SCB STATISTISKA CENTRALBYRÅN

1985-05-30

PROMEMORIOR FRÅN P/STM

NR 15

GRANSKNING OCH EVALVERING AV SURVEYMODELLER, TIDEN FÖRE 1960.

AV GÖSTA FORSMAN

INLEDNING

TILL

Promemorior från P/STM / Statistiska centralbyrån. – Stockholm : Statistiska centralbyrån, 1978-1986. – Nr 1-24.

Efterföljare:

Promemorior från U/STM / Statistiska centralbyrån. – Stockholm : Statistiska centralbyrån, 1986. – Nr 25-28.

R & D report : research, methods, development, U/STM / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1987. – Nr 29-41.

R & D report : research, methods, development / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1988-2004. – Nr. 1988:1-2004:2.

Research and development : methodology reports from Statistics Sweden. – Stockholm : Statistiska centralbyrån. – 2006-. – Nr 2006:1-.

Promemorior från P/STM 1985:15. Granskning och evalvering av surveymodeller, tiden före 1960 / Gösta Forsman.
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

urn:nbn:se:scb-PM-PSTM-1985-15

STATISTISKA CENTRALBYRÅN 1985-05-30

PROMEMORIOR FRÅN P/STM

NR 15

GRANSKNING OCH EVALVERING AV SURVEYMODELLER, TIDEN FÖRE 1960.

AV GÖSTA FORSMAN

Förord

Föreliggande rapport, som utarbetats inom projektet KVIST (Kvalitet i statistiken) vid P/STM, utgör den första delen i en planerad sammanställning som syftar till att historiskt belysa utvecklingen av felbegrepp, mätmetoder och felmodeller inom surveyområdet. Denna del omfattar tidsperioden fram till 1960 och bygger på ett 50-tal arbeten som behandlar dels surveymodeller, dels specifika felkällor som svarsfel, intervjuarfel, bortfallsfel, ramfel etc. Den samtida intensiva utvecklingen inom stickprovsteorin ligger däremot utanför ramen för framställningen. Ett speciellt syfte med projektet är att studera surveymodeller ur följande perspektiv:
- vilket syfte författarna haft med modellbyggnaderna
- vilken nytta modellerna haft i praktisk statistikproduktion.

Den fortsatta rapporteringen planeras att ske för ytterligare två epoker separat, dels för tiden 1960-1970, dels för tiden efter 1970.

För det mycket omfattande arbetet med utskrift av olika versioner av denna rapport vill jag tacka Elseliv Lindfors, P/STM.

Gösta Forsman

INNEHÅLLSFÖRTECKNING

GRANSKNING OCH EVALVERING AV SURVEYMODELLER, TIDEN FÖRE 1960

1 INLEDNING

1.1 Rapportens syfte och uppläggning

Föreliggande rapport utgör den första delen av en planerad sammanställning av litteratur om surveymodeller. På grund av ämnets omfattning kommer rapporteringen att ske för tre olika epoker separat

(1) Tiden före 1960, som redovisas här;

(2) Tiden 1960-1970;

(3) Tiden efter 1970.

Sammanställningen syftar till att historiskt belysa utvecklingen av felbegrepp, mätmetoder och felmodellbyggnad inom surveyområdet. Till grund för denna första del ligger en mängd vetenskapliga artiklar från tidsperioden fram till 1960, som behandlar dels surveymodeller dels viktigare arbeten om specifika felkällor som svarsfel, intervjuarfel, bortfallsfel, ramfel etc. Den samtidiga intensiva utvecklingen inom stickprovsteorin ligger däremot utanför ramen för denna framställning. Stickprovsfelet behandlas inte separat utan endast i samband med surveymodellerna, där det ingår som en av flera felkomponenter.

Ett speciellt syfte med projektet är att studera surveymodellerna ur följande perspektiv:

(1) Vilket syfte författarna haft med modellbyggnaderna

(2) Vilken nytta modellerna haft i praktisk statistikproduktion.

Som bakgrund till den fortsatta framställningen diskuteras kortfattat begreppen "survey" och "surveyfel" i avsnitt 1.2 och 1.3. Avsnitt 2 behandlar översiktligt den metodologiska utvecklingen inom området

"non-sampling errors" (non-samplingfel) fram till 1960. I avsnitt 3 redovisas felbegreppets utveckling under tidsperioden. Avsnitt 4 innehåller en översikt av mätmetoder för olika felkomponenter samt exempel på praktiska tillämpningar av dessa metoder. De surveymodeller som presenterades under tidsperioden beskrivs i avsnitt 5 och i avsnitt 6 sammanfattas var forskningen kring surveymodeller stod vid 1950-talets slut. Även om framställningen är omfattande gör den naturligtvis inga anspråk på att vara heltäckande.

1.2 Begreppet survey

Begreppet survey har beskrivits på olika sätt i den statistiska litteraturen. Vi skall ansluta oss till framställningen i Dalenius (1969a) där följande fem karakteristiska egenskaper hos en survey anges:

(1) En survey är en undersökning av en mängd av "enheter" av något slag. Dessa enheter kan vara fysiska objekt (i vid mening; "individ", "hushåll", "företag", är tre exempel) eller händelser ("trafikolycka", "födelse").

(2) Dessa enheter observeras med avseende på en mätbar egenskap. Detta innebär, att någon väl definierad mätmetod tillämpas; mätmetoden kan beskrivas i termer av mätinstrument, en eller flera mätoperationer, ordningen mellan dessa, och de villkor, under vilka de utförs.

(3) Det finns en enumerationsregel, med vars hjälp det kan entydigt fastslås, om en viss enhet tillhör eller icke tillhör den mängd av enheter, som skall studeras vid undersökningen. Man brukar ibland säga, att denna enumerationsregel specificerar målpopulationen ("the target population").

(4) För att kunna tillämpa mätmetoden, dvs genomföra datainsamling, måste observationskontakt skapas med populationen. Denna kontakt skapas med hjälp av en ram ("frame"). Denna specificerar den observerade populationen.

(5) Med utnyttjande av de sålunda insamlade data beräknas ett mått på den egenskap, som studeras. I det enklaste fallet är detta mått ett tal, t ex ett procenttal. I andra fall är det en vektor eller en matris av tal, t ex en statistisk tabell.

Denna beskrivning täcker ett brett spektrum av undersökningar, t ex rymmer den både stickprovsundersökningar och totalundersökningar. Gemensamt för surveyer är att insamlade data genererats "av naturen". Således räknas inte t ex experimentella undersökningar till surveyer, eftersom statistikern i dessa på ett avgörande sätt kontrollerar datagenereringsprocessen.

1.3 Felbegreppet, olika fel och felkomponenter

Den tidiga utvecklingen av surveyteorin var inriktad på mätning och kontroll av enskilda felkällor. Ett viktigt exempel var studiet av stickprovsfelet, dvs den del av det totala felet i urvalsundersökningar som härrör från det faktum att man endast studerar ett stickprov av populationen. Denna forskning har varit mycket framgångsrik och idag finns ett stort antal tekniker tillgängliga för att kontrollera samplingfelet.

Sedan 1930-talet har man i surveyer blivit allt mer medveten om problemen med andra typer av fel än stickprovsfelet, s k non-samplingfel (i brist på vedertagen svensk benämning kommer vi att använda denna engelska term). Non-samplingfel kan förekomma i samband med såväl datainsamlingen som databearbetningen och bero på t ex defekter i ramen, frågeblankettens utformning, avsiktliga eller oavsiktliga felsvar från uppgiftslämnarna, bortfall eller fel i kodnings- och stansningsprocedurerna. Även fel som uppstår pga misstag i urvalsförfarandet räknar vi till non-samplingfelen.

Begreppet mätfel (measurement error) används ofta när non-samplingfel behandlas. Hit hör fel som uppstår vid själva mätförfarandet, alltså inte feltyper som ram- och stickprovsfel. Bortfallsfel och bearbetningsfel räknas ibland till mätfelen, ibland inte.

Som beskrivits i Dalenius (1962) har studiet av non-samplingfelet följt två linjer:

1) Utvecklandet av teori och metoder för behandling av specifika typer av non-samplingfel.

2) Utvecklandet av en omfattande teori för en integrerad behandling av non-samplingfel (hit hör också utvecklingen av s k "blandade felmodeller", i vilka också urvalsfelet ingår).

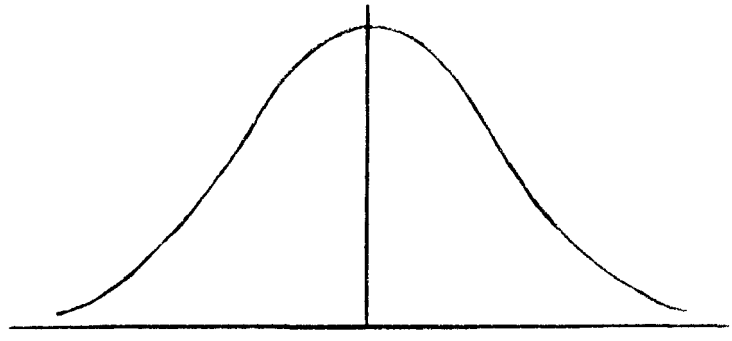
Under de senaste decennierna har stora ansträngningar gjorts för att konstruera vad man kallat "Total survey models". Dessa surveymodeller karakteriseras bl a av att de postulerar närvaron av non-samplingfel samt att de anger metoder att under den angivna modellen skatta det totala felet i ett surveyresultat.

2 MÄTNING AV NON-SAMPLINGFEL. METODOLOGISK UTVECKLING

2.1 Förhistorik

Mätfel, eller observerationsfel, studerades tidigt inom de experimentella naturvetenskaperna och då speciellt inom astronomin, eftersom sjömännen behövde korrekta mätmetoder för sin navigering. Att mätfelen kunde vara av olika slag var tidigt bekant. Redan Tycho Brahe (1546-1601), den experimentella astronomins grundläggare, delade in observationsfelen i två typer, de systematiska och de tillfälliga. Vid särskilt viktiga observationer försökte han eliminera de förra genom att variera försöksvillkoren. För naturvetenskapliga tillämpningar utvecklade Gauss omkring 1810 en statistisk teori för slumpmässiga mätfel enligt vilken upprepade observationer fördelade sig kring det sanna värdet som en s k klockkurva, senare kallad täthetsfunktion för en normalfördelad stokastisk variabel.

Figur 2.1: Gauss klockkurva.



Gauss utförde bl a också beräkningar av de sannolika felgränserna för en viss serie av observationer av samma storhet. Det viktigaste resultatet är här att medelfelet för det aritmetiska medelvärdet är omvänt proportionellt mot kvadratroten ur antalet observationer. Gauss kunde därmed konstatera att den sannolika precisionen för medelvärdet ökar med fler observationer men att den ökar långsammare ju större antal observationer man gör.

Karl Pearson, som också arbetade med en naturvetenskaplig referensram, fann vid en serie experiment (se Pearson, 1902) att korrelerade mätfel förekommer mycket mer än man tidigare trott.

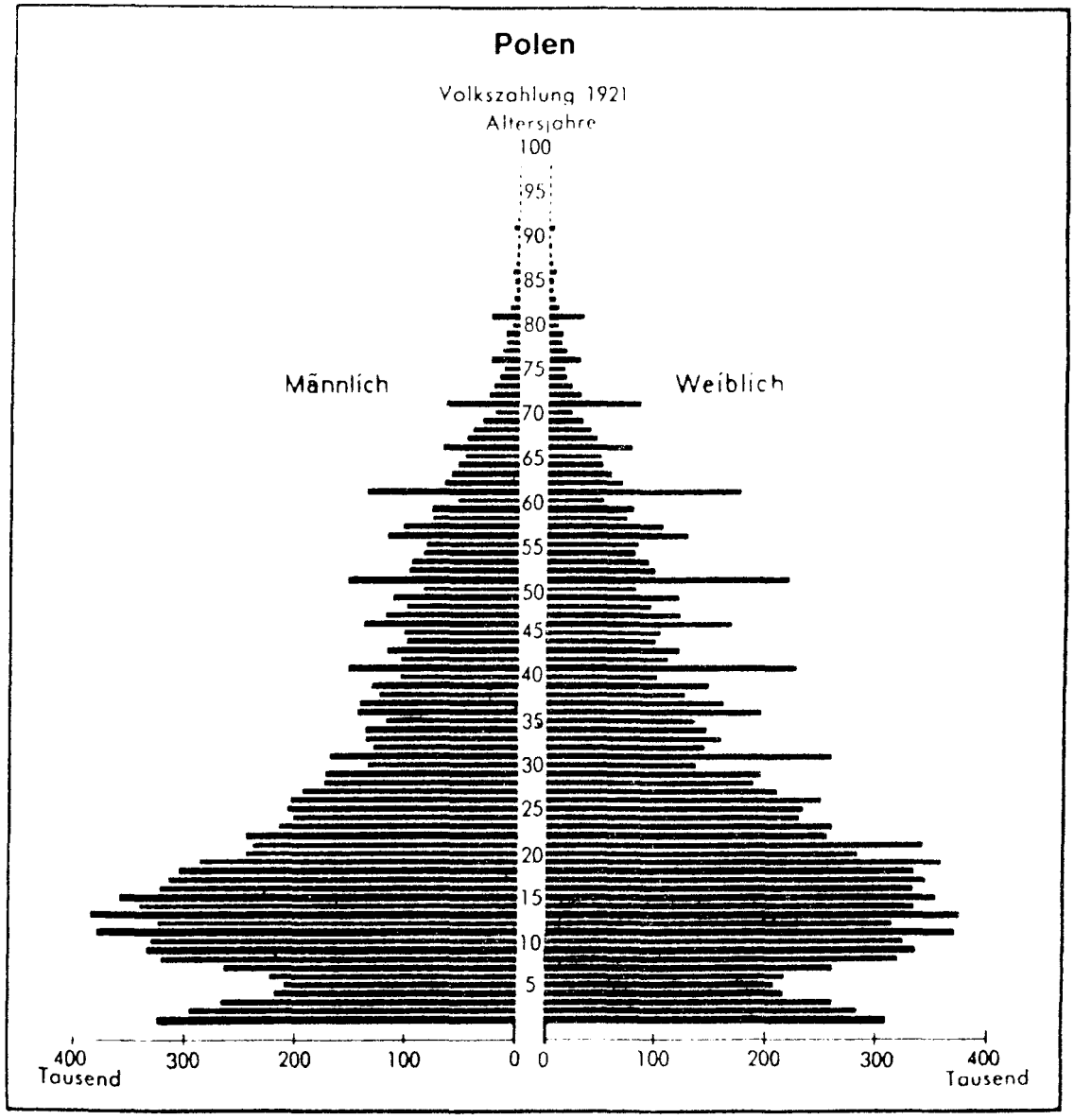
Ett av Pearsons experiment kan tjäna som exempel på hur han arbetade. Tre bedömare fick bedöma halva längden av 500 linjer. Varje bedömning skulle göras raskt och enbart med hjälp av ögonmått. Linjerna varierade i längd och i sin relativa placering på papperet. De sanna värdena var kända i samtliga fall och Pearson och hans kollegor beräknade de tre observatörernas "absoluta" och "relativa personliga ekvationer", dvs medelavvikelsen från det sanna värdet för varje bedömare resp medeltalet av differensen mellan två bedömares observationer. Till sin förvåning fann Pearson att det var en positiv korrelation mellan de tre observatörernas bedömningsfel. Han ansåg sig inte kunna förklara denna, eftersom observatörerna arbetat oberoende av varandra, och talade om "någon speciell form av mental eller fysisk likhet" som orsak till denna korrelation mellan bedömare.

Från tidiga undersökningar av den typ vi numera kallar surveyer hämtar vi följande två exempel på förekomst av mätfel (ur Strecker och Wiegert, 1981):

Exempel 2.1: Ålderspyramiden i Polen enligt folkräkningen 1921.
Figur 2.2 visar en typisk befolkningspyramid från länder med dåligt utvecklad kyrkobokföring: Åldersuppgifterna är i stor utsträckning avrundade till "jubileumstal", särskilt i högre åldersgrupper. I detta exempel gäller detta i högre grad för kvinnor än för män. Denna typ av mätfel beror på att folk inte känner till sin rätta ålder och kan i dag uppträda i befolkningsstatistik från u-länder. Redan Wargentin uppmärksammade f ö fenomenet i sin uppsats från 1766 om mortaliteten i Sverige.

Figur 2.2:

Ålderspyramiden i Polen enligt folkräkningen 1921



Ur Strecker och Wiegert (1981).

Slut på exemplet.

Exempel 2.2: Statistik över hundraåringar i Bulgarien.
Resultaten av de första folkräkningarna i Bulgarien 1887 och 1892 väckte uppseende i omvärlden pga det stora antalet personer i mycket hög ålder. 1887 var antalet 100-åringar 123 per 100 000 invånare och 1892 var det 102. Vid senare räkningar avtog dessa tal men förblev trots det osannolikt höga, som framgår av tabell 2.1.

Tabell 2.1:

Folkräkning	Personer i åldersklassen "100 år och äldre"	
	Totalt antal	Antal per 100 000 invånare
1887	3883	123
1892	3372	102
1900	2719	73
1905	2407	60
1910	2067	48
1920	2161	44
1926	1756	33
(Kontrollunder-		
sökning 1926	158	2,9)
1934	1280	21

Som jämförelse nämner Strecker och Wiegert att vid folkräkningen 1890 i Tyskland befanns endast 0.16 personer av 100 000 vara över 100 år (sammanlagt 78 personer av 49 428 470). Vid folkräkningen 1925 i Tyskland var motsvarande siffra 0.12 på 100 000 invånare. Dessa skillnader mellan länder föranledde Internationella Statistiska Institutet att på 1920-talet ta initiativ till en kontrollundersökning av alla påstådda 100-åringar i Bulgarien, vilken omfattade dels läkarundersökning, dels utfrågning om historiska händelser som krig, hungersnöd, bränder etc, som åldringarna borde ha upplevt.

Denna konroll gav till resultat att högst 158 av de 1756 100-åringarna från folkräkningen 1926 ansågs kunna vara rätt klassificerade. Förutom ofullständig kyrkobokföring under 1800-talet ansågs orsakerna till överskattningarna vara en låg utbildningsnivå (särskilt hos den äldre befolkningen) samt ett visst "ålderskoketteri".

Slut på exemplet.

Ovanstående tidiga exempel visar hur man kunde upptäcka förekomsten av non-samplingfel i undersökningsresultat. Med få undantag (som i exempel 2.2) negligerades emellertid länge problemen med mätfel i surveyer. Den allt starkare medvetenheten om felens förekomst ledde dock så småningom fram till att en metodologisk utveckling startade, som syftade till kontroll av i första hand specifika felkällor.

Det bör här också nämnas att en viktig grund för utvecklingen av teorin för mätfel i surveyer (och för surveyteorin i allmänhet) var den forskning om experimentplanering och -analys som inleddes under 1920-talet. Den mest berömda institutionen i detta sammanhang är Rothamstead Experimental Station i England, där arbetet först leddes av Fisher och från början av 1930-talet av Yates. Det finns många paralleller mellan teorierna för surveyer och experiment och det kan noteras att många framstående statistiker, som Cochran, Yates, Finney, Hartley och Mahalanobis, arbetade inom båda områdena. De tre grundläggande begreppen för experimentella undersökningar enligt Fishers teori, randomisering, lokal kontroll och replikering, har sålunda direkta motsvarigheter inom teorin för mätfel, nämligen sannolikhetsurval, stratifiering respektive upprepad mätning (t ex återintervju). Hur dessa begrepp används framgår av fortsättningen av denna rapport.

2.2 Studium av specifika felkällor 1930-1960

Utveckling av metoder för kontroll och mätning av non-samplingfel inleddes under 1930-talet och var särskilt stark i Indien och USA. I Indien skedde den i samband med undersökningar av skördarnas storlek och kvalitet. Fram till 1930-talet hade dessa mätts med en subjektiv metod, med s k eye-estimation. Sedan man visat att denna metod ledde till skattningar med allvarlig bias utvecklades en objektiv mätmetod som innebar att man för ett stickprov av provytor skördade och mätte avkastningen. Senare visade det sig att även denna metod var utsatt för mätfel - provytornas storlek visade sig ha ett stort inflytande på resultaten. Detta nya mätfel blev sedan föremål för en intensiv forskning rörande aspekter som felets mekanism, mätning och kontroll. I Mahalanobis (1946) och Sukhatme (1947) beskrivs detta arbete.

Under senare delen av 1930-talet inleddes också i Indien studium av det fel som berodde på den enskilde intervjuarens (undersökarens, observatörens) beteende i mätsituationen. Den teknik som utvecklades bygger på s k interpenetrerande delurval och tillgår i sin enklaste form på följande sätt: Urvalet till undersökningen dras i form av ett antal oberoende delurval, som alla är dragna enligt samma kriterier, och från vilka det är möjligt att skatta populationsparametrar. Genom att tilldela varje intervjuare ett sådant delurval och sedan jämföra de skattningar av samma parameter som baseras på olika intervjuares resultat, kan man mäta variationen mellan dessa. Man kan också beräkna mätresultatens variation "inom" delurvalen och med variansanalysmetoder testa om intervjuarvariationen är signifikant eller estimera varianskomponenterna.

Interpenetrerande delurval användes även för att jämföra olika grupper av intervjuare med varandra. Det var vanligt i Indien att två olika intervjuarorganisationer samlade in data från vilka oberoende skattningar beräknades, detta för att man bättre skulle kunna bedöma kvaliteten hos den hopvägda skattningen. Om de två skattningarna skilde sig mycket från varandra, samlade man i allmänhet in mer data innan resultatet publicerades. (Se t ex Sukhatme och Seth, 1952).

Interpenetrerande delurval (som också kallats "replicated samples" eller "interpenetrating checks") kan dras på flera olika sätt, beroende på vad de skall användas till, t ex

(1) Oberoende och disjunkta delurval som beskrivits ovan.
(2) Helt eller delvis överlappande delurval.
(3) S k "linked samples" där urvalsproceduren är sådan att delurvalen inte är oberoende men dragna så att man från varje delurval kan beräkna en väntevärdesriktig skattning av populationsparametern.

Förutom för kontroll av datainsamlingen användes interpenetrerande delurval också i samband med kvalitetsbedömning av databearbetning och analys.

Tekniken med interpenetrerande delurval utvecklades vid Indian Statistical Institute (I.S.I.) under 1930- och 1940-talen under ledning av P.C. Mahalanobis. En utförlig beskrivning av detta arbete finns i Mahalanobis (1946). Vid Indian Council of Agricultural Research (I.C.A.R.) användes tekniken under 1940-talet i flera undersökningar. Detta har beskrivits i Sukhatme och Seth (1952), som också innehåller en utförlig diskussion om användbarheten av interpenetrerande delurval.

I USA skedde forskningen kring non-samplingfel i första hand vid Bureau of the Census, dels i samband med de stora totalräkningarna t ex 1940 och 1950 års "Censuses of Population and Housing", 1946 års "Census of Manufacturing" och 1948 års "Census of Business", dels i anslutning till större stickprovsundersökningar som "Current Population Survey". Dessutom bedrevs arbete inom området även vid andra byråer, t ex Bureau of Labor Statistics samt vid universitet och fristående forskningsinstitut. Man studerade olika typer av felkällor såsom frågeblankettens utformning, svarsvariabilitet, bortfall, intervjuareffekter och fel som uppstår vid databearbetningen. Tidiga referenser inom dessa områden är Blankenship (1940), Palmer (1943), Hansen och Hurwitz (1946), Rice (1929) respektive Deming och Geoffery (1941). Mätfelen behandlades ofta i termer av "coverage errors" och "content errors", varmed avsågs fel i antalsräkningar (0-1-variabler) respektive fel i andra typer av variabler. Den metod som kom till användning kallas "quality check" och tillgår på följande sätt:

(1) Ett stickprov av den undersökta populationen återintervjuas.

(2) Återintervjun har två inslag:
- en "coverage check", som är utformad för att mäta graden av "under-counting" bland de enheter som skall räknas, t ex individer, lägenheter och företag.
- en "content check", som avser att mäta felet för studerade variabler som t ex ålder, antal rum i en lägenhet och omsättning.

(3) Ett antal åtgärder vidtages för att göra återintervjun så effektiv som möjligt, bl a:

- De bästa intervjuarna i organisationen väljs ut för att genomföra återintervjun.
- Speciella frågeblanketter utformas. Man sätter en övre gräns för antalet frågor i originalintervjun som skall kontrolleras eftersom man för varje sådan fråga använder en hel uppsättning av kontrollfrågor.
- Speciellt utbildningsmaterial används. Utbildningstiden för intervjuarna är också längre vid återintervjun än vid originalintervjun.
- Vid återintervjun används en längre fältarbetsperiod och en grundligare övervakning och kontroll av intervjuarbetet än vid originalintervjun. Man väljer också uppgiftslämnare noggrannare såtillvida att man om möjligt intervjuar den mest välinformerade i en familj eller i ett företag.
- Ibland görs återintervjuerna så att intervjuarna känner till resultaten av originalintervjun, vilket ger dem möjlighet att direkt utreda orsaker till eventuella skillnader.

En sammanfattning av olika "quality checks" som genomfördes vid Bureau of the Census under senare delen av 1940-talet finns i Eckler och Pritzker (1951). I denna referens beskrivs också hur man mätt noggrannheten i en totalskattning med olika evalveringsmetoder, t ex jämförelse med annan statistik och - vid folkräkningar - med förväntade åldersfördelningar som beräknats med utgångspunkt från tidigare folkräkningar.

Användandet av återintervjuer fortsatte och utvecklades vid Bureau of the Census under 1950-talet. I Eckler och Hurwitz (1957) presenteras en teori för användandet av återintervjuer när man har klassindelade variabler. De använder mått som "gross difference" och "net difference" (brutto- och nettofel) för att skatta svarsvariansen vid skattningar av totaler, dvs den slumpmässiga variation i en skattning som beror på att varje individ kan avge olika svar i en viss, bestämd intervjusituation. I artikeln visas också i några exempel vilken betydelse studium av svarsfel vid totalräkningar haft vid planerandet av nya undersökningar.

Det finns stora likheter mellan arbetet vid U.S. Bureau of the Census och det som bedrevs i Indien (I.S.I. och I.C.A.R.) under 1940-talet. Ett citat från Hansen och Madow (1974) kan belysa detta:

> "On one of his early visits to the Census Bureau, Mahalanobis surprised and impressed us with the validity of his statement that the Census Bureau and the Indian Statistical Institute were, perhaps, the two organizations in the world that were most similar in the way they dealt with the total systems-design aspects and operational approaches to sample surveys. The similarity became more apparent as Mahalanobis described the nature of the studies done to improve the efficiency of design through empirical studies, to resolve theoretical problems as needed, but especially to achieve effective operational control in the actual execution of surveys through design, control, and feedback procedures in data collection and data processing."

Trots dessa grundläggande likheter kan man notera vissa skillnader i inriktningen av det praktiska arbetet med kontroll och mätning av non-samplingfel mellan de båda länderna. Medan man i Indien genomgående använde tekniken med interpenetrerande delurval var arbetet vid Bureau of the Census i lika hög grad inriktat på "quality checks". Behovet av en teori för mätning, tolkning och kontroll av non-samplingfel ledde dock till liknande åtgärder i båda länderna, nämligen till utvecklandet av modeller för det totala felet, se Hansen et al (1951) och Sukhatme och Seth (1952). Dessa modeller, ehuru presenterade på mycket olikartade sätt, uppvisar stora likheter med varandra, vilket närmare diskuteras i kapitel 5 i denna rapport.

Speciellt artikeln av Hansen et al (1951), som återkom i en omarbetad version i Hansen et al (1953), vol 2, kapitel 12, skulle visa sig få stor betydelse för kommande arbeten om non-samplingfel. Där definieras och diskuteras ett flertal begrepp som "the essential conditions of a survey", sanna värden, individuellt svarsfel, individuell svarsbias mm. Det synsätt på non-samplingfel som kommer till uttryck i artikeln hade utvecklats vid Bureau of the Census under 1940-talet och fortsatte att användas där under 1950-talet, (se Eckler och Hurwitz, 1957 och Hanson

och Marks, 1958). Även Kish och Lansing (1954), som arbetade vid Survey Research Center, Ann Arbor, Michigan, använde terminologin från Hansen et al (1951). Senare arbeten, skrivna efter 1960 och därmed utanför ramen för framställningen i denna rapport, har också gjort det, t ex Kish (1965), Des Raj (1968) och O'Muircheartaigh (1977).

Under 1950-talet presenterades i USA ett antal studier av intervjuarvariation som byggde på principerna för interpenetrerande delurval och som belyste olika faktorers inverkan på intervjuarvariation. Den vanligaste faktorn som studerades var "typ av fråga". Tre exempel på sådana rapporter är Stock och Hochstim (1951) från Bureau of Labor Statistics, Feldman et al (1951) från National Opinion Research Center vid University of Chicago och den tidigare nämnda Hanson och Marks (1958).

Utanför USA och Indien förekom studium av specifika felkällor även i Europa och Kanada. I samband med den franska företagsräkningen 1946 genomfördes en "quality check" enligt samma metoder som i USA. Därvid uppdagades så stora fel att man tog det ovanliga beslutet att inte publicera vissa resultat av företagsräkningen (se Chevry, 1949).

Vid London School of Economics studerade man variationen mellan såväl enskilda intervjuare som grupper av intervjuare under det att olika förutsättningar som frågeblankettens omfattning, intervjuarnas erfarenhet och utbildning, uppgiftslämnarnas kön och ålder m m varierades (se Durbin och Stuart, 1951 och Gales och Kendall, 1957).

Arbetet med non-samplingfel i Kanada under 1950-talet sammanfattas i Goldberg (1957). Där redovisas studium av svarsvariation med återintervjuer och beräkningar av brutto- och nettodifferenser inom i första hand arbetskraftsundersökningarna. Metoderna påminner starkt om dem som samtidigt tillämpades i USA (Eckler och Hurwitz, 1957).

2.3 Blandade felmodeller

Fram till mitten av 1940-talet studerade man praktiskt taget alltid samplingfel och non-samplingfel oberoende av varandra. Oftast studera-

des dessutom endast en typ av non-samplingfel i taget. Under senare delen av 1940-talet började man arbeta efter en ny utvecklingslinje som syftade till att skapa metoder för en integrerad kontroll av flera feltyper, s k blandade felmodeller (mixed error models). Resultat av denna forskning började publiceras i början av 1950-talet. I Stock och Hochstim (1951) presenteras sålunda olika experiment där man beräknar hur stora andelar av den totala variansen som beror på samplingvariation respektive variation mellan intervjuare. I Hansen et al (1951) anges en matematisk modell för hur variansen för en medelvärdesskattning kan delas upp i olika varianskomponenter, som hänför sig dels till samplingvariationen dels till svarsvariationen. Modellen gäller för en specificerad undersökningsplan under vilken formler för skattning av varianskomponenterna också anges. Sukhatme och Seth (1952) presenterar en allmän linjär mätfelsmodell som tillåter multipla observationer på samma objekt och diskuterar utifrån den ett antal undersökningsscheman som möjliggör estimation av komponenter som svarsvarians, intervjuarvarians och samspelseffekter. Kish och Lansing (1954) anger en generell modell för effekten av individuell svarsbias på surveyestimat, vilken använder "bättre" värden snarare än sanna värden för att evalvera felet i en survey. De ger också formler för skattning av felkomponenterna under antagande om obundet slumpmässigt urval.

I några av de läroböcker som gavs ut i början av 1950-talet förekommer också surveymodeller. I Hansen et al (1953) och Sukhatme (1954) redovisas de modeller som samma författare tidigare presenterat (i Sukhatmes bok är den redovisade modellen ett specialfall av den ursprungliga). I Cochran (1953) beskrivs några linjära mätfelsmodeller och hur den totala variansen för en medelvärdesestimator kan delas upp i olika varianskomponenter under dessa modeller. Cochran presenterar emellertid inte skattningsmetoder för alla de ingående komponenterna.

De olika surveymodellerna kommer att presenteras närmare i kapitel 5.

3 FELBEGREPPET I EN SURVEY

3.1 Allmänt

Indelningen av det totala felet i en survey i olika komponenter och diskussionen av dessas karaktär skiljer sig helt naturligt mellan olika framställningar beroende på dessas syften. Deming (1944) och Hansen et al (1951) innehåller sålunda mycket utförliga feldiskussioner medan artiklar som behandlar studiet av en enskild felkomponent ofta endast innehåller ett konstaterande att felen kan indelas i stickprovsfel och non-samplingfel samt - som ett exempel på det senare - en diskussion av karaktären hos den felkomponent som diskuteras i artikeln. I detta avsnitt sammanfattas och diskuteras hur olika författare behandlat felbegreppet. Först redovisas olika uppdelningar av felet. Därefter diskuteras begreppet "sant värde" och slutligen behandlas de viktigaste felkomponenterna var för sig, förutom stickprovsfelet, som ju inte studeras explicit i denna rapport. För att illustrera den något annorlunda terminologi som användes på 1930- och 1940-talen jämfört med i dag och för att illustrera författarnas olika språkbruk, har vi valt att inte översätta felbeskrivningarna till svenska.

3.2 Indelningar av det totala felet i olika komponenter

Vi skall i detta avsnitt behandla olika sätt att dela upp det totala felet i en survey. Beroende på syftet med uppdelningen kan vi urskilja fyra huvudgrupper:

(1) "Ekonomisk" indelning. Med detta avses en mycket detaljerad uppdelning av fel som görs i syfte att underlätta en optimal resursallokering mellan olika produktionsmoment i undersökningen. Denna typ av uppdelning kan närmast betraktas som en checklista över alla förekommande felkällor.

(2) "Mätteknisk" indelning. Denna typ av indelning görs i samband med att metoder för mätning och kontroll av fel presenteras. Den styrs av dessa metoder och är mycket grövre än den ekonomiska indelningen, eftersom mätmetoderna som regel fångar upp effekter från flera felkällor samtidigt.

(3) "Pedagogisk" indelning. Denna indelningstyp kan vara relativt detaljerad men inte hänförbar till (1) eller (2). Den kan förekomma i t ex läroböcker.

(4) Övriga indelningar. Hit hör de mycket enkla indelningar som nämndes i inledningen. Vi skall inte beakta dem vidare i detta avsnitt.

Vi skall nu redovisa hur några olika författare gjort felindelningar enligt metoderna (1) - (3):

Ekonomisk indelning

W. Edwards Deming, som i början av 1940-talet arbetade vid Bureau of the Census, presenterade i en artikel (Deming, 1944) tretton olika faktorer som han ansåg påverkade en surveys användbarhet. Dessa var

"(1) Variability in response;
(2) Differences between different kinds and degrees of canvass:
    a) Mail, telephone, telegraph, direct interview;
    b) Intensive vs. extensive interviews;
    c) Long vs. short schedules;
    d) Check block plan vs. response;
    e) Correspondence panel and key reporters;
(3) Bias and variation arising from the interviewer;
(4) Bias of the auspices;
(5) Imperfections in the design of the questionnaire and tabulation plans;
    a) Lack of clarity in definitions; ambiguity; varying meanings of same word to different groups of people; eliciting an answer liable to misinterpretation;
    b) Omitting questions that would be illuminating to the interpretation of other questions;
    c) Emotionally toned words; leading questions; limiting response to a pattern;
    d) Failing to perceive what tabulations would be most significant;
    e) Encouraging nonresponse through formidable appearance;

(6) Changes that take place in the universe before tabulations are available;
(7) Bias arising from nonresponse (including omissions);
(8) Bias arising from late reports;
(9) Bias arising from an unrepresentative selection of data for the survey, or of the period covered;
(10) Bias arising from an unrepresentative selection of respondents;
(11) Sampling errors and biases;
(12) Processing errors (coding, editing, calculating, tabulating, tallying, posting and consolidating);
(13) Errors in interpretation;
    a) Bias arising from bad curve fitting; wrong weighting; incorrect adjusting;
    b) Misunderstanding the questionnaire; failure to take account of the respondents' difficulties (often through inadequate presentation of data); misunderstanding the method of collection and the nature of the data;
    c) Personal bias in interpretation."

I artikeln diskuteras varje feltyp ingående. En del av denna diskussion återges i avsnitten 3.3 - 3.8. Deming anger som huvudsyfte med artikeln att peka på behovet att ta hänsyn till alla felkällor när man planerar en survey och att fördela resurserna på de olika felkällorna så att man får största möjliga reduktion av det totala felet. Ett annat syfte var att påtala behovet av att utveckla teorier för svarsbias och svarsvariabilitet, vilket skulle underlätta en god resursallokering. Detta utvecklingsarbete inleddes som tidigare nämnts vid Bureau of the Census i mitten av 1940-talet och ledde så småningom fram till utvecklandet av surveymodeller.

Ovanstående fellista finns även i en utförligare version i Deming (1950).

Mätteknisk indelning

Mahalanobis (1946) refererar till Demings artikel men väljer en enklare klassificering som bättre passar hans syften att beskriva, mäta och kontrollera felen vid skördeuppskattningar. Han bortser från feltyper

som svarsvariabilitet och förändringar i populationen, vilka inte utgör något stort problem i skördeuppskattningar, samt från misstag i planering och analys och gör istället indelningen

(1) "Sampling variation";

(2) "Recording mistakes", som är av två typer:
   a) Misstag som beror på observatören
   b) Misstag som begås i databearbetningsprocessen (till största delen manuell i Indien vid denna tid) samt i samband med analysen och presentationen av materialet;

(3) "Physical fluctuations". Med detta avses en fysikalisk osäkerhet som finns hos de objekt som studeras. Som exempel nämner Mahalanobis att vikten på en skörd kan variera 4-5 % beroende på fuktighetsgrad.

Ovanstående tre feltyper kunde mätas och kontrolleras med olika metoder. Förutom samplingvariationen, som matematiskt kunde härledas för en given urvalsplan, kan båda typerna av "recording mistakes" mätas med hjälp av interpenetrerande delurval. Feltypen "physical fluctuations" kunde i viss utsträckning minska i betydelse genom en noggrann specificering av den parameter som skulle skattas, t ex genom att ange vid vilken fuktighetsgrad en skörd skulle vägas. Även efter en sådan specificering ansåg Mahalanobis att det återstod en fluktuation på 1-2% och han påpekade att det därmed var meningslöst att försöka sänka stickprovsfelet under denna nivå.

Ett annat exempel på "mätteknisk" indelning av fel tillämpades vid US Bureau of the Census under utvecklandet av en teori för mätning och kontroll av non-samplingfel. Indelningen, som beskrivs i flera artiklar omkring 1950, t ex Marks och Mauldin (1950) och Hansen et al (1951), består av tre felkomponenter:

(1) "Sampling errors";
(2) "Response errors";
(3) "Processing errors".

Svarsfelet ges här en vid mening. Det innefattar alla typer av non-samplingfel som uppträder i samband med datainsamlingen. Det kan t ex bero på frågeblankettens utformning och intervjuarens och uppgiftslämnarens beteende i intervjusituationen. Speciellt räknas "coverage error" till svarsfelet. Med "coverage error" avses då det fel som uppkommer vid klassindelade data då man felaktigt inkluderar eller exkluderar en enhet i en klass. Enligt Hansen et al (1951) räknas även bortfallsfelet som svarsfel. Vi skall närmare diskutera dessa olika feltyper senare i detta kapitel.

I de läroböcker som publicerades på 1950-talet ingick i regel avsnitt som handlade om mätning av fel. Indelningen av det totala felet i olika feltyper blev i dessa avsnitt "mätteknisk":

Cochran (1953) gör följande indelning (förutom urvalsfelet):

(1) "Failure to measure some of the units in the chosen sample";
(2) "Errors of measurement on a unit";
(3) "Errors introduced in editing and tabulating the results".

Denna indelning är anpassad till den fortsatta framställningen i lärobokens kapitel 13, som ingående behandlar mätning av bortfallsfel och mätfel, vilka tillsammans tycks motsvara det som Bureau of the Census kallar svarsfel.

Sukhatme (1954) anger följande typer av fel i sitt inledningskapitel:

(1) "Sampling errors";
(2) "Lack of precision i reporting observations";
(3) "Incomplete or faulty canvassing of a designated random sample";
(4) "Faulty methods of estimation".

Det förefaller som om feltyperna (2) och (3) här i stort sett motsvarar Cochrans mätfel respektive bortfallsfel. Tillsammans med stickprovsfelet omfattar de i princip alla de feltyper som Deming (1944, 1950) räknar upp. De är här ordnade i tre grupper som var för sig kan mätas och kontrolleras under vissa förutsättningar.

Pedagogisk indelning

Som exempel på en pedagogiskt motiverad indelning av fel nämner vi indelningen i Hansen et al (1953), vol. I, kap. 2. Författarna påpekar att en fullständig förteckning över alla sorters fel ligger utanför ramen för deras framställning, men att några av de vanligaste felen i en survey kan hänföras till följande kategorier (förutom stickprovsfelet):

(1) "Errors due to faulty planning and definitions";
(2) "Response errors";
(3) "Errors in coverage";
(4) "Errors in classification";
(5) "Compiling errors";
(6) "Publication errors".

"Response errors" har här en snävare mening än i t ex Hansen et al (1951). Emellertid påpekar Hansen et al (1953) att både "errors in coverage" och "errors in classification" (dvs kodningsfel) kan räknas som "response errors". Hansen et al räknar även ramfel till "errors in coverage" vilket vi skall diskutera närmare i avsnitt 3.7.

3.3 Begreppet "sant värde"

Med felet i en mätning, x, avser vi differensen $x - \mu$ mellan mätvärdet och det s k sanna värdet, μ, för den storhet vi mäter. Detta väcker omedelbart frågan: Hur är det sanna värdet definierat? Det har visat sig vara svårt att entydigt besvara denna fråga och det har utvecklats olika sätt att se på existensen av sanna värden. Vi skall här sammanfatta denna diskussion och redovisa hur den fördes i den statistiska litteraturen före 1960.

Låt oss betrakta följande variabler:

(1) Avstånd mellan bostad och närmaste livsmedelsbutik

(2) Intelligens

Det kan för den första variabeln vara lätt att tänka sig ett objektivt sant värde. Begreppet "avstånd" verkar - åtminstone i förstone - enkelt att definiera eftersom längdenheter är väldefinierade enligt internationella överenskommelser. Däremot kan det vara svårare att tänka sig ett sant värde för begreppet "intelligens" eftersom vi för denna variabel inte har något objektivt mått att använda för att formulera en definition.

Vi skall nu mot bakgrund av ovanstående exempel diskutera två synsätt som har utvecklats angående existensen av sanna värden.

(1) Objektiva sanna värden existerar oberoende av mätproceduren.

(2) Det existerar endast operationella sanna värden, dvs de sanna värdena måste definieras i termer av de arbetsmetoder (survey procedures) som används i den aktuella undersökningen.

Synsätt (1) innebär alltså att det inte bara för variabeln "avstånd mellan bostad och närmaste livsmedelsaffär" utan också för "intelligens" existerar ett objektivt sant värde. Observera att detta värde inte behöver vara möjligt att mäta med kända metoder.

Enligt synsätt (2), å andra sidan, tänker man sig att det varken för variabeln "intelligens" eller ens för en variabel som "avstånd mellan bostad och närmaste livsmedelsaffär" existerar ett sant värde. Man pekar i det senare fallet på svårigheterna att definiera detta avstånd exakt: Skall man mäta fågelvägen eller närmaste promenad- eller cykelväg? Den som cyklar kanske tar en annan väg än den som går eller åker bil. Vilken sträcka skall man då mäta? Skall man mäta från dörr till dörr eller mellan de "närmaste punkterna" i respektive lokaler? Eller kanske mellan lokalernas "mittpunkter" om sådana över huvud taget går att definiera. Enligt detta synsätt får man i stället för objektiva sanna värden definiera operationella sanna värden som definieras i termer av arbetsmetoderna. Detta innebär i korthet att det sanna värdet är ett "förväntat värde" av den bästa tänkbara (och praktiskt genomförbara) mätprocedur som genomförs upprepade gånger under idealiska förutsättningar. Denna procedur, som ofta kallas "preferred procedure", antages vara utsatt för vissa slumpmässiga störningar.

Vi skall återkomma till detta i litteraturgenomgången nedan. Först kan vi dock konstatera att skillnaden i synsätt inte innebär att undersökningar i praktiken genomförs på olika sätt. Även med synsätt (1) måste man formulera en operationell definition av sant värde. Till skillnad från fallet vid synsätt (2) anser man sig då emellertid endast ha definierat en approximation till det sanna värdet.

Vi skall nu se hur de båda beskrivna synsätten framträtt i några artiklar från tiden före 1960.

Hansen, Hurwitz, Marks och Mauldin (1951) som beskriver det synsätt man hade vid Bureau of the Census, utgår från begreppet "individual true value" och definierar det som "a characteristic of the individual quite independent of the survey conditions which affects the individual response".

Författarna ansluter sig alltså till synsätt (1) ovan. De föreslår tre kriterier för definition av sant värde (de första två väsentliga (essential), det tredje användbart men inte väsentligt):

(1) Det sanna värdet som måste vara entydigt definierat;

(2) Det sanna värdet måste vara definierat på sådant sätt att syftena med surveyen är tillgodosedda;

(3) När det är möjligt att göra i överensstämmelse med de första två kriterierna, bör det sanna värdet definieras i termer av operationer som verkligen kan genomföras (även om det kan vara svårt eller dyrt att genomföra dessa).

Författarna påpekar att det - om man utelämnar det andra kriteriet - är möjligt att definiera sant värde så att en survey inte behäftas med något (eller med ett negligerbart) svarsfel (svarsfel definieras inledningsvis som non-samplingfel som uppkommit i samband med datainsamlingen). T ex kan man definiera en persons födelseplats som det svar man får från denne via en intervjuare som är instruerad att fråga "Var föddes ni?". Den definitionen svarar mot första och tredje kriteriet. I de flesta praktiska situationer skulle emellertid en sådan definition

av sant värde inte accepteras och det andra kriteriet blir därmed inte uppfyllt.

Författarna konstaterar att det för ovanstående exempel kan vara omöjligt att definiera ett sant värde som svarar mot alla tre kriterierna. Man kan dock ofta definiera ett sant värde som stämmer för de två första (väsentliga) kriterierna och sedan definiera en operation vars "förväntade värde" ger en tillfredsställande approximering av det sanna värdet, vilket skulle innebära att det tredje kriteriet uppfylldes approximativt. Som exempel på en sådan operation anger man återintervjuer med bättre utbildade och bättre tränade intervjuare, vars resultat visserligen inte får betraktas som "perfekta" men som kan anses bättre än resultaten i den ordinarie intervjun. Man introducerar också synsättet att man tänker sig att varje individ i populationen intervjuas ett stort antal gånger under exakt samma villkor som återintervjun. Denna process skulle generera en population av svar för varje individ. Man tänker sig vidare att urvals- och mätproceduren tillgår så att man drar ett stickprov av individer ur individpopulationen och för var och en av dem drar ett av deras möjliga svar. Väntevärdet av en skattning från detta stickprov skulle då kunna betraktas som en approximation av det sanna värdet.

Deming (1950) ansluter sig till synsätt (2) ovan och påtalar att varje karakteristika, vare sig det är ålder, arbetskraftsstatus, inkomst, storlek på skörd eller något annat, måste definieras i termer av en operation eller procedur för mätningen av denna karakteristika. För vissa variabler finns det en s k preferred procedure, som leder till resultat som är "nearest to what are needed for a particular end". En "preferred procedure" är dyrare och mer tidskrävande än en "unpreferred procedure" och ibland t o m omöjlig att genomföra. Det mätvärde som är resultatet av en "preferred procedure" kallas ibland sant värde.

Deming anser att en "preferred procedure", om en sådan existerar, kan vara slumpmässig och därmed ha ett matematiskt förväntat värde, men den behöver inte vara slumpmässig. Detsamma kan sägas om en "unpreferred procedure". Han skriver att det vanligen antas att väntevärden existerar för de procedurer som används i surveyer och att han också gör sådana antaganden i sin bok trots att alltför litet är känt om riktig

heten i dessa antaganden och om vad man skall göra när dessa inte är uppfyllda.

Deming definierar bias för en "unpreferred procedure" som skillnaden mellan det resultat den ger och det resultat som man fått om man i stället använt en "preferred procedure". Om båda procedurerna är sådana att de har matematiska väntevärden, är definitionen av bias speciellt enkel: den är lika med skillanden mellan de två väntevärdena.

Uppenbarligen har Hansen et al (synsätt (1)) och Deming (synsätt (2)) i princip samma syn på begreppet "operationellt sant värde". Skillnaden är, som vi nämnt, att Hansen et al betraktar detta som en approximation till det objektiva "sanna värdet" medan Deming ser det som det verkliga sanna värdet.

Eckler och Hurwitz (1957), som redovisar arbeten med non-samplingfel vid Bureau of the Census, tycks uppfatta "sant värde" på samma sätt som Deming:

> "We may sometimes wish to refer to a survey procedure to define a "true value", even though the result would vary over repeated trials. For example, we may wish to use a survey procedure with highly trained and well-paid interviewers as the measuring rod for other survey procedures purporting to measure the same concepts. In this case, we may define the expected value of this survey procedure as the "true value".

Själva uttrycket "true value" uppfattas av många författare som missvisande eller filosofiskt osunt. Därför används ofta andra uttryck som "target value", "intrinsic value" och "value to be estimated" i stället. Cochran, Mosteller och Tukey (1954) använder uttrycken "target quantity" och "quantity of interest":

> "We admit the existence of systematic error of a difference between the quantity measured (the measured quantity) and the quantity of interest (the target quantity). We ask the observations about the measured quantity. We ask our subject matter knowledge, intuition, and general information about the relation between the measured quantity and the target quantity".

Detta synsätt, med subjektiva sanna värden som existerar oberoende av mätprocedurerna, går uppenbarligen inte att hänföra till något av de synsätt som redovisades inledningsvis.

Problemet att definiera ett sant värde för en variabel antyds ibland av författare till artiklar om non-samplingfel men diskuteras ur andra perspektiv. T ex påtalar Mahalanobis (1946), som vi såg i avsnitt 3.2, att vikten på en skörd inte är ett väldefinierat begrepp - den är beroende på fuktighetsgraden - men detta diskuteras endast som ett exempel på en viktig felkälla, physical fluctuations, och leder inte till någon diskussion om vilket värde som är det "sanna". Övriga författare till artiklar om non-samplingfel från den aktuella tidsperioden som studerats i denna översikt har i regel undvikit helt att diskutera begreppet "sant värde"; man förutsätter förmodligen att läsaren har någon form av intuitiv uppfattning om det.

3.4 Intervjuarfel

Det gäller för intervjuarfel, som för praktiskt taget alla felkällor, att de inte går att avgränsa helt från de övriga. En felaktigt ifylld blankett kan, förutom på intervjuarens egna misstag, bero på felaktiga instruktioner, dåliga frågeformuleringar, avsiktliga felsvar från uppgiftslämnaren och andra faktorer som intervjuaren inte kan påverka och som ofta räknas till andra felkällor. När intervjuarfelet skall beskrivas skiljer man i regel på dessa faktorer medan man i samband med mätning av intervjuarfelet tvingas att i princip definiera det som det fel som finns i ett ifyllt frågeformulär, eftersom det annars inte blir mätbart. I detta avsnitt skall vi redovisa hur den förra ansatsen behandlats i litteraturen före 1960. Den omfattande forskningen kring intervjuarfelet, och särskilt då dess slumpmässiga del, intervjuarvariansen, är anledningen till att vi behandlar detta separat och inte som en del av andra felkällor, vilket annars är vanligt. Det vid Bureau of the Census utvecklade synsättet, nämligen att intervjuarvariationen är en slags korrelerad svarsvariation, behandlar vi i avsnitt 3.5.

Rice (1929) diskuterar "bias in the interview" som beroende av både fel som härrör från uppgiftslämnaren och från intervjuaren. Han konstaterar att det alltid finns en risk för att intervjuaren inte bara gör misstag i samband med urvalet och nedteckningen av informationen från uppgiftslämnaren, utan även missförstår innehållet i informationen. Denna risk finns också om villkoren för intervjun är noggrannt kontrollerade. Intervjuarfelet kan dock minskas genom mycket specificerade instruktioner om hur intervjun skall utföras. Rice understryker vikten av att intervjuaren fungerar som ett objektivt medium genom vilket information från de intervjuade kan omvandlas till sådan form att den kan ligga till grund för statistiska bearbetningar.

Deming (1944) konstaterar att olika intervjuare beskriver samma situation på olika sätt och gör olika tolkningar av identiska påståenden från en uppgiftslämnare. Han anger intervjuarens politiska, religiösa och sociala uppfattning samt dennes utbildning och ekonomiska status som exempel på faktorer som påverkar intervjuarfelet. De flesta intervjuare kan kanske heller inte hjälpa att de blir påverkade av sina arbetsgivares intressen.

Som orsaker till intervjuarfel anger Deming också brist på förståelse för undersökningens syfte och för dess genomförande samt att olika intervjuare leder in uppgiftslämnarna i olika sinnesstämningar. Variabiliteten som beror på det senare skälet är svår att skilja från svarsfelet.

Deming påpekar att mindre grupper av intervjuare kan tränas till en hög grad av homogenitet. I mindre undersökningar med relativt få intervjuare kan man på detta sätt minska intervjuarfelet på ett sätt som av kostnadsskäl inte är möjligt i större undersökningar och allra minst i folkräkningar med tusentals intervjuare. Det visade sig emellertid senare, när metoder för mätning av intervjuarvarians utvecklades, att intervjuarnas bidrag till det totala felet minskar när antalet intervjuare ökar på ett liknande sätt som precisionen ökar med större urval. Det är därför inte självklart om man skall sträva mot ett stort eller ett litet antal intervjuare. Hansen, Hurwitz, Marks och Mauldin (1951) visade att man under vissa villkor kan beräkna ett optimalt antal intervjuare vilket vi återkommer till i avsnitt 5.2.

Mahalanobis (1946) behandlar intervjuarfelet framförallt i undersökningar som avser att mäta storleken av skördar i Indien (i dessa fall är naturligtvis beteckningar som "observatörsfel" eller "fältarbetarfel" att föredra). Felet beskrivs som misstag som inträffar endera på grund av observatörernas "bias or personal equation" eller grov nonchalans eller till och med överlagd oärlighet från deras sida.

Problemet att intervjuarna (observatörerna) är oärliga och fyller i blanketterna med påhittade värden diskuteras mer av Mahalanobis än av andra författare. En orsak till detta är de ytterst svåra förhållanden under vilka de indiska skördeundersökningarna genomfördes. Till exempel måste arbetet under sommartid utföras under mycket stark värme, under regnperioder kunde vägarna bli oframkomliga och många intervjuare drabbades under undersökningsperioderna av malaria och andra sjukdomar.

I den klassiska intervjuarvariansstudien som genomfördes vid Bureau of Labor Statistics av Stock och Hochstim (1951) anges ett antal faktorer som författarna anser kan påverka uppgiftslämnaren på ett sätt som återspeglas i intervjusvaren:

(1) Den modulering, hastighet och artikulering med vilken intervjuaren läser upp frågor;
(2) Vad intervjuaren säger till uppgiftslämnaren före och under intervjun vid sidan av instruktionerna;
(3) Intervjuarens personlighet, kläder, uppträdande och utseende.

Författarna diskuterar också de speciella problemen vid kvoturval, där det finns en typ av intervjuarvariabilitet som hänger samman med det icke-slumpmässiga sätt på vilket varje intervjuare väljer sina uppgiftslämnare.

Gales och Kendall (1957) anger i sin studie av intervjuarfelet vid London School of Economics en uppdelning av intervjuarfelet i tre komponenter. Man utgår från att en frågeblankett som ifyllts av en intervjuare, efter en intervju med rätt person, kan vara defekt av flera skäl. Några av dessa defekter, som felaktigt formulerade frågor, inadekvata eller missledande instruktioner, tillfälliga eller avsiktliga felsvar från uppgiftslämnarens sida, kan man naturligtvis inte lasta intervjuaren för. Men om man bortser från sådana faktorer återstår ändå

en variation som beror på den enskilde intervjuaren. Man kan urskilja tre bidragande element:

(1) Noggrannhet. Detta begrepp är kopplat till hur intervjuaren följer sina instruktioner och hur korrekt denne noterar den information som ges av uppgiftslämnaren. Författarna påpekar att det fel som hör till denna situation vanligen kallas intervjuarbias men anser att "interviewer inaccuracy" skulle vara bättre. En intervjuare kan nämligen vara "inaccurate" utan att vara det systematiskt, även om bias är en vanlig form av detta fel (inaccuracy) i praktiken.

(2) Reliabilitet. Detta begrepp kan ses som den grad i vilken en intervjuare producerar lika resultat under lika villkor. En intervjuare kan, utan att brista i noggrannhet (enligt ovanstående definition), variera sin teknik medvetet eller omedvetet, och skulle kunna få något olika svar från samma uppgiftslämnare vid två olika tillfällen, förutsatt att det var möjligt att göra oberoende intervjuer. Det fel som består i brister i reliabilitet ("unreliability") kan betraktas som en variabel komponent av intervjuarfelet.

(3) Variabilitet. Två intervjuare kan få olika mätresultat av samma intervjuuppgift utan att brista i noggrannhet eller reliabilitet. "Framlockandet" av svar i en intervjusituation är inte en rent mekanisk process och det finns helt legitima variationer mellan de svar som olika intervjuare producerar.

3.5 Svarsfel

Det är väl känt att, vid upprepade observationer på ett och samma objekt, det finns en mer eller mindre uttalad tendens att observationerna blir olika vid olika tidpunkter. Genom att göra allt mer noggranna instruktioner för hur observationerna skall genomföras kan man minska dessa avvikelser, men helt kommer man inte ifrån dem. Denna variabilitet kallas vanligen svarsfel i intervju- och enkätsituationer. Ibland kallas den också observationsfel eller mätfel.

Vi berörde begreppet svarsfel redan i avsnitt 3.2 och fann att det i regel omfattar flera enskilda felkomponenter som påverkar resultatet av datainsamlingen, t ex felaktigheter i frågeformuläret, felaktigt beteende hos intervjuaren och de tillfälliga faktorer som stör respondenten i en intervjusituation. Ofta räknar man till svarsfel även komponenter som bortfallsfel och ramfel. Det visar sig lämpligt ur mätsynpunkt att slå samman denna stora grupp av feltyper till en enda eftersom de enskilda felkomponenterna inte var för sig kan mätas och kontrolleras med de metoder man förfogar över, t ex interpenetrerande delurval och "quality checks".

Svarsfelet kan vara både slumpmässigt och systematiskt. Det är i första hand den slumpmässiga delen - svarsvariabiliteten - som är behandlad i litteraturen. Denna del av svarsfelet speglar effekten av de olika tillfälliga faktorer som påverkar en intervju eller en observation: väder, tid på dygnet, intervjuarens och uppgiftslämnarens tillfälliga sinnesstämning etc. Ofta tänker man sig att uppgiftslämnaren väljer sitt svar ur en population av möjliga svar med olika (och okända) sannolikheter. Det systematiska svarsfelet - svarsbiasen - är ett fel som är lika stort och går i samma riktning för alla uppgiftslämnare, vilket t ex kan inträffa om en fråga i blanketten är felformulerad. För att mäta detta fel krävs i regel tillgång till "sanna värden" eller att man har råd och möjlighet att genomföra en ny mätning (t ex återintervju) med förbättrade metoder. Sådana studier är dock sällsynta.

En viss begreppsförvirring kan uppstå på grund av att en del författare kallar den avvikelse $(x-\mu)$ som ett enskilt intervjuarsvar, x, har från det i någon mening "rätta" svaret μ, för en individuell bias. Om man antar att dessa individuella avvikelser - sett över alla individer i populationen - tar ut varandra utgör emellertid den sammanlagda effekten av dem just svarsvariansen.

Utan att använda ordet "svarsfel" diskuterar redan Rice (1929) förekomsten av en individuell bias hos uppgiftslämnarna ("the informants") vid en enskild mätning. Dessa individuella biastermer kan ibland antas ta ut varandra när många uppgiftslämnare är inblandade i undersökningen:

> "when a considerable number of informants are uniformly involved, it is often legitimately assumed that the influences upon the data attributable to bias in these informants will in large measure cancel out."

Som framgått av ovanstående är svarsfel och intervjuarfel mycket nära besläktade med varandra och omöjliga att hålla helt isär, eftersom det sker ett samspel mellan intervjuare och uppgiftslämnare. Då många författare, t ex Mahalanobis, Stock och Hochstim samt Durbin och Stuart, har studerat intervjuarfelet separat utan att blanda in de fel i intervjusvar, som beror på andra faktorer än intervjuaren, behandlade vi sådana intervjuarfel i ett särskilt avsnitt (3.4).

Vi skall nu redovisa det synsätt på begreppet "response error" som utvecklades vid U S Bureau of the Census under 1940- och 1950-talen. Vi utgår i första hand från framställningarna i Palmer (1943), Deming (1944), Hansen, Hurwitz, Marks och Mauldin (1951) och Eckler och Hurwitz (1957). Samtliga dessa författare arbetade vid Bureau of the Census när artiklarna skrevs.

Palmer (1943) konstaterar att man i enumerativa studier måste räkna med en svarsvariabilitet även om intervjuarna är välutbildade och har noggranna instruktioner. Dåligt formulerade frågor och otillräcklig utbildning leder dels till att denna variabilitet ökar, dels till ett systematiskt svarsfel. Palmer anser emellertid att, i avsaknad av en procedur som ger sanna värden och som skulle medge att bias kunde beräknas, det antagligen är lämpligare att tala om variabilitet i svar mellan tänkta upprepningar av en undersökning än att tala om bias i de enskilda realiseringarna av undersökningen.

Deming (1944) diskuterar svarsfelets uppdelning i en slumpmässig och en systematisk komponent. Den slumpmässiga komponenten, som beror på t ex inflytande av vädret eller tid på dagen när intervjun utförs, kan slå åt olika håll vilket kan leda till att de individuella felen kompenserar varandra så att nettofelet blir litet. De slumpmässiga svarsfelen har större chans att kompensera varandra om skattningar görs för stora redovisningsgrupper.

Däremot är det inte säkert att individuella systematiska svarsfel kompenserar varandra. Detta innebär att det totala svarsfelet också kan vara biased och Deming varnar för misstaget att tro att fel som beror på dåligt utformade frågeblanketter och felaktigt sätt att intervjua, tar ut varandra om man gör många intervjuer.

Ovanstående diskussioner om begreppet "response error" visar att man tidigt betraktade intervjusvar som utsatta för både systematiska och slumpmässiga fel och att resultatet från en survey kunde ses som ett utfall från en stokastisk variabel. För utvecklandet av en teori för mätning och kontroll av svarsfelet krävdes emellertid exakta definitioner av begrepp som individuell svarsvarians och svarsbias och individuellt sant värde. Denna teori började utvecklas vid Bureau of the Census under senare delen av 1940-talet. Den nödvändiga begreppsapparaten formulerades i Hansen et al (1951).

I anslutning till denna artikels diskussion om svarsfel definieras flera begrepp som

- "Error in a survey result":
Skillnaden mellan ett surveyestimat och det värde som skall estimeras.

- "Non-sampling errors":
Fel som skulle finnas även i en totalräkning.

- "Response errors":
Non-samplingfel som kan hänföras till datainsamlingen.

Hansen et al diskuterar också om bortfall skall ses som en speciell typ av svarsfel och drar slutsatsen att det bör göra det, eftersom varje skattningsprocedur innebär att man tillordnar även bortfallsobjekten mätvärden, antingen explicit eller implicit.

Författarna definierar "individuellt svarsfel" som skillnaden mellan en observation på en individ och det sanna värdet för denne.

Författarna påpekar vidare att variabiliteten i individuella svar vanligen har behandlats i termer av slumpmässig variation. De anser att

denna ansats har vissa nackdelar (de nämner dock inte vilka dessa är), men använder den ändå för den aktuella framställningens syften. Som en konsekvens av slumpmässigheten kommer svarsfelet att ha ett förväntat värde. Detta förväntade värde utgör den individuella svarsbiasen. På motsvarande sätt kommer aggregatet eller medelvärdet av en mängd svar från olika individer att ha en svarsbias och en svarsvarians som bestäms av individernas i populationen svarsbias och svarsvarians.

Det är dock inte tillräckligt att säga att ett individuellt svar är en observation på en stokastisk variabel, utan det behövs en specificering av den population av svar som är tillordnad varje individ. Författarna avgränsar denna population som alla svar som kan fås under vissa "essential conditions" vilka bestäms av undersökningsplanen. Som ett minimum måste denna plan specificera ämnet för undersökningen, datainsamlingsmetod och sättet att registrera informationen. Dessa specifikationer kan vara generella eller speciella. Enskilda undersökningar kan innehålla ytterligare specifikationer, t ex att en survey skall genomföras under en viss period. Det finns också "essential conditions" för en survey som uppkommer implicit som nödvändiga konsekvenser av de explicit specificerade villkoren.

Å andra sidan finns det faktorer som påverkar ett svar och som varken är specifierade surveyvillkor eller omedelbara konsekvenser av sådana: T ex kan undersökningsplanen innehålla instruktioner till intervjuaren att ställa vissa frågor, men det är inte säkert att frågorna alltid ställs på precis likadant sätt. Planen kan också specificera ett visst sätt att ta kontakt med uppgiftslämnarna, men den kan inte fastställa hur intervjuaren blir mottagen av en uppgiftslämnare som t ex blir avbruten under matlagning.

Författarna diskuterar också hur surveyvillkoren påverkar svarsvariationen. Allmänt sett begränsar surveyvillkoren storleken på svarsvariationen, men denna elimineras ingalunda fullständigt. Under vissa villkor blir storleken på svarsvariationen stor, under andra villkor blir den liten. Också beroende på villkoren, kan svarsfelet bli kompenserande till sin karaktär eller mer eller mindre systematiskt. Det förväntade värdet av svarsfelen och den slumpmässiga variationen kring detta förväntade värde kan betraktas som bestämt av surveyvillkoren.

Man diskuterar också den positiva korrelation som finns mellan svar från uppgiftslämnare som haft samma intervjuare. Man konstaterade att, i t ex en arbetskraftsundersökning, en intervjuare som med sitt beteende visar att han inte väntar sig att finna kvinnor sysselsatta i förvärvsarbete tenderar att registrera färre sysselsatta kvinnor än en intervjuare som verkar ha uppfattningen att alla vuxna skall vara fullt sysselsatta. Denna effekt, som tidigare diskuterats som ett rent intervjuarfel, presenteras nu för första gången som en korrelerad svarsvarians. Analogin med klusterurval blir här uppenbar. I det enklaste fallet kan man tänka sig att de svar en viss intervjuare skulle få från alla individer i populationen utgör ett kluster. Inom varje sådant kluster dras ett obundet slumpmässigt urval av individer som utgör denna intervjuares tilldelning. Korrelationen mellan svaren från urvalet i detta kluster blir då en skattning av intraklusterkorrelationskoefficienten.

Analogin med klusterurval kommer vi att behandla närmare i kapitlen 4 och 5. Vi skall här konstatera att detta synsätt har haft stort inflytande på senare tiders forskning kring mätning av svarsfel. Följande citat av Madow ur Krewski et al (1981) säger något om idéns ursprung:

> "In 1945, Fred Stephan and I tried to begin work in this area [non-sampling errors] and we felt we were failing because we didn't feel that we had arrived at an appropriate model for non-sampling error. Some time later, Billy Hurwitz, and I am sure it was with Morris Hansen, came up with the very bright idea that an interviewer created a cluster of up to a thousand, as in the census, so that if the interviewer was creating a non-sampling error it could have an enormous effect on the variance. This was one of the precursors of the census getting up the nerve to go on a mail basis in the 1950 census."

Eckler och Hurwitz (1957) beskriver arbetet med svarsfel vid U S Bureau of the Census under 1950-talet. Det framgår att man där vid den tiden använde det synsätt som Hansen et al beskrev 1951. Man utgår från en folkräkningssituation där man kan genomföra återintervjuer med en förbättrad metod vars resultat kan antas ligga närmare de sanna värdena än originalintervjuns. Den variabel som skall skattas är antalet personer i en redovisningsgrupp.

Författarna diskuterar förutsättningarna för definition av begreppen "response variability" och "response bias". De påpekar att urvalssituationen när en uppgiftslämnare väljer ett svar slumpmässigt är annorlunda än den urvalssituation man har när man drar ett slumpmässigt urval av individer för en undersökning. I det senare fallet har undersökaren kontroll över urvalssituationen och urvalet blir verkligen slumpmässigt om en procedur för sannolikhetsurval använts. Man kan emellertid inte kontrollera hur en individ väljer sitt svar bland de möjliga svar han eller hon skulle ha kunnat avge under samma "essential conditions". Därmed kan vi inte vara säkra på att detta urval faktiskt är slumpmässigt.

Eckler och Hurwitz definierar svarsvariation och svarsbias för antalet personer i en klass, givet en viss folkräkningsprocedur. Svarsvariationen är den variation i antalet personer i klassen man skulle få om man upprepade undersökningen många gånger med samma procedur. Svarsbiasen är skillnaden mellan medelvärdet av skattningarna av antalet personer i klassen över alla upprepningar och det "sanna" värdet på antalet personer i klassen.

Författarna utvidgar så begreppsapparaten för svarsvariabilitet och svarsbias genom att definiera dessa begrepp för differenser mellan surveyprocedurer: De betraktar alla möjliga par av undersökningar, där den ena genomförs med den traditionella folkräkningsproceduren och den andra med en förbättrad procedur (som används vid återintervjuer). Om man beräknar skillnaden mellan undersökningsresultaten (skattningar av antalet personer i en klass) i varje par av undersökningar, får man en serie nettodifferenser. Variabiliteten mellan dessa differenser över succesiva par av undersökningar definieras som svarsvariabiliteten för differenserna. Medelvärdet av differenserna definieras som svarsbiasen för dessa.

Eckler och Hurwitz angav mått på svarsvariansen i termer av "gross differences" och "net differences", vilka vi kommer att studera i avsnitt 4. Vi skall där också se exempel på hur mycket svarsvariansen kan öka när man tar hänsyn till den tidigare nämnda klustereffekten.

3.6 Bortfallsfel

Bortfall uppkommer när man misslyckas att mäta en del av enheterna i ett stickprov eller i en population. Det kan i individundersökningar uppkomma av följande orsaker som anges i t ex Deming (1950):

i) Uppgiftslämnaren anträffas inte, trots upprepade försök;

ii) Uppgiftslämnaren vägrar att medverka i undersökningen;

iii) Uppgiftslämnaren glömmer eller gör sig inte besvär att återsända en enkät eller att komma till ett avtalat möte med en intervjuare;

iv) Ren oförmåga hos uppgiftslämnaren att avge informationen t ex sjukdom.

Deming nämner också fem orsaker som starkt påverkar bortfallsstorleken i en survey där uppgiftslämnarnas medverkan är frivillig.

i) Frågeblanketten;
ii) Intervjuaren;
iii) Träning och handledning av intervjuaren;
iv) Introduktionsbrev;
v) Publicitet i massmedia.

Bortfallet leder till ett systematisk fel, vars storlek beror på bortfallets storlek och den genomsnittliga skillnaden i mätvärden mellan de som svarat och bortfallsenheterna. Denna skillnad kan i lyckliga fall vara liten men kan också i många undersökningar vara så stor att bortfallsfelet allvarligt snedvrider undersökningsresultaten. Bortfallsfelet intar en särställning bland felkällorna genom att man endast undantagsvis är kapabel, ens med en massiv resursinsats, att införskaffa mätvärden för det slutliga bortfallet. Regelrätta evalveringsundersökningar kan därför genomföras endast i speciella fall där man har tillgång till annan information om bortfallsobjektens mätvärden.

Bortfallet leder, förutom till systematiska fel även till sämre precision i undersökningsresultaten, eftersom antalet observationer blir mindre än planerat.

En närliggande åtgärd man vidtar när bortfall uppkommer är att återkontakta uppgiftslämnaren. I detta läge står man inför ett antal valmöjligheter. Hur stor andel skall återkontaktas, hur många kontaktförsök skall göras, skall man använda effektivare (och dyrare) metoder etc? Till grund för ett optimalt val av metod enligt effektivitetskriteriet "högsta precision till given kostnad" ligger alltid en modell för bortfall. Sådana modeller bygger på antagandet att varje individ i populationen svarar med en viss sannolikhet q_k för individ k. Vi skall här redovisa tre modeller som alla behandlades i bortfallslitteraturen på 1940- och 1950-talet:

(1) Bortfallsstratumansatsen. I denna tänker man sig populationen indelad i två strata. Det ena, svarsstratum, består av alla objekt för vilka mätvärden skulle erhållas om de blev utvalda. Det andra, bortfallsstratum, består av de objekt för vilka vi inte skulle erhålla något mätvärde. En enhets stratumtillhörighet är här uppenbarligen beroende av de använda fältarbetsmetoderna. En undersökningsplan som innehåller ett omfattande program för bortfallsuppföljning ger ett mycket mindre bortfallsstratum än en som inte innehåller någon bortfallsuppföljning alls. Denna modell, som alltid varit dominerande inom bortfallshanteringen kan också karakteriseras av att svarssannolikheterna, q_k, är 1 i svarsstratum och 0 i bortfallsstratum. Modellen ligger till grund för den kanske viktigaste teknik som tagits fram för att eliminera bortfallsbias, nämligen subsampling i bortfallsstratum, vilken persenteras av Hansen och Hurwitz (1946). Se avsnitt 4.4.

(2) Svarssannolikheterna, q_k, kan variera, $0 \leq q_k \leq 1$. Denna ansats, som alltså är generellare än bortfallsstratumsatsen, anses t ex av Cochran (1953) vara en bättre beskrivning av verkligheten eftersom slumpen spelar en roll när man försöker hitta och intervjua en person vid ett givet antal försök.

Modeller med varierande svarssannolikheter, q_k, leder också de i praktiken till en uppdelning av populationen i olika grupper eller strata, där man inom strata har samma svarssannolikhet. Uppdelningen leder dock till fler än de två grupper som finns i modell (1). Ansatsen användes först av Politz och Simmons (1949, 1950) och därefter av Deming (1953). Även till dessa metoder återkommer vi i avsnitt 4.4.

(3) Hansen, Hurwitz, Marks och Mauldin (1951) inkorporerar som vi tidigare nämnt bortfallet i sin modell för svarsfel. Enligt denna väljer varje respondent i en intervjusituation under vissa "essential conditions" sitt svar ur en population av möjliga svar. Varje möjligt svar har en viss, okänd, sannolikhet att bli utvalt. Genom att till denna population av svar lägga alternativet "nonresponse", som väljs med sannolikheten $1-q_k$, inlemmas bortfallet i svarsmodellen. Det mätvärde som utgör svar om bortfallsalternativet väljs, bestäms av undersökningsplanen, eftersom varje estimationsprocedur medför att man tillordnar mätvärden till bortfallet implicit eller explicit. Mätmetoder för svarsfel enligt denna modell redovisas i avsnitt 5.2.

3.7 Ramfel

Ramfel berördes i samband med begreppet "coverage error" i avsnitt 3.2. Vi skall i detta avsnitt diskutera detta begrepp något samt gå igenom de olika typer av ramfel som behandlades i litteraturen fram till 1960.

Kring begreppet "coverage" råder en viss oklarhet. Det används både vid i) beskrivning av det fel som uppstår när man räknat för många eller för få personer i en klass i en totalräkning och vid ii) beskrivning av det fel som beror på att den population som bildar ramen i en undersökning ej överensstämmer med den population som man avser att undersöka. Detta kan förklaras med att ramar ofta konstrueras med "census"-liknande metoder. Man upprättar t ex en lista över alla hushåll eller alla människor i ett visst område. Det täckningsfel man får när man upprättar en sådan lista blir ett ramfel när man skall undersöka denna population av hushåll eller människor.

I Hansen et al (1953) låter man båda feltyperna enligt i) och ii) ingå i begreppet "errors in coverage":

> "Errors in coverage. One can conceive of a true number of residents in an area, according to some definition, at a point in time. When we take a census of population we will miss some persons and duplicate or enumerate others in error... Similarly, in an agriculture or business census, there is a question of under- or overcoverage ... Samples drawn from such populations may be subject to coverage limitations, also, if the sample is so designed that certain classes of the population have no probability of being included or inappropriate probabilities of being included."

Däremot skiljer Zarkowich (1955) på felen enligt i) och ii) genom att med "net underenumeration" beteckna antalet personer som missats i en totalräkning och med "net undercoverage" antalet enheter som saknas i en ram.

Tre olika typer av ramfel har behandlats i de studerade referenserna från 1940- och 1950-talen:

(1) Enheter som tillhör undersökningspopulationen saknas i ramen och kan därmed inte deltaga i undersökningen. Här förekommer som nämnts ovan begreppen "net undercoverage" och "undercoverage".

(2) Ramen innehåller enheter som inte tillhör undersökningspopulationen. Hansen et al (1953) kallar sådana enheter som "out of scope-units" och feltypen för "overcoverage".

(3) Samma enhet förekommer på flera ställen i ramen. Hansen et al (1953) och Marks et al (1953) använder för detta fel termen "duplicate listings".

Metoder för mätning och reducering av dessa fel presenteras i avsnitt 4.5.

3.8 Bearbetningsfel

Bearbetningsfel kan vara av många olika slag. Deming (1944) räknar här in fel som uppkommer vid "coding, editing, calculation, tabulating,

tallying, posting and consolidating". Varje sådan operation är en potentiell felkälla som kan leda till såväl systematiska som slumpmässiga fel. T ex stansar även den skickligaste stansare ibland hål i fel position eller kolumn på ett kort, hoppar över en kolumn eller stansar två hål i en kolumn där det bara skall stansas ett. En kodare kan också göra tillfälliga fel i enkla situationer och kan dessutom ställas inför oklara kodningsuppgifter där man kan tänka sig olika tolkningar som ger olika koder, etc. Vi kommer emellertid att ägna bearbetningsfelen ett begränsat utrymme i denna rapport eftersom man, som Hansen et al (1951) påpekar, kan kontrollera dem på ett liknande sätt som man kontrollerar svarsfel. Detta innebär att det synsätt på svarsfel (avsnitt 3.5) som Hansen et al introducerar kan överföras till t ex kodnings- eller stansningsfel. En kodare kan sålunda, inför en viss kodningsuppgift, antas välja sin kod ur en population av tänkbara koder som var och en kan väljas med en viss, okänd, sannolikhet. Man kan med detta synsätt definiera "kodarvarians" och "kodarbias" på motsvarande sätt som svarsvarians och svarsbias. En motsvarighet till återintervju blir då att kodaren genomför samma kodningsuppgift vid två olika tillfällen. Möjligheten att beräkna kodarvarians får anses vara bättre än att beräkna svarsvarians, eftersom det rimligen är lättare att undvika problem som t ex minneseffekter vid upprepning av en kodningssituation än vid upprepning av en intervju.

Bearbetningsfelen kan alltså behandlas analogt med svarsfelen. Man har emellertid också för kontroll av dessa använt metoder som utvecklats för industriell kvalitetskontroll eftersom, som Deming och Geoffrey (1941) påpekar, en statistikbyrå kan betraktas som en statistisk fabrik, där huvudprodukten är statistiska tabeller. De ifyllda blanketter som intervjuarna skickar in måste bearbetas innan produkten är färdig. Denna bearbetning påminner mycket om ett löpande band-system i en fabrik, där olika operationer utföres och varje operation avslutas innan produkten överflyttas till nästa.

Vi återkommer till kontrollmetoder för bearbetningsfel i avsnitt 4.6.

4 MÄTNING AV FEL

4.1 Allmänt

I detta avsnitt skall vi för varje feltyp enligt indelningen i kapitel 3 gå igenom metoder för mätning, kontroll och reducering av felen. Huvudsyftet med framställningen är att presentera de mätmetoder (t ex interpenetrerande delurval, återintervjuer) som ligger till grund för surveymodellerna i kapitel 5. Därvid beskrivs också hur dessa metoder kan användas för kontroll av olika feltyper. För främst bortfallsfelet gäller att detta i praktiken är mycket svårt att mäta och kontrollera och avsnittet om bortfallsfelet handlar därför mycket om metoder att reducera det. Vad beträffar reducering av fel finns det också en rad metoder som i princip går ut på att man "anstränger sig mer" i olika produktionsmoment, t ex intensivare bortfallsuppföljning, noggrannare metoder att testa blanketten och bättre utbildning av intervjuare. Sådana åtgärder kommer dock inte att behandlas här.

4.2 Intervjuarfel

Variationen mellan intervjuare är ett kvalitetsproblem som uppmärksammades tidigt. Redan i en studie från 1914 upptäckte Rice (1929) att intervjuarna påverkade uppgiftslämnarna mycket starkt i deras svar. Den teknik för studium av intervjuarfelet som - uppenbarligen av en tillfällighet - kunde användas av Rice var interpenetrerande delurval, som har blivit den dominerande tekniken för kvantifiering av intervjuarfelet. I detta avsnitt skall vi presentera denna teknik med ett antal exempel. Vi kommer också att behandla två andra metoder som har betydelse framförallt för att kontrollera intervjuarfelet, nämligen återintervju och "supervision".

4.2.1 Rice's tidiga studie

Rice (1929) rapporterar om en studie som genomfördes 1914 under ledning av the Commission of Public Charities vid New York Municipal Lodging House avseende fysiska, mentala och sociala karakteristika hos omkring 2000 hemlösa personer. Dessa intervjuades av en grupp på 12 utbildade intervjuare, som tilldelades var sin grupp uppgiftslämnare "wholly

without selection". Intervjun var 20-30 minuter lång och genomfördes enskilt med var och en av uppgiftslämnarna. En blankett om fyra sidor fylldes i för varje intervju.

Vid en senare genomgång av blanketterna slogs Rice av hur svaren från olika intervjuares uppgiftslämnare skilde sig från varandra. Han sammanställde därför för vissa variabler de svar som varje intervjuare fått och redovisade i sin artikel de mest uppseendeväckande skillnaderna. I tabell 4.1 redovisas svaren på frågan om vad intervjuaren ansåg vara orsaken till uppgiftslämnarens situation, för de två intervjuare, A och B, som skilde sig mest från varandra. Svarsalternativet "liquor" avser alkoholproblem och alternativet "industrial" innefattar orsaker som stod utanför uppgiftslämnarens kontroll, som avskedad från arbetet, fabriken nedlagd, säsongarbetare etc. Intervjuarna skulle i denna klassificering skilja på huvudorsaker och bidragande orsaker.

Tabell 4.1: Orsaker till uppgiftslämnarnas belägenhet enligt intervjuarnas bedömning. Jämförelse mellan intervjuare A:s och intervjuare B:s tilldelningar.

Intervjuare	Huvudorsak Procentandel		Bidragande orsak Procentandel		Procentandel uppgiftslämnare för vilka huvudorsak eller bidragande orsak var	
	Liquor	Industrial	Liquor	Industrial	Liquor	Industrial
A	62	7	16	22	78	29
B	22	39	15	34	37	73

(Ur Rice, 1929)

Medan intervjuare A såg alkohol som huvudorsak eller bidragande orsak till hemlösheten hos 78 % av uppgiftslämnarna, såg intervjuare B detta endast hos 37 %. Medan B angav orsaken "industrial" i 73 % av fallen gjorde A det endast i 29 %. Sedan Rice upptäckt dessa skillnader gjorde

han efterforskningar om de berörda intervjuarna och fann att A var en nitisk anhängare till "prohibiton", spritförbud i USA, medan B av sin omgivning ansågs vara socialist.

Mot bakgrund av intervjuarnas åsikter ansåg Rice att resultaten i tabell 4.1 inte var överraskande. Han fann emellertid resultaten i tabell 4.2, som visar uppgiftslämnarnas egna svar på samma fråga om orsakerna till deras belägenhet, mera anmärkningsvärda:

Tabell 4.2: Orsaker till uppgiftslämnarnas belägenhet enligt deras egen bedömning. Jämförelse mellan intervjuare A:s och intervjuare B:s tilldelningar.

Intervjuare	Procentandel uppgiftslämnare som tillskriver sina problem	
	Liquor	Industrial
A	34	42.5
B	11	60

(Ur Rice, 1929)

Här framkommer samma skillnad som i tabell 4.1 och det är enligt Rice uppenbart att felet hos intervjuarna hade överförts till uppgiftslämnarna och dykt upp i deras egna svar.

Rice kände ej till hur många intervjuer som låg till grund för procenttalen, men ansåg ändå att dessa var tillräckligt många för att jämförelserna skulle vara säkra:

> "The data that I have found give percentages only, without the number upon which they were based. It is obvious, however, that each investigator interviewed a sufficiently large number of applicants to provide a fair sample of the latter. A period of several hour each evening for several weeks was devoted to the work."

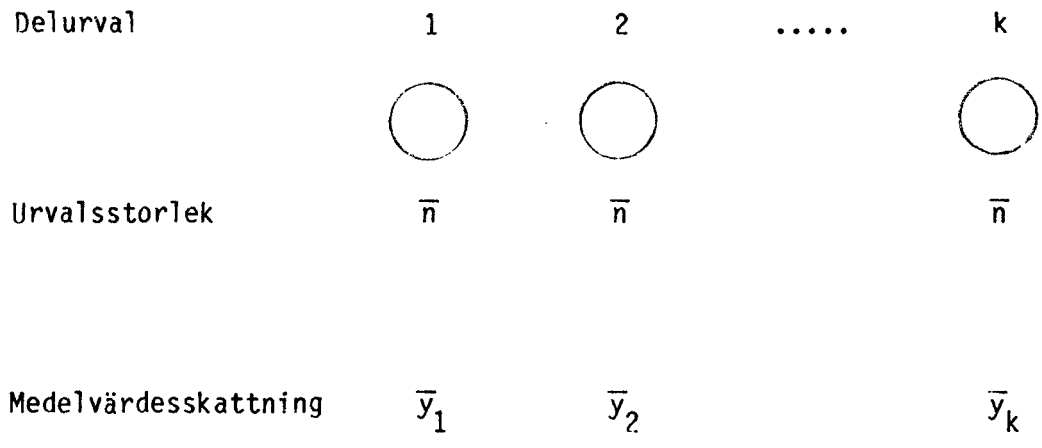
Uppenbarligen ansåg Rice också att intervjuartilldelningen var helt slumpmässig.

Denna studie, som enbart redovisar ovanstående resultat har blivit klassisk och är den första studie där intervjuarfel och intervjuarnas påverkan på uppgiftslämnaren kvantifieras.

4.2.2 Interpenetrerande delurval

I sin enklaste form tillgår tekniken med interpenetrerande delurval på följande sätt: Urvalet till undersökningen dras i form av ett antal, säg k st, oberoende disjunkta delurval, som alla är av storleken $\bar{n}=n/k$, där n är storleken på hela urvalet till undersökningen. Varje delurval är draget från samma ram och med samma design, vilket medger att man kan skatta populationsparametrar, t ex medelvärdet $\bar{Y}$, från vart och ett av dem. Om vi tilldelar k st intervjuare ett sådant delurval vardera kan vi få k st skattningar av t ex medelvärdet, som kan antas vara oberoende även om vi postulerar ett beroende mellan svar eller observationer "inom" varje intervjuare. Det är detta som är den stora fördelen med metoden : Vi erhåller k st oberoende skattningar av samma parameter. Dessa kan inte ens under idealiska förhållanden antas vara helt lika, de är ju utsatta för en stickprovsvariation, men om de varierar mer än man kan förvänta sig, så beror detta på att det också finns en variation mellan intervjuarna.

Figur 4.1:



Uppenbarligen uppfattade Stuart Rice sitt material som insamlat enligt i huvudsak dessa principer, fastän detta uppenbarligen inte gjorts under kontrollerade former. Hans resultat för intervjuare A och intervjuare B betraktade han som två oberoende skattningar av samma parameter, som borde skilja sig betydligt mindre åt än de faktiskt gjorde om inga intervjuarfel förelåg.

Den i figur 4.1 skisserade planen för interpenetrerande delurval medger att man också beräknar variationen "inom" intervjuare och med variansanalysmetoder testar om intervjuarvariationen är signifikant. Man kan också, som Stock och Hochstim (1951) visade, beräkna intervjuarvariationens inverkan på det totala felet. Vi skall nu illustrera detta utförligt i ett exempel, där vi också noterar en parallell med klusterurval, som ofta är uppenbar i samband med användandet av interpenetrerande delurval. Nedanstående modell diskuteras också på ett liknande sätt i Hyman et al (1954), kap 7.

Exempel 4.1: Stock och Hochstims modell för intervjuarfelet. Analogi mellan interpenetrerande delurval och klusterurval.

Antag att vi står inför en situation som skisserats ovan, dvs vi skall göra en intervjuundersökning med k st intervjuare av en population ur vilken ett urval om sammanlagt n individer dras i form av k st oberoende interpenetrerande delurval om $\bar{n} = n/k$ individer. Varje intervjuare tilldelas ett sådant delurval. Vi antar också att de k intervjuarna som deltar är valda slumpmässigt ur en stor population av intervjuare. Vi betraktar nu de svar som skulle kunna erhållas av en viss intervjuare från alla uppgiftslämnare i populationen som ett kluster av svar. Undersökningssituationen är då att ett stickprov om k av dessa kluster, eller k intervjuare, har valts ut och inom varje kluster har dragits ett delurval av $\bar{n}$ svar. Antag att vi undersöker en variabel y och avser att skatta dess medelvärde $\bar{Y}$. Skattningen $\bar{y}$ av $\bar{Y}$, som baserar sig på stickprovet, har variansen $\sigma^2_{\bar{y}}$, vilken kan beräknas som den van-

liga variansen vid klusterurval. Om vi antar att populationerna av intervjuare och individer är mycket stora, kan denna varians approximativt skrivas

$$\sigma^2_{\bar{y}} = \frac{\sigma^2_I}{k} + \frac{\sigma^2_R}{\bar{n}}$$

där

σ^2_I är variationen mellan de medelvärden som skulle erhållas av olika intervjuare om alla intervjuare i intervjuarpopulationen intervjuade alla individer i individpopulationen

σ^2_R är den genomsnittliga variationen mellan alla möjliga uppgiftslämnare för varje intervjuare.

Låt nu y_{ij} vara mätvärdet när intervjuare i intervjuar individ j, $\bar{y}_{i\cdot} = \frac{1}{\bar{n}} \sum_{j=1}^{\bar{n}} y_{ij}$ och $\bar{y}_{\cdot\cdot} = \frac{1}{k\bar{n}} \sum_{i=1}^{k} \sum_{j=1}^{\bar{n}} y_{ij}$. Följande variansanalystabell kan då sättas upp:

Tabell 4.3: Variansanalystabell för Stocks och Hochstims mätfelsmodell

Typ av variation	Kvadratsumma	Antal frihetsgrader	Medelkvadrat
Total	$Q_T = \sum_{i=1}^{k} \sum_{j=1}^{\bar{n}} (y_{ij} - \bar{y}_{\cdot\cdot})^2$	n - 1	
Mellan intervjuare	$Q_I = \sum_{i=1}^{k} \sum_{j=1}^{\bar{n}} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$	k - 1	$S^2_I = \frac{Q_I}{k-I}$
Mellan uppgiftslämnare (inom intervjuare)	$Q_R = \sum_{i=1}^{k} \sum_{j=1}^{\bar{n}} (y_{ij} - \bar{y}_{i\cdot})^2$	n - k	$S^2_R = \frac{Q_R}{n-k}$

Vi kan nu beräkna följande estimatorer av σ_I^2, σ_R^2 och $\sigma_{\bar{y}}^2$:

$$\hat{\sigma}_I^2 = \frac{S_I^2 - S_R^2}{\bar{n}}$$

$$\hat{\sigma}_R^2 = S_R^2$$

$$\hat{\sigma}_{\bar{y}}^2 = \frac{\hat{\sigma}_I^2}{k} + \frac{\hat{\sigma}_R^2}{\bar{n}\, k}$$

Den första termen i högra ledet är en skattning av det tillskott till variansen för $\bar{y}$, som beror på intervjuarvariationen. Vi skall nu se hur Stock och Hochstim använde denna metod på ett empiriskt material för att beräkna intervjuarnas bidrag till den totala variansen:

Tre intervjuare tilldelades var sitt interpenetrerande delurval. Tre frågor ställdes - en faktafråga, en kunskapsfråga och en attitydfråga med följande resultat:

Tabell 4.4:
Variation mellan tre olika intervjuares resultat för tre olika frågor.

Fråga	Andel uppgiftslämnare som avgivit det specificerade svaret			
	Alla 1015 uppgifts-lämnare	Inter-vjuare A	Inter-vjuare B	Inter-vjuare C
Faktafråga. Kan Du köra bil? (Svar-Ja)..	66.1	66.9	63.3	68.2
Informationsfråga: Känner Du till om staten X har några lagar som begränsar storleken på lastbilar? (Svar-Ja).....	67.4	64.4	65.0	72.6
Attitydfråga: Tycker Du att större lastbilar skulle tillåtas eller är de tillräckligt stora nu? (Svar-Tillräckligt stora nu)	74.0	73.3	71.1	77.6

Man ser i tabellen att intervjuare C har ett starkt avvikande resultat på informationsfrågan jämfört med de övriga intervjuarna. Genom att för svaren på denna fråga använda den beskrivna mätmetoden med $k = 3$, $\bar{n} = 338$ och $n = 1014$ fås följande variansanalystabell (vi har förenklat Stocks och Hochstims ursprungliga exempel något eftersom intervjuarna i detta inte hade exakt lika stora tilldelningar. I det ursprungliga exemplet var tilldelningarna 326, 346 och 343 intervjupersoner, dvs sammanlagt 1015. Förenklingen har lett till en del marginella förändringar i nedanstående beräkningar):

Tabell 4.5:
Variansanalystabell för informationsfrågan

Typ av variation	Kvadrat-summa	Antal fri-hetsgrader	Medelkvadrat
Total	223.0581	1013	
Mellan intervjuare.....	1.4101	2	0.7051
Mellan uppgiftslämnare (inom intervjuare).....	221.6480	1011	0.2190

Den totala kvadratsumman kan också beräknas från tabell 4.4 med formeln np(1-p) (eftersom $\bar{y}$ här är ett procenttal använder vi i stället beteckningen p) = 1014 x 0.674 x 0.326 = 222.80 (skillnaden från värdet i tabellen beror på att p = 67.4 är avrundat och på den inledningsvis nämnda förenklingen av exemplet).

Följande estimat erhålls enligt ovanstående formler:

$$\sigma_I^2 = 0.001438$$

$$\sigma_R^2 = 0.2192$$

$$\sigma_p^2 = \frac{0.001438}{3} + \frac{0.2192}{1014} = 0.000479 + 0.000216 = 0.000695$$

Första termen (0.000479) i uttrycket för σ_p^2 representerar intervjuarbidraget till den totala varainsen. Standardavvikelsen för p är $\sqrt{0.000695} \approx 2.6\%$. Om man inte tagit hänsyn till intervjuarvariabilitetn skulle standardavvikelsen för p beräknas som $\sqrt{\frac{p(1-p)}{n}} \approx 1.5\%$. Nettoeffekten av att ta hänsyn till intervjuarvariansen var alltså att variansen tredubblades och standardavvikelsen nästan fördubblades.

I Stocks och Hochstims siffermaterial var alla villkor för att använda mätmetoden inte uppfyllda, eftersom intervjuarna använde sig av s k kvoturval vilket medför risken att den enskilde intervjuaren påverkar urvalsprocessen och att urvalen därigenom inte blir ekvivalenta. Eftersom syftet med siffermaterialet endast var att illustrera mätmetoden bortser vi emellertid från detta.

Slut på exemplet

Hittills har vi betraktat en ganska enkel undersökningsplan med interpenetrerande delurval. I anslutning till praktiska undersökningar har det ofta visat sig vara lämpligt att utveckla denna plan, dels för att göra den praktiskt genomförbar, dels för att göra den effektivare. Vi kan urskilja fyra sådana möjligheter:

(1) I stället för att tilldela varje delurval till en enda intervjuare (observatör) kan man tilldela det en hel grupp intervjuare. Man kan då utröna om någon skillnad föreligger mellan grupperna. Det var, som Mahalanobis (1946) och Sukhatme och Seth (1952) påpekar, vanligt vid skördeundersökningar i Indien att två olika grupper av observatörer samlade in data från vilka oberoende skattningar beräknades. Dessa grupper kunde tillhöra olika organisationer. Man kunde med de båda oberoende skattningarna göra en bedömning av kvaliteten i den hopvägda skattningen. Om de två skattningarna skilde sig mycket från varandra, samlade man i allmänhet in mer data innan resultatet publicerades.

(2) Ett av de svåraste problemen med att genomföra den enklaste varianten av interpenetrerande delurval i en landsomfattande undersökning är att intervjuarnas restider blir orimligt långa. Detta kan man komma till rätta med genom en geografisk stratifiering av intervjuar- och individpopulationerna, där man inom varje stratum tillämpar metoden med interpenetrerande delurval. Denna teknik var till exempel nödvändig att tillämpa i Indien, där skördeundersökningarna omfattade stora distrikt och resor var förenade med stora svårigheter. Även i den surveymodell som utvecklades vid U S Bureau of the Census (Hansen et al, 1951, se avsnitt 5.2) tillämpades denna teknik.

(3) Delvis överlappande delurval. Denna teknik innebär att vissa enheter ingår i två eller flera delurval och alltså observeras mer än en gång. Mahalanobis ansåg detta vara ett effektivt sätt att kontrollera observatörernas arbete. Dessa informerades om att det förekom en överlappning men inte vilka enheter det gällde, vilket förmodades leda till en bättre kvalitet i samtliga observationer.

En typisk undersökningsplan vid Indian Statistical Institute beskrives i nedanstående exempel, där samtliga nämnda urvalsplaner (1), (2) och (3) förenas:

Exempel 4.2: Bengal Crop Survey

Datainsamlingen vid skördeundersökningarna i Bengalen gjordes i stora drag på följande sätt i mitten av 1940-talet: Provinsen Bengalen delades in i ca 1100 "celler", som vardera hade en yta av ungefär 64 kvadratmiles. Dessa celler kan betraktas som strata. Inom varje cell drogs ca 100 urvalsenheter på så sätt att punkter lokaliserades slumpmässigt på en karta och förseddes med ett löpnummer. Vid varje punkt skulle ett område på 2.25 acres (0.9 ha) undersökas av en observatör. Två interpenetrerande delurval bildades nu inom varje cell genom att man lät punkterna med udda löpnummer bilda det ena och punkterna med jämna löpnummer bilda det andra. Båda delurvalen tillsammans täckte alltså hela cellen. De båda urvalen tilldelades slumpmässigt var sin grupp av intervjuare. Ett antal slumpmässigt valda punkter tillordnades dessutom båda delurvalen. I t ex 1946 års undersökning var 14 punkter av sammanlagt 54 i varje delurval gemensamma. För att ingen risk skulle föreligga att observatörerna med gemensamma undersökningspunkter kom i kontakt med varandra såg man till att de olika grupperna av intervjuare hela tiden arbetade i olika områden.

Den typ av resultat man tog fram illustrerades av Mahalanobis med följande tabell, som anger skördeuppskattningen av jute, "Aus (monsoon) rice" och "Aman (winter) rice", 1943, 1944 och 1945. Undersökningsplanen ändrades något i fråga om detaljer mellan åren, men var i princip av den typ som skisserats ovan. De två oberoende urvalen (A och B) kallades "half-samples" eftersom de tillsammans utgjorde hela urvalet.

Tabell 4.6:
Bengal Crop Survey. Jämförelser mellan "half-samples", 1943-45.

År	Antal undersökningsytor			Skördad areal i tusental acres			Differens (5)-(6)		Medelfel för differensen	Fishers t
	Urval (A)	Urval (B)	Totalt	Urval (A)	Urval (B)	Sammanvägt	Total	Procentuell		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
				Jute crop						
1943	29676	29676	59352	2759	2757	2758	2	0,1	100	0,02
1944	30487	30037	60524	2150	2056	2106	94	4,45	50	1,87
1945	53504	52623	106127	2512	2528	2520	-16	0,8	*	-
				Aus (monsoon) rice						
1943	29676	29676	59352	6807	6923	6865	-116	-1,7	446	0,26
1944	30457	30037	60524	7815	7942	7873	-127	-1,6	90	1,41
1945	53504	52623	106127	6966	6805	6864	161	2,3	*	-
				Aman (winter) rice						
1943	31216	31215	62431	23840	24044	23942	-204	-0,8	703	0,29
1944	50501	50107	100608	22491	21903	22201	588	2,7	550	1,07
1945	46148	45971	92119	20970	21202	21087	-232	-1,1	103	2,26 ÷

Ur Mahalanobis (1946)

* Medelfelsberäkningen var inte klar för 1945 vid tryckningen.

÷ Signifikant på 5-procentnivån.

De två skördeestimaten i 1000-tal acres baserade på de två oberoende delurvalen (A och B) visas i kolumnerna 5 och 6 respektive, och den sammanvägda skattningen i kolumn 7. Differensen mellan de två skattningarna i 1000-tal acres visas i kolumn 8 och samma differens uttryckt i procent av den sammanvägda skattningen i kolumn 9. Standardfelet för differensen, beräknad från urvalen, anges i kolumn 10 och Fishers t i kolumn 11.

Vi ser att i 1945 års undersökning är Fishers t signifikant på 5%-nivån för "Aman rice". Mahalanobis nämner inte vilka åtgärder just detta ledde till i undersökningen, men han påpekar på flera ställen i artikeln de kontrollmöjligheter man får med denna typ av undersökningsplan. T ex skulle man för "Aman rice" 1945 kunna studera skattningar från de olika cellerna och se om differensen (A-B) var speciellt stor någonstans. I så fall kunde man specialundersöka dessa celler och med hjälp av de dubbelobserverade undersökningspunkterna reda ut vad som varit orsaken till felet. Mahalanobis anger som exempel på hur oegentligheter kan upptäckas med denna metod, att man ibland upptäckt att medelhöjden på plantorna skattats lägre vid andra undersökningstillfället än vid första även om observationerna gjorts med 2-3 veckors mellanrum. Sådana fel berodde enligt Mahalanobis antingen på att fel fält undersökts eller på slarv eller fusk från observatörens sida.

Slut på exemplet

(4) I exempel 4.2 användes inte variansanalysmetoder för att studera intervjuarfelet. Mahalanobis konstruerade emellertid avancerade undersökningsplaner som medgav att man kunde använda tvåvägs variansanalys, där både intervjuare och intervjuområde varierades, vilket möjliggjorde en mer noggrann analys av intervjuarvariationens orsaker än vad den enklare undersökningsplanen i exempel 4.1 medgav. Vi illustrerar även detta med ett exempel.

Exempel 4.3: Bengal Labour Inquiry

I Jaggadalområdet undersöktes fabriksarbetares förhållanden avseende bostäder och ekonomi 1941, 1942 och 1945. Undersökningsområdet delades upp i fem geografiska delområden (kvarter). Inom varje kvarter delades urvalsenheterna upp i fem lika stora delurval, där vart och ett var

ett oberoende slumpmässigt urval från hela kvarteret. Därmed hade man inom varje kvarter fem interpenetrerande delurval. Varje delurval tilldelades en intervjuare och samtliga fem intervjuare arbetade i alla fem områdena. Denna plan, som kallats "replicated interpenetrating samples", medger att en tvåvägs variansanalys görs för att studera effekten av variationen mellan områden och intervjuare samt samspelseffekter. Resultaten av en sådan analys för 1942 års undersökning visas i tabellen nedan.

Tabell 4.7:
Bengal Labor Inquiry, Jagaddalområdet, 1942 - variansanalys

Typ av variation	Antal frihetsgrader	Medelkvadratsummor för Ålder	Kostnader i rupees per hushåll och månad		Konsumtion av säd i pounds per person och månad
			Totalt	Mat	Säd	
Mellan områden	2	62.13	805.52	36.5	0.07	79.6
Mellan intervjuare	4	304.84	275.78	22.3	1.25	114.7
Område x intervjuare	8	78.47	129.38	9.6	1.47	152.8
Mellan delurval	14	140.81	267.80	16.8	1.21	132.0
Inom delurval	510	127.74	168.12	9.9	0.49	100.3
Total	524	128.09	170.78	10.1	0.51	100.8
F-kvoter (medelkvadratsummor ovan dividerade med respektive medelkvadratsumma inom delurval)						
Mellan områden		0.49	4.79÷	3.69*	0.14	0.80
Mellan intervjuare		2.39*	1.64	2.26	2.55*	1.15
Område x intervjuare		0.61	0.77	0.98	3.00÷	1.53
Mellan delurval		1.10	1.59	1.70	2.47÷	1.32

Ur Mahalanobis (1946)

* Signifikant på 5-procentsnivån.
÷ Signifikant på 1-procentsnivån.

Endast tre av de fem delområdena användes i denna analys eftersom antalet observationer i de övriga två var för få. Genom slumpmässig reducering i vissa av de femton återstående cellerna såg man till att antalet observationer i dessa blev lika (35 stycken).

De olika varianskomponenterna jämfördes med medelkvadratsumman inom delurval. F-test visade att för ålder och månatliga kostnader för säd var intervjuarvariansen signifikant skild från 0. För den senare variabeln var även samspelseffekten (område x intervjuare) signifikant, vilket indikerar att intervjuareffekterna var olika i olika områden. En närmare analys visade att det höga värdet på testvariabeln berodde på en enda intervjuare i ett speciellt område. När replikerade interpenetrerande urval av denna sort används kan man alltså i vissa fall lokalisera felet inte bara till en intervjuare utan också till ett område, vilket kraftigt förenklar kontrollen av felet.

Slut på exemplet

Mahalanobis' syfte med användningen av interpenetrerande delurval var i första hand att kontrollera intervjuarfelet i undersökningarna. Detta kan jämföras med Rices syfte att undersöka förekomsten av intervjuarfel. Stock och Hochstim hade som vi kunnat konstatera ett tredje syfte, nämligen att mäta intervjuarvariansens storlek i förhållande till den totala variansen.

Om vi undantar Bengal Crop Survey har de studier av intervjuarfelet som presenterats i detta avsnitt det gemensamt att de genomförts med mycket få intervjuare. Rice använde endast två vilket emellertid kan betraktas som tillräckligt för hans syfte. Däremot har de ovan refererade studierna av Mahalanobis och Stock och Hochstim kritiserats t ex av Feldman et al (1951) just för att för få intervjuare använts. Anledningen är att ett litet antal intervjuare ger få frihetsgrader vid signifikanstest så att man riskerar att inte upptäcka verkliga intervjuareffekter och att man får dålig precision vid skattningar av intervjuarvariansen. Dessutom kvarstår frågan huruvida frånvaron eller närvaron av intervjuareffekter beror på intervjuare i allmänhet eller på de speciella intervjuare som använts i den aktuella undersökningen. Behovet av studier som medgav generella slutsatser om intervjuarvariationens orsaker ledde därför på 1950-talet till experiment med större antal intervjuare. Durbin och Stuart (1951) och Gales och Kendall (1957) använde 228 respektive 48 intervjuare vid flerfaktorförsök där man studerade hur faktorer som organisationstillhörighet, frågeblankettens utformning, intervjuarutbildning samt intervjuarnas och intervjupersonernas ålder och kön påverkade intervjuarvariansen. Hanson och Marks (1958) rapporterade om den s k Enumerator Variance Study som gjordes i anslutning

till 1950 års folkräkning i USA. I denna studie, där man undersökte för vilka typer av frågor som intervjuarvariansen var störst, användes omkring 1000 intervjuare.

Flertalet av de ovannämnda studierna med interpenetrerande delurval är experiment eller speciella kontrollstudier genomförda i samband med någon större undersökning. Särskilt vid Indian Statistical Institute användes emellertid interpenetrerande delurval i planen för reguljära undersökningar. Denna ansats ger som vi sett i exemplen 4.2 och 4.3 möjlighet till dels kontroll av intervjuarfelet genom att man jämför skattningar från olika delurval eller genom att man med F-test undersöker om signifikanta intervjuareffekter föreligger, dels mätning av intervjuarvariationens bidrag till den totala variansen. I praktiska undersökningar är kontrollen av intervjuarbetet det viktigaste och i Mahalanobis (1946) används metoden enbart i detta syfte. Emellertid finns invändningar mot detta sätt att i kontrollsyfte integrera interpenetrerande delurval i en reguljär undersökningsplan. Sukhatme och Seth (1952) framhåller att interpenetrerande delurval även i en stratifierad design är en dyr metod pga intervjuarnas höga reskostnader, och att den blir ineffektiv i delgrupper där urvalsstorlekarna är små. De framhåller att det är betydligt billigare att kontrollera intervjuarbetet genom direkt handledning och övervakning (supervision) och anger en rad skäl till att satsa resurserna på detta i stället för på interpenetrerande delurval (se även avsnitt 4.2.4). De anser att interpenetrerande delurval kan vara lämpliga att använda om non-samplingfelen kan förväntas bli stora, men i så fall i en provundersökning för att resultaten skall kunna leda till förbättringar i t ex frågeblanketten och intervjuarutbildningen i samband med huvudundersökningen.

4.2.3 Återintervju

Återintervjuer är ett kraftfullt instrument för att skatta vissa felkomponenter i surveymodeller och för att kontrollera rutiner i fältarbetet. Vid sidan av interpenetrerande delurval är återintervjuer den viktigaste metoden att mäta och kontrollera intervjuarfelet. Vi skall i detta avsnitt med "återintervju" avse olika typer av upprepade observationer, även i situationer där datainsamlingen inte sker med intervju.

Återintervju kan göras med samma metod som i originalintervjun. Man kan då genom att jämföra återintervjun med originalintervjun få en uppfattning om den slumpmässiga variation som i allmänhet förekommer i intervjusvar. Återintervjuer kan också göras med en förbättrad metod, då man får möjlighet att bedöma storleken på det totala svarsfelet i originalintervjun. Jämförelserna mellan original- och återintervju kan kompletteras med beroende eller oberoende reconciliering. Reconciliering innebär att man med hjälp av uppgiftslämnaren härleder orsaker till eventuella avvikelser mellan intervjuerna. Beroende reconciliering innebär att återintervjuaren har tillgång till resultaten från originalintervjun medan detta inte är fallet vid oberoende reconciliering. Återintervjun är också användbar som ett löpande inslag i kontrollen av intervjuare och av rutiner i intervjuarbetet. Exempel på sådana inslag kan vara:

- utvärdering av individuella intervjuarinsatser;
- kontroll av att intervju utförts;
- utvärdering av intervjuarnas instruktioner och arbetsprocedurer;
- kontroll av bortfallsklassificeringen.

Vid US Bureau of the Census användes tidigt återintervjuer i samband med s k quality checks i totalräkningar. I dessa utvärderas intervjuarbetet genom att man återintervjuar ett stickprov av den undersökta populationen med en bättre metod än i originalintervjun. Principen för "quality checks" har beskrivits i avsnitt 2 och principen för användning av återintevjuer i samband med mätning av svarsfel för klassindelade variabler återkommer vi till i avsnitt 4.3. Vi skall här i stället uppmärksamma användandet av upprepade observationer på samma enhet i samband med helt eller delvis interpenetrerande delurval, vilket tillämpades i Indien. Vi har redan konstaterat detta i exempel 4.2 ovan. Mahalanobis rapporterar ett flertal resultat av sådana undersökningsplaner av vilka vi återger följande exempel:

Exempel 4.4: Bengal Crop Survey 1943

Att Mahalanobis fäste stor vikt vid resultat från vad han kallar "duplicated enumeration" är förståeligt om man betraktar tabell 4.8, som redovisar observationer på 332 fält i byn Kazikandi genomförda av

observatör A den 14 augusti 1943 och av observatör B den 2 september 1943. Observatörerna noterade - helt oberoende av varandra - namnet på grödan eller vilken blandning av grödor som växte på varje fält:

Tabell 4.8:

Bengal Crop Survey, 1943. Jämförelse mellan två oberoende observationer på vardera av 332 fält vid byn Kazikandi i Goalundo-området.

	B-undersökning 2 september, 1943				
A-undersökning 14:e augusti, 1943	Jute	Aman rice	Jute/ aman	Ingen skörd	Totalt
Jute	4	15	4	3	26
Aus rice	4	12	1	4	21
Jute/aus	17	66	2	9	94
Jute/aus/aman	-	2	-	-	2
Aus/aman	1	-	-	-	1
Ingen skörd	37	45	4	102	188
Totalt	63	140	11	118	332

Endast för 106 av 332 fält (eller 31.9 procent) hade observatörerna gjort samma bedömning av vilken gröda eller blandning av grödor som växte på fältet.

En mer tillfredsställande bild uppvisar tabell 4.9 för ett annat par av observatörer som arbetade i byn Monirampur med samma undersökning.

Tabell 4.9:

Bengal Crop Survey, 1943. Jämförelse mellan två oberoende observationer på vardera av 334 fält vid byn Monirampur i Chuadanga-området.

		B-undersökning, 10 augusti, 1943			
		Aus	Jute/ aus	Ingen skörd	Totalt
	Jute	1	-	1	2
A-undersökning	Aus	123	2	8	133
28:e juni, 1943	Jute/aus	-	3	-	3
	Ingen skörd	7	-	189	196
	Totalt	131	5	198	334

Här överensstämde bedömningarna för hela 315 av 334 fält (=94.3 %).

De mycket olika resultaten i tabell 4.8 och 4.9 visar den stora variation på fältarbetets kvalitet som fanns mellan olika intervjuare och förklarar de stora ansträngningar man i Indien lade ned på att kontrollera intervjuarfelet.

Slut på exemplet

4.2.4 Övervakning av intervjuarbetet (supervision)

Direkt övervakning och handledning av intervjuare utövad av speciellt kvalificerade förmän (supervisors) är en gammal metod att nedbringa intervjuarfelet. Den är tillämpbar och bör förekomma vid alla former av intervjuarbete och kan alltså kombineras med de tidigare beskrivna metoderna interpenetrerande delurval och återintervju. Supervision (i brist på vedertagen svensk benämning använder vi den engelska termen) är, som den normalt bedrivs, inte tillämpbar om man skall mäta intervjuarfelets storlek i en undersökning men när det gäller kontrollen av intervjuarbetets kvalitet är supervision, enligt Sukhatme och Seth (1952), bättre än interpenetrerande delurval. De räknar upp ett antal fördelar med supervison t ex:

- Supervision leder till omedelbara förbättringar av fältarbetsresultaten i den pågående undersökningen medan metoden med interpenetrering vanligen leder till att förbättningar föreslås först när undersökningen är genomförd.

- Interpenetrering avslöjar inte smärre felaktigheter i en observatörs arbete och definitivt inte fel som är lika för alla observatörer. Båda dessa feltyper kan emellertid upptäckas med supervision.

- Enheter som skall tas ut för supervision behöver inte väljas slumpmässigt (till skillnad från interpenetrerande delurval). Om de ändå väljs slumpmässigt kan detta urval användas för skattning av non-samplingfel och för att korrigera skattningarna som baserar sig på den ordinarie datainsamlingen.

4.3 Svarsfel

4.3.1 Allmänt

Med svarsfel (som ibland benämnes observationsfel eller mätfel) avses ofta, som vi tidigare nämnt, flera olika enskilda felkomponenter som "coverage error", intervjuarfel och bortfallsfel förutom det fel som beror på uppgiftslämnaren själv. Vi behandlar därför ett antal metoder för att mäta delar av svarsfelet i andra avsnitt, t ex evalveringsmetoder i avsnitt 4.5 (ramfel), interpenetrerande delurval i avsnitt 4.2 (intervjuarfel) och subsampling i bortfallsstratum i avsnitt 4.4. I detta avsnitt skall vi bortse från bortfallsfelet och täckningsfelet. Däremot är intervjuarfelet så nära förknippat med svarsfelet att vi inte kan skilja dessa helt från varandra. I jämförelse med framställningen av intervjuarfelet i avsnitt 4.2 kommer dock detta avsnitt att bygga på en annan svarsmodell: I avsnitt 4.2 (speciellt exempel 4.1) antogs att intervjuarvariansen och urvalsfelet för skattningar som tar hänsyn till intervjuarvariation var samma som vid klusterurval, där svaren från alla individer i populationen till en viss intervjuare utgjorde kluster. Emellertid behöver svaren från en given intervjuperson till en given intervjuare inte betraktas som fixa. De kan ses som utfall av en stokastisk variabel, dvs det finns ett antal möjliga svar vilka avges med vissa (okända) sannolikheter. Vi skall studera svarsfelet under denna modell.

Svarsfelet kan delas upp i

i) En slumpmässig del, svarsvarians, som i sin tur kan delas upp i två typer, "korrelerad" och "enkel" svarsvarians.

ii) En systematisk del, svarsbias.

De viktigaste metoderna för att mäta svarsfelet är interpenetrerande delurval ("korrelerad" svarsvarians) och återintervju ("enkel" svarsvarians och svarsbias).

4.3.2 Svarsvarians

Svarsvarians uppdelas vanligen i två komponenter, korrelerad och enkel svarsvarians. Den korrelerade svarsvariansen uppstår på grund av den korrelation som finns mellan svar som avges till samma intervjuare och som i sin tur beror på att varje intervjuare tenderar att påverka uppgiftslämnaren i en viss riktning . Den korrelerade svarsvariansen kan mätas med metoder som påminner om klusterurval och som är analoga med dem som vi under en annan modell studerade i avsnitt 4.2. I Hansen et al (1951) presenteras metoder för mätning av korrelerad svarsvarians under svarsmodellen med slumpmässiga svar och med användande av interpenetrerande delurval. Dessa metoder är i Hansen et al's artikel integrerade i en surveymodell för mätning av det totala felet i en survey och kommer därför att behandlas i kapitel 5, "surveymodeller".

Vi övergår nu till att betrakta den enkla svarsvariansen som beror på uppgiftslämnarens tillfälliga sinnesstämning och andra faktorer i intervjusituationen som intervjuaren inte kan påverka. Då dessa faktorer är olika för olika uppgiftslämnare talar man ibland om okorrelerade svarsfel. Hansen et al visade att den enkla svarsvariansen kommer med som en del av den varians man får när man på "vanligt" sätt skattar variansen för en medelvärdesestimator vid obundet slumpmässigt urval, under en modell för enbart stickprovsfel. Detta kan matematiskt visas på följande sätt:

Antag att vi står inför en situation där svarsfelen är okorrelerade för varje par av individer i populationen och där de kompenserar varandra, så att nettoeffekten är = 0 och ett obundet slumpmässigt urval av storlek n dras på vilket man gör observationer (en på varje individ). Antag också att det "sanna värdet" är x_i för enhet i men att våra observationer är behäftade med ett slumpmässigt svarsfel $r_{i\alpha}$ för enhet i vid upprepning α (vi tänker oss alltså att undersökningen kan upprepas under exakt samma förutsättningar ett stort antal gånger). För mätvärdet y_i gäller alltså $y_{i\alpha} = x_i + r_{i\alpha}$. Variansen för en skattning

($\bar{y}$) av ett populationsmedelvärde är under dessa antaganden:

$$\sigma^2_{\bar{y}} = \frac{\sigma^2_y}{n} , \qquad (4.3.1)$$

där σ^2_y är den totala variansen mellan alla mätvärden från alla individer till alla intervjuare. Eftersom $y_{i\alpha} = x_i + r_{i\alpha}$ kan vi skriva

$$\sigma^2_y = \sigma^2_x + \sigma^2_{r_y} + 2\rho_{xr_y}\sigma_x\,\sigma_{r_y}, \qquad (4.3.2)$$

där

σ^2_x = variansen mellan de sanna värdena,

$\sigma^2_{r_y}$ = svarsvariansen (enligt förutsättningarna ovan är detta alltså den "enkla" eller "okorrelerade" "svarsvariansen). Denna är komponerad dels av variansen mellan svarsfel för en och samma uppgiftslämnare kring dennes förväntade svarsfel och dels av variansen av de förväntade svarsfelen mellan olika uppgiftslämnare,

ρ_{xr_y} = korrelationen mellan x_i och $\bar{r}_i$, som är det förväntade svarsfelet för den i:te uppgiftslämnaren.

(För en definition av begreppet "förväntat svarsfel" hänvisar vi till avsnitt 5.2.)

Vi ser i formel (4.3.2) ovan att σ^2_y innehåller alla effekter från såväl svarsfelet som variansen mellan sanna värden. På samma sätt gäller om vi skattar σ^2_y med

$$s_y^2 = \frac{\sum_{i=1}^{n} (y_{i\alpha} - \bar{y})^2}{n-1}$$

att effekten av svarsfelet ryms i den skattade variansen s_y^2, eftersom $E[s_y^2] = \sigma_y^2$.

4.3.3 Mätning av svarsvariansen med återintervju

För att skatta den enkla svarsvariansen separat behöver man en undersökningsplan med återintervjuer. Under 1950-talet utvecklades US Bureau of the Census' metodik för behandling av svarsfel, som redovisats i bl a Hansen et al (1951), till att omfatta även sådana undersökningsplaner. I Eckler och Hurwitz (1957) presenterades sålunda en teori för mätning av svarsvarians i samband med klassindelade data. Denna teori grundar sig på återintervjuer med antingen samma metod som i originalintervjun eller med en s k preferred procedure. De skillnader som uppstår mellan intervjuomgångarna mäts med följande två mått:

(i) Gross number of errors (bruttofel), som är antalet felaktigt inkluderade plus antalet felaktigt exkluderade enheter i en klass

(ii) Net number of errors (nettofel), som är antalet felaktigt inkluderade minus antalet felaktigt exkluderade enheter i en klass.

Om man gjort återintervjuerna med förbättrade metoder betraktar man resultaten av dem som sanna vilket förklarar användningen av ordet "felaktigt" i definitionerna. Om återintervjuerna gjorts med samma metoder som originalintervjuerna talar man istället om "fel i förhållande till återintervjun". Ibland används också begreppen "gross difference" och "net difference". Vi skall illustrera begreppen med ett exempel.

Exempel 4.5: Post-enumeration Survey (I)

I tabell 4.10 redovisas resultat från en återintervjustudie i samband med 1950 års folkräkning i USA.

Tabell 4.10:

Resultat från en återintervjustudie (Post-enumeration Survey) i samband med 1950 års folkräkning i USA. (Eckler och Hurwitz, 1957)

Åldersklass	Antal personer enligt folkräkning	Felklassificerade i folkräkningen			Nettofel i folkräkningstotaler		Antal personer enligt återintervju (2)-(6)
		Felaktigt inkluderade	Felaktigt exkluderade	Bruttofel (3) + (4)	Antal (3)-(4)	Standardavvikelse $\sqrt{(5)}$	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total	26980	1611	1611	3222	14	±14	26966
0-4	2897	40	88	128	-48	±11	2945
5-9	2377	71	50	121	21	±11	2356
10-14	2023	65	62	127	3	±11	2020
15-19	1907	86	55	141	31	±12	1876
20-24	1997	98	93	191	5	±14	1992
25-29	2160	123	124	247	- 1	±16	2161
30-34	2062	150	140	290	10	±17	2052
35-39	2010	147	132	279	15	±17	1995
40-44	1835	163	138	301	25	±17	1810
45-49	1632	154	155	309	- 1	±18	1633
50-54	1495	161	156	317	5	±18	1490
55-59	1309	108	150	258	-42	±16	1351
60-64	1083	82	93	175	-11	±13	1094
65-69	899	76	81	157	- 5	±13	904
70-74	606	57	56	113	1	±11	605
75 och äldre	688	30	38	68	- 8	± 8	696

I tabellen visar kolumn 2 det totala antalet fall för vilket åldersinformation var tillgänglig i både folkräkningen och återintervjustudien, den senare baserad på ett stickprov. Kolumn 3 visar antalet personer som (i förhållande till återintervjun) av intervjuaren var felaktigt inkluderade i varje åldersgrupp i originalintervjun. Kolumn 4 visar hur många som felaktigt exkluderades i varje åldersgrupp och kolumn 5, som är summan av kolumnerna 3 och 4, anger det s k "gross number of errors" i varje åldersgrupp. Kolumn 6, skillnaden mellan kolumnerna 3 och 4, anger det återstående nettofelet (net error) för varje åldersklass. I t ex åldersgruppen 25-29 år noterades en "gross total" på 247 men eftersom de felaktigt inkluderade och de felaktigt exkluderade var ungefär lika många blev nettofelet endast 1. Nettofelet (kolumn 6) är en skattning av biasen i folkräkningen i förhållande till återintervjustudien. Antalet "gross errors" (kolumn 5) är enligt Eckler och Hurwitz en skattning av svarsvariansen för nettofelet och kvadratroten av antalet "gross errors" (kolumn 7) är en skattning av standardavvikelsen för nettofelet. Detta gäller om man använt samma procedur i båda intervjuomgångarna och om svaren från olika individer är oberoende. Det kan emellertid också gälla om man använt olika procedurer så länge differenserna är uppmätta oberoende för varje uppgiftslämnare och summerar sig till noll. Om de inte summerar sig till noll gäller att skillnaden mellan bruttofelet och kvadraten på nettofelet är en skattning av variansen för nettofelet.

Slut på exemplet

Svarsvariansen för en total (i stället för differensen mellan två totaler) i en folkräkning kan enligt Eckler och Hurwitz skattas med hälften av det totala antal "gross differences" som skulle erhållas vid återintervju med samma procedur. En förutsättning för detta förfarande är att svaren är oberoende både inom och mellan intervjuomgångarna. Under samma förutsättning anger författarna följande metod att beräkna en övre gräns för svarsvariansen för antalet personer i en klass:

Om P är andelen av den totala populationen som tillhör klassen och N är totala antalet enheter som ingår i undersökningspopulationen, kan svarsvariansen inte vara större än $NP(1-P)$. Analogin med obundet slumpmässigt urval är här uppenbar, eftersom variansen för en stickprovstotal är approximativt $nP(1-P)$, där n är stickprovsstorleken.

På motsvarande sätt blir övre gränsen för svarsvariansen när man skattar en proportion lika med P(1-P)/N.

Om man tillåter beroende mellan vissa svar, t ex låter samma intervjuare intervjua flera uppgiftslämnare och alltså inför en korrelerad svarsvarians, ökar den totala svarsvariansen och dess övre gräns blir vid skattning av en proportion P(1-P)/m, där m är antalet intervjuare. Denna gräns är alltså N/m gånger så stor som den gräns som gäller under antagande om oberoende observationer. I USAs folkräkning 1950 var N/m ca 1000 vilket illustrerar den stora potentiella effekt på svarsvariationen som intervjuarnas insatser har.

De ovan redovisade övre gränserna för den enkla svarsvariansen kan överskrida den aktuella svarsvariabiliteten kraftigt, som följande exempel visar. (Exemplet är hämtat ur Eckler och Hurwitz, 1957).

Exempel 4.6: Post-enumeration Survey (II)

Tabell 4.11 visar variansberäkningar baserade på återintervjuer av ett urval av personer i 1950 års folkräkning i USA. Datainsamlingen skedde med en "self-enumeration procedure" vilket gör att man kan anta att observationerna är oberoende inom varje undersökningstillfälle. Man har också antagit oberoende mellan undersökningstillfällena.

Tabell 4.11:

Redovisningsgrupp	Område med 6500 invånare			Område med 100000 invånare		
	Antal pers.	Standardavvikelse Antal	%	Antal pers.	Standardavvikelse Antal	%
Total population	6500	-	-	100000	-	-
Ålder, män						
Under 5 år	376	3	0,9	6016	13	0,2
15 och äldre	2314	3	0,12	37024	11	0,03
35 och äldre	1360	4	0,3	21760	15	0,07
55 och äldre	554	4	0,7	8864	15	0,17

Tabellen visar, på basis av resultat från återintervjustudien, den svarsvarians man kan förvänta sig för två hypotetiska områden omfattande 6 500 och 100 000 individer. Dessa varianser befinner sig långt under de övre gränserna. Till exempel indikerar tabellen att kvadratroten ur svarsvariansen för antalet män i klassen 55 år och äldre är 4 personer i ett område med 6 500 individer. Den övre gränsen för svarsstandardavvikelsen i denna klass beräknad enligt formeln $\sqrt{NP(1-P)}$, är:

$$\sqrt{6500 \frac{554}{6500} \times \frac{5946}{6500}} \approx 22$$

Slut på exemplet.

4.3.4 Svarsbias

Svarsbiasen för t ex ett medelvärde kan, under den svarsmodell vi utgår från i detta avsnitt, beskrivas som skillnaden mellan förväntade värdet av medelvärdet av de observerade svaren och medelvärdet av de "sanna" värdena i populationen. Svarsbias uppstår t ex om en fråga i intervjublanketten är felformulerad eller om alla eller många intervjuare gör ett likadant fel. Svarsbiasen kan mätas med återintervju om man vid denna använder en förbättrad metod som kan antas ge en god approximation till det sanna värdet. Hansen et al (1951) har i anslutning till sin surveymodell angett en sådan metod som kan utnyttjas till att minska svarsbiasen i de reguljära skattningarna. Då denna metod är nära kopplad till surveymodellen beskriver vi den emellertid inte här utan i avsnitt 5.2. Vi skall i stället ge ett exempel (från Eckler och Hurwitz, 1957) på hur man beräknar svarsbias med hjälp av brutto- och nettofel:

Exempel 4.7: Post-enumeration Survey (III)

Även detta exempel är hämtat från den s k Post-enumeration Survey som genomfördes i samband med 1950 års folkräkning i USA. I denna återintervjuades ett urval om 30 000 individer med en förbättrad metod, som bl a innebar att man rekryterade de bästa intervjuarna, gav dem omfattande utbildning, intervjuade den "bäst" informerade personen i varje hushåll i stället för den man först anträffade och ställde mer detaljerade och inträngande frågor. Man använde sig också av beroende reconciliering (se avsnitt 4.2). Totalt innebar den förbättrade metoden en fördyring av intervjukostnaden per intervju med 20 gånger motsvarande kostnad för originalintervjun.

I tabell 4.12 nedan ur Eckler och Hurwitz (1957) visas hur man mätte svarsbiasen med hjälp av Post-enumeration Survey. Den första raden av siffror är en summering av tabell 4.10. Övriga rader är summeringar av motsvarande tabeller för variablerna yrke, näringsgren, inkomst och antal rum i lägenheten. I tabell 4.12 anger kolumn 2 antalet klasser, vilket ger en grov uppskattning av möjligheterna till felklassificering.

Kolumn 5 anger andelen felklassificerade i folkräkningen i förhållande till återintervjun. I kolumn 7 anges det genomsnittliga nettofelet per klass (oavsett tecken). Kolumn 9 anger detta fel i procent av det genomsnittliga antalet personer i en klass. Den skattade standardavvikelsen av det genomsnittliga nettofelet i absoluta tal och procentuellt ges i kolumnerna 8 respektive 10. Detta är beräknat enligt den i avsnitt 4.3.3 angivna metoden, som kvadratroten ur bruttofelet. Storleken på det genomsnittliga nettofelet dividerat med dess standardavvikelse är, liksom procenten felklassificerade (kolumn 5), ett grovt mått på den genomsnittliga kvaliten på folkräkningsresultaten för en variabel. Av tabell 4.12 drog Eckler och Hurwitz slutsatsen att för variablerna ålder, yrke, näringsgren och antal rum ger de ordinarie folkräkningsprocedurerna likartade skattningar som de förbättrade procedurerna i återintervjun. Däremot ger uppenbarligen återintervjuprocedurerna resultat för variabeln inkomst, som signifikant skiljer sig från den ordinarie folkräkningen. Då skattningarna från återintervjustudien kan antas vara de bästa har vi här ett exempel på svarsbias i undersökningsresultaten.

Tabell 4.12: Brutto- och nettofel i 1950 års folkräkning i USA enligt återintervjustudie i Post-enumeration Survey.

Variabel	Antal klasser	Antal personer enligt folkräkningen	Felklassificerade i folkräkningen		Genomsnitt antal personer per klass	Nettofel		Nettofelets andel av hela klassen	
			Antal	% (4)/(3)		Antal	Standardavvikelse $\sqrt{2x(4)/(2)}$	Andel (7)/(6)	Standardavvikelse (8)/(6)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ålder (5-årsklasser)	16	26980	1611	6 %	1686	14	±14	0,8 %	±0,8 %
Yrke	10	9502	1402	15	950	17	±17	1,8	±1,8
Näringsgren	14	9464	1069	11	676	13	±12	1,9	±1,8
Inkomst	15	9012	2827	31	601	65	±19	10,8	±3,2
Antal rum	9	19300	3528	18	2144	54	±28	2,5	±1,3

4.4 Bortfall

Bortfallsfelet intar, som vi tidigare nämnt, en särställning bland felkällorna eftersom man endast i undantagsfall är kapabel att införskaffa mätvärden för bortfallsobjekten. Metoder för mätning av bortfallsfelet kan därför inte konstrueras på samma sätt som för övriga felkällor. Vi skall i stället betrakta en metod som leder till att bortfallsfelet, som i regel är systematiskt, omvandlas till en ökning av variansen för den aktuella skattningen. Denna metod, som är den enda som existerar som leder till unbiased skattningar vid bortfall, presenterades av Hansen och Hurwitz (1946), men den hade skisserats redan av Cornfield (1942). Den citeras fortfarande ytterst frekvent och används ofta i undersökningar även om det är sällan den i praktiken fungerar helt i enlighet med teorin. Metoden bygger på bortfallsstratumansatsen (se avsnitt 3.6) och innebär att man utnyttjar tvåfasurval i kombination med två olika mätförfaranden, ett relativt billigt (t ex enkät) och ett relativt dyrt (t ex telefonintervju), på följande sätt:

- I den första fasen dras ett slumpmässigt urval från populationen. Data samlas in för detta med den billigare metoden varvid ett bortfall uppkommer.

- I den andra fasen dras ett slumpmässigt urval från bortfallet. Data samlas in för detta med den dyrare metoden som antages vara så effektiv att inget bortfall uppkommer. Därmed kan unbiased skattningar fås för bortfallsstratum och bortfallsproblemet är i denna mening löst. Dock har detta skett till priset av en variansökning jämfört med om inget bortfall förekommit.

Metoden med subsampling i bortfallsstratum bygger alltså på antagandet att data kan samlas in för hela subsamplet . Detta antagande är sällan realistiskt men även om ett visst bortfall uppträder i subsamplet kan metoden leda till att bortfallsfelet reduceras.

Vi skall här beskriva metoden matematiskt i idealfallet när inget bortfall uppträder i andra fasen:

Beteckningar:

	Svars-stratum	Bortfalls-stratum	Hela populationen
Populationsstorlek	N_s	N_b	$N(=N_s+N_b)$
Urvalsstorlek	n_s	n_b	$n(=n_s+n_b)$
Medelvärde	$\bar{Y}_s$	$\bar{Y}_b$	$\bar{Y}$
Medelvärdesskattning	$\bar{y}_s$	$\bar{y}_b$	$\bar{y}$
Populationsvarians	S_s^2	S_b^2	S^2
Populationsvikter	$W_s = \frac{N_s}{N}$	$W_b = \frac{N_b}{N}$	
Stickprovsvikter	$w_s = \frac{n_s}{n}$	$w_b = \frac{n_b}{n}$	

Antag att vi skall skatta populationsmedelvärdet $\bar{Y}$ som kan skrivas

$$\bar{Y} = W_s\bar{Y}_s + W_b\bar{Y}_b$$

Från undersökningspopulationen dras ett obundet slumpmässigt urval utan återläggning (OSU-UÅ) av storleken n. De som svarar utgör då ett OSU-UÅ från svarsstratum och bortfallet utgör ett OSU-UÅ från bortfallsstratum.

En väntevärdesriktig skattning av $\bar{Y}$ blir då

$$\bar{y} = w_s \bar{y}_s + w_b \bar{y}_b$$

där $E(\bar{y}_s) = \bar{Y}_s$ och $E(\bar{y}_b) = \bar{Y}_b$

Då $\bar{y}_b$ inte kan beräknas eftersom data för bortfallet inte är tillgängliga, dras ett urval omfattande r_b objekt med OSU-UÅ från de n_b objekten i urvalet från bortfallsstratum. Stickprovsmedelvärdet $\bar{y}_{r_b}$ i detta subsample är en väntevärdesriktig skattning av $\bar{Y}_b$. Om vi ersätter $\bar{y}_b$ med $\bar{y}_{r_b}$ i formeln för $\bar{y}$ ovan får vi en ny väntevärdesriktig skattning, $\hat{\bar{y}}$, av $\bar{Y}$:

$$\hat{\bar{y}} = w_s \bar{y}_s + w_b \bar{y}_{r_b}$$

Variansen för $\hat{\bar{y}}$ blir:

$$V(\hat{\bar{y}}) = (1 - \frac{n}{N}) \frac{S^2}{n} + (\frac{n_b}{r_b} - 1) \frac{w_b S_b^2}{n} \quad .$$

Hansen och Hurwitz (1946) beräknar också det optimala värdet på kvoten $\frac{n_b}{r_b}$ med avseende på kostnaderna för de båda insamlingsmetoderna.

Hansen och Hurwitz' metod utvidgades av El-Badry (1956) till fallet när $m \geq 1$ successiva enkätpåminnelser görs i fas 1 följda av telefon- eller besöksintervjuer i fas 2. De enheter som svarar vid den i:te påminnelsen med inte vid de föregående bildar stratum i, som omfattar en andel P_i av populationen. Detta förfarande förutsätter att det finns m + 1 olika nivåer på svarssannolikheterna och metoden bygger alltså inte, som Hansens och Hurwitz metod, på den enklaste svarssannolikhetsmodellen med endast ett bortfallsstratum och ett svarsstratum.

Vi skall nu betrakta andra metoder som utvecklats för mätning, kontroll och reduktion av bortfallsfelet.

Man kan beräkna övre och under gränsen för bortfallsfelet genom att tillordna samtliga bortfallsobjekt likadana extremvärden. Detta utnyttjade Birnbaum och Sirken (1950) till att (under bortfallsstratummodellen) göra tabeller över vilken stickprovsstorlek som är lämplig under olika antaganden om bortfallets storlek och givet ett visst precisionskrav.

De första som använde en modell med varierande svarssannolikheter i samband med bortfallsfelet var Politz och Simmons (1949, 1950). De utvecklade en teknik som föreslagits redan av Hartley (1946) och som innebar att varje intervjuperson endast besöks en gång, vid en slumpvis vald tidpunkt. Intervjupersonerna ombeds då ange hur många gånger under de fem närmast föregående dagarna som de varit hemma vid samma tidpunkt. Denna extra information om sannolikheten att vara hemma kan sedan utnyttjas för att reducera den bortfallsbias som annars skulle erhållas vid endast ett besök. Simmons (1954) utvidgade planen så att den tillåter flera återbesök. Även Deming (1953) använde implicit svarssannolikheter för att definiera strata i vilka uppgiftslämnarna blir allt mindre benägna att svara. Deming introducerade också begreppet "hard core nonrespondents", som beteckning för den grupp från vilken svar inte kan erhållas med någon metod .

Kish och Hess (1959) anvisar en originell metod för att reducera bortfallsbias vilken går ut på att bortfallsenheter i en pågående undersökning ersätts med bortfallsobjekt från en tidigare undersökning. De båda undersökningarna förutsättes då vara så lika varandra vad beträffar t ex variabelinnehåll och datainsamlingsmetoder, att bortfallet genereras på samma sätt och består av likartade grupper av individer. Den stora fördelen med metoden är att man kan börja bearbeta bortfallet redan vid den tidpunkt då datainsamlingen för den ordinarie under sökningen börjar. Metoden bygger alltså på antagandet att det går att få in mätresultat för en stor del av bortfallet bara man får tid på sig. Vi skall illustrera den med ett exempel ur Kishs och Hess' artikel.

Exempel 4.8:

Antag att vi skall undersöka ett urval på 1111 individer. Av erfarenhet vet vi att vi kommer att få in svar från 90 % (efter normal bortfallsuppföljning), dvs vi beräknas få 1111 x 0.90 = 1000 intervjuer. Antag nu att vi har ett "gammalt" bortfall om 103 individer tillgängligt från en annan undersökning och att vi kan räkna med att få in svar från 70 % av dessa med fortsatt uppföljning, dvs vi kan erhålla 103 x 0.70 = 72 svar från denna grupp. För att få in det antal, 1000, svar som vi kunde räkna med i en konventionell undersökning, behöver vi nu endast få in 1000-72 = 928 svar från det ordinarie urvalet och alltså kan detta minskas till 928/0.90 = 1031 individer. Om fältarbetet går som planerat får vi alltså in 928 + 72 = 1000 svar. Sammanlagt får vi bearbeta 1031 + 103 = 1134 personer, vilket är något fler än i en konventionellt genomförd undersökning, men osäkerheten pga bortfall har reducerats från 10 % till omkring 10 % x 30 % = 3 %. Effekten av bortfall kan ytterligare reduceras genom uppvägning av de 72 svaren i bortfallet med 1/0.70 till 103 svar även om biasen inte kan elimineras helt eftersom de 31 individer som inget mätvärde erhållits för kan skilja sig från de insamlade 72.

Slut på exemplet.

Kishs och Hess' metod hade med gott resultat använts vid Survey Research Center vid University of Michigan (där författarna fortfarande är verksamma) på den del av bortfallet som bestod av "ej anträffade". Den hade där visat sig praktisk och enkel att tillämpa och relativt billig. Författarna ansåg också att den borde prövas på bortfallsgruppen vägrare.

4.5 Ramfel

Vi skall här beskriva de metoder som angetts för att mäta och reducera ramfel av de tre olika typer som nämns i avsnitt 3.7.

4.5.1 Enheter som tillhör undersökningspopulationen saknas i ramen

För denna situation, som alltid leder till underskattning av en total, diskuterar Eckler och Pritzker (1951) fem metoder att kontrollera ramen:

i) Om totalen Y har mätts i en annan undersökning kan jämförelser mellan de båda skattningarna indikera att ramen innehållit för få enheter. Denna metod kan naturligtvis endast användas som ett första steg i en utredning om ramfelet. En svårighet är att avgöra vilken av skattningarna som är den "bästa". Vi skall illustrera denna metod med ett exempel (ur Eckler och Pritzker, 1951).

Exempel 4.9:

För att bedöma kvaliteten i 1940 års folkräkning i USA, jämfördes för olika redovisningsgrupper skattningar av antalet män i åldersklassen 21-35 år med antalet män i dessa grupper som registrerades för "Selective Service" i oktober 1940, 6 månader efter folkräkningen. Jämförelsen visade att folkräkningssiffran för totala antalet män 21-35 år var 5.5 % lägre än den från "Selective Service". I klassen "Negro males aged 21 to 35" var motsvarande siffra omkring 15 %. Den förra skillnaden trodde man berodde på en underskattning i folkräkningen eftersom unga män allmänt sett är den mest rörliga gruppen i befolkningen och därmed svår att räkna. Den senare var svårare att förklara och man bedömde närmast att dess storlek berodde på felaktigheter i registreringen av "ras" i Selective Service.

Slut på exemplet

ii) Inflow-outflow-teknik. För vissa typer av skattningar är totalerna lika med summor eller differenser av vissa kvantiteter, vilket ger möjligheter till kontroll av en ram. T ex är populationen vid tiden t_2 lika med populationen vid tiden t_1 plus antalet födda minus antalet döda plus nettomigrationen. Även här är det viktigt att kunna avgöra vilken siffra som är bäst.

Exempel 4.10: Inflow-outflow-teknik

Inflow-outflow-teknik användes vid 1940 års folkräkning i USA för att beräkna undertäckningen för barn 0-4 år. Undersökningen genomfördes i följande fyra steg:

1) Antalet födslar som registrerats under de fem åren närmast före folkräkningen räknades.

2) Antalet av de födda enl 1) som dött räknades.

3) Antalet i 1) och 2) korrigerades för ofullständig rapportering.

4) Antalet överlevande jämfördes med antalet barn i åldersgruppen 0-4 år som räknats i folkräkningen.

Med denna teknik kunde man skatta felet i folkräkningsresultaten för variabeln "antal barn 0-4 år". Detta fel innebar en underskattning av det verkliga antalet med totalt 7.6 % och med följande andelar i några redovisningsgrupper:

	%
Vita	6,4
Icke vita	15,2
Boende i städer	5,0
Boende på landsbygden, ej bondgård	5,8
Boende på landsbygden, bondgård	13,2
Totalt	7,6

(Ur Eckler och Pritzker, 1951)

Slut på exemplet.

iii) Förväntad åldersfördelning. Man kan upptäcka att enheter saknas i ramen för vissa åldersgrupper genom att jämföra den förväntade åldersfördelningen baserad på data från en tidigare undersökning med den åldersfördelning man fått i den aktuella undersökningen. Denna metod kan liknas vid inflow-outflow-teknik applicerad på flera åldersgrupper samtidigt.

iv) Reverse record check. Reverse record check innebär att man jämför ett register, vars samtliga enheter ingår i undersökningspopulationen, med ramen och därigenom kan bedöma hur fullständig denna är. Denna metods begränsning ligger framförallt i att man i regel endast kan kontrollera en mindre del av ramen med den.

v) Quality check. Med quality check-tekniken görs ramen, eller snarare en del av den, om med bättre metoder såsom utnyttjande av bättre kartor, mer kvalificerad personal etc. En jämförelse mellan de båda ramarna kan sedan leda till en skattning av undertäckningen i den ursprungliga ramen.

4.5.2 Ramen innehåller enheter som inte tillhör undersökningspopulationen

Om ramen, förutom undersökningspopulationen, omfattar ytterligare enheter överskattas alltid en total, såvida det inte är möjligt att identifiera övertäckningen i urvalet. Fyra av de metoder som angavs i föregående avsnitt kan användas även i denna situation.

"Reverse record check" är dock inte lämplig för att beräkna övertäckningen (såvida inte det jämförda registret innehåller den fullständiga urvalsramen men i så fall skulle rimligen detta register ha använts från början). Marks, Mauldin och Nisselson (1953) anger olika varianter av "quality checks" som alla bygger på att man för ett urval av enheter gör en noggrannare klassificering och därmed kan skatta både övertäckningens storlek och dess inflytande på totalskattningar.

4.5.3 Vissa enheter förekommer flera gånger i urvalsramen

När vissa enheter i undersökningspopulationen förekommer flera gånger i urvalsramen leder detta till en bias i undersökningsresultaten om man genomför undersökningen som om detta problem inte existerade, eftersom de duplicerade enheterna får för hög urvalssannolikhet. Hansen et al (1953) diskuterar detta problem i anslutning till det välkända exemplet när man har "familjer med skolbarn" som undersökningsenhet och drar ett stickprov av sådana familjer från en ram som består av alla skolbarn i det aktuella området. Sannolikheten för en familj att komma med i urvalet beror då på hur många skolbarn som finns i den. Hansen et al (1953) anger tre metoder att komma till rätta med problemet:

i) Bortsortering av övertaliga enheter innan urvalsdragning sker. I exemplet med skolbarnsfamiljer skulle detta innebära att man utifrån den ursprungliga ramen av skolbarn upprättade en ny ram av skolbarnsfamiljer. Denna metod förutsätter att man på ett enkelt sätt kan identifiera alla poster med samma enhet.

ii) Vägning vid estimationsproceduren. Om man kan ta reda på de rätta urvalssannolikheterna för varje undersökningsenhet kan man konstruera väntevärdesriktiga skattningar även från en ram med duplicerade enheter. I exemplet ovan kan detta ske genom att man i anslutning till datainsamlingen frågar om antalet skolbarn som finns i familjen.

iii) Definition av en entydig inkluderingsregel. Om man t ex kan identifiera den första av de poster i ramen som associeras till samma enhet kan man formulera inkluderingsregeln att en enhet kommer med i urvalet endast om den första posten dras (en sådan post finns ju för samtliga enheter i populationen). I vårt exempel kan man använda inkluderingsregeln att en familj kommer med i urvalet endast om dess äldsta skolbarn dras.

4.6 Bearbetningsfel

Som nämnts i avsnitt 3.8 kan bearbetningsfelen mätas och kontrolleras med metoder som är analoga med dem som används i samband med svars-

felen. Därutöver har man för kontroll av bearbetningsfelen tillämpat metoder som utvecklats inom den industriella kvalitetskontrollen. Med dessa metoder har man kunnat kontrollera dels enskilda kodares och stansares arbetsinsatser, dels hela partier ("lots") av material med syfte att snabbt kunna sätta in åtgärder som att arbetet (delvis) görs om eller att personal byts ut om kvaliten på databearbetningen är för låg. Denna typ av kontroller har ingen motsvarighet i datainsamlingen beroende dels på att fältarbetstiden normalt är för kort, dels på det i regel stora antalet intervjuare som är inblandade i en undersökning.

I samband med stickprovsteorins utveckling övergick man successivt från 100-procentig kontroll till stickprovskontroll. I USA infördes sådan kontroll först i samband med stansningen i 1940 års folk- och bostadsräkning (se Deming och Geoffrey, 1941 och Deming et al, 1942). I Hansen et al (1953) vol I, kap. 12 beskrivs kvalitetskontrollarbetet i 1950 års folk- och bostadsräkning. Ett typiskt kontrollschema för stansning var då uppbyggt så att man inledningsvis kontrollstansade 100 % av hålkorten för en stansare. Om dennes felprocent (ev efter en tid) var tillräckligt låg enligt något specificerat kriterium övergick man för denna person till urvalskontroll. Om stansaren inte höll måttet omplacerades denne, eventuellt till kontrollstansning. Det visade sig nämligen vara effektivast att använda de bästa stansarna i den ordinarie stansningen och de mindre kvalificerade i kontrollstansningen.

Kontrollen av kodningen i 1950 års folk- och bostadsräkning i USA gick till på följande sätt. Blanketterna kodades först av en "reguljär" kodare. Därefter kontrollkodades ett urval av blanketterna på så sätt att kontrollkodaren granskade den reguljäre kodarens kodsättning. De sålunda insamlade observationerna rörande kodningens kvalitet låg till grund för vissa kontrollåtgärder som t ex omkodning av dåligt kodade blanketter. Denna form av kontroll har kallats beroende kontroll till skillnad från senare utvecklade kontrollscheman som bygger på oberoende kodningar och jämförelser mellan dessa.

Effektiviteten i kontrollmetoder för databearbetning kan mätas genom att man planterar in fel i det material som skall bearbetas och noterar hur många av dessa fel som upptäcks. Vid studium av effektiviteten av beroende kontrollkodning i USA fann man att endast ungefär hälften av felen upptäcktes vilket ledde till att man utvecklade andra metoder. Metoden med felplantering användes också på andra håll t ex vid Indian Statistical Institute för att utvärdera kontrollstansningens effektivitet.

5 SURVEYMODELLER

5.1 Allmänt

Med surveymodell avser vi en teori för det totala felet i en survey. Denna teori beaktar fler än en felkälla och anvisar mätmetoder för både det totala felet och de specificerade komponenterna. En surveymodell skall innehålla

. Postulat;

. Definition av ingående felkomponenter;

. Scheman för skattning av felkomponenterna och av det totala felet.

Vi kan urskilja olika användningsområden för en surveymodell:

i) En surveymodell ger möjlighet till en integrerad behandling av olika felkällor, vilket leder till att det totala felet kan skattas.

ii) Surveymodeller kan användas till att skatta de enskilda felkällornas inbördes storlek, vilket ger möjlighet till en effektiv fördelning av resurser vid behandling av olika felkällor.

iii) Surveymodeller kan tillämpas på enskilda felkällor för studium av hur felet fördelar sig på olika felkomponenter. Exempelvis kan man ansätta en viss surveymodell på svarsfelet och mäta det totala svarsfelet samt de ingående komponenterna svarsbias och svarsvarians.

Surveymodeller kan indelas efter de feltyper de är ämnade att behandla, t ex surveymodeller för svarsbias, surveymodeller för "measurement variability" etc. De kan också indelas i linjära och kvadratiska, beroende på hur sambandet mellan felkomponenterna presenteras. Slutligen kan de indelas i generella och icke generella modeller. De förra avser då modeller som är allmänt tillämpliga i surveysituationer medan de senare avser modeller som är framtagna för en enskild undersöknings speciella mätsituation.

Vi skall i detta avsnitt studera de surveymodeller som presenterades före 1960. Stocks och Hochstims redan beskrivna modell för beräkning av det totala felet med hänsyn till både stickprovsvarians och intervjuarvarians, kan med ovanstående definition betraktas som en surveymodell. Ungefär samtidigt med denna presenterades två andra surveymodeller av Hansen et al (1951) och Sukhame och Seth (1952) vilka behandlar svarsfel, intervjuarfel och stickprovsfel. Senare presenterade Kish och Lansing (1954) en modell för effekten av svarsbias. Dessa tre modeller behandlas i de följande avsnitten. Därutöver kommer vi att i korthet beröra behandlingen av surveymodeller i läroböckerna av Cochran (1953) och Sukhatme (1954).

5.2 Surveymodellen enligt Hansen, Hurwitz, Marks och Mauldin

5.2.1 Postulat

Vi har i avsnitt 3.5 och 4.3 beskrivit den begreppsapparat som ligger till grund för denna surveymodell. Vi repeterar först i korthet:

"True Value"

För varje individ i populationen finns ett "sant värde", som är oberoende av de aktuella undersökningsprocedurerna.

"Individual response"

Ett "individuellt svar" är ett värde som erhålles vid en given intervju med en given intervjuare, en given uppgiftslämnare och vid en given tidpunkt. Det individuella svaret betraktas som valt med en viss sannolikhet från en population av möjliga svar och antas därför vara ett utfall av en stokastisk variabel.

"Individual response error"

Skillnaden mellan ett "individuellt svar" och det "sanna värdet" för denna individ.

"Individual response bias"

Förväntade värdet av det "individuella svarsfelet". Kring detta väntevärde finns alltså en slumpmässig variation.

"Response bias" och "response variance" för en skattning.

Om parametern som skall skattas är t ex ett medelvärde av de "sanna" värdena för individerna i populationen definieras "svarsbiasen" för skattningen av medelvärdet som skillnaden mellan förväntade värdet av medelvärdet av de observerade individuella svaren och medelvärdet av de "sanna" värdena i populationen.

Skattningen kommer också att ha en svarsvarians kring det förväntade värdet av medelvärdet av de observerade individuella svaren.

5.2.2 Design

Den matematiska modellen förutsätter att vi har en population av N individer och en population av K intervjuare. Vidare antages att korrelationen mellan svar från två olika uppgiftslämnare till två olika intervjuare är noll, vilket innebär att man bortser från den eventuella korrelation som kan bero på att intervjuarna har samma supervisor eller på att de genomgått utbildning tillsammans. Svar från uppgiftslämnare som intervjuats av samma intervjuare antages däremot kunna vara korrelerade.

Eftersom intervjuare av praktiska och ekonomiska skäl vanligen endast kan intervjua en viss, geografiskt avgränsad, del av populationen tillåter modellen att de delas in i L st grupper med K_A intervjuare i den A:te gruppen vilka är tillgängliga för intervju av N_A individer ur individpopulationen (och inga andra). För den speciella situation där alla intervjuare kan intervjua alla individer är $L = 1$, $K_A = K$ och $N_A = N$.

Urval av intervjuare och individer och fördelning av individer på intervjuare sker nu på följande sätt:

1) n av de N individerna väljs ut slumpmässigt utan restriktioner. Obundet slumpmässigt urval förutsättes i modellen men författarna påpekar att resultaten kan utvidgas till klusterurval genom att varje kluster behandlas som en individ.

2) k_A intervjuare väljs ut slumpmässigt, utan restriktioner, från A:te gruppen för att intervjua de n_A individer som är tillgängliga för att intervjuas av denna intervjuargrupp. Det totala antalet utvalda intervjuare är

$$k = \sum_{A=1}^{L} k_A$$

3) Samma antal individer, $\bar{n}$, tilldelas var och en av de k_A intervjuarna. Varje sådan grupp av $\bar{n}$ individer är ett slumpmässigt delurval från hela den del av individurvalet (omfattande n_A individer) som är tillgängliga för att intervjuas av intervjuargrupp A.

Det kan noteras speciellt att n_A är en stokastisk variabel och att $\bar{n}$ bestäms i förväg, vilket betyder att antalet intervjuare varieras efter hur stort n_A visar sig bli. Hansen et al påpekar dock att man skulle få mycket likartade variansskattningar om man i stället fixerade k_A och lät $\bar{n}$ variera, varför den gällande restriktionen inte innebär någon allvarlig inskränkning i modellens användbarhet.

5.2.3 Den totala variansen för stickprovsmedelvärdet

Om y_{Abc} = mätvärdet för uppgiftslämnare c och intervjuare b i grupp A så kan stickprovsmedelvärdet skrivas

$$\bar{y} = \frac{\sum_{A}^{L} \sum_{b}^{k_A} \sum_{c}^{\bar{n}} y_{Abc}}{n} .$$

Med den beskrivna undersökningsplanen kan $\bar{y}$ användas som skattning av populationsmedelvärdet $\bar{X}$ med medelkvadratfelet

$$MSE(\bar{y}) = B_Y^2 + \sigma_{\bar{y}}^2 \quad ,$$

där $B_Y = E\bar{y} - \bar{X}$,

$E\bar{y}$ = väntevärdet av $\bar{y}$

och

$$\sigma_{\bar{y}}^2 = \frac{\sigma_Y^2 - \sigma_{YI}}{n} + \frac{\sigma_{YI}}{k} = \frac{\sigma_Y^2}{n} + \frac{(n-k)}{n}\,\frac{\sigma_{YI}}{k} \quad . \qquad (5.2.1)$$

σ_Y^2 representerar här den "totala variansen" för alla individuella svar kring medelvärdet av alla individuella svar i populationen, som man skulle få om alla intervjuare intervjuade samtliga individer i populationen. Varje möjligt svar från varje individ till varje intervjuare är då viktat med sannolikheten för detta svar.

σ_{YI} representerar kovariansen mellan svar som är erhållna från olika individer till samma intervjuare. Denna kovarians är beräknad inom intervjuargrupper eftersom oberoende urval av intervjuare dragits från varje intervjuargrupp.

Hansen et al påpekar urvalsplanens likhet med klusterurval:

Låt σ_{wY}^2 beteckna variansen mellan svar inom intervjuargrupper, beräknad på alla tänkbara svar från alla (N_A) individer till alla (K_A) intervjuare i gruppen,

σ_{bY}^2 beteckna variansen mellan förväntade svar (medelvärden) från intervjuargrupperna, och

ρ beteckna intraklasskorrelationen, dvs korrelationen mellan svar från olika individer till samma intervjuare.

Då gäller följande samband

$$\sigma_Y^2 = \sigma_{wY}^2 + \sigma_{bY}^2$$

och

$$\rho = \frac{\sigma_{YI}}{\sigma_{wY}^2}$$

Man kan betrakta urvalsplanen som en där kluster av svar har dragits. Ett kluster består då av alla svar en intervjuare skulle ha kunnat få om denne intervjuat samtliga N_A individer i den del av individpopulationen som är tillgänglig för intervju.

Inom de dragna klustren dras därefter urval av svar på så sätt att högst ett svar kommer från varje individ. Denna restriktion skiljer urvalsplanen från strikt klusterurval, men de grundläggande principerna är analoga.

Likheten med klusterurval blir tydligare med skrivsättet

$$\sigma_{\bar{y}}^2 = \frac{\sigma_{wY}^2}{n} [1+ \rho(\bar{n} - 1)] + \frac{\sigma_{bY}^2}{n}$$

där den sista termen representerar variansen som beror på att individer dragits oberoende av intervjuargrupp. Med endast en intervjuargrupp (L = 1) blir denna term = 0 och $\sigma_{wY}^2 = \sigma_Y^2$ så att

$$\sigma_{\bar{y}}^2 = \frac{\sigma_Y^2}{n} [1 + (\bar{n} - 1)\rho].$$

Denna formel är identisk med formeln för variansen för ett stickprovsmedelvärde när man drar k kluster om vardera $\bar{n}$ enheter eller drar k kluster av lika storlek och inom varje sådant ett delurval om $\bar{n}$ enheter.

Hansen et al presenterar också väntevärdesriktiga skattningar av σ_{YI} och σ_Y^2. Dessa är, respektive:

$$S_{YI} = \frac{\sum_{A=1}^{L} \frac{k_A}{k_A-1} \sum_{b=1}^{k_A} (\bar{y}_{Ab}-\bar{y}_A)^2}{k} - \frac{\sum_{A=1}^{L} \sum_{b=1}^{k_A} \sum_{c=1}^{\bar{n}} (y_{Abc}-\bar{y}_{Ab})^2}{n(\bar{n}-1)} \qquad (5.2.2)$$

och

$$S_Y^2 = \frac{\sum_{A=1}^{L} \sum_{b=1}^{k_A} \sum_{c=1}^{\bar{n}} (y_{Abc}-\bar{y})^2}{n-1} + \frac{(n-k)}{(n-1)} \frac{S_{YI}}{k}, \qquad (5.2.3)$$

där

$$\bar{y}_{Ab} = \frac{\sum_{c=1}^{\bar{n}} y_{Abc}}{\bar{n}}$$ = medelvärde för intervjuare b i intervjuarurvalet från grupp A

och

$$\bar{y}_A = \frac{\sum_{b=1}^{k_A} \sum_{c=1}^{\bar{n}} y_{Abc}}{k_A \bar{n}}$$ = stickprovsmedelvärde för grupp A

Med ovanstående skattningar erhålles även en väntevärdesriktig skattning, $S^2_{\bar{y}}$, av $\sigma^2_{\bar{y}}$:

$$S^2_{\bar{y}} = \frac{\sum_{A}^{L} \sum_{b}^{k_A} \sum_{c}^{\bar{n}} (y_{Abc}-\bar{y})^2}{n(n-1)} + \frac{(n-k)}{(n-1)} \frac{S_{YI}}{k} . \qquad (5.2.4)$$

Den första termen i (5.2.4) är den vanliga formeln för skattning av variansen för ett stickprovsmedelvärde vid OSU om n enheter från en oändlig population. Den andra termen representerar approximativt den variansökning som beror på inom-intervjuarkorrelationen.

Man kan med följande testvariabel testa om signifikant mellanintervjuarvarians finns i ett material:

$$\frac{\left[\sum_{A=1}^{L} \frac{k_A}{k_A-1} \sum_{b=1}^{k_A} (\bar{y}_{Ab}-\bar{y}_A)^2\right]/k}{\left[\sum_{A=1}^{L} \sum_{b=1}^{k_A} \sum_{c=1}^{\bar{n}} (y_{Abc}-\bar{y}_{Ab})^2\right]/(n-k)}, \qquad (5.2.5)$$

som är F-fördelad med (k,n-k) frihetsgrader under antagandet att urvalet är draget från en normalfördelad stokastisk variabel. Kvadratsumman i täljaren kan sägas vara ett viktat medelvärde av medelkvadrater mellan intervjuare inom intervjuargrupper och kvadratsumman i nämnaren är medelkvadraten mellan uppgiftslämnare inom intervjuare.

5.2.4 Beräkning av optimalt antal intervjuare

Om man antar att intervjuarvariabiliteten är oberoende av antalet intervjuare som används, minskar dess effekt på skattningarna om man ökar antalet intervjuare. Fler intervjuare innebär dock ökade kostnader för utbildning och resor vilket kan leda till att man, med en fast totalkostnad för undersökningen, måste spara genom att minska urvalet. Hansen et al har därför beräknat optimala värden på stickprovsstorleken n och antalet intervjuare k för att få minimal varians till given kostnad. Låt oss anta att kostnaden är

$$C = n\,C_R + k\,C_I,$$

där C = total budget för fältarbetet
C_R = kostnaden per uppgiftslämnare
C_I = kostnaden per intervjuare

C_R och C_I antages oberoende av storleken på n och k.

De optimala värdena på n och k kan nu beräknas till

$$n_{opt} = A\sqrt{(\sigma_Y^2 - \sigma_{YI})/C_R} \quad \text{och} \quad k_{opt} = A\sqrt{\sigma_{YI}/C_I} \ ,$$

där

$$A = \frac{C}{\sqrt{C_R(\sigma_Y^2 - \sigma_{YI})} + \sqrt{C_I \cdot \sigma_{YI}}} \ .$$

Av olika skäl är det emellertid svårt att i praktiken använda denna metod för att bestämma antalet intervjuare i en survey. Hyman et al (1954) anger fyra sådana skäl:

i) De flesta surveyer har fler än en viktig variabel, vilka ger olika värden på det optimala antalet intervjuare. Det valda antalet måste därmed bli en kompromiss, vilket begränsar värdet av metoden.

ii) Antalet intervjuare som används bestäms normalt av administrativa och operationella överväganden, t ex måste i regel fältarbetet vara färdigt inom en viss tid. I de exempel som Hansen et al redovisar är optimala antalet intervjuare lika med två i tre fall av fem, vilket i de flesta undersökningar torde vara otänkbart av bl a just tidsskäl.

iii) Metoden att bestämma optimalt antal intervjuare bygger på antagandet att intervjuarvariansen inte ändras om antalet intervjuare varierar. Detta antagande kan ifrågasättas.

iv) I undersökningar som är icke-återkommande är det svårt att utnyttja metoden eftersom man då inte har någon grund för att bedöma storleken på kostnader och varianskomponenter i förväg. Att ta reda på detta med provundersökningar är dyrt.

5.2.5 En metod att minimera effekten av både bias och varians

Hansen et al anger en metod att, med utgångspunkt från modellen, samtidigt reducera både en skattnings totala bias och totala varians. En förutsättning är att den ordinarie mätmetoden ger en bias och att

det finns en bättre metod som, vanligen till högre kostnad, ger en betydligt mindre bias. Genom att återintervjua ett delurval av det ordinarie urvalet med den dyrare metoden kan man använda en kvotskattning som ger mindre bias och varians än den ovan redovisade medelvärdesskattningen $\bar{y}$.

Antag att delurvalet omfattar n' individer som är lika fördelade på ett urval av k' intervjuare. Dessa k' intervjuare kan ha dragits från de ursprungliga K intervjuarna. Emellertid är det ur biasreducerande synpunkt lämpligare att dra dem ur en population av bättre kvalificerade intervjuare.

Låt y_{Abc} = svaren vid originalintervjun,

$z_{A'b'c}$ = svaren vid återintervjun, som anses avgivna under andra "essential conditions" och behäftade med mindre bias än svaren från originalintervjun,

$\bar{y}$ = medelvärdet av y_{Abc}-värdena för hela urvalet om n individer,

$\bar{y}'$ = medelvärdet av y_{Abc}-värdena för delurvalet om n' individer,

$\bar{z}'$ = medelvärdet av $z_{A'b'c}$-värdena för delurvalet om n' individer.

Vi kan nu skatta populationsmedelvärdet $\bar{X}$ med kvotskattningen

$$\hat{\bar{z}} = \bar{y} \frac{\bar{z}'}{\bar{y}'} . \qquad (5.2.6)$$

$\hat{z}$ har medelkvadratfelet

$$MSE(\hat{z}) \doteq \bar{R}_z^2 + \bar{Z}^2 \left\{ \frac{u}{n} + \frac{v}{n'} + \frac{w}{k'} \right\}, \qquad (5.2.7)$$

där

$$\bar{Z} = E(\bar{z}') \doteq E(\hat{z}),$$

$$\bar{Y} = E(\bar{y}) = E(\bar{y}'),$$

$$\bar{R}_z = \bar{Z} - \bar{X},$$

$$u = \frac{2\rho_{YZ}\sigma_Y\sigma_Z}{\bar{Y}\ \bar{Z}} - \frac{\sigma_Y^2 - \sigma_{YI}}{\bar{Y}^2},$$

$$v = \frac{\sigma_Z^2 - \sigma_{ZI}}{\bar{Z}^2} - u,$$

$$w = \frac{\sigma_{ZI}}{\bar{Z}^2},$$

och

ρ_{YZ} = korrelationen mellan de förväntade y- och z-värdena för samma individ.

Formel (5.2.7) är en approximation där moment av tredje och högre ordningar har uteslutits. Approximationen kan vara dålig om värdena på k, n, k' eller n' är små.

$MSE(\hat{\bar{z}})$ är mindre än $MSE(\bar{y})$, eftersom återintervjusvaren antas vara behäftade med betydligt mindre svarsbias än originalintervjusvaren, vilket uppväger den tekniska bias som finns i kvotskattningen.

Hansen et al anger följande skattningsformler för $\rho_{YZ}\rho_Y\rho_Z$ och $\bar{R}_z$:

$$r_{YZ}s_Y s_Z = \frac{\sum_{c=1}^{n'} y_{Abc} z_{A'b'c} - n'\bar{y}'\bar{z}'}{n'-1}$$

respektive

$$r'_y = \bar{y}-\bar{z}' \text{ (eller } \hat{r} = \bar{y}-\hat{\bar{z}})$$

5.2.6 Sammanfattning och diskussion

Sammanfattningsvis kan sägas att surveymodellen enligt Hansen, Hurwitz, Marks och Mauldin är en matematisk modell för svarsfel under vilken man kan skatta den totala variansen för en medelvärdesskattning. Den totala variansen delas upp i två komponenter som vardera kan skattas för sig. Den ena speglar stickprovsvariansen och den s k enkla svarsvariansen (den del av svarsvariansen som inte beror på att intervjusvar till samma intervjuare är korrelerade) medan den andra speglar den korrelerade svarsvariansen (som alltså är det varianstillskott som beror på inomintervjuarkorrelationen). Den enkla svarsvariansen kan däremot inte skattas separat under modellen, eftersom detta kräver återintervjuer.

En viktig faktor när det gäller modellens praktiska användbarhet är att intervjuarna tillåts vara indelade i grupper. Eftersom det behövs många intervjuare för att få god precision i skattningar och tillräckligt många frihetsgrader i test, och eftersom många intervjuare normalt endast behövs i landsomfattande undersökningar skulle reskostnaderna bli orimligt höga om man tvingades sprida varje intervjuares tilldelning över hela undersökningsområdet. En faktor som å andra sidan kan försvåra användandet av modellen i praktiken är att alla intervjuare

skall ha lika stor tilldelning av intervjupersoner. Tilldelningen måste alltid göras med hänsyn till kostnads- och tidsvillkor, vilket gör det svårt att uppfylla dessa krav.

Modellen utvecklades som ett led i de strävanden vid US Bureau of the Census som inleddes omkring 1950, nämligen att utveckla en teori för mätning, tolkning och kontroll av svarsfel i surveyer. Syftet var att förbättra kvalitetsredovisningen i olika undersökningar och att förbättra evalveringsmöjligheterna av dessa inför planeringen av nya undersökningar. Ett exempel på hur modellen använts för det senare syftet redovisas i Hanson och Marks (1958). De genomförde en omfattande s k Enumerator Variance Study i anslutning till 1950 års folkräkning i USA där de för ett stort antal frågor av olika slag testade (med F-test) om intervjuarvariansen var signifikant. Redovisningar av surveyer eller experiment där man tillämpat modellen för att skatta felkomponenternas storlek saknas emellertid under denna tidiga modellutvecklingsperiod. Den begreppsapparat som definierades i Hansens et al artikel har i alla händelser haft stort inflytande på senare arbeten inom surveyfelområdet.

5.3 Surveymodellen enligt Sukhatme och Seth

Ungefär samtidigt som Hansen, Hurwitz, Marks och Mauldin utvecklade sin modell utvecklades en surveymodell vid Indian Council of Agricultural Research, presenterad i Sukhatme och Seth (1952). Syftet med detta arbete var att föreslå mätmetoder för de olika komponenterna i nonsamplingfelet som var baserade på en modell som var tillräckligt generell för att täcka de situationer man vanligen mötte i lantbrukssurveyer och socio-ekonomiska surveyer. Modellen är mycket generell och tillåter skattning av olika varianskomponenter som svarsvarians och intervjuarvarians, förutom den totala variansen, för ett antal olika situationer med fixa eller slumpmässigt valda intervjuare och med en eller flera observationer på samma enhet. Vi skall återge modellen och presentera skattningar av varianskomponenter för i första hand situationer som liknar dem som Hansen et al behandlade.

Sukhatme och Seth utgår från en poplation som är ändlig eller oändlig med medelvärde μ och varians σ^2. Från denna dras ett slumpmässigt urval omfattande ℓ enheter för vilka mätningar görs. Med

$$Y_{ijk} \quad i=1, 2,\ldots, \ell \; ; \; j=1,\ldots, m; \quad k=1,\ldots, n_{ij}$$

betecknas det mätvärde som av den j:te intervjuaren (observatören) erhålles för en variabel på den i:te enheten i stickprovet och vid den k:te mätningen. Antalet intervjuare som medverkar i undersökningen är m och det antages att intervjuare j gör n_{ij} olika mätningar på enhet i.

y_{ijk} kan delas upp i fyra komponenter enligt följande

$$y_{ijk} = x_i + \alpha_j + \delta_{ij} + \varepsilon_{ijk}, \qquad (5.3.1)$$

där

x_i betecknar det "sanna" värdet på den i:te enheten (Sukhatme och Seth använder uttrycket "intrinsic value" men diskuterar inte närmare vad detta innebär),

α_j är den j:te intervjuarens bias (genomsnittliga avvikelse från x_i-värdena) vid upprepade observationer på alla enheter i populationen,

$\alpha_j + \delta_{ij}$ är den j:te intervjuarens bias vid upprepade observationer på den i:te enheten (δ_{ij} representerar samspelseffekten mellan den j:te intervjuaren och den i:te enheten),

$\alpha_j + \delta_{ij} + \varepsilon_{ijk}$ representerar den totala avvikelsen från x_i när intervjuare j mäter enhet i vid tillfälle k.

ε_{ijk} och δ_{ij} betraktas som stokastiska variabler med de betingade väntevärdena

$$E(\varepsilon_{ijk} \mid ij) = 0$$

och

$$E(\delta_{ij} \mid j) = 0 \ .$$

Vi kan redan här notera ett par skillnader mellan denna surveymodell och den vi redovisat ovan av Hansen et al. De senare diskuterar ej explicit ett linjärt samband mellan felkomponenterna som Sukhatme och Seth gör utan baserar sina beräkningar av den totala variansen på principerna för klusterurval. (Det går dock att formulera ett linjärt samband mellan felkomponeterna i Hansens et al modell som leder fram till deras resultat). Vi ser också att Sukhatmes och Seths surveymodell tillåter återintervjuer varför man i denna modell kan skatta en variansterm (längre fram definierad och betecknad med σ_ε^2) som uppenbarligen är densamma som den enkla svarsvariansen.

För att skatta felkomponenterna i Sukhatmes och Seths modell betraktas först situationen när undersökningspopulationen består av endast ett stratum. Låt

$y_{ij\cdot}$ vara medelvärdet av de n_{ij} observationerna som intervjuare j gjort på enhet i;

$y_{\cdot j\cdot}$ medelvärdet av alla observationer som gjorts av intervjuare j ($n_{\cdot j}$ stycken där $n_{\cdot j} = \sum_{j=1}^{\ell} n_{ij}$);

$y_{i\cdot\cdot}$ medelvärdet av alla observationer som gjorts på enhet i ($n_{i\cdot}$ stycken där $n_{i\cdot} = \sum_{j=1}^{m} n_{ij}$);

$y_{\cdot\cdot\cdot}$ medelvärdet av alla observationer som gjorts på alla enheter i stickprovet (n stycken där $n = \sum_{j=1}^{m} \sum_{i=1}^{\ell} n_{ij}$);

Med en analog notation för ε_{ijk}:na kan vi skriva

$$y_{ij\cdot} = x_i + \alpha_j + \delta_{ij} + \varepsilon_{ij\cdot} \ ,$$

$$y_{\cdot j\cdot} = \frac{\sum_{i=1}^{\ell} x_i n_{ij}}{n_{\cdot j}} + \alpha_j + \frac{\sum_{i=1}^{\ell} \delta_{ij} n_{ij}}{n_{\cdot j}} + \varepsilon_{\cdot j\cdot} \ ,$$

$$y_{i\cdot\cdot} = x_i + \frac{\sum_{j=1}^{m} \alpha_j n_{ij}}{n_{i\cdot}} + \frac{\sum_{j=1}^{m} \delta_{ij} n_{ij}}{n_{i\cdot}} + \varepsilon_{i\cdot\cdot} \qquad (5.3.2)$$

och

$$y_{\cdots} = \frac{\sum_{i=1}^{\ell} x_i n_{i\cdot}}{n} + \frac{\sum_{j=1}^{m} \alpha_j n_{\cdot j}}{n} + \frac{\sum_{i=1}^{\ell} \sum_{j=1}^{m} \delta_{ij} n_{ij}}{n} + \varepsilon_{\cdots} \ .$$

Om vi nu antar att de m intervjuarna dragits slumpmässigt ur en population av M intervjuare, får vi följande betingade väntevärden och varianser för variablerna ovan:

$$E(y_{ij\cdot}|j) = \mu + \alpha_j,$$

$$E(Y_{\cdot j\cdot}|j) = \mu + \alpha_j, \qquad (5.3.3)$$

$$E(y_{i\cdot\cdot}|m) = \mu + \alpha \ ,$$

$$E(y_{\cdots}|m) = \mu + \alpha \ , \quad \text{där} \quad \alpha = \sum_{j=1}^{M} \alpha_j / M$$

och (om man bortser från termen $-\frac{\sigma^2}{N}$ om undersökningspopulationen är ändlig och består av N enheter)

$$V(y_{ij\cdot}|j) = \sigma^2 + \sigma_\alpha^2 (1 - \frac{1}{M}) + \sigma_{\delta j}^2 + \frac{\sigma_\varepsilon^2}{n_{ij}},$$

$$V(y_{\cdot j\cdot}|j) = \sigma^2 \sum_{i=1}^{\ell} n_{ij}^2/n_{\cdot j}^2 + \sigma_\alpha^2(1 - \frac{1}{M}) + \sum_{i=1}^{\ell} n_{ij}^2 \sigma_{\delta j}^2/n_{\cdot j}^2 + \frac{\sigma_\varepsilon^2}{n_{\cdot j}},$$

(5.3.4)

$$V(y_{i\cdot\cdot}|m) = \sigma^2 + \{ \frac{\sum_{j=1}^{m} n_{ij}^2}{n_{i\cdot}^2} - \frac{1}{M} \} \sigma_\alpha^2 + \frac{\sum_{j=1}^{m} n_{ij}^2 \sigma_{\delta j}^2}{n_{i\cdot}^2} + \frac{\sigma_\varepsilon^2}{n_{i\cdot}},$$

$$V(y_{\cdot\cdot\cdot}|m) = \frac{\sum_{i=1}^{\ell} n_{i\cdot}^2}{n^2} \sigma^2 + \{ \frac{\sum_{j=1}^{m} n_{\cdot j}^2}{n^2} - \frac{1}{M} \} \sigma_\alpha^2 + \frac{\sum_{i=1}^{\ell} \sum_{j=1}^{m} n_{ij} \sigma_{\delta j}^2}{n^2} + \frac{\sigma_\varepsilon^2}{n},$$

där $\sigma^2 = E(x_i - \mu)^2 = \sum_{i=1}^{N} (x_i - \mu)^2/(N-1)$,

$$\sigma_\alpha^2 = \sum_{j=1}^{M} (\alpha_j - \alpha)^2/(M-1), \qquad (5.3.5)$$

$$\sigma_{\delta j}^2 = E[\{\delta_{ij} - E(\delta_{ij}|j)\}^2|j],$$

och

$$\sigma_\varepsilon^2 = E[\{\varepsilon_{ijk} - E(\varepsilon_{ijk}|i,j)\}^2|i\ j].$$

Sukhatme och Seth anger följande kvadratsummor för skattning av varianskomponenterna:

Inom celler (en "cell" bestäms av en intervjuare j och en stickprovsenhet i):

$$S_{\omega}^2 = \frac{1}{n-c} \sum_i \sum_j \sum_k (y_{ijk} - y_{ij\cdot})^2 ,$$

där c = antal n_{ij} (i=1,..., ℓ; j=1,..., m) som är skilda från noll;

Mellan celler:

$$S_c^2 = \frac{1}{c-1} \sum_i \sum_j n_{ij} (y_{ij\cdot} - y_{\cdots})^2;$$

Mellan intervjuare:

$$S_e^2 = \frac{1}{m-1} \sum_j n_{\cdot j} (y_{\cdot j\cdot} - y_{\cdots})^2 ;$$

Mellan stickprovsenheter:

$$S_o^2 = \frac{1}{\ell-1} \sum_i n_{i\cdot} (y_{i\cdot\cdot} - y_{\cdots})^2;$$

Mellan stickprovsenheter inom intervjuare j:

$$S_{jo}^2 = \frac{1}{c_j-1} \sum_i n_{ij}(y_{ij\cdot} - y_{\cdot j\cdot})^2 ;$$

Mellan stickprovsenheter inom intervjuare:

$$S_{co}^2 = \frac{1}{c-m} \sum_i \sum_j n_{ij}(y_{ij\cdot} - y_{\cdot j\cdot})^2;$$

Genom att lägga olika restriktioner på n_{ij}:na kan modellen anpassas till realistiska surveysituationer. Sukhatme och Seth diskuterar fyra sådana restriktioner:

Fall 1: $n_{i\cdot} = 1$ och $n_{\cdot j} = \frac{n}{m}$ för alla i och j, vilket innebär att en enhet observeras endast en gång och tillordnas en intervjuare slumpmässigt på så sätt att alla intervjuare får lika stor tilldelning $(\frac{n}{m})$. Detta motsvarar den situation som Hansen, Hurwitz, Marks och Mauldin (1951) behandlar för fallet med L = 1, dvs med endast en intervjuargrupp.

Sukhatme och Seth beräknar följande variansanalystabell (deras artikel är behäftad med ett antal tryckfel vad beträffar medelkvadratsummorna som vi här har korrigerat):

	Frihetsgrader	Medelkvadratsumma (S^2)	$E(S^2)$
Mellan intervjuare	m-1	S_e^2	$\sigma^2 + \frac{n}{m}\sigma_\alpha^2 + \bar{\sigma}_\delta^2 + \delta_\varepsilon^2$
Inom intervjuare	n-m	S_{co}^2	$\sigma^2 + \bar{\sigma}_\delta^2 + \sigma_\varepsilon^2$
Total	n-1	S_c^2	$\sigma^2 + \frac{n(m-1)}{m(\ell-1)}\sigma_\alpha^2 + \bar{\sigma}_\delta^2 + \sigma_\varepsilon^2$

där $\sigma_\alpha^2 = \dfrac{\sum_{j=1}^{M} (\alpha_j - \frac{\sum \alpha_j}{M})^2}{M-1}$ och $\bar{\sigma}_\delta^2 = \dfrac{\sum_{j=1}^{M} \sigma_{\delta j}^2}{M}$.

(Om alla $\sigma_{\delta j}^2$ är lika används beteckningen $\sigma_{\delta j}^2 = \sigma_\delta^2 \; (= \bar{\sigma}_\delta^2)$)

I detta fall kan inte varianskomponenterna σ^2, σ_δ^2 och σ_ε^2 skattas var för sig, men summan av dem skattas väntevärdesriktigt med S_{co}^2. Komponenten σ_α^2 skattas väntevärdesriktigt med $\frac{m}{n}(S_e^2 - S_{co}^2)$. Stickprovsvariansen för $y_{...}$ skattas med S_e^2/n.

Enligt Sukhatme och Seth täcks den situation som behandlas av Hansen, Hurwitz, Marks och Mauldin (1951) av "the general formulae in our paper when σ_{YI} is replaced by σ_α^2 and $\sigma_Y^2 - \sigma_{YI}$ by $\sigma^2 + \sigma_\delta^2 + \sigma_\varepsilon^2$ the interviewers being selected at random out of a given pool of M". Detta innebär att "fall I" ovan, med $\sigma_{\delta j}^2 = \sigma_\delta^2$, j=1,..., m, är detsamma som fallet i Hansen et al med antal intervjuargrupper, L, lika med 1. Vi illustrerar detta genom att jämföra den totala variansen för stickprovsmedelvärdet enligt de båda modellerna:

I Sukhatmes och Seths modell fås variansen för ett stickprovsmedelvärde ur formel (5.3.4), eftersom

$$V(y_{...}) = E_I(V_R(y_{...}|m)) + V_I(E_R(y_{...}|m)).$$

(Index I avser här intervjuardimensionen och index R respondentdimensionen). Den andra termen i högra ledet är = 0, eftersom $E(y_{...}|m)$ enligt (5.3.3) är en konstant.

Enligt (5.3.4) är den betingade variansen för $y_{...}$, under förutsättningarna i fall I, dvs $n_{i\cdot} = 1$, $n_{\cdot j} = \frac{n}{m}$ och $\sum_i \sum_j n_{ij} = n$, samt antagandena att $\sigma_{\delta j}^2 = \sigma_\delta^2$ för alla j och att σ_δ^2/M är försumbart litet, lika med

$$V_R(y_{...}|m) = \frac{\sigma^2 + \sigma_\delta^2 + \sigma_\varepsilon^2}{n} + \frac{\sigma_\alpha^2}{m} \qquad (5.3.6)$$

Men under dessa antaganden gäller att $V(y_{...}|m)$ är konstant i intervjuardimensionen, varför

$$V(y_{...}) = E_I(V_R(y_{...}|m)) = V_R(y_{...}|m),$$

dvs (5.3.6) anger den totala variansen för ett stickprovsmedelvärde under Sukhatmes och Seths modell, fall I. Detta är samma formel som (5.2.1) med de skillnader i beteckningar som anges i citatet ovan. (Hansen et al betecknar antalet intervjuare med k i stället för m.)

Övriga fall med restriktioner på n_{ij}:na som Sukhatme och Seth behandlar är

Fall II: $n_{i.} = p$ och $n_{ij} = 0$ eller p, dvs en enhet observeras p gånger av samma intervjuare, eller inte alls. Samma antal enheter tillordnas slumpmässigt varje intervjuare. Fördelen med denna design jämfört med fall I är att σ_ε^2 kan skattas. Däremot kan inte heller här σ^2 och $\bar{\sigma}_\delta^2$ skattas separat.

Fall III: $n_{i.} = p$ och $n_{ij} = 0$ eller 1, dvs en enhet observeras p gånger som i fall II men av p olika intervjuare. Enheterna fördelas på intervjuarna slumpmässigt på så sätt att varje intervjuare obseverar samma antal enheter. Antalet gånger som två intervjuare observerar samma enhet är dessutom lika för alla par av intervjuare. I detta fall kan σ^2 och σ_α^2 skattas separat men inte $\bar{\sigma}_\delta^2$ och σ_ε^2.

Fall IV: $n_{i.} = 1$ eller 2 och $n_{ij} = 0$, 1 eller 2. Här observeras en del enheter en gång och övriga två gånger. I det senare fallet kan observationerna göras av en intervjuare eller av två. Observationerna tilldelas intervjuarna slumpmässigt under bivillkoren:

- antal enheter som observeras endast en gång är samma för varje intervjuare;

- antal enheter som observeras två gånger av samma intervjuare är samma för alla intervjuare;

- antal enheter som observeras två gånger av olika intervjuare är samma för varje par av intervjuare.

Med denna design kan alla varianskomponenterna σ^2, σ_α^2, σ_δ^2 och σ_ε^2 skattas separat.

Fall I - IV ovan utreds också för den situation där man har fixa intervjuare, dvs där intervjuarna inte betraktas som slumpmässigt dragna ur en intervjuarpopulation. Författarna anger detta synsätt som lämpligt när intervjuarna - vilket var vanligt i Indien - kommer från olika organisationer eller från "different schools using different methods of enumeration".

Sukhatme och Seth betraktar också det stratifierade fallet med L strata inom vilka det dras stickprov om ℓ_s enheter, s = 1, ..., L slumpmässigt. Dessa intervjuas av m intervjuare inom varje stratum, som kan vara valda fixt eller slumpmässigt. I det senare fallet kan de antingen anses dragna från en population av M_s intervjuare i stratum s eller kan m L intervjuare ha dragits slumpmässigt från den totala populationen av M intervjuare och sedan fördelats slumpmässigt på L strata.

För alla dessa situationer anges medelkvadratsummor för skattning av varianskomponenter och designernas effektivitet jämförs på basis av medelkvadratfelet.

Det tidigare citerade påståendet av Sukhatme och Seth att deras modell omfattar den situation som Hansen et al (1951) behandlar, stämmer uppenbarligen inte för det stratifierade fallet eftersom de senare efterstratifierar undersökningspopulationen, vilket medför att antalet urvalsenheter i varje efterstratum blir stokastiskt.

Sukhatme och Seth beskriver i ett antal exempel hur olika designer i deras modell tillämpats vid undersökningar som utförts vid Indian Council of Agricultural Research under 1940-talet. I de flesta av dessa exempel skattas σ_α^2 och $\sigma^2 + \sigma_\delta^2 + \sigma_\varepsilon^2$ eftersom detta låter sig göras med endast en observation per enhet. I ett exempel har fler än en mätning gjorts på vissa enheter vilket medger separata skattningar av σ^2 och $\sigma_\delta^2 + \sigma_\varepsilon^2$ förutom av σ_α^2.

Sukhatme och Seth anför i slutet av sin artikel ett antal kritiska synpunkter på de undersökningsplaner som deras modell behandlar. Mot principen att i reguljära undersökningar slumpmässigt tillordna urvalspersoner till intervjuare så att dessa får var sitt interpenetrerande delurval eller "replicated sample", som författarna betecknar dem, anges i huvudsak två argument:

- Höga kostnader. Även om intervjuarna stratifieras så blir reskostnaderna avsevärt högre med interpenetrerande delurval än om intervjuarna tilldelats urvalsenheter som ligger geografiskt nära varandra.

- Då intervjuarna och urvalsenheterna stratifieras geografiskt, vilket normalt är nödvändigt ur kostnadssynpunkt, blir urvalen relativt små inom strata vilket gör metoden med interpenetrering ineffektiv när det gäller att upptäcka skillnader mellan intervjuare i fältarbetet. Verkliga skillnader kan befinnas vara icke-signifikanta och tvärtom.

En ytterligare nackdel med metoden med interpenetrerande delurval är att den aldrig kan avslöja fel i fältarbetet som är lika för alla intervjuare (dvs bias). För detta behövs en annan typ av kontroll av fältarbetet och författarna anför en rad argument för "supervision" och "supervisory checks" (se avsnitt 4.2.4). Om syftet med undersökningen är att skatta storleken på non-samplingfel är dock metoden med interpenetrerande urval lämplig. Den är också lämplig när syftet med undersökningen är att jämföra olika metoder eller olika grupper av undersökare, t ex i en provundersökning inför en stor survey. Den är dock enligt författarna endast effektiv om den används i förening med noggrann supervision.

I Sukhatme (1954) behandlas en surveymodell som kan ses som ett specialfall av Sukhatmes och Seths modell. I denna antages att endast en observation görs per urvalsenhet av samma intervjuare och att $\delta_{ij} = 0$, vilket ger modellen

$$y_{ij} = x_i + \alpha_j + \varepsilon_{ij}$$

Storheterna x_i och α_j är här definierade som i Sukhatme och Seth (1952) och ε_{ij} representerar avvikelsen mellan det rapporterade värdet y_{ij} och $x_i + \alpha_j$, och bestäms av tillfälliga faktorer som påverkar mätsituationen. Under denna modell presenteras skattningsformler för observationsfel för fallet med slumpmässigt valda intervjuare.

5.4 Sammanfattning av surveymodeller för svarsvariation och jämförelser mellan dem

Vi har i det föregående studerat olika modeller för beräkning av den totala variationen i medelvärdesskattningar i surveyer. Modellerna uppvisar på viktiga punkter stora likheter, t ex när det gäller scheman för skattning av varianskomponenter, men bygger på olika synsätt när det gäller hur mätresultaten genereras. I detta avsnitt skall vi jämföra dessa surveymodeller översiktligt. Vi utgår därvid från en undersökningsplan baserad på principen för interpenetrerande delurval. Varje delurval antages draget med obundet slumpmässigt urval från en population av N st individer och tillordnas slumpmässigt en intervjuare. Intervjuarna (k st) är i sin tur dragna från en population av K st intervjuare.

Respondenter indiceras med i och intervjuare (delurval) med j. Delurvalens storlek är n_j, $j=1,\ldots,k$ och den totala urvalsstorleken är n, dvs $n = \sum_{j=1}^{k} n_j$.

Den beskrivna situationen behandlas i alla de studerade modellerna. Därutöver behandlar vissa författare andra situationer, så t ex betraktar Sukhatme och Seth fallet när alla K intervjuare deltar i undersökningen och både dessa författare och Hansen et al behandlar situationen med stratifiering av intervjuarurvalet.

Stock och Hochstim (1951) utgår från synsättet att en uppgiftslämnare endast har ett möjligt svar på en fråga om denne intervjuas av en given intervjuare. Däremot kan han eller hon ha olika svar till olika intervjuare.

Om man tänker sig att varje intervjuare intervjuar alla individer i populationen kan mätresultaten, y_{ij}, sägas genereras på följande sätt:

$$y_{ij} = \mu + I_j + R_{ij} , \quad i=1,\ldots,N \quad J=1,\ldots,K$$

där μ är medelvärdet av alla N × K mätvärden;

I_j är den individuella intervjuarbiasen för intervjuare j, dvs avvikelsen mellan μ och medelvärdet för intervjuare j över dennes N mätvärden;

R_{ij} är avvikelsen mellan mätresultatet och $\mu + I_j$ när intervjuare j intervjuar respondent i.

Det antages att I_j och R_{ij} är okorrelerade för alla i och j, både inom och mellan komponenterna. Observera att denna modell inte tar hänsyn till om det finns sanna värden eller ej, utan beskriver enbart slumpmässiga mätfel. Modellen är ett exempel på en deterministisk felmodell där det stokastiska inslaget är begränsat till urvalsförfarandet (som beskrevs inledningsvis) och fördelningen av delurval på intervjuarna.

Den modell Hansen, Hurwitz, Marks och Mauldin (1951) redovisar bygger på ett annat synsätt beträffande genereringen av mätresultat, nämligen att svaren väljs slumpmässigt av uppgiftslämnaren och alltså kan variera om intervjun upprepas med samma "essential conditions", t ex med samma intervjuare. Vi markerar med index t att man tänker sig denna variation över upprepade försök. Modellen kan presenteras på följande sätt:

$$y_{ijt} = x_i + B_i + \varepsilon_{ijt}\,, \qquad i = 1,\ldots,N \quad j=1,\ldots,K$$

där x_i är det sanna värdet för enhet i;

B_i är en genomsnittlig bias för mätvärdet av enhet i över alla tänkbara mätningar av alla intervjuare. Vi kan för enkelhets skull anta att den är noll i detta sammanhang, där vi diskuterar modeller för slumpmässiga surveyfel. (Alternativt skulle vi kunna betrakta $x_i' = x_i + B_i$ som ett operationellt sant värde och studera modellen $y_{ijt} = x_i' + \varepsilon_{ijt}$).

ε_{ijt} är den slumpmässiga avvikelsen från det sanna värdet som erhålles när intervjuare j intervjuar respondent i vid tillfälle t.

Modellen förutsätter att det inte finns någon korrelation mellan det sanna värdet x_i och mätfelet ε_{ijt}. Däremot kan det finnas en korrelation, ρ, mellan olika mätfel, nämligen i de mätningar som görs av samma intervjuare, dvs

$$\mathrm{Cov}(y_{ijt}, y_{i'j't}) = \begin{cases} 0 & \text{om } j \neq j' \\ \mathrm{Var}(y) & \text{om } i=i',\ j=j' \\ \rho\mathrm{Var}(y) & \text{om } i \neq i',\ j=j' \end{cases}$$

Denna korrelation speglar samma typ av surveyfel som komponenten I_j i Stocks och Hochstims modell ovan. Vi kan också konstatera att den variation som härrör från ε_{ijt} innehåller såväl den enkla som den korrelerade svarsvariansen.

Cochran (1953) diskuterar i ett kapitel olika sätt att bygga upp linjära surveymodeller för svarsvariation. Han utgår därvid från det för-

enklade antagandet att individ- och intervjuarpopulationerna är oändliga och beräknar variansen för en medelvärdesskattning under olika modeller. En av dessa är mycket snarlik US Bureau of the Census' modell ovan och kan beskrivas linjärt på samma sätt:

$$y_{ijt} = x_i + \varepsilon_{ijt},$$

där ε_{ijt} liksom hos Hansen et al är korrelerade inom given intervjuartilldelning. Cochran antar dessutom att det finns en korrelation, $\rho_{x\varepsilon}$, mellan x och ε. Vi behandlar emellertid inte denna modell utförligare eftersom den inte är anpassad till en praktisk undersökningssituation, så t ex saknas skattningsschema för $\rho_{x\varepsilon}$.

Sukhatmes och Seths (1952) modell presenteras av författarna själva på linjär form. Med en lätt modifiering av framställningen i avsnitt 5.3 för att få konsistenta beteckningar kan modellen skrivas:

$$y_{ijt} = x_i + \alpha_j + \delta_{ij} + \varepsilon_{ijt},$$

där x_i är "sanna" värdet för enhet i;

α_j är den j:te intervjuarens genomsnittliga avvikelse från x_i-värdena. Denna term förefaller mycket snarlik intervjuareffekten I_j i Stock och Hochstims modell;

δ_{ij} är samspelseffekten mellan den j:te intervjuaren och den i:te enheten;

ε_{ijt} är en komponent som varierar vid upprepade observationer av intervjuare i på respondent j.

Alla x_i, α_j, δ_{ij} och ε_{ijt} antages vara okorrelerade både mellan och "inom" komponenterna, vilket alltså innebär att variationen i ε_{ijt} endast motsvaras av den okorrelerade svarsvariationen. Den feltyp som leder till den korrelerade svarsvariansen är hos Sukhatme och Seth representerad av komponenten α_j.

Cochran (1953) och Sukhatme (1954) behandlar förenklade versioner av denna modell. Båda antar att $\delta_{ij} = 0$ och Cochran dessutom oändliga populationer och en korrelation mellan x och ε. Av skäl som tidigare angivits tillskriver vi dock inte dessa framställningar någon självständig betydelse.

5.5 Surveymodell för svarsbias enligt Kish och Lansing

Kish och Lansing (1954) presenterade en modell för effekten av individuell svarsbias (definierad enligt Hansen et al (1951)) på surveyskattningar. Modellen använder "sanna" värden, mätvärden från ordinarie intervjuer samt "bättre" värden dvs värden som erhållits med en mätmetod som anses bättre än den ordinarie. Syftet med modellen var att evalvera felet i en undersökning av marknadsvärdet på småhus, vilken ingick i 1950 års Survey of Consumer Finances i USA. Husägare ombads i denna undersökning att skatta marknadsvärdet på sina hus, varefter motsvarande skattningar gjordes för samma hus av experter. Modellen uttrycker, i termer av medelkvadratfelet, skillnaden i noggrannhet i skattningar med den ordinarie mätmetoden jämfört med den "bättre" metoden. Modellen kan sammanfattas på följande sätt:

Låt r_i beteckna mätvärdet för enhet i med ordniarie metod,
a_i beteckna mätvärdet för enhet i med den "bättre" metoden och
y_i beteckna det "sanna", men okända, värdet för enhet i.

Populationsmedelvärdena betecknas

$$\bar{R} = E(r), \quad \bar{A} = E(a) \quad \text{och} \quad \bar{Y} = E(y)$$

och varianserna

$$V(r) = E(r-\bar{R})^2, \quad V(a) = E(a-\bar{A})^2 \quad \text{och} \quad V(y) = E(y-\bar{Y})^2 \quad \text{respektive.}$$

Det individuella svarsfelet med den ordinarie metoden är $r_i - y_i$ och med den "bättre" metoden $a_i - y_i$. Skillnaden mellan de två svarsfelen är $d_i = (r_i - y_i) - (a_i - y_i) = r_i - a_i$

Författarna definierar också

$\bar{R} - \bar{Y}$ = svarsbiasen med ordinarie metod;

$\bar{A} - \bar{Y}$ = svarsbiasen med "bättre" metod;

$\bar{D} = (\bar{R} - \bar{Y}) - (\bar{A} - \bar{Y}) = \bar{R} - \bar{A}$ = skillnaden mellan ovanstående biastermer;

samt medelkvadratdifferensen av mätningarna

$$M.S.(d) = E(d^2) = E(r - a)^2$$

och kovariansen mellan d och a

$$Cov(d,a) = E(d-\bar{D})\,(a - \bar{A}) = Cov(r,a) - V(a).$$

Med ovanstående definitioner kan man beräkna följande ekvation som uttrycker hur mycket större det totala medelkvadratfelet blir om man använder den ordinarie mätmetoden jämfört med om man använder den bättre.

$$V(r) + \bar{D}^2 - V(a) = M.S.(d) + 2Cov(d,a)\ . \qquad (5.5.1)$$

Om resultaten som erhållits med den bättre metoden betraktas som "sanna" uttrycker ekvationen den ökning i totalt medelkvadratfel, som beror på mätfel. Om de "sanna" värdena y_i varit tillgängliga så skulle ekvationen fått utseendet:

$$V(r) + (\bar{R} - \bar{Y})^2 - V(y) = E(r - y)^2 + 2Cov\,[(r - y),y\,].$$

Författarna anger formler för väntevärdesriktiga skattningar (nedan betecknade med små bokstäver) av de fem termer som ingår i ekvation (5.5.1) under antagandet att obundet slumpmässigt urval använts:

$$v(r) = \frac{1}{n-1} \sum_{1}^{n} (r_i - \bar{r})^2, \text{ där } \quad \bar{r} = \frac{1}{n} \sum_{1}^{n} r_i$$

$$v(a) = \frac{1}{n-1} \sum_{1}^{n} (a_i - \bar{a})^2, \text{ där } \quad \bar{a} = \frac{1}{n} \sum_{1}^{n} a_i$$

$$\text{cov}(r,a) = \frac{1}{n-1} \sum_{1}^{n} (r_i - \bar{r})\,(a_i - \bar{a}),$$

$$\text{cov}(d,a) = \text{cov}(r,a) - v(a),$$

$$\text{m.s.}(d) = \frac{1}{n} \sum_{1}^{n} d^2,$$

$$\bar{d}^2 = (\bar{r} - \bar{a})^2 - \frac{1}{n} [v(r) + v(a) - 2\text{cov}(r,a)].$$

En nackdel med skattningen $\bar{d}^2$ är att den kan anta negativa värden. För att undvika detta kan man trunkera fördelningen för $\bar{d}^2$ vid noll, genom att substituera med noll när man får negativa värden. Man kan också använda estimatorn $(\bar{r} - \bar{a})^2$ som har en positiv bias vars storlek kan beräknas.

6 VAR STOD MAN 1960?

Vi har beskrivit det tidiga studiet av non-samplingfel över de två utvecklingslinjer som nämndes i kapitel 1, studium av specifika felkällor och av blandade felmodeller. Hur långt hade då denna utveckling nått vid 1950-talets slut? För att besvara den frågan kan det vara lämpligt att specificera en målsättning med forskning angående surveyfel. Målsättningsfrågor har diskuterats i flera artiklar och rapporter under 1960- och 1970-talen, t ex i Hansen, Hurwitz och Pritzker (1967) och i Dalenius (1969b, 1974). Den målsättning som där formuleras innebär en strävan mot kontroll av det totala felet och har vanligen kallats Total Survey Design. Före 1960 var den inte så uttalad men en kortfattad beskrivning av den kan här ändå ha ett intresse då den ger ett vidare perspektiv på diskussionen i detta avsnitt. Vi skall följa framställningen i Hansen et al (1967).

En survey kan betraktas ur följande tre synvinklar.

(1) Behoven
(2) Specifikationerna
(3) Surveyoperationerna

Behoven bestäms av det problem, till vars lösning surveyen förväntas bidra. Mot bakgrund av dessa behov finns det en idealisk survey som, i princip om än inte i praktiken, skulle ge det idala målet nämligen en uppsättning statistikor. Vi betecknar dessa med $\bar{Z}$.

Specifikationerna bestäms i undersökningsplanen. Där identifieras en population, det fastställs hur datainsamlingen skall gå till, hur bearbetningen av data skall genomföras etc. Det mål för surveyen som därmed definieras betecknas med $\bar{X}$. I praktiken kommer $\bar{X}$ och $\bar{Z}$ att skilja sig åt eftersom specifikationer som överensstämmer helt med behoven normalt inte kan genomföras operationellt.

Surveyoperationerna, den praktiska realiseringen av undersökningen leder till statistikorna $\hat{\bar{y}}$ som har det förväntade värdet $\bar{Y}$. Felet relativt det ideala målet kan då skrivas $\hat{\bar{y}} - \bar{Z}$.

Vi kan nu dela upp väntevärdet av kvadraten på detta totala fel i olika komponenter:

$$E(\hat{\bar{y}} - \bar{Z})^2 = E(\hat{\bar{y}} - \bar{X})^2 + (\bar{X} - \bar{Z})^2 + 2(\bar{Y} - \bar{Z})(\bar{X} - \bar{Z}) \qquad (6.1)$$

Den första termen är medelkvadratfelet av $\bar{y}$ relativt det definierade målet $\bar{X}$. De övriga termerna speglar den grad i vilken undersökningen leder till relevanta resultat: $(\bar{X} - \bar{Z})^2$ är kvadraten på biasen av surveyspecifikationerna och $2(\bar{Y} - \bar{Z})(\bar{X} - \bar{Z})$ är en interaktionsterm.

Man kan säga att det är en skiss av en generaliserad surveymodell som här har beskrivits. Den visar att man behöver besvara frågor som:

(1) Vilka är de relevanta statistikorna?
(2) Hur skall relevansen i surveyspecifikationerna mätas?
(3) Vilka metoder är tillgängliga för kontroll av det totala medelkvadratfelet $E(\hat{\bar{y}} - \bar{Z})^2$?

Om vi nu återgår till tiden före 1960 kan vi konstatera att arbetet med surveymodeller under denna tid utgjorde de första stegen mot utvecklandet av metoder för kontroll av den första termen $E(\hat{\bar{y}} - \bar{X})^2$ i (6.1). Det är helt inom ramen för denna term som surveymodeller utvecklades under epoken.

Var stod då utvecklingen 1960? En rad avgörande insatser inom teoribildningen hade gjorts. De båda grundläggande principerna för mätning och kontroll av non-samplingfel, replikering (t ex återintervju) och interpenetrering, var välkända och hade tillämpats och utvecklats under många år framförallt i Indien och i USA. Den begreppsmässiga grunden för modellutvecklingen var framtagen inom US Bureau of the Census.

I detta sammanhang framstår än i dag artikeln av Hansen et al (1951) som den som kanske haft starkast inflytande på utvecklingen. Dess begreppsapprat tillämpas fortfarande vid US Bureau of the Census och har använts vid studium av totalfel i många andra sammanhang genom åren. Det matematiska problemet att beräkna en total varians som tar hänsyn till både stickprovsvarians och slumpmässigt mätfel var, under vissa specificerade förutsättningar, sedan länge löst vid periodens slut. I situationer där förutsättningarna kunnat uppfyllas (åtminstone approximativt) hade modellerna kommit till praktisk användning, t ex i den omfattande Post Enumeration Survey som genomfördes i samband med 1950 års folk- och bostadsräkning i U.S.A. (se Eckler och Hurwitz, 1957 och Hanson och Marks, 1958). Sukhatme och Seth (1952) redovisar också tillämpningar av sin modell. Utveckling av surveymodeller är emellertid en mycket långsiktig process eftersom teoribildning måste varvas med praktiska undersökningar och det kan ta många år att utvärdera en metod. De grundläggande insatser som gjordes under 1940- och 1950-talen innebar därför bara början på en utveckling som ännu 1985 är långt ifrån avslutad. Vi skall nedan i ett antal punkter ta upp faktorer som begränsar de tidiga surveymodellernas praktiska användbarhet. Dessa punkter ger samtidigt en bild av vilka problem som vid 1950-talets slut återstod att lösa för att modellerna skulle kunna användas i vanliga surveysituationer. Vi kommer i första hand att kommentera de mycket genomarbetade modeller som utvecklades vid US Bureau of the Census och vid Indian Council of Agricultural Research. Som vi tidigare noterat bygger dessa på mycket likartade antaganden.

1) Utvecklingen var under 1940-och 1950-talen till övervägande delen inriktad på en integrerad behandling av stickprovsfel och mätfel (svarsfel, observationsfel). Speciellt var det den slumpmässiga delen av mätfelet som då behandlades (ett undantag var Kishs och Lansings modell). Detta beror på de speciella svårigheterna att beräkna storleken på systematiska fel eftesom detta kräver tillgång till mätvärden som kan betraktas som sanna. Likafullt kvarstår det faktum att de tidiga surveymodellerna inte tar hänsyn till t ex bortfallsfel och ramfel. (I US Bureau of the Census modell "medverkar" visserligen bortfallet i undersökningen med sina ersatta värden, t ex imputeringar, men detta kan endast påverka svarsvariationen. Modellen tillåter inte mätning av storleken på den eventuella snedvridande effekt de ersatta

värdena har på undersökningsresultaten.)

2) Vid tillämpning på stickprovsundersökningar är en genomgående förutsättning i de tidigare modellerna att urvalsenheterna dras med obundet slumpmässigt urval. Något annat urvalsförfarande är inte utrett i dessa modeller även om Hansen et al (1951) antyder att resultaten i deras artikel kan utvidgas till klusterurval. Detta innebär en påtaglig begränsning av modellernas användbarhet eftersom bruket av urvalsmetoder som stratifierat urval, systematiskt urval och PPS-urval i praktiken är mycket vanligare än obundet slumpmässigt urval.

3) Det redan på 1940-talet vid US Bureau of the Census förekommande synsättet att individuella svar kan ses som utfall av stokastiska variabler, sågs inte som något självklart av Hansen et al (1951). De hävdade i sin diskussion av modellen att det behövdes mycket fler experimentella bevis för att fastställa om så verkligen var fallet. Denna tveksamhet berodde på att det inte går att kontrollera hur en individ väljer sitt svar bland de möjliga svaren. Man känner t ex inte till urvalssannolikheterna och ibland inte heller de möjliga svarsalternativen. 1951 förfogade man inte över metoder att studera den enkla svarsvariansen separat utan endast tillsammans med stickprovsfelet. Sedan man utvecklat den nödvändiga teorin (översiktligt presenterad i Eckler och Hurwitz, 1957) gavs också möjligheter till omfattande studier av den enkla svarsvariansen. Detta arbete, som var inriktat på 0-1-variabler (kända svarsalternativ) och återintervjuer (från vilka det teoretiskt går att skatta urvalssannolikheter för olika svarsalternativ) kom dock huvudsakligen att ske efter 1960.

4) För många surveyorganisationer kan det vara svårt att anpassa en undersökningsplan till modellernas förutsättning att intervjuarna ses som ett urval från en population, eller "pool", av intervjuare som ska arbeta inom ett visst område. Modellerna föreskriver för denna situation att varje individ kan tillordnas endast en pool av intervjuare och att varje intervjuare tillhör en enda pool. I praktiken har två intervjuare ofta arbetsområden som överlappar varandra, men överlappningen är vanligen inte total. Ofta är också fältarbetsledningen i den situationen att den önskar att ett fall skall tas om hand av en speciell intervjuare, som kanske tillhör en annan pool.

5) En ytterligare svårighet att anpassa modellerna till en praktisk situation ligger i förutsättningen att intervjuarna dras slumpmässigt från en pool, att urvalspersonerna tillordnas dem slumpmässigt och att individ- och intervjuarurvalen är oberoende av varandra. Även om alla urvalspersoner i ett område skulle kunna intervjuas av varje intervjuare i den pool som hör till området, är det rimligt att fältarbetsledningen vill göra tilldelningen så att resekostnaderna minimeras, i vilket fall tilldelningen alltså inte sker slumpmässigt.

Det kan synas som om en slumpmässig tilldelning skulle vara enklare att genomföra vid telefonintervjuer än vid besöksintervjuer, eftersom problemet med resekostnader försvinner. Emellertid kvarstår då ett annat problem, nämligen att spårningsarbete kräver god lokalkännedom. En intervjuare riskerar alltså att få ett större bortfall om uppgiftslämnarna slumpas ut, än vad han eller hon annars skulle ha fått. (Detta problem kan även uppträda vid besöksintervjuer om den första kontakten tas per telefon.)

6) Modellerna förutsätter att alla intervjuare får en tilldelning, vars storlek är fastställd i förväg. Detta kan vara svårt att genomföra i praktiken av bl a tidsskäl. Exempelvis kan intervjuare som av någon anledning blivit försenade behöva avlastas genom att någon annan intervjuare tar över en del uppgiftslämnare.

7) Den teori för mätning av den enkla svarsvariansen som presenterades av Eckler och Hurwitz (1957) bygger på återintervjuer. Ett grundläggande antagande var här att återintervjuerna kan göras oberoende av originalintervjuerna. Detta antagande är svårt att uppfylla i praktiken eftersom uppgiftslämnaren vid återintervjun kan komma ihåg vad han eller hon svarade vid originalintervjun. Det bör därför dröja en tid innan återintervjun äger rum, dock inte allt för länge eftersom glömskeeffekter kan uppträda. Hur den optimala tidpunkten för återintervju skall bestämmas är ett grundläggande problem vid användndet av denna teori. Tillämpningen av teorin kan däremot vara betydligt enklare i andra mätsituationer, t ex vid studium av bearbetningsfel.

Vi påminner också om de kritiska synpunkter på Hansens et al survey-modell som redovisas på sid 89.

REFERENSER

Birnbaum, Z.W. och Sirken, M.G. (1950). Bias Due to Non-availability in Sampling Surveys. Journal of the American Statistical Association, vol. 45, pp. 98-111.

Blankenship, A.B. (1940). The Influence of the Question Form Upon the Response in a Public Opinion Poll. The Psychological Record, pp. 349-422.

Chevry, G. (1949). Control of a General Census by Means of an Area Sampling Method. Journal of the American Statistical Association, vol. 44, pp. 373-379.

Cochran, W.G. (1953). Sampling Techniques. Wiley.

Cochran, W.G., Mosteller, F. och Tukey, J.W. (1954). Principles of Sampling. Journal of the American Statistical Association, vol. 49, pp. 13-35.

Cornfield, J. (1942). On Certain Biases in Samples of Human Populations. Journal of the American Statistical Association, vol. 37, pp. 63-68.

Dalenius,T.E. (1962). Recent Advances in Sample Survey Theory and Methods. The Annals of Mathematical Statistics, vol 33, pp.325-349.

Dalenius, T.E. (1969a). Modern surveymetodik. Intern handbok nr 8. Statistiska Centralbyrån.

Dalenius, T.E. (1969b). Towards a Survey Measurement System. Zastosowania Matematyki. Hugo Steinhaus Jubilee Volume, X, pp. 289-303.

Dalenius, T.E. (1974). Ends and Means of Total Survey Design. Forskningsprojektet FEL I UNDERSÖKNINGAR. Stockholms Universitet, Statistiska Institutionen.

Deming, W.E. (1944). On Errors in Surveys. American Sociological Review, vol. 9, no. 4, pp.359-369.

Deming, W.E. (1950). Some Theory of Sampling. Wiley.

Deming, W.E. (1953). On a Probability Mechanism to Attain an Economic Balance Between the Resultant Error of Nonresponse and the Bias of Nonresponse. Journal of the American Statistical Association, vol. 48, pp. 743-772.

Deming, W.E. och Geoffrey, L. (1941). On Sample Inspection in the Processing of Census Returns. Journal of the American Statistical Association, vol. 36, pp. 351-360.

Deming, W.E., Tepping, B.J. och Geoffrey, L. (1942). Errors in Card Punching. Journal of the American Statistical Association, vol. 37, pp.525-536.

Des Raj (1968). Sampling Theory. Wiley.

Durbin, J. och Stuart, A. (1951). Differences in Response Rates of Experienced and Inexperienced Interviewers. Journal of the Royal Statistical Society (A), vol. 114, pp. 163-184.

Eckler, A.R. och Hurwitz, W.N. (1957). Response Variance and Biases in Censuses and Surveys. Bulletin of the International Statistical Institute, vol. 36, pt. 2, pp.12-35.

Eckler, A.R. och Pritzker, L. (1951). Measuring the Accuracy of Enumerative Surveys. Bulletin of the International Statistical Institute, vol. 33, pt 4, pp. 7-24.

El-Badry, M.A. (1956). A Sampling Procedure for Mailed Questionnaires. Journal of the American Statistical Association, vol. 51, pp.209-227.

Feldman, J.J., Hyman, H.H. och Hart, C.W. (1951). A Field Study of Interviewer Effects on the Quality of Survey Data. Public Opinion Quarterly, vol. 15, pp. 734-761.

Gales, K. och Kendall, M.G. (1957). An Inquiry Concerning Interviewer Variability. Journal of the Royal Statistical Society (A), vol. 120, pt.2, pp. 121-138.

Goldberg, S.A. (1957). Non-sampling Error in Household Surveys. Bulletin of the International Statistical Institute, vol. 36, pt. 2, pp. 44-59.

Hansen, M.H. och Hurwitz, W.N. (1946). The Problem of Non-response in Sample Surveys. Journal of the American Statistical Association, vol. 41, pp. 517-529.

Hansen, M.H. och Madow, W.G. (1974). Some Important Events in the Historical Development of Sample Surveys. I Owen. D.B. (ed.): On the History of Statistics and Probability, pp. 75-102. Marcel Dekker.

Hansen, M.H., Hurwitz, W.N. och Madow, W.G. (1953). Sample Survey Methods and Theory. Vols. I and II. Wiley.

Hansen, M.H., Hurwitz, W.N., Marks, E.S. och Mauldin, W.P. (1951). Response Errors in Surveys. Journal of the American Statistical Association, vol. 46, pp. 147-190.

Hanson, R.H. och Marks, E.S. (1958). Influence of the Interviewer on the Accarucy of Survey Results. Journal of the American Statistical Association, vol. 53, pp. 635-655.

Hartley. H.O. (1946). Diskussion av Yates: A Review of Some Recent Statistical Developments in Sampling and Sample Surveys. Journal of the Royal Statistical Society (A), vol. 109, pp. 36-38.

Hyman, H.H., Cobb, W.J., Feldman, J.J., Hart, C.W. och Stember, C.H. (1954). Interviewing in Social Research. The University of Chicago Press.

Kish, L. (1965). Survey Sampling. Wiley

Kish, L. och Hess, I. (1959). A "Replacement" Procedure for Reducing the Bias of Non-Response. The American Statistician, October 1959, pp. 17-19.

Kish, L. och Lansing, J.B. (1954). Response Errors in Estimating the Value of Homes. Journal of the American Statistical Association, vol. 49, pp. 520-538.

Krewski, D., Platek, R. och Rao, J.N.K. (1981). Current Topics in Survey Sampling. Academic Press.

Mahalanobis, P.C. (1946). Recent Experiments in Statistical Sampling in the Indian Statistical Institute. Journal of the Royal Statistical Society, vol. 109, pt. 4, pp. 326-370.

Marks, E.S. och Mauldin, W.P. (1950). Response Errors in Census Research. Journal of the American Statistical Association, vol. 45, pp. 424-438.

Marks, E.S., Mauldin, W.P. och Nisselson, H. (1953). The Postenumeration Survey of the 1950 Census: A Case History in Survey Design. Journal of the American Statistical Association, vol. 48, pp. 220-243.

O'Muircheartaigh, C.A. (1977). Response Errors. I O'Muircheartaigh, C.A. och Payne, C. (ed.): The Analysis of Survey Data, vol. 2, Model Fitting, pp. 193-239. Wiley.

Palmer, G.L. (1943). Factors in the Variability of Response in Enumerative Studies. Journal of the American Statistical Association, vol. 38, pp. 143-152.

Pearson, K. (1902). On the Mathematical Theory of Errors of Judgment, with Special Reference to the Personal Equation. Philosophical Transactions of the Royal Society of London, ser. A, vol. 198, pp 235-299.

Politz, A. och Simmons, W. (1949, 1950). An Attempt to get the "Not at Homes" Into the Sample Without Callbacks. Journal of the American Statistical Association, vol. 44, pp. 9-31 and vol. 45, pp. 136-137.

Rice, S.A. (1929). Contagious Bias in the Interview. American Journal of Sociology, vol. 35, pp. 420-423.

Simmons, W.R. (1954). A Plan to Account for "Not-at-homes" by Combining Weighting and Callbacks. Journal of Marketing, vol. 11, pp. 42-53.

Stock, J.S. och Hochstim, J.R. (1951). A Method of Measuring Interviewer Variability. Public Opinion Quarterly, vol. 15, pp. 322-334.

Strecker, H. och Wiegert, R. (1981). Fehler in Statistischen Erhebungen, Darstellung anhand von Beispielen, in: Wirdschaftstheorie und Wirdschaftspolitik, Gedenkschrift für Erich Preiser, Hrsg. von Mückl, W.J. und Ott, A.E., Passau 1981, pp. 439-458.

Sukhatme, P.V. (1947). The Problem of Plot Size in Large-Scale Yield Surveys. Journal of the American Statistical Association, vol. 42, pp. 297-310.

Sukhatme, P.V. (1954). Sampling Theory of Surveys With Applications. U.N. Food and Agriculture Organization, Rome. Published by: The Iowa State College Press and The Indian Society of Agricultural Research.

Sukhatme, P.V. och Seth, G.R. (1952). Non-Sampling Errors in Surveys. Journal of the Indian Society of Agricultural Statistics, vol. 4, pp. 5-41.

Zarković, S.S. (1955). Sampling Methods in the Yugoslav 1953 Census of Population. Journal of the American Statistical Association, vol. 50, pp. 720-737.

Tidigare nummer av Promemorior från P/STM:

NR

1 Bayesianska idéer vid planeringen av sample surveys. Lars Lyberg (1978-11-01)

2 Litteraturförteckning över artiklar om kontingenstabeller. Anders Andersson (1978-11-07)

3 En presentation av Box-Jenkins metod för analys och prognos av tidsserier. Åke Holmén (1979-12-20)

4 Handledning i AID-analys. Anders Norberg (1980-10-22)

5 Utredning angående statistisk analysverksamhet vid SCB: Slutrapport. P/STM, Analysprojektet (1980-10-31)

6 Metoder för evalvering av noggrannheten i SCBs statistik. En översikt. Jörgen Dalén (1981-03-02)

7 Effektiva strategier för estimation av förändringar och nivåer vid föränderlig population. Gösta Forsman och Tomas Garås (1982-11-01)

8 How large must the sample size be? Nominal confidence levels versus actual coverage probabilities in simple random sampling. Jörgen Dalén (1983-02-14)

9 Regression analysis and ratio analysis for domains. A randomization theory approach. Eva Elvers, Carl Erik Särndal, Jan Wretman och Göran Örnberg (1983-06-20)

10 Current survey research at Statistics Sweden. Lars Lyberg, Bengt Swensson och Jan Håkan Wretman (1983-09-01)

11 Utjämningsmetoder vid nivåkorrigering av tidsserier med tillämpning på nationalräkenskapsdata. Lars-Otto Sjöberg (1984-01-11)

12 Regressionsanalys för f d statistikstuderande. Harry Lütjohann (1984-02-01)

13 Estimating Gini and Entropy inequality parameters. Fredrik Nygård och Arne Sandström (1985-01-09)

14 Income inequality measures based on sample surveys. Fredrik Nygård och Arne Sandström (1985-05-20)

Kvarvarande exemplar av ovanstående promemorior kan rekvireras från Elseliv Lindfors, P/STM, SCB, 115 81 Stockholm, eller per telefon 08 7834178

☆Garnisonstryckeriet Stockholm 1985☆