och iakttagelser finns ett särskilt protokoll. Med ett sannolikhetsurval och större stickprovsstorlekar kan man mäta felens effekt på skattningarna. Summerat över sex månader och med totalt 2154 svarande visade studien att 7 % för anknytningsgrad till arbetsmarknaden var felklassificerade och 5 % för status.

9.2.4 Medlyssning

Vid medlyssning sitter intervjuaren och "medlyssnaren" i samma rum, så att medlyssnaren ser dataskärmen och tangentbordet och kan följa allt som händer i intervjun. Medlyssnaren är enbart observatör och får inte ingripa i intervjun. De har var sitt headset, och med hjälp av en omkopplare hör båda vad intervjupersonen (IP) säger. Innan intervjun startar, måste intervjuaren upplysa IP om att ytterligare en person lyssnar på intervjun. Det är sällan någon som reagerar negativt på detta.

I och med att medlyssnaren hör vad både intervjuaren och intervjupersonen säger, kan medlyssnaren studera hur blanketten fungerar för båda parter. Man får höra hur intervjuarna uppfattar frågorna, hur de läser dem och vilka ord de betonar, om de utelämnar eller lägger till ord och vilka fraser eller meningar, frågor, termer, begrepp, instruktioner osv. som ställer till problem för dem i kommunikationen med uppgiftslämnarna. Exempel på observationer vid medlyssning är att introduktioner varit för långa, att instruktioner har varit svårbegripliga eller att frågor kommit i ologisk följd, att en fråga som förefaller entydig har uppfattats på olika sätt av olika intervjupersoner.

Medlyssning kan användas vid en *dress rehearsal* eller i löpande produktion dvs. i faser där blanketten redan är mättekniskt färdig. I första hand är det intervjuarens hantering av frågor och information samt samspelet med uppgiftslämnaren man behöver skaffa ökad kunskap om.

Medlyssning är ett effektivt hjälpmedel för att identifiera intervjuarnas behov av grund- och vidareutbildning. Revidering av instruktioner till intervjuarna, ändring av frågeordning och förbättring av anvisningar till uppgiftslämnarna kan förkorta intervjutiden och därmed minska både uppgiftslämnarbörda och kostnader.

Medlyssning i någon form är relativt vanligt, när intervjuare arbetar samlade i en central telefongrupp. Jämfört med beteendekodning är medlyssning mer informativt om missförstånd och felbeteenden, men också mer tids- och kostnadskrävande både att genomföra och att analysera.

9.2.5 Kvalitetsbevakning genom intervjuare via e-post

Metoden är en starkt förenklad variant av "observationer i fält", utformad för att kunna rapporteras inom 14 dagar efter att den startat och göras till en låg kostnad. Den bygger på medverkan från intervjuarna, med specialutbildning av Mättekniska laboratoriet. Intervjuarna är vana att samarbeta med ML:s övriga personal och att rapportera systematiskt. Arbetet går till så, att ML-intervjuarna

får anvisningar om vilka problem som är särskilt viktiga att bevaka. Efter att ha genomfört ett begränsat antal intervjuer - normalt fem till tio - rapporterar de sina iakttagelser per e-post. Cirka två timmar används för rapporteringen. De får tala om vilka frågor som bereder störst problem. Intervjuaren börjar med det viktigaste och fortsätter så länge tiden räcker. Förbättringsförslag är också välkomna. Svaren ställs samman och en mycket kortfattad översikt och analys görs. Metoden används för att få underlag till undersökningens kvalitetsdeklaration och om undersökningen är löpande till att förbättra nästa insamling.

Kvalitetsbevakning via e-post upptäcker vissa av de frågor som uppgiftslämnarna har svårigheter att förstå och svara på. Den kan dock inte upptäcka när uppgiftslämnarna missförstår frågorna men svarar snabbt ändå. Det finns inte tid för fördjupningsfrågor, så metoden kan inte upptäcka orsakerna till svårigheterna på samma sätt som kognitiva test. Resultaten är inte kvantifierbara.

9.3 Revidera blanketter i löpande undersökningar

Löpande undersökningar behöver emellanåt revidera sin blankett. Flera typer av skäl kan finnas:

1. Den verklighet man mäter har ändrats.
Kan studeras med observationer, djupintervjuer eller fokusgrupper
2. Informationsbehovet har ändrats.
Kan påpekas av beställaren, uppmärksammas fortlöpande av producenten eller studeras systematiskt med fokusgrupper
3. Ord och begrepp i frågorna har fått ändrad innebörd.
Studeras bäst med kognitiva test
4. Man har upptäckt brister i blanketten.
Studeras med kognitiva test inriktade med hjälp av felindikatorer

De två första skälen är nära sammanhängande. Ibland förändras den omvärld som den löpande undersökningen vill beskriva. De variabler som en gång gav en relevant beskrivning av en verklighet gör det inte längre några år senare. En uppsättning frågor anpassade för att mäta ekonomiska eller sociala förhållanden 1990 är säkert inte lika relevanta 2000 och senare. Företeelser, förväntningar, attityder, organisationsformer, varor och tjänster förändras med tiden - men även människors inställning till dem. Klassificeringar som näringsgren, utbildning och socioekonomisk grupp revideras. I takt med att t.ex. arbetsmiljön förbättras och jämställdheten ökar, ökar också kraven. En situation som en gång betraktats som godtagbar räknas tio år senare som ett arbetsmiljöproblem.

När t.ex. näringslivet struktureras om, när rörligheten på och anknytningen till arbetsmarknaden ändras, när värderingar och vanor förändras, när miljöfrågorna får allt större vikt, måste blanketterna anpassas till den nya situationen. I annat fall kommer de varken att tillgodose statistikanvändarnas informationsbehov eller låta uppgiftslämnarna ge relevanta svar. Att innehållet i vissa (främst företags-) undersökningar är regelstyrkt får inte innebära att man bortser från att frågorna blir allt mindre relevanta för uppgiftslämnaren. KPI (Konsumentprisindex) är ett tydligt exempel på en undersökning som arbetar systematiskt med att förändra sitt innehåll allteftersom utbudet förändras.

En tumregel borde vara att varje löpande undersökning grundligt ser över datainsamlingsmetod och blankett minst vart femte år. Vid en revision av en blankett i produktion är man i samma situation som vid en ny undersökning och

behöver använda samma metoder för studium av relevans och innehåll. Arbetet måste börja i fas 1. Arbete i faserna 5 och 6 kan ge signaler om att en fråga inte längre är relevant för uppgiftslämnaren, t.ex. genom ökande andel partiellt bortfall, ”vet ej” eller ”ej aktuellt”.

Tidsplanen för en revidering styrs av att en löpande produkt inte kan byta blankett annat än vid årsskiptena. Man vill kunna summera till årsnivå över månads- eller kvartalsundersökningar. Revideringen kan därmed planeras och utsträckas över ett kalenderår, vilket ger tillfälle till att testa även den reviderade blanketten.

I panelundersökningar är det troligt att de uppgiftslämnare som deltar för första gången ger starkare signaler om svårigheter i blanketten. Man bör se på denna grupp separat. Till de följande undersökningsomgångarna kan en inläring ha skett om hur man ska undvika att ge svar som utlöser förfrågningar och återkontakter från producenten. Att en uppgift passerar granskningen innebär inte att den är riktig. Att lägga samman nya och tränade uppgiftslämnare kan därmed dölja felrisker.

Förslag till förbättring av blanketter i löpande undersökningar bromsas ibland med argumentet att ”då skulle tidsserierna förstöras”. Man hävdar gärna att trenden ändå ”är den rätta” - ett argument som inte har något generellt stöd. Med en fråga som ger systematiskt felaktiga eller icke relevanta svar kommer man att mäta trenden för en helt annan variabel än den som verkligen är intressant. När systematiska fel i uppgifterna uppstår genom exkludering av en komponent i en summa, kan man inte förutsätta att just den komponenten följer samma trend som de övriga följer en trend. Mätfelsvariation eller bortfall till följd av dåliga frågor ger dessutom sämre tillförlitlighet i skattningar av differenser och tidsserier.

9.4 Exempel

I detta avsnitt redovisas några exempel på hur undersökningar framgångsrikt använt felsignaler för att upptäcka mätproblem. Den följande revisionen av blanketten har medfört kvalitetshöjningar av olika slag.

Felsignaler via granskning

Granskningen ger utslag för värden som synes vara orimliga, inkonsistenta eller av fel storleksordning. Granskningen riktar uppmärksamheten på värden som är **outliers**. I en löpande undersökning ska **resultatet av granskningen** användas för att spåra sannolika felkällor i datainsamling och blankett så att felkällor kan undanröjas till kommande omgångar av undersökningen. Många gånger är oklarheter i blanketten orsaken till dessa fel. Stora partiella bortfall uppstår ibland pga. olika brister i blanketten.

Granskningen upptäcker inte ens systematiska mätfel så länge uppgifterna ligger inom kontrollgränserna. Den typen av fel kallas **inliers**. Inliers inträffar t.ex. när det är skillnader mellan definitionen av begreppet i undersökningen och definitioner i företagens administrativa system.

Extern evalvering

I det följande exemplet upptäcktes bristande överensstämmelser med en annan statistikprodukt som användes vid granskningen. Bristen i överensstämmelse berodde på att uppgiftslämnaren inte fått information om hur en storhet var definierad. Problemet löstes genom att producenten frågade efter en mer välkänd storhet och själv gjorde omräkningar.

Exempel 9.4.1

Byggproduktionsstatistiken på ES/IN är en årlig urvalsundersökning som genomförts sedan 1993. Insamling sker genom postenkät till de utvalda byggföretagen. Under de tre första undersökningsåren lade man i granskningsarbetet märke till att en fråga ofta missuppfattades av uppgiftslämnarna. Variabeln heter i blanketten "SUMMA prod. och adm. kostnader" och ska spegla "under året nedlagda kostnader". I granskningsarbetet har man tillgång till företagets skriftliga årsredovisningar där uppgift om Rörelsekostnad finns. Rörelsekostnaden för ett år är alla kostnader som fakturerats under året och kan innehålla arbeten som avslutats året innan men sakna arbeten som inte avslutats under året. Nedlagda kostnader är däremot de utgifter företaget haft under året, oavsett om arbetena har hunnit faktureras eller inte. När man jämförde svaren i Byggblanketten med Rörelsekostnad i årsredovisningen, märkte man att uppgifterna alltför ofta var identiska. Rörelsekostnad ska inte vara identisk med enkätsvaret när det finns pågående arbeten. Det visade sig att ca 90 procent av uppgiftslämnarna inte svarat med Nedlagda kostnader, eftersom den definitionen var obekant för dem. En tiondel beräknade och/eller uppskattade den efterfrågade variabeln. Det blev ett mycket omfattande gransknings- och rättningsarbete, eftersom man var tvungen att titta på alla objekt för att avgöra om uppgiftslämnaren svarat med Nedlagda kostnader eller Rörelsekostnad. De felaktiga objekten rättades. Rättningen var ett omfattande manuellt arbete. Det innebar att man från Rörelsekostnaden drog avslutade fjolårsarbeten och lade till icke avslutade arbeten för att få "nedlagda kostnader". Numera samlas Rörelsekostnader in i stället för Nedlagda kostnader. En omräkning till "Nedlagda kostnader" görs fortfarande, men i dag maskinellt. Omräkningen görs för alla företag. Någon verifiering behöver inte längre göras av vilken variabel som uppgiftslämnaren besvarat frågan med, eftersom en bekant variabel numera efterfrågas.

Bortfall och partiellt bortfall

I exemplet nedan signalerade ett stort objektsbortfall, stora partiella bortfall samt att många uppgiftslämnare ringde producenten tillsammans att uppgiftslämnarbördan i undersökningen var för hög. Användning av administrativa uppgifter och beräkningsalgoritmer vid skattning tillät att blanketter för små företag kunde revideras och förenklas.

Exempel 9 4.2

Under de första två åren som undersökningen Statistiken över Byggproduktion genomfördes, hade de små byggföretagen problem att lämna så detaljerade uppgifter om sin verksamhet som efterfrågades i postenkäten. En stor del av gransknings-, återkontakts- och rättningsarbetet ägnades därför åt små företag. Med en simuleringsstudie på 1994 års material visades att det arbete som lagts ned på småföretagen, av både uppgiftslämnare och SCB:s personal, inte hade någon märkbar påverkan på slutresultatets tillförlitlighet. En reducerad uppsättning svar för småföretagen med endast en handfull uppgifter, mot tidigare ett femtiotal, konstruerades med hjälp av 1994 års siffror. De uppgifter som inte skulle inhämtas från småföretagen skattades utifrån vad de övriga företagen svarat, med en modellberoende skattning. Skattningarna från den simulerade undersökningen jämfördes sedan med skattningarna i den ordinarie undersökningen. Ingen betydande skillnad påvisades. Slutsatsen blev att den förenklade blanketten går att använda på företag med upp till fyra anställda utan några större försämringar av tillförlitligheten. Den förenklade blanketten kunde användas av drygt 400 företag (36 procent av urvalet). Som väntat minskade arbetet vid granskning och rättning både för uppgiftslämnarna och för SCB:s personal. Revisionen av blanketten fick svarsfrekvensen att öka kraftigt. Bland företag med noll anställda har den ökat från 69 procent till 81 och bland dem med 1-4 anställda från 66 procent till 76.

Variabelns fördelning

I exemplet nedan kunde man vid **jämförelser av en variabels fördelning vid två tillfällen** se att antal större avrundningar ökat i svaren på en fråga. Ökningen signalerade att tillförlitligheten hade försämrats till följd av att undersökningen bytt insamlingsmetod från postenkät till telefonintervju.

Exempel 9.4.3

Bostads- och Hyresundersökningens (BHU) blankett frågar efter låneskuld på eget småhus. Uppgifterna samlades in med postenkät i 1991 års undersökning men med telefonintervju i 1993 års undersökning. En räkning gjordes av hur många uppgiftslämnare som svarade i jämnt hundratusentals kronor vid respektive insamlingsmetod. Det var 10 procent som svarade med jämnt hundratusental i postenkäten 1991. I telefonintervjuerna 1993 var det 22 procent som gjorde det. Skuld på lån är tydligen en sådan fråga som det är lättare att svara exakt på när uppgiftslämnaren får ta tid på sig och svara skriftligt.

9.5 Checklistor och resultat

De vanligaste måtten på tillförlitligheten i mätningarna samt indikatorer och signaler om mätfel som möjlig förklaring till störningar kan grovt delas in i tre grupper.

A: Skattningarnas svarsbias och svarsvarians

För att uppskatta storleken av svarsbias och svarsvarians måste man använda kvantitativa metoder baserade på sannolikhetsurval. Några vanligen användbara metoder är:

Skattningar erhållna genom:

1. Evalveringar av bias mot annan statistikprodukt - urval eller totalregister.
2. Uppskattning av mätfel genom inbyggt eller fristående experiment.

Dessutom vid intervjuundersökningar:

3. Återintervjustudier - med eller utan reconciliering ev med fördjupningsfrågor.
4. Mätningar av intervjuarvarians.

Kvantitativa studier har blivit relativt ovanliga sedan 1980-talet. Större evalveringsstudier har gjorts i samband med de senaste FoBarna. Några undersökningar har gjort återintervjustudier - studier med och utan reconciliering förekommer.

Att här gå in även på kvantitativa metoder skulle göra denna handbok onödigt omfattande, särskilt som metoderna är väl etablerade och kända. En metodöversikt finns hos Groves (1989) och en översikt av empiriska studier vid SCB hos Lindholm (se sid. 82)

B: Indikatorer och observationer ur produktionsprocessen

Om behovet av utvärdering, test och revidering gäller för hela blanketten eller för enskilda variabler signaleras av olika indikatorer.

B.1. Exempel på indikatorer med problem hos blanketten i sin helhet:

1. Hög nivå på objektbortfallet speciellt då det orsakas av vägran.
2. Långsam inströmningstakt av blanketter, hög andel sent inkomna svar
3. Många klagomål, negativa spontana kommentarer, frågor eller synpunkter från uppgiftslämnarna allmänt om undersökningen.
4. Högt antal återkontakter och kompletteringar p.g.a. ofullständigt ifyllda blanketter.
5. Hög tidsåtgång för att ta fram underlag och fylla i blanketten.

B.2. Exempel på indikatorer med problem hos en enskild fråga:

1. Hög andel partiellt bortfall.
2. Många "vet ej" - svar, ändrade svar och orimliga svar.
3. Oregelbundenheter i variabelns svarsfördelningen t.ex. "spikar", stor snedhet eller många outliers.
4. Hög frekvens av felmeddelanden vid logiska kontroller och storlekskontroller ev. även med stort behov av rättningar och återkontakter.
5. Låg kodbarhet vid automatisk kodning.
6. Många spontana meddelanden från uppgiftslämnarna om oklarheter och svårigheter i specifika frågor.
7. Många spontana iakttagelser av svårigheter som den egna personalen gjort
8. Hög arbetsinsats krävs av uppgiftslämnaren för att räkna om från befintliga uppgifter till den uppgift som efterfrågas i rutan.
9. I löpande undersökningar - tidsseriebrott eller stor variation i resultaten mellan undersökningstillfällena.

C: Särskilda evalverande kognitiva och andra kvalitativa studier

1. Inbyggda relevansstudier (t.ex. djupintervjuer med vissa uppgiftslämnare)
 2. Spontana observationer och debriefing, dvs. samma metoder som används vid kvalitetsbevakning vid undersökningens start
- Dessutom i intervjuundersökningar
3. Mätning av intervjutid och annan tidsåtgång
 4. Återintervjuer för kvalitetskontroll med hjälp av fördjupningsfrågor
 5. Beteendekodning
 6. Kvalitetsbevakning utförd av intervjuarna
 7. Medlyssning

Tolkningen av signalerna är i mycket en fråga om omdöme. Det finns ännu inga tumregler för hur stark en signal ska vara innan den ska leda till åtgärd. Indikatorerna säger inte säkert när en störning orsakats av fel i mätprocessen eller när den beror på något annat. Vilken åtgärd som är den effektivaste måste övervägas och ibland även testas. Andra åtgärder än en generell revision av blanketten kan ibland vara effektivare. T.ex. kan ett nytt informationsmaterial behövas till vissa grupper, intervjuarna kan behöva en kompletterande utbildning eller ett förstärkt program behöva sättas in för att säkra att alla produktionsprocesser fungerar som tänkt. I löpande undersökningar kan fördjupade kontakter med särskilt felbenägna uppgiftslämnare vara en effektivare åtgärd än blankettrevision, t.ex. om ett företag just ha bytt uppgiftslämnare och denne behöver kompletterande anvisningar om blanketten.

Resultat:

I alla undersökningar: Underlag till kvalitetsdeklaration av undersökningen.

I löpande undersökningar dessutom: Indikationer/mått på svagheter i insamlingsförfarande och blankett och förslag på åtgärder som kan genomföras direkt eller efter test.

Referenser

1. Andersson, C., Lindström, H.L., och Polfeldt, T. *Att mäta statistikens kvalitet*. R&D Report 1999:3.
2. Grankvist, L. *Granska effektivt*. SCB-rapport från granskningsgruppen mars 1997.
3. Groves, R. M, *Survey Errors and Survey Costs*. John Wiley & Sons 1989
4. *Evalvering av FoB 85. Sysselsättningsdata*. Statistiska meddelanden BE 42 SM 9001
5. *Evalvering av FoB 85. Hushållsdata och lägenhetsdata*. Statistiska meddelanden BE 42 SM 9002
6. *Journal of Official Statistics* , 1992:1, sid 63
7. Lindholm P. (1995) *Mättekniska studier vid SCB - en översikt*.

10 Fas 7 - Kvalitetsredovisning

10.1 Innehåll

Arbetsuppgift: Att sammanställa kvalitetsinformationen systematiskt och göra den begriplig, tillgänglig och användbar för beställare och andra användare.

- Det är producentens uppgift att göra en systematisk kvalitetsdeklaration. Mätteknikern lämnar underlag från sina studier. Målgrupp är inte bara beställaren utan även andra kommande användare.

Resultat: En kvalitetsredovisning enligt anvisningarna för SOS. I den ska ingå uppgifter om hur blanketten utvecklats och testats, indikationer från insamling och bearbetning samt skattningar av eller indikatorer på svarsvarians och -bias.

10.2 Anvisningar och förutsättningar

I de riktlinjer för kvalitetsredovisning som fastställdes 1999-10-11, definieras det nya begreppet för statistisk kvalitet för Sveriges officiella statistik (SOS) i MIS 94:3 och dess efterföljare. Den består av fem huvudkomponenter, som var och en består av ett antal delkomponenter.

Primärt är det mättekniska arbetet inriktat på förbättring av huvudkomponenternas innehåll och tillförlitlighet. Det bidrar också till en utförligare redovisning av osäkerhetsmått och höjer tillförlitligheten totalt.

Kvalitetsbegrepp för officiell statistik

Innehåll

- Statistiska målstorheter
 - Objekt och population
 - Variabler
 - Statistiska mått
 - Redovisningsgrupper
 - Referenstider
- Fullständighet

Tillförlitlighet

- Tillförlitlighet totalt
- Osäkerhetskällor
 - Urval
 - Ramtäckning
 - Mätning
 - Svartsbortfall
 - Bearbetning
 - Modellantaganden
- Redovisning av osäkerhetsmått

Aktualitet

- Frekvens
- Framställningstid
- Punktlighet

Jämförbarhet och sammanvändbarhet

- Jämförbarhet över tiden
- Jämförbarhet mellan grupper
- Sammanvändbarhet med annan statistik

Tillgänglighet och förståelighet

- Spridningsformer
- Presentation
- Dokumentation
- Tillgång till primärmaterial
- Upplysningstjänster

Eftersom kvalitetsdeklarationen är användarorienterad, ska den innehålla sådan information som hjälper användaren att bedöma mätkvaliteten hos olika variabler och avgöra vilka bearbetningar som kan göras med godtagbar tillförlitlighet samtidigt som den är översiktlig. Översiktligheten kräver att kvalitetsdeklarationen hänvisar till utförligare dokumentation och till en person

som man kan kontakta för att få fördjupad kunskap. Detaljerna i arbetet med blankettutveckling och produktionsanpassning rapporteras alltså inte så länge det inte ger kunskap av betydelse för att bedöma skattningarnas kvalitet. Merparten av detta arbete redovisas i metodrapporter.

10.2.1 Huvudkomponent: Innehåll

Mätteknikernas bidrag till urvalet och definitionen av variablerna består i att genom fokusgrupper och/eller djupintervjuer hjälpa till att identifiera frågeområden som uppgiftslämnarna vet är relevanta men som beställaren/producenten inte hade uppmärksammat under arbetet ”vid sitt skrivbord”. Det består också i att identifiera storheter som är mätbara, dvs. för vilka uppgiftslämnarna kan skaffa fram sådan information som man frågar om. Information om regler och överväganden som styr valet av variabler och åtgärder för att kvalitetssäkra innehållet bör redovisas. Användaren kan då bedöma kvaliteten i förarbetet och förstå motiven för att innehållet blivit vad det är.

Exempel

Innehållet i undersökningen har bestämts sedan man genomfört tre fokusgrupper bestående av arbetsgivare, arbetsförmedlare respektive arbetssökande.

Innehållet i undersökningen är i sin helhet fastlagt av EU-direktiv no xxxx .

Frågorna 7 - 19 är kopierade från en liknande undersökning genomförd 1995, för vilken de redan utvärderats.

Kontakter med företagen visar att flertalet inte kan redovisa transporter i volym, utan att man enbart kan få svar från alla om man frågar om transporterens vikt.

Av hänsyn till uppgiftslämnarbördan måste frågor om ärendet med resan, som kommittén från början önskade sig, prioriteras bort.

10.2.2 Huvudkomponent: Tillförlitlighet

I 1999 års reviderade anvisningar för kvalitetsdeklaration inom SOS står under rubriken **2.2.3 Mätning:**

”Beskriv det använda mätförfarandet. När frågeformulär använts bör detta redovisas i sin helhet eller i lämpligt sammandrag. Redovisa mätsvårigheter som uppträtt vid insamlingen av uppgifter och deras troliga konsekvenser för statistikens tillförlitlighet.”

”Om redovisade konfidensintervall fångar upp också osäkerhet från slumpmässiga mätfel, nämn detta och ge beskrivning här eller under Urval. Om justeringar för systematiska mätfel gjorts, beskriv dessa här eller under Urval.”

Mätkvalitet redovisas dels deskriptivt och konkret, dels i form av råd för användningen av statistikunderlaget.

Den deskriptiva delen redovisar t.ex. bias, svarsvarians och bruttofel hos centrala skattningar. Den rapporterar också vilka specifika åtgärder man vidtagit för att uppnå god mätkvalitet i undersökningen. Bedömningen om skattningarnas kvalitet är god nog överlämnas till användaren.

Den rådgivande delen anger, utan att kunna göra några preciseringar, vilka skattningar som är mer osäkra. För att veta att en fråga inte fungerar bra behöver man inte ha gjort en kvantitativ evalvering. En kognitiv studie räcker oftast för detta och kan dessutom ha indikerat orsaken till svårigheterna. I andra fall har beteendekodning eller observationer visat problem hos vissa frågor. Producenten har då tagit på sig ett ansvar och gjort en bedömning att skattningen inte är ”alltför dålig” och dessutom göra användaren uppmärksam på detta.

10.3 Mätteknisk kvalitet – exempel på redovisning

Blanketten i sin helhet samt en beskrivning av datainsamlingsmetoden och dess egenskaper

- *Undersökningen är ursprungligen en postenkät, men TDE (touchtone data entry) har införts som alternativ från och med 1997. TDE användes av 51 % av urvalet i den aktuella undersökningen. 22 % svarade med posten och 12 % faxade in sina svar. Bortfallet uppgår till 15 %. Telefonpåminnelse gjordes 21 dagar efter utskick. Sju procent av urvalet svarade först efter påminnelsen. Vid registrering av svar med TDE sker omedelbar storleksgranskning med möjlighet för uppgiftslämnaren att korrigera eller kommentera sina uppgifter. Metodstudier har visat att svaren på TDE har minst samma tillförlitlighet som svaren på pappersblanketter.*
- *Undersökningen är en intervjuundersökning och använder fortfarande pappersblankett. Intervjutiden är i genomsnitt 67 minuter. För att hålla svarsfrekvensen uppe accepteras telefonintervjuer i 11 % av svaren. För 4 % av alla svarande görs indirekta intervjuer med någon familjemedlem. I dessa är frågor av kunskaps- och attitydkaraktär (en femtedel av samtliga) undantagna. I telefonintervjuerna ser en trötthetseffekt ut att uppträda under den sista fjärdedelen av intervjun, genom att det partiella bortfallet genomgående är mer vanligt där.*

Systematiska och icke systematiska fel

Felen kan uppmärksammas på flera sätt. Evalvering, dvs. skattningar av tillförlitlighet hos nyckelvariabler och andra kvantitativa mått som beteendekodning, eftersträvas. Dessutom kan kvalitativ information genom spontana påpekanden från uppgiftslämnarna, genom debriefing av egen personal, genom spontana observationer vid mottagning och bearbetning av inkommande material t.ex. medlyssning visa på uppenbara svagheter i mätningen av vissa variabler.

- *Genom en evalveringsundersökning kunde man visa att antalet sysselsatta i tjänstenäringarna underskattades med tre procent. Det gällde särskilt tillfälligt anställda i storstadsområdena.*
- *Vid en återintervjustudie besvarades vissa frågor olika vid de båda tillfällena: Om Ip var "långvarigt sjuk" 7 % av urvalet, "hemmavarande barn" av 2 %, facklig tillhörighet av 8 % osv.*
- *Uppgiftslämnarnas anteckningar på de insända blanketterna visade att de hade haft svårigheter att välja bland svarsalternativen för fråga Qx*

Påpekanden om generella felmekanismer

Sådana kan t.ex. gälla hur glömskeeffekter beror av erinringstidens längd och variabelns betydelse, om det kan finnas egenintresse av att underrapportera, om uppgiftslämnarbördan i någon mening varit särskilt hög osv.

Äkta exempel hämtat från MIS 94:3 sid. 51.

- *”Priser ska vara fakturerade priser men i vissa branscher ger många uppgiftslämnare listpriser. Rabatter ska vara fråndragna men är det inte alltid...I en situation då minskad efterfrågan leder till ökande inslag av rabatter så överskattar ett index baserat på listpriser den verkliga prisutvecklingen. Motsatsen gäller när rabatterna minskar.”*

Påpekanden om svagheter hos enskilda variabler

- *Det partiella bortfallet på frågan var 7 procent.*
- *Vid återkontakter föranledda av granskningsutfallet visade det sig att 37 procent av transportörerna känner inte till volymen på lasten utan kan bara redovisa vikten.*
- *Spontan återrapportering från intervjuarna visade att frågan om försäljningspris för bostadsrätt var känslig för många och ledde till undvikande svar.*
- *Uppgiften att beräkna vilket våningsplan lägenheten låg på krävde ofta förklaringar och hjälp från intervjuaren.*

Äkta exempel från Levnadsförhållanden, appendix 15 sid 23.

- *”Frågorna i ULF-blanketten är mer eller mindre lättbesvarade. Det är ganska säkert att vissa frågor uppfattas olika av olika intervjupersoner. Speciell försiktighet bör iaktas vid tolkningen av svar på attitydfrågor, avståndsfrågor och frågor om hur ofta man gjort något, t.ex. motionerat eller träffat släkt och vänner.”*

Åtgärder för att nå god mätkvalitet

Det ska finnas en redogörelse för hur blanketten har utvecklats genom expertgranskning, kognitiva test och revision samt vilka standarder för frågor och klassificering som använts. Användaren får hjälp av redogörelsen att avgöra om producenten på ett rimligt sätt försäkrat sig mot misstag eller ej i formuleringar, begrepp, frågeordning osv.

- *Blanketten har ursprungligen utvecklats av uppdragsgivaren. Blanketten har senare expertgranskats vid SCB och reviderats. Den har därefter genomgått ett kognitivt test, vilket ledde till att drygt en tredjedel av frågorna samt introduktionen skrevs om. Förbättringarna ansågs innebära så uppenbara förbättringar att inget nytt test gjordes. Eftersom standardiserade rutiner för urval utsändning och påminnelser användes, ansågs det inte nödvändigt att göra något produktionstest. Testet visade dock att frågorna om föregående års inköp av konsumentkapitalvaror är svåra att besvara med säkerhet hur de än formuleras.*

Experiment som gjorts för att välja mellan metoder motiverar valet och ger relevant kunskap om den valda metodens tillförlitlighet.

- *Två versioner för bokföring av utgifter jämfördes i ett experiment med ett slumpmässigt urval omfattande 300 objekt i vardera gruppen. Utfallet blev att kassabok med förtryckta spalter för olika utgiftslag gav i genomsnitt 15 % fler noteringar än kassaböcker med löpande bokföring. Versionen med förtryckta spalter valdes. Det ansågs att de förtryckta spalterna fungerade som påminnelse och att denna minskade glömskeffekterna.*

Processkontroll för att uppmärksamma och åtgärda störningar

- *Undersökningen låter göra 40 återintervjuer varje månad för att identifiera och korrigera eventuella felaktigheter vid intervjun. Återintervjuerna är oberoende och reconciliering görs.*
- *Tolv procent av alla svar rättades efter granskningen. I vart femte fall togs en återkontakt med företaget. Återkontakterna fungerar som en utbildning av uppgiftslämnaren och tillförlitligheten är i genomsnitt bättre vid nästa undersökningstillfälle.*
- *Beteendekodning under undersökningens inledningskede visade att frågorna 12, 23 och 37 ofta krävde förklaringar. Information och frågeformulering sågs över. Eftersom det var en undersökning med CATI, kunde blanketten snabbt korrigeras. 58 intervjuer hade genomförts före korrigeringen. 42 av dessa måste kodas om som partiellt bortfall.*

Mätfel eller annat fel?

Klassificeringen av fel är inte alltid entydig. T.ex. kan man under fas 1 ha upptäckt att man inte kan mäta just den variabel man är intresserad av. I stället väljer man att fråga om ”näraliggande” variabel som man i fas 2 och 3 hittat ett sätt att mäta. I det senare fallet har man en relevansbrist. Hade man däremot ställt den omöjliga frågan hade felet setts som ett mätfel.

När ett företag rapporterar med brutet räkenskapsår kan man beskriva detta som ett Innehållsfel (*referenstid*). Alternativt kan man beskriva det som mätfel orsakat av tidsförskjutning (*teleskopering*), eftersom det ändå är en och samma variabel fast med olika referenstid.

När uppgiftslämnaren glömmer bort att rapportera t.ex. en kostnad som ska ingå i en summa, kan man beskriva detta som partiellt bortfall i variabeln men som mätfel i summan.

Man kan ibland ha olika uppfattningar om hur det är korrekt att klassificera ett fel. Det viktiga är att alla tas med och redovisas tydligt i kvalitetsdeklarationen.

Referenser:

Kvalitetsbegrepp och riktlinjer för kvalitetsdeklaration av officiell statistik. MIS 1994:3

Levnadsförhållanden appendix 15. Teknisk rapport avseende 1990-91 års och 1992-93 års undersökningar av levnadsförhållanden. SCB 1995 (Sid. 23)

Bilaga

Mätfelsmodell för effektivare resursallokering

Håkan L Lindström

Storleken och betydelsen av de mätfel, som följer av att använda dåliga blanketter underskattas av många användare och köpare av statistiska undersökningar. De inser inte nyttan av systematisk blankettutveckling utan tror att undersökningen blir billigare att genomföra om de låter bli att testa sina blanketter. Med ett modellresonemang kan man enkelt visa att det med små omfördelningar av givna resurser går att genomföra ett blanketttest. Undersökningens tillförlitlighet kan avsevärt förbättras under mycket allmänna förutsättningar. Modellresonemanget är med nödvändighet tekniskt. Det är däremot inte nödvändigt att läsa det för att förstå och tillämpa principerna för blankettutveckling. Sådana resonemang är viktiga för att visa hur skattningarnas tillförlitlighet i högsta grad påverkas av att man satsar på att reducera systematiska och icke systematiska mätfel.

Med en statistisk modell som innefattar effekterna av systematiska och slumpmässiga mätfel kan man visa hur mycket sämre tillförlitligheten blir av att använda bristfälliga blanketter. Inom ramen för givna undersökningsresurser kan en satsning på blankettutveckling förbättra undersökningsresultaten avsevärt. I denna modell är urval och mätning de enda felkällor som tillåts påverka skattningens tillförlitlighet. Med rimliga antaganden om kostnader och hur mätfelen reduceras kan man också visa hur effektivt det är att prioritera fram resurser till blankettutveckling.

Den statistiska modellen innebär att man undersöker en ändlig population, som består av N objekt. För varje objekt finns ett sant värde. Objekt nr i ($i = 1, 2, \dots, N$) har värdet x_i . Om detta mäts utan fel i undersökningen, har skattningar baserade på x_i -värden en samplingvariation men ingen mätfelsbias eller -variation. När blanketten och/eller datainsamlingsprocessen leder till missförstånd och felaktiga svar, blir inte svaret x_i utan ett annat värde y_i .

I modellen består mätfelet av en systematisk komponent b_i och en slumpmässig komponent ε_i . ε_i är 0 i förväntan och har variansen $\sigma_{\varepsilon_i}^2$. Därmed blir mätvärdet för objekt nr i : $Y_i = x_i + b_i + \varepsilon_i$; (1)

Förväntan, bias och varians hos en medelvärdeskattning \bar{y} härleds i två steg - över sannolikheterna i urvalet (p) och sannolikhetsfördelningen i mätfelsmodellen (m). Biasen b_i hos varje objekt är konstant över mätfelsmodellen varför biasen vid skattningen av medelvärdet i den ändliga populationen är:

$$E_{pm}(\bar{y}) - \bar{x} = \bar{x} + \bar{b} - \bar{x} = \bar{b} \text{ där } \bar{b} = \sum_U b_i / N. \quad (2)$$

Vid ett obundet slumpmässigt urval (o.s.u.) och med oberoende mellan mätningarna inbördes och mellan objektens sanna värde och deras mätfel är medelkvadratfelet för skattningen \bar{y} baserad på n observationer med mätfel från bristfällig blankett

$$MSE_{pm}(\bar{y}) = \left(\frac{S_{x+b,U}^2}{n} + \frac{\sigma_\varepsilon^2}{n} \right) \left(1 - \frac{n}{N} \right) + \left(\frac{\sigma_\varepsilon^2}{N} \right) + (\bar{b})^2 ; \quad \text{där} \quad (3)$$

$S_{x+b,U}^2$ populationsvariansen för sant värde + systematiskt mätfel.

$\sigma_\varepsilon^2 = \frac{1}{N} \sum_U \sigma_{\varepsilon_i}^2$ den genomsnittliga mätfelsvariansen, när mätningarna är

- b okorrelerade, som de kan antas vara i en postenkät.
- b en bias, som beror på att man svarar systematiskt fel t.ex. genom att missledas av formuleringar och anvisningar i en ej testad och reviderad blankett.
- N populationsstorlek
- n urvalsstorlek

För att minimera \bar{b} och σ_ε^2 testar man och reviderar den felbemängda blanketten. Om kostnadsramen är given, kan man finansiera test och revidering av blanketten t.ex. genom att minska antalet undersökningsobjekt med n_t så att stickprovsstorleken vid den reviderade blanketten blir $n - n_t$.

Om test och revision av blankett skulle lyckas att helt eliminera både slumpmässiga och systematiska mätfel, blir variansen för medelvärdesestimatoren \bar{x} (baserad på den testade och reviderade blanketten) väntevärdesriktig. Medelkvadratfelet är då:

$$MSE_{pm}(\bar{x}) = Var_p(\bar{x}) = \frac{S_{x,U}^2}{n - n_t} \left(1 - \frac{n - n_t}{N} \right) \quad (4)$$

När medelkvadratfelet (3) för \bar{y} jämförs med (4) för \bar{x} , ser man att så snart stickprovsstorleken inte är försumbar är det åtgärderna för att reducera biasen, som snabbast ökar tillförlitligheten. Men även en reduktion av slumpmässiga mätfel kan förbättra en skattnings precision avsevärt.

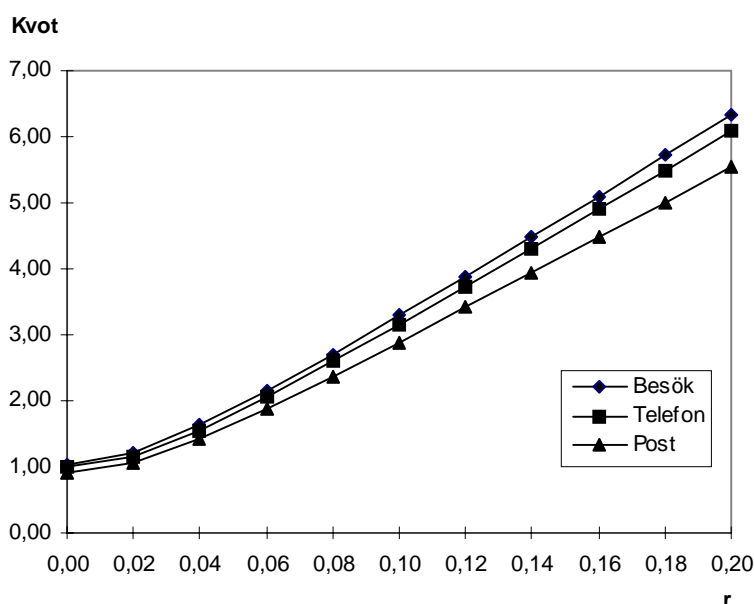
Genom att sätta in empiriska data eller rimlighetsantaganden i formlerna ser man hur lönsam en satsning på systematisk blankettutveckling blir. Vi redovisar här ett räkneexempel med förutsättningarna:

- Ett enkelt standardiserat blanketttest kostar lika mycket som 25 besöksintervjuer, 100 telefonintervjuer eller 250 postenkäter (dvs. $n_t = 25, 100$ respektive 250). Sambandet är grundat på erfarenheter vid SCB.
- Stickprovsstorleken n sätts till 1000, en vanlig nivå i många engångsundersökningar.
- Komponenterna i medelkvadratfelet har sambanden $S_{x+b,U}^2 + \bar{\sigma}_\varepsilon^2 = kS_{x,U}^2$ och $\bar{b} = r * S_{x,U}$.
- koefficienten k sätts = 1.1, dvs. en variansökning med 10 % till följd av att slumpmässiga mätfel förekommer.
- koefficienten r får gå från 0.00 till 0.20. Högsta värdet $r = 0.20$ svarar mot att 10 av 100 svarar rätt i stället för fel i en fråga som gäller en dikotom variabel där 50% har den sökta egenskapen.

[(För att diskutera om r-värdena är realistiska kan man göra enkla hypotetiska kalkyler för en dikotom variabel. Antag t.ex. att man mäter en företeelse som inträffar i hälften av alla fall. Då är $p = 0,5$ och populationsvariansen har den välkända formen $p(1-p)$ dvs. är här $(0,5)^2$. Då svarar värdet $r = 0,02$ mot att 1 av 100 ($\bar{b} = 0,01 = r \cdot 0,50$) fått svara rätt i stället för fel och $r = 0,20$ att 10 av 100 fått svara rätt. (Inga slumpmässiga mätfel antas ha förelegat utan enbart systematiska).]

Diagrammet nedan visar kvadratroten ur kvoten (3)/(4) dvs. medelkvadratfelet vid otestad respektive reviderad blankett, för de tre vanliga datainsamlingsmetoderna besöksintervju, telefonintervju och postenkät.

Kvadratroten ur kvoten mellan MSE vid otestad (3) och testad och reviderad (4) blankett, då $k= 1,10$ och $n=1000$.



Vi ser från ovanstående bild att så fort en bias kan avlägsnas, blir vinsten i tillförlitlighet stor. Det är enbart vid postenkäter och då ingen bias föreligger som man kan vinna något på att låta bli att testa. Effekter i form av högre svarsfrekvens, mindre behov av granskning och återkontakter, snabbare svar som kan komma till på den positiva sidan är inte medräknade i denna modell. I praktiken lönar sig de standardiserade testen alltid - om man inte har små urval, högst ett par hundra undersökningsobjekt och en osannolikt rutinerad blankettkonstruktör.

Det kan vid första anblicken tyckas som om det är viktigast att minska de systematiska mätfelen. Men när man gör sambandsstudier och använder variabler för klassindelning eller selektioner, har också slumpmässiga mätfel en starkt störande effekt. I praktiskt arbete prioriterar man inte utan inriktar sig samtidigt på att reducera båda typerna av mätfel.

I en konkret situation kan det finnas andra faktorer som kan påverka jämförelsen i båda riktningarna. Man kan t.ex. ha en bortfallsbias eller inte lyckas avlägsna hela mätfelsbiasen. Ändå finns så stora marginaler till fördel för den reviderade

blanketten, att det krävs mycket ogynnsamma förutsättningar för att den inte ska vara det bättre alternativet, så länge inte urvalen är mycket små.

Modellresonemangets främsta uppgift är att ge en generell bild av betydelsen att använda en utprövad blankett och inte att ge underlag för en effektivitets- och kostnadskalkyl. Vågar man ta risken att inte testa? Följden kan bli att svaren på ett antal variabler blir felaktiga. Alla missförstånd uppmärksammas inte ens vid granskning, vilket leder till att man publicerar felaktiga och missledande resultat.

Referenser

- Särndal, C-E.; Swensson, B.; Wretman, J. (1991) *Model Assisted Survey Sampling*. New York osv. :Springer-Verlag
- Akkerboom Hans and Håkan L Lindström. (1997)*Terminology and Abbreviations for the Workshop - Minimum Standards in Questionnaire Testing*
- SCPR *Proposal for improving question testing at SCPR*. By the PILOTING WORKING GROUP (forthcoming).
- Nastasha Rees, Forms Consultancy Group, Australian Bureau of Statistics (1997) *Questionnaire Testing Concepts and their Definitions*. Prepared for Workshop on Minimum Standards in Questionnaire Testing,

Schema över mättekniskt arbete

Undersökningsfas	Exempel på mättekniska metoder/verktyg, som kan användas i respektive fas
1 Definiera undersökningen	Fokusgrupper Djupintervjuer Frågebanker, administrativa register
2 Blankettkonstruktion	Tidigare undersökningar Facklitteratur och erfarenhet Standarder för klassificering Expertgranskning Expertpanel Variabellista och tabellplan
3 Kognitiva test	Fält- och "laboratorie"test med <ul style="list-style-type: none"> • fördjupningsfrågor • tänka högt metoder • observationer • debriefing
4 Experiment	Inbyggda experiment Fristående experiment
5 Anpassning till produktionen	Teknikanpassning av blanketten Provintervjuer Produktionstest Beteendekodning Spontana och planerade observationer
6 Utvärdering	Evalvering genom experiment eller jämförelse med annan statistikprodukt Återintervjustudier Återintervjustudier med reconciliering Debriefing Beteendekodning Indikatorer från produktionsprocessen t.ex.: <ul style="list-style-type: none"> • behovet av editering • nivån på partiellt bortfall • egenheter i fördelningen
7 Kvalitetsredovisning	Systematisk sammanställning av kunskaper från främst faserna 3 och 6 enligt gällande anvisningar för SOS