SCB STATISTISKA CENTRALBYRÅN 1982-11-01

PROMEMORIER FRÅN P/STM

NR 7

EFFEKTIVA STRATEGIER FÖR ESTIMATION AV FÖRÄNDRINGAR OCH NIVÅER VID FÖRÄNDERLIG POPULATION

AV GÖSTA FORSMAN OCH TOMAS GARÅS

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 1982:7. Effektiva strategier för estimation av förändringar och nivåer vid föränderlig population / Gösta Forsman; Tomas Garås.
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

urn:nbn:se:scb-PM-PSTM-1982-7

SCB STATISTISKA CENTRALBYRÅN

1 9 8 2 - 1 1 - 0 1

PROMEMORIOR FRÅN P/STM

NR 7

EFFEKTIVA STRATEGIER FÖR ESTIMATION AV FÖRÄNDRINGAR OCH NIVÅER VID FÖRÄNDERLIG POPULATION

AV

GÖSTA FORSMAN OCH TOMAS GARÅS

I N N E H Å L L S F Ö R T E C K N I N G

INLEDNING

Denna rapport är utarbetad vid Enheten för statistiska metoder (P/STM) på Avdelningen för planering och samordning, inom projektet Urval och estimation (URVEST), delprojektet Förändringsskattningar.

Syftet med utredningen är att ange urvals- och estimationsplaner (strategier) som är effektiva för nivå- och förändringsskattningar då föränderlig population föreligger.

Inom SCBs statistikgrenar produceras en mängd statistik vid successiva tidpunkter vars periodicitet kan variera från en månad upp till flera år. Vanligen publiceras skattningar av totaler eller medelvärden (nivåer) som så småningom bildar tidsserier. Statistikkonsumenterna kan då vid analysen av undersökningsresultaten notera ökningar, minskningar eller oförändrat resultat mellan intilliggande tidpunkter samt studera trender över längre perioder. Ofta är emellertid inte undersökningarna i första hand gjorda för att ge en god precision för förändringsskattningarna. Undersökningsmetodiken är i stället inriktad på nivåskattningarnas precision. På senare år har man på olika håll vid SCB uppmärksammat denna situation, t ex vid Utredningsinstitutet (I/UI) och vid Avdelningen för areell statistik.

Ett annat problem är att undersökningspopulationerna är föränderliga över tiden, dvs, populationsobjekt avgår och tillkommer mellan successiva tidpunkter. Den traditionella samplingteorin behandlar endast fall med konstanta populationer dvs, inget in- och utflöde av populationsobjekt över tiden. Detta beror troligen på att denna teori till stor del är utvecklad i USA och att man där oftast använder sig av areasampling och därmed erhåller konstanta populationer.

Slutligen uppstår det en konflikt eftersom de straterier är olika som ger optimala nivå- respektive förändringsskattningar. Man kan då välja en kompromiss som beror på hur man prioriterar mellan de båda skattningstyperna.

I rapportens avsnitt 1 ges en allmän översikt om förändringsskattningar. Avsnitt 2 beskriver en urvalsplan som tar hänsyn till att populationen förändras mellan två undersökningstillfällen. Två användbara satser för härledning av LMVU-estimatorer (Linear Minimum Variance Unbiased) samt dess varianser är presenterade i avsnitt 3. I avsnitt 4 beskrivs härledningsmetodik för LMVU-estimatorer (nivå- och förändringsskattningar) samt dess varianser. Resultat av utförda härledningar presenteras samt en variansjämförelse mellan den här studerade strategin och en strategi kännetecknad av oberoende urval.

Minsta varians för en förändringsskattning uttryckt som en kvot presenteras i avsnitt 5.

Avsnitt 6 innehåller en sammanfattning av rapporten samt kommentarer.

I avsnitt 7 återfinns en litteraturförteckning med skrifter som behandlar estimation vid successiva urval i tiden.

1 ALLMÄNT OM FÖRÄNDRINGSSKATTNINGAR

1.1 Beteckningar och definitioner

Förändringsskattningar förekommer vanligen då man vid olika tidpunkter, t, upprepar en undersökning. Vi betecknar populationen vid tidpunkten t med U_t. För en undersökningsvariabel y_t vill vi skatta totalen Y_t med stickprovsundersökningar vid successiva tidpunkter t (t = 0, 1, 2,...). Låt nu $\hat{Y}_t$ beteckna en estimator av Y_t. Vi får alltså en serie skattningar $\hat{Y}_0$, $\hat{Y}_1$, $\hat{Y}_2$,... Förändringen för totalen Y_t mellan tidpunkterna t-1 och t kan uttryckas med två olika parametrar

$$R_t = \frac{Y_t}{Y_{t-1}} \qquad (1.1.1)$$

och

$$D_t = Y_t - Y_{t-1} \qquad (1.1.2)$$

R_t och D_t skattas vid successiva tidpunkter med

$$\hat{R}_t = \frac{\hat{Y}_t}{\hat{Y}_{t-1}} \qquad (1.1.3)$$

respektive

$$\hat{D}_t = \hat{Y}_t - \hat{Y}_{t-1} \qquad (1.1.4)$$

Variansen för kvotestimatorn $\hat{R}_t$ är

$$V(\hat{R}_t) \doteq \frac{1}{Y_{t-1}^2}\left[V(\hat{Y}_t) + R_t^2\, V(\hat{Y}_{t-1}) - 2R_t \mathrm{Cov}(\hat{Y}_t, \hat{Y}_{t-1})\right] \qquad (1.1.5)$$

och för $\hat{D}_t$:

$$V(\hat{D}_t) = V(\hat{Y}_t) + V(\hat{Y}_{t-1}) - 2\ Cov(\hat{Y}_t,\ \hat{Y}_{t-1}) \tag{1.1.6}$$

Den enklaste situationen har vi när populationen U_t består av samma objekt oberoende av tiden, dvs $U_t = U$ för alla tidpunkter t. En sådan population kallas i det följande för en konstant population.

I samband med successiva mätningar över tiden talar man om longitudinella studier. I Goldstein (1979) görs följande definitioner:

a. Rena longitudinella studier.
Samtliga utvalda objekt mäts vid alla väl definierade tidpunkter.

b. Blandade longitudinella studier.
Minst ett objekt mäts vid två eller flera väl definierade måttillfällen.

c. Icke longitudinella studier.
Inga objekt mäts vid fler än en tidpunkt.

Longitudinella studier innebär alltså ej enbart att man gör successiva mätningar över tiden, utan också att urvalen mellan olika tidpunkter på något sätt är hopkopplade och att mättillfällena är väl definierade. I praktiken är det nästan alltid blandade longitudinella studier det är fråga om vid successiva mätningar i tiden.

Begrepp som förekommer i dessa sammanhang är paneler och panelundersökningar (se t ex Kish (1965) eller Dahmström (1971)):

- En <u>panel</u> är ett urval där samtliga objekt mäts vid två eller flera tillfällen.

- Med en panelundersökning menas en undersökning där minst en panel förekommer.

1.2 Olika urvalsplaner vid successiva mätningar över tiden

Gemensamt för de urvalsplaner som leder till bättre förändringsskattningar än när man har oberoende urval vid varje tidpunkt, är att man återkommer till en bestämd del av tidigare urval. På så sätt kan man utnyttja populationskorrelationen mellan två eller flera mättillfällen. Vi ska i detta avsnitt behandla två olika typer av sådana urvalsplaner.

1.2.1 Matchade urval

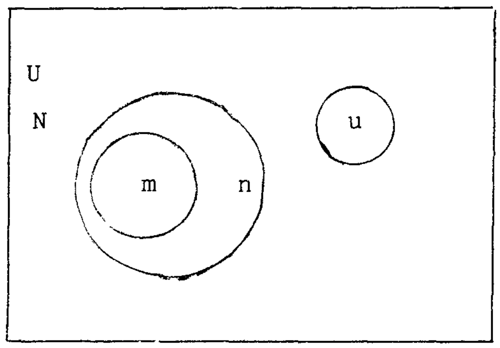
Ambitionen att framställa både nivå- och förändringsskattningar med lägsta möjliga varians kan leda till en konflikt när man ska bestämma graden av överlappning mellan stickproven. Detta illustreras i nedanstående urvals- och estimationsplan som först beskrevs av Jessen (1942):

Man har en konstant population U som innehåller N objekt.

Vid tidpunkt 1 drages ett slumpmässigt urval av n objekt utan återläggning (OSU). Vi betecknar det $s_1(n)$.

Vid tidpunkt 2 drages ett suburval av m (matched) objekt med OSU från $s_1(n)$ samt ett OSU av u (unmatched) objekt från de N-n objekten i $U-s_1(n)$. Undersökningsvariabeln kallas vid tidpunkt 1 för x och vid tidpunkt 2 för y.

Följande figur illustrerar urvalsplanen:



Det antages gälla att:

a) Urvalsfraktionerna $\frac{n}{N}$, $\frac{m}{N}$ och $\frac{u}{N}$ är försumbart små.

b) $S_x^2 = S_y^2 = S^2$, dvs populationsvarianserna är lika mellan tidpunkterna. Då kan korrelationskoefficienten ρ mellan y och x skrivas $\rho = \frac{\text{Cov}(x,y)}{S^2}$ vilket medför att $\text{Cov}(x,y) = \rho S^2$, $(-1<\rho<1)$

c) m och u väljes så att m+u=n. Vi inför beteckningen $\theta = \frac{u}{n}$, vilket leder till att urvalsöverlappningen är $1-\theta = \frac{m}{n}$.

d) $\hat{X}_m$, $\hat{X}_n$, $\hat{Y}_m$ och $\hat{Y}_u$ utgör enkla skattningar av X och Y, baserade på respektive urval.

Vid tidpunkt 1 skattas totalen X med $\hat{X}_n$. Dess varians är $V(\hat{X}_n) = N^2 \frac{S^2}{n}$

Totalen Y vid tidpunkt 2 skattas med

$$\hat{Y} = a(\hat{X}_m - \hat{X}_n) + b\,\hat{Y}_m + (1-b)\,\hat{Y}_u \tag{1.2.1}$$

där a och b är godtyckliga konstanter.

Skattningen $\hat{X}_n$ kan förbättras vid tidpunkt 2 i efterhand genom

$$\hat{X} = c\hat{X}_m + (1-c)\hat{X}_n + d(\hat{Y}_m - \hat{Y}_u) \qquad (1.2.2)$$

där c och d är godtyckliga konstanter.

Om varianserna för $\hat{X}$ och $\hat{Y}$ beräknas och minimeras med avseende på konstanterna a, b, c och d erhålles följande varianser:

$$\min_{a,b} V(\hat{Y}) = \min_{c,d} V(\hat{X}) = N^2 \frac{S^2}{n} \left[\frac{1-\rho^2\theta}{1-\rho^2\theta^2}\right] \quad (-1<\rho<1) \qquad (1.2.3)$$

Minimeras detta uttryck nu med avseende på θ fås:

$$\min_{a,b,\theta} V(\hat{Y}) = \min_{c,d,\theta} V(\hat{X}) = N^2 \frac{S^2}{n} \left[\frac{1+\sqrt{1-\rho^2}}{2}\right] \text{ där } \theta_{min} = \frac{1}{1+\sqrt{1-\rho^2}} \qquad (1.2.4)$$

Vid dessa minimeringar erhålles bl a sambandet $a = -\rho b$.

Insättning av detta i (1.2.1) ger

$$\hat{Y} = (1-b)\hat{Y}_u + b(\hat{Y}_m + \rho(\hat{X}_n - \hat{X}_m)) \qquad (1.2.5)$$

Strukturen av (1.2.5) innebär en sammansatt estimator som består av en OSU-skattning och en 2-fas regressionsskattning.

Vi betraktar nu differensskattningen $\hat{D} = \hat{Y}-\hat{X}$ med $\hat{Y}$ enligt (1.2.1) och $\hat{X}$ enligt (1.2.2). Minimeringar av variansen för $\hat{D}$ ger:

$$\min_{a,b,c,d} V(\hat{D}) = 2N^2 \frac{S^2}{n} \frac{1-\rho}{1-\rho\theta} \quad ; -1<\rho<1 \qquad (1.2.6)$$

$$\min_{a,b,c,d,\theta} V(\hat{D}) = 2N^2 \frac{S^2}{n} (1-\rho) \text{ där } \theta_{min} = 0 \; ; \; 0 \leq \rho < 1 \qquad (1.2.7)$$

$$\min_{a,b,c,d,\theta} V(\hat{D}) = 2N^2 \frac{S^2}{n} \quad \text{där } \theta_{min} = 1 \; ; -1<\rho<0 \qquad (1.2.8)$$

Som synes får vi här andra optimala θ-värden än för $\min V(\hat{Y})$ och $\min V(\hat{X})$. Detta illustrerar konflikten att med samma urvalsplan effektivt skatta både nivå- och förändringsskattningar.

Det bör påpekas att samma θ_{min} ej alltid ger optimal varians för olika typer av förändringsskattningar. Detta gäller t ex $\hat{D}=\hat{Y}-\hat{X}$ och $\hat{R}=\hat{Y}/\hat{X}$ (med $\hat{Y}$ enligt (1.2.1) och $\hat{X}$ enligt (1.2.2)) vilket framgår av (1.2.7) samt tabellen på sidan 42.

Följande exempel visar hur man kan kompromissa för att få bra skattningar för både nivåer och förändringar med samma urvalsplan.

Exempel:

(Beteckningar enligt sid 5-7)

Antag att man vill välja ett värde på $\theta = \frac{u}{n}$ som tar hänsyn till att man vill ha så bra skattningar som möjligt för både nivåer och förändringar. Antag också att korrelationskoefficienten ρ mellan x och y är 0.9. En förutsättning för resonemanget i detta exempel är att man har en hygglig uppfattning om ρ före tidpunkt 2.

Följande beteckningar införs:

$$\min_{a,b} V(\hat{Y}) = V_{\hat{Y}}(\theta) = V_{\hat{X}}(\theta)$$

$$\min_{a,b,c,d} V(\hat{D}) = V_{\hat{D}}(\theta)$$

Den tidigare beskrivna strategin ska nu jämföras med en enklare strategi. Denna innebär att man drar oberoende OSU av n objekt från populationen U vid de båda tidpunkterna. Man får då $\hat{X}_n$ och $\hat{Y}_n$ som enkla totalskattningar

vid tidpunkt 1 respektive 2 samt differansskattningen $\hat{D}_n = \hat{Y}_n - \hat{X}_n$.
Varianserna för dessa estimatorer blir:

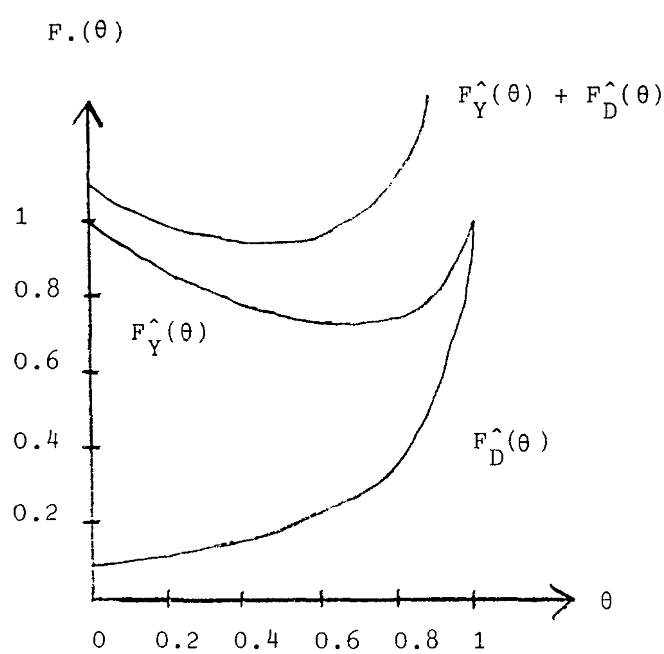
$$V(\hat{Y}_n) = V(\hat{X}_n) = N^2\frac{S^2}{n} \quad \text{samt} \quad V(\hat{D}_n) = V(\hat{Y}_n - \hat{X}_n) = 2N^2\frac{S^2}{n}$$

Varianskvoterna $F.(\theta)$ definieras nu på följande sätt:

$$F_{\hat{Y}}(\theta)=F_{\hat{X}}(\theta)= \frac{V_{\hat{Y}}(\theta)}{V(\hat{Y}_n)} = \frac{V_{\hat{X}}(\theta)}{V(\hat{X}_n)} = \frac{1-\rho^2\theta}{1-\rho^2\theta^2} \quad \text{och} \quad F_{\hat{D}}(\theta) = \frac{V_{\hat{D}}(\theta)}{V(\hat{D}_n)} = \frac{1-\rho}{1-\rho\theta} \quad \text{då } 0\leq\theta\leq 1.$$

Funktionerna $F.(\theta)$ uttrycker alltså varianserna för den matchade strategin relativt varianserna för den oberoende strategin, för olika värden på θ. Eftersom varianserna $V(\hat{Y}_n)$, $V(\hat{X}_n)$ och $V(\hat{D}_n)$ är konstanta med avseende på θ, har $F.(\theta)$ samma profiler (och θ_{min}) som variansfunktionerna $V.(\theta)$.

• Då $\rho = 0.90$ fås följande funktionskurvor $F.(\theta)$:



Beteckna $\min\limits_{a,b,\theta} V(\hat{Y}) = \min\limits_{\theta} V_{\hat{Y}}(\theta)$

$$\min\limits_{a,b,c,d,\theta} V(\hat{D}) = \min\limits_{\theta} V_{\hat{D}}(\theta)$$

Av figuren framgår att valet av θ är självklart om man ensidigt önskar bästa förändringsskattning eller bästa nivåskattning.

Från (1.2.4) fås att $\theta_{min} = \dfrac{1}{1+\sqrt{1-0.81}} = 0.70$ minimerar $V_{\hat{Y}}(\theta)$ och $V_{\hat{X}}(\theta)$. Dessa varianser är 72% av motsvarande varianser vid oberoende urval. Av (1.2.7) framgår att $\theta_{min} = 0$ leder till att $V_{\hat{D}}(\theta)$ minimeras och att storleken av denna är 10% jämfört med då man har oberoende urval. Man kan t ex välja $\theta = 0.5$ om nivå- och förändringsskattningar anses lika viktiga. Detta θ-värde sammanfaller här med strategin att välja det θ som minimerar $F_{\hat{Y}}(\theta) + F_{\hat{D}}(\theta)$. Detta illustreras i nedanstående tabell:

θ / $F.(\theta)$	0	0.50	0.70
$F_{\hat{Y}}(\theta) = F_{\hat{X}}(\theta)$	1	0.75	0.72
$F_{\hat{D}}(\theta)$	0.10	0.16	0.27

Beroende på prioriteringen mellan nivå- och förändringsskattningen kan man naturligtvis välja andra θ-värden som motsvarar den vikt man lägger vid respektive skattning.

(slut på exemplet)

Den ovan behandlade urvals- och estimationsmetodiken beskrevs först av Jessen (1942), som tidigare nämnts. Fallet då man vid flera successiva mätningar utnyttjar hjälpinformation från föregående mättidpunkt behandlades av Yates (1949) och Pattersson (1950). Ett avsnitt med motsvarande stoff brukar ingå i de vanligast förekommande samplinghandböckerna, t ex Cochran (1977), Hansen-Hurwitz-Madow (1953) och Des Raj (1968). I Sverige har Kulldorff behandlat ämnet ingående och bl a ansatt kostnadsfunktioner, se t ex Kulldorff (1963). På senare tid har indiern Arnab arbetat vidare med dessa urvalsplaner och behandlat tvåstegsurval och dragning med godtyckliga sannolikheter, se Arnab (1980, 1981). Det har även förekommit modellbaserade ansatser vid successiva mätningar där man antagit att totalen Y_t förändras enligt någon stokastisk process, se t ex Scott-Smith (1974). I S.M. Woodruff (1980) beskrivs en klass av sammansatta kvotestimatorer.

1.2.2 Roterande panelurval

Den tidigare beskrivna urvalsplanen är ett panelurval eftersom delurvalet m enligt definitionen på sidan 4 är en panel. En annan typ av panelurval som utnyttjar hjälpinformation från tidigare tidpunkter är ett s k roterande panelurval, som kan beskrivas på följande sätt.

Antag att en konstant population föreligger. Man drager t ex med OSU k olika urval, vilka är parvis icke överlappande. De är också oftast av samma storlek.

Vid en given tidpunkt deltager r urvalsgrupper och mellan två tidpunkter byts en urvalsgrupp ut mot en ny, vilket utgör rotationen. Detta kan illustreras med följande exempel där k=8 och antalet rotationsgrupper som mäts vid varje tidpunkt är $r = 3$.

Tidpunkt \ Rotationsgrupp	1	2	3	4	5	6	7	8
1	x	x	x					
2		x	x	x				
3			x	x	x			
4				x	x	x		
5					x	x	x	
6						x	x	x

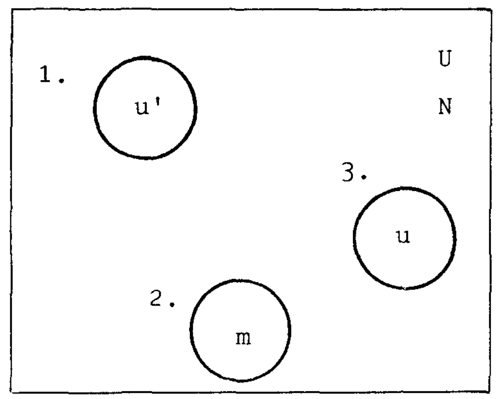
I detta fall utgör rotationsgrupp 2-7 paneler eftersom dessa mäts vid minst 2 tidpunkter.

I Dahmström (1971) och Woodruff (1963) ges översikter av roterande panelurval. Beroende på önskemål om nivå- och förändringsskattningar kan man bestämma andelen av stickprovet som mäts vid varje tillfälle.

Skattningarna leder som tidigare till kombinerade estimatorer. Ett enkelt fall av roterande panelurval är när r=2 och k=3. Se Raj (1968).

Tidpunkt \ Rotationsgrupp	1	2	3
1	x	x	
2		x	x

Urvalsstorlekarna för rotationsgrupperna 1, 2 och 3 är u', m respektive u:



Vid tidpunkt 1 mäts variabeln x på rotationsgrupperna 1 och 2. Vid tidpunkt 2 mäts variabeln y på rotationsgrupperna 2 och 3.

Enligt samma förutsättningar som vid den tidigare urvalsplanen (redovisad i avsnitt 1.2.1) kan man välja urvalsstorlekarna så att de blir lika vid tidpunkterna 1 och 2, dvs m + u' = n och m + u = n vilket medför att u' = u.

Estimatorerna för Y, X och Y-X blir:

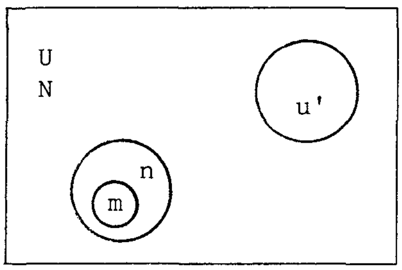
$$\hat{Y} = a(\hat{X}_m - \hat{X}_{u'}) + b\hat{Y}_m + (1-b)\hat{Y}_u$$

$$\hat{X} = c(\hat{Y}_m - \hat{Y}_u) + d\hat{X}_m + (1-d)\hat{X}_{u'}$$

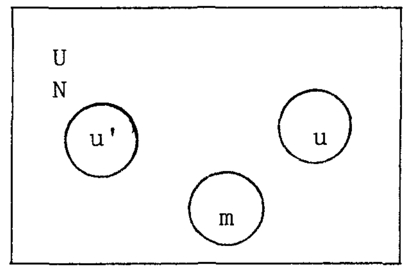
$$\hat{D} = \hat{Y} - \hat{X}$$

Optimala varianser för dessa blir samma som för den tidigare beskrivna urvalsplanen. Detta är inte förvånande eftersom urvalsplanerna är identiska i detta fall:

Matchat urval:



Roterande panelurval:



För det roterande panelurvalet utgör m + u' = n ett OSU-urval där m är ett delurval av m + u'. Detta visar att roterande panelurval och det tidigare panelurvalet är starkt besläktade.

På SCB används roterande panelurval i vissa individ- och hushållsundersökningar, t ex Arbetskraftsundersökningarna (AKU), Hushållens energianvändning (HEA) och Hushållens inköpsplaner (HIP) (8,2 respektive 5 rotationsgrupper). AKU är dock den enda undersökningen som använder sig av sammansatt estimation. Inom företagsstatistiken drages urval enligt ett system kallat samordnade urval inom företagsstatistiken (SAMU), som ger viss möjlighet att samordna urval i tiden. Inom jordbruksstatistiken används roterande panelurval i Deklarationsundersökningen för jordbrukare (DU) och i Jordbruksekonomiska undersökningen (JEU) (3 respektive 4 rotationsgrupper). Här används inte sammansatt estimation.

De register som används som urvalsramar på SCB är främst Registret över totalbefolkningen (RTB), Centrala företagsregistret (CFR) samt Lantbruksregistret (LBR).

Det största praktiska problemet vid tillämpandet av urvalsmetoder beskrivna ovan är att populationerna inte är konstanta utan föränderliga över tiden. Med detta menas att populationsobjekt tillkommer och avgår mellan olika tidpunkter.

I vissa fall kan man komma runt problemen med en föränderlig population av undersökningsobjekt genom att tillgripa areasampling. Man får då en konstant population av urvalsenheter. Ett exempel på detta är Current Population Survey (CPS) som görs på Bureau of the Census i USA. Areasampling är då ett clusterurval som på klassiskt sätt löser ramproblem då register ej finns på undersökningspopulationen. Denna lösning ger dock oftast sämre precision för skattningarna än då man har ramar som överensstämmer hyggligt med undersökningspopulationen. På SCB förekommer ej areasampling då det finns register över undersökningsobjekt att tillgå. Ramproblemen vid användande av dessa register är dock ej obetydliga.

Problemen med föränderliga populationer i samband med urvals- och estimationsplaner för successiva mätningar i tiden är mycket litet beskrivet i den statistiska litteraturen i motsats till fallet med konstanta populationer. På SCB har dock en del arbete gjorts i detta sammanhang med föränderlig population:

I Anderson (1979) analyseras vad som händer vid roterande panelurval samt ges ett antal förslag till nivå- och förändringsskattningar samt deras varianser.

Cassel (1979) studerar vilken variansreducerande effekt roterande panelurval har på olika nivå- och förändringsskattningar i samband med estimation i delgrupper, s k domains of study. I Råbäck (1980) presenteras tabeller med variansjämförelser som utgår från Cassels resultat. Bäcklund (1972) har angett den allmänna kovarianstermen i variansen för kvotestimatorn enligt (1.1.5) då urval sker med SAMU i företagsstatistiken.

Vid successiva mätningar då föränderliga populationer föreligger dyker givetvis alla vanliga undersökningsproblem upp som t ex ram-, bortfalls- och mätproblem. Dessa problem är dessutom oftast förstärkta jämfört med en engångsundersökning. Se Lindström et al (1981). Sådana undersökningsproblem behandlas dock ej i denna rapport.

I det följande av rapporten behandlas effektiva förändrings- och nivåskattningar samt kompromisser mellan dessa då nedanstående förutsättningar gäller:

a Föränderlig population

b Ändlig populationsstorlek

c Två successiva tidpunkter.

2 EN URVALSPLAN SOM TAR HÄNSYN TILL FÖRÄNDERLIG POPULATION

Betrakta en population U(t) som innehåller N(t) objekt och som är föränderlig över tiden t. I nedanstående urvalsplan betecknas undersökningsvariabelns värde vid tidpunkt t=1 med x och vid tidpunkt t=2 med y.

Man önskar skatta medelvärdet $\bar{X} = \frac{\sum^{N} x_i}{N}$ och totalen $X = N\bar{X}$ i populationen U=U(1) och motsvarande parametrar $\bar{Y}$ och Y i populationen U'=U(2). Dessutom vill man skatta differenserna $\bar{Y}-\bar{X}$ samt Y-X. Samtliga skattningar ska vara minimum-variance-unbiased inom en klass av linjära estimatorer (LMVU). Även minsta varians (MV) för en skattning av kvoten Y/X studeras.

Mängden av urvalda objekt vid tidpunkt t och som omfattar n objekt betecknas $s_t(n)$.

Begreppet strategi innebär, som tidigare nämnts, kombinationen av en bestämd urvals- och estimationsplan.

Urvalsplan

Tidpunkt t = 1:

Populationen U består av N objekt. n objekt drages med slumpmässigt urval utan återläggning (OSU) av N. Mätningar görs på variabeln x. Vid denna tidpunkt känner man ej populationen U'.

Tidpunkt t = 2:

Populationen U' består av $N' = N_{12} + N_2$ objekt där N_{12} är det gemensamma antalet objekt i U och U' ($U_{12} = U \cap U'$).

Mellan tidpunkterna 1 och 2 har N_1 objekt fallit bort ur populationen U och N_2 har kommit till.

För urvalet vid tidpunkt 1 gäller att n_1 objekt hamnar i delpopulationen U_1=U-U_{12} och n_{12} i U_{12}. n_1 och n_{12} blir därigenom stokastiska variabler. Estimatorer och dess varianser kommer dock fortsättningsvis att vara betingade av n_1 och n_{12}.

Följande urvalsplan gäller vid tidpunkt 2:

n_{12m} av n_{12} objekt drages ur $s_1(n_{12})$ med OSU (m = matched)

n_{12u} av $N_{12}-n_{12}$ objekt drages ur $U_{12} - s_1(n_{12})$ med OSU (u = unmatched)

n_2 av N_2 objekt drages ur U_2 med OSU

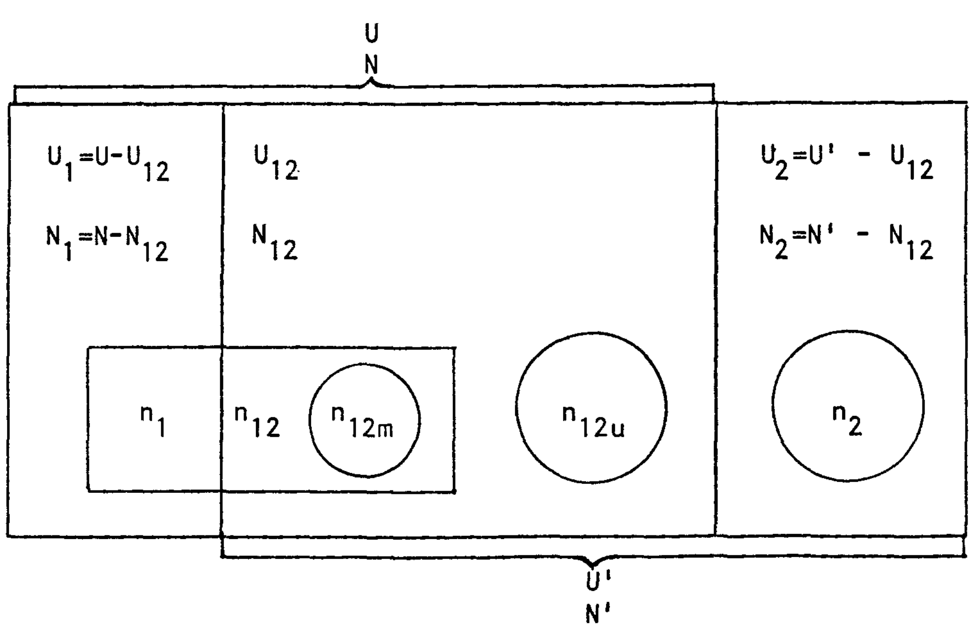
Följande samband gäller för urvalsstorlekarna:

$n = n_1 + n_{12}$

$n' = n_{12m} + n_{12u} + n_2$ (n' är totala urvalsstorleken vid tidpunkt 2)

$n_{12} = n_{12m} + n_{12u}$

Figur över urvalsplanen



Populationsmedelvärden, totaler samt varianser för x och y i delpopulationerna U_1, U_{12} samt U_2 betecknas:

$$\bar{X}_1 = \frac{\overset{N_1}{\Sigma} x_i}{N_1} , \quad X_1 = N_1\bar{X}_1 , \quad S_1^2 = \frac{\overset{N_1}{\Sigma} (x_i-\bar{X}_1)^2}{N_1-1}$$

$$\bar{X}_{12} = \frac{\overset{N_{12}}{\Sigma} x_i}{N_{12}} , \quad X_{12} = N_{12}\bar{X}_{12} , \quad S_{12x}^2 = \frac{\overset{N_{12}}{\Sigma} (x_i-\bar{X}_{12})^2}{N_{12}-1}$$

$$\bar{Y}_{12} = \frac{\overset{N_{12}}{\Sigma} y_i}{N_{12}} , \quad Y_{12} = N_{12}\bar{Y}_{12} , \quad S_{12y}^2 = \frac{\overset{N_{12}}{\Sigma} (y_i-\bar{Y}_{12})^2}{N_{12}-1}$$

$$\bar{Y}_2 = \frac{\overset{N_2}{\Sigma} y_i}{N_2} , \quad Y_2 = N_2\bar{Y}_2 , \quad S_2^2 = \frac{\overset{N_2}{\Sigma} (y_i-\bar{Y}_2)^2}{N_2-1}$$

Fortsättningsvis antages att

$$S_{12x}^2 = S_{12y}^2 = S^2$$

Korrelationskoefficienten ρ mellan x och y i delpopulationen U_{12} kan då skrivas som

$$\rho = \frac{\mathrm{Cov}(x\ ,y\ )}{S^2} \Rightarrow \mathrm{Cov}(x\ ,y\ ) = \rho S^2$$

där $\mathrm{Cov}(x\ ,y\ ) = \dfrac{\overset{N_{12}}{\Sigma} (x_i-\bar{X}_{12})(y_i-\bar{Y}_{12})}{N_{12}-1}$

3 TVÅ ANVÄNDBARA SATSER

Följande två satser kommer att användas vid härledning av LMVU-estimatorer samt deras varianser. I ett härledningsexempel i avsnitt 4.1 framgår det hur dessa satser används.

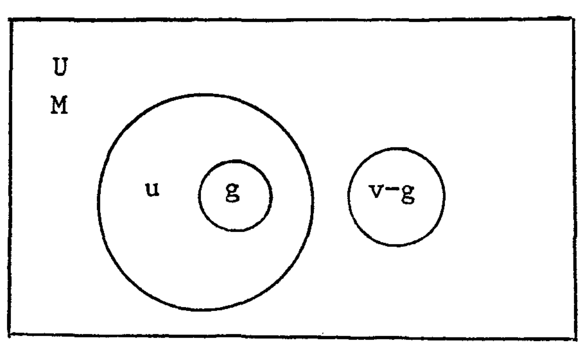
Sats 1:

Betrakta en konstant population U som innehåller M objekt. Följande urvalsplan gäller:

Vid tidpunkt 1 drages med OSU u av M objekt ur U.

Vid tidpunkt 2 drages med OSU g objekt från $s_1(u)$ och ett OSU av v-g objekt från $U-s_1(u)$. (Observera att $s_2(v) = s_2(g) \cup s_2(v-g)$ är ett OSU från U.)

Figur över urvalsplanen:



Undersökningsvariablerna x och y mäts vid tidpunkterna 1 respektive 2. Det gäller också att $\bar{x}_u$ och $\bar{y}_v$ är enkla medelvärdesestimatorer samt att $Cov(x,y)=\rho S^2$

Då är $Cov\ (a\bar{x}_u,\ b\bar{y}_v|g\ ) = ab(\frac{g}{uv} - \frac{1}{M})\rho S^2$

där $g = \#\ s_1(u) \cap s_2(v)$ samt a och b är godtyckliga konstanter.

Bevis av sats 1

Först visas hjälpsatsen

$$\mathrm{Cov}(x_i, y_j) = -\frac{\mathrm{Cov}(x_i, y_i)}{M} \quad \text{då} \quad i \neq j.$$

Bevis:

$$\mathrm{Cov}(x_i, y_j) = E(x_i-\bar{X})(y_j-\bar{Y}) =$$

$$= E\left[(x_i-\bar{X})\ E(y_j-\bar{Y}\,|\,x_i-\bar{X})\right] =$$

$$= \frac{E\left[(x_i-\bar{X})\left(\sum_{j=1}^{M}(y_j-\bar{Y}) - (y_i-\bar{Y})\right)\right]}{M-1} =$$

$$= -\frac{E(x_i-\bar{X})(y_i-\bar{Y})}{M-1} = -\frac{\mathrm{Cov}(x_i, y_i)}{M} \quad \text{V.S.B.}$$

Hjälpsatsen kommer till användning vid bevis av sats 1.

Bevis:

$$\mathrm{Cov}(a\bar{x}_u, b\bar{y}_v\,|\,g) = a\cdot b\ \mathrm{Cov}\left(\frac{\sum^{u} x_i}{u}, \frac{\sum^{v} y_i}{v}\right) =$$

$$= \frac{a\cdot b}{u\cdot v}\, E\left[\sum^{u}(x_i-\bar{X})\sum^{v}(y_i-\bar{Y})\right] = \frac{a\cdot b}{u\cdot v}\, E\left[\sum\sum^{uv}(x_i-\bar{X})(y_i-\bar{Y})\right]$$

$$= \frac{a\cdot b}{u\cdot v}\, E\left[\sum^{g}(x_i-\bar{X})(y_i-\bar{Y}) + \sum\sum_{i\neq j}^{u\cdot v-g}(x_i-\bar{X})(y_j-\bar{Y})\right]$$

$$= \frac{a\cdot b}{u\cdot v}\left[\frac{g\sum^{M}(x_i-\bar{X})(y_i-\bar{Y})}{M} - \frac{(u\cdot v-g)\sum^{M}(x_i-\bar{X})(y_i-\bar{Y})}{M(M-1)}\right] =$$

$$= a\cdot b\left[\frac{(M-1)g}{u\cdot v\cdot M} - \frac{u\cdot v-g}{u\cdot v\cdot M}\right]\mathrm{Cov}(x_i, y_i) = a\cdot b\left(\frac{g}{u\cdot v} - \frac{1}{M}\right)\mathrm{Cov}(x_i, y_i) =$$

$$= a\cdot b\left(\frac{g}{u\cdot v} - \frac{1}{M}\right)\rho\, S^2 \quad \text{då} \quad S_x^2 = S_y^2 = S^2 \qquad \text{V.S.B.}$$

Man får t ex följande specialfall av sats 1:

a) $a=b=M \Rightarrow Cov(\hat{X}_u, \hat{Y}_v | g) = M^2 (\frac{g}{u \cdot v} - \frac{1}{M}) \rho S^2$

b) $s_2(v) \subseteq s_1(u) \wedge a=b=1 \Rightarrow Cov(\bar{x}_u, \bar{y}_v | g=v) = (\frac{1}{u} - \frac{1}{M}) \rho S^2$

c) $s_1(u) \cap s_2(v) = \phi \wedge a=b=1 \Rightarrow Cov(\bar{x}_u, \bar{y}_v | g=0) = -\frac{1}{M} \rho S^2$

d) $Cov(a\bar{x}_u, b\bar{x}_v | g) = a \cdot b (\frac{g}{u \cdot v} - \frac{1}{M}) S^2$

e) $s_2(v) \subseteq s_1(u) \wedge a=b=1 \Rightarrow Cov(\bar{x}_u, \bar{x}_v | g=v) = V(\bar{x}_u) = (\frac{1}{u} - \frac{1}{M}) S^2$

I vårt fall motsvaras u, v och M av n_{12}, n_{12m}, n_{12u} samt N_{12}.
Med hjälp av sats 1 kan man lätt få fram erforderliga covarianstermer.

Följande sats är bevisad av C.R. Rao (1952). Se också Raj (1968).

Sats 2:

Ett nödvändigt och tillräckligt villkor för att T_0 ska vara en MVU-estimator (MVU=minimum-variance-unbiased) för en parameter är att

$Cov\ (T_0, z) = 0$ för alla z

där z är en nollfunktion, dvs $E(z) = 0$.

Följdsats 2.1:

Om T_0 är en MVU-estimator av en parameter och T_1 är en unbiased estimator av parametern så gäller $V(T_0) = Cov(T_0, T_1)$.

Detta följer ur sats 2 eftersom

$Cov(T_0,\ T_0 - T_1) = 0 \Rightarrow Cov(T_0,\ T_0) = V(T_0) = Cov(T_0,\ T_1)$.

Följdsats 2.2:

Låt α och β vara godtyckliga konstanter. Om T_0'' och T_0' är MVU-estimatorer för parametrarna P'' respektive P', då är $D_0 = \alpha T_0'' - \beta T_0'$ en MVU-estimator av $\alpha P'' - \beta P'$.

Detta följer från sats 2 eftersom

$\mathrm{Cov}(T_0'',z) = \mathrm{Cov}(T_0', z) = 0 \Rightarrow$

$\Rightarrow \mathrm{Cov}(D_0,z) = \alpha\, \mathrm{Cov}(T_0'', z) - \beta\, \mathrm{Cov}(T_0', z) = 0$

Sats 2 med följdsatser kommer alltså till användning för att få fram önskade LMVU-estimatorer samt deras varianser eftersom satsen gäller allmänt för MVU-estimatorer.

4 LMVU-ESTIMATORER OCH DESS OPTIMALA VARIANSER

LMVU-skattningar samt dess varianser har härletts för parametrarna $X, \bar{X}$, Y, $\bar{Y}$, $Y-X$ samt $\bar{Y}-\bar{X}$. Dessutom har optimala varianser beräknats genom att minimera LMVU-skattningarnas varianser med avseende på n_{12m} och n_{12u}. Dessa resultat redovisas i avsnitt 4.2.

För att illustrera ovannämnda härledningar genomförs dessa för parametern $\bar{Y}$ i nästa avsnitt.

4.1 Härledning av LMVU-estimator för $\bar{Y}$, dess varians samt optimala varians

Man vill skatta parametern

$$\bar{Y} = \frac{\sum^{N'} y_i}{N'} = \frac{N_{12}\bar{Y}_{12} + N_2\bar{Y}_2}{N'} = Q_2\bar{Y}_{12} + (1-Q_2)\bar{Y}_2 \tag{4.1}$$

där $Q_2 = \frac{N_{12}}{N'}$ och $(1-Q_2) = \frac{N_2}{N'}$

med $\hat{\bar{Y}}_{LMVU}$ enligt tidigare angivna förutsättningar i avsnitt 2.

Ansätt linjärkombinationen

$$L(\underset{\sim}{\bar{x}};\underset{\sim}{\bar{y}}) = a\bar{x}_1 + b\bar{x}_{12} + c\bar{x}_{12m} + d\bar{y}_{12m} + e\bar{y}_{12u} + f\bar{y}_2 \tag{4.2}$$

där $\underset{\sim}{\bar{x}} = (\bar{x}_1, \bar{x}_{12}, \bar{x}_{12m})$ och $\underset{\sim}{\bar{y}} = (\bar{y}_{12m}, \bar{y}_{12u}, \bar{y}_2)$

utgörs av enkla medelvärdesskattningar,

$$\bar{x}_1 = \frac{\sum^{n_1} x_i}{n_1}, \ \bar{x}_{12} = \frac{\sum^{n_{12}} x_i}{n_{12}}, \ldots \text{ osv, inom } U_1, U_{12}, \text{ och } U_2.$$

a, b, c, d, e och f är godtyckliga konstanter.

Se figur och beteckningar i avsnitt 2.

Då $L(\underset{\sim}{\bar{x}};\ \underset{\sim}{\bar{y}})$ ska vara väntevärdesriktig gäller

$$
\begin{aligned}
E(L(\underset{\sim}{\bar{x}};\underset{\sim}{\bar{y}})) &= a\bar{X}_1 + (b+c)\bar{X}_{12} + (d+e)\bar{Y}_{12} + f\,\bar{Y}_2 = \\
&= Q_2\,\bar{Y}_{12} + (1-Q_2)\,\bar{Y}_2 = \bar{Y}
\end{aligned}
\tag{4.3}
$$

(4.3) medför följande restriktioner på konstanterna a, b, c, d, e och f:

$$
\begin{cases}
a = 0 \\
b + c = 0 \\
d + e = Q_2 \\
f = 1 - Q_2
\end{cases}
\tag{4.4}
$$

Insättning av (4.4) i linjärkombinationen (4.2) ger

$$
\begin{aligned}
L(\underset{\sim}{\bar{x}};\underset{\sim}{\bar{y}}) &= \hat{\bar{Y}} = c(\bar{x}_{12m} - \bar{x}_{12}) + d\bar{y}_{12m} + \\
&+ (Q_2 - d)\bar{y}_{12u} + (1-Q_2)\bar{y}_2
\end{aligned}
\tag{4.5}
$$

(4.5) utgör i detta fall klassen av väntevärdesriktiga linjära estimatorer.

För att $\hat{\bar{Y}}$ ska ha minsta varians ska den enligt sats 2 vara okorrelerad med varje nollfunktion, t ex

$$\bar{y}_{12m} - \bar{y}_{12u} \quad \text{och} \quad \bar{x}_{12m} - \bar{x}_{12}$$

Därav följer att

$$\begin{cases} Cov(\hat{\bar{Y}}, \bar{y}_{12m}) = Cov(\hat{\bar{Y}}, \bar{y}_{12u}) \\ Cov(\hat{\bar{Y}}, \bar{x}_{12m}) = Cov(\hat{\bar{Y}}, \bar{x}_{12}) \end{cases} \tag{4.6}$$

Med hjälp av sats 1 kan ekvationssystemet (4.6) med c och d som obekanta skrivas upp.

Ekvationssystemet (4.6) har lösningen:

$$\begin{cases} c = -\rho d \\ d = \dfrac{Q_2(1-\Theta)}{1-\rho^2\Theta^2}, \quad \text{där } \Theta = \dfrac{n_{12u}}{n_{12}} \end{cases} \tag{4.7}$$

Sätts (4.7) in i $\hat{\bar{Y}}$ enligt (4.5) fås LMVU-estimatorn för $\bar{Y}$:

$$\hat{\bar{Y}}_{LMVU} = (1-Q_2)\bar{y}_2 + Q_2\bar{y}_{12u} + d\left[(\bar{y}_{12m}-\bar{y}_{12u}) + \rho(\bar{x}_{12}-\bar{x}_{12m})\right] =$$

$$= \left[d = \frac{Q_2(1-\Theta)}{1-\rho^2\Theta^2}\right] = (1-Q_2)\bar{y}_2 + Q_2\bar{y}_{12u} +$$

$$+ \frac{Q_2(1-\Theta)}{1-\rho^2\Theta^2}\left[(\bar{y}_{12m} - \bar{y}_{12u}) + \rho(\bar{x}_{12} - \bar{x}_{12m})\right] \tag{4.8}$$

Enligt hjälpsats 2.1 är $Cov(\hat{\bar{Y}}_{LMVU}, T_1) = V(\hat{\bar{Y}}_{LMVU})$.

T_1 väljes så att $Cov(\hat{\bar{Y}}_{LMVU}, T_1)$ blir så lätt som möjligt att räkna ut.

Vi väljer t ex $T_1 = Q_2\, \bar{y}_{12u} + (1-Q_2)\, \bar{y}_2$.

Med hjälp av sats 1 får man då

$$V(\hat{\bar{Y}}_{LMVU}) = (1-Q_2)^2\, s_2^2\left(\frac{1}{n_2} - \frac{1}{N_2}\right) + Q_2^2\, s^2\left(\frac{1}{n_{12u}} - \frac{1}{N_{12}}\right) - \frac{dQ_2 s^2}{n_{12u}} =$$

$$= \left[d = \frac{Q_2(1-\theta)}{1-\rho^2\theta^2}\right] = (1-Q_2)^2\, s_2^2\left(\frac{1}{n_2} - \frac{1}{N_2}\right) + Q_2^2\, \frac{s^2}{n_{12}}\left[\frac{1-\rho^2\theta}{1-\rho^2\theta^2} - \frac{n_{12}}{N_{12}}\right] \qquad (4.9)$$

Alternativt kan dessa resultat erhållas utan att använda sats 2:

Man skriver då upp $V(\hat{\bar{Y}})$ med hjälp av sats 1 och minimerar $V(\hat{\bar{Y}})$ med avseende på c och d.

$$V(\hat{\bar{Y}}) = (1-Q_2)^2 s_2^2 \left(\frac{1}{n_2} - \frac{1}{N_2}\right) +$$

$$+ \frac{s^2}{n_{12}}\left[\frac{\theta^2(c^2 + 2cd\rho) + \theta(2d-1) + (Q_2-d)^2}{\theta(1-\theta)} - \frac{n_{12}}{N_{12}}\right] \text{där } \theta = \frac{n_{12u}}{n_{12}},\ \underline{0 \le \theta \le 1}$$

$$\begin{cases} \dfrac{\partial V(\hat{\bar{Y}})}{\partial c} = 0 \\[2ex] \dfrac{\partial V(\hat{\bar{Y}})}{\partial d} = 0 \end{cases}$$

Man löser detta ekvationssystem och kontrollerar att lösningen på c och d ger $\min_{c,d} V(\hat{\bar{Y}})$.

Detta förfarande ger också resultatet (4.9) men tar längre tid.

Optimala variansen för $V(\hat{\bar{Y}}_{LMVU})$ enligt (4.9) fås genom

$$V_{opt}(\hat{\bar{Y}}_{LMVU}) = \min_{n_{12u}, n_{12m}} V(\hat{\bar{Y}}_{LMVU}) = \min_{\theta} V(\hat{\bar{Y}}_{LMVU})$$

där $\theta = \dfrac{n_{12u}}{n_{12}}$; $\underline{0 \le \theta \le 1}$

Lösningen ges av $\theta_{min} = \dfrac{1}{1 + \sqrt{1-\rho^2}}$; $-1 < \rho < 1$ (4.10)

sedan man kontrollerat randpunkterna $\theta=0$ och $\theta=1$.

Resultatet blir då

$$V_{opt}(\hat{\bar{Y}}_{LMVU}) = (1-Q_2)^2\, S_2^2\left(\frac{1}{n_2} - \frac{1}{N_2}\right) + \frac{Q_2^2\, S^2}{n_{12}}\left[\frac{(1+\sqrt{1-\rho^2})}{2} - \frac{n_{12}}{N_{12}}\right] \quad (4.11)'$$

Eftersom $Y = N'\,\bar{Y} = N_{12}\,\bar{Y}_{12} + N_2\,\bar{Y}_2 = Y_{12} + Y_2$

följer nedanstående resultat för totalen Y:

$\hat{X}_1 = N_1\bar{x}_1$, $\hat{X}_{12} = N_{12}\bar{x}_{12}$ osv för delpopulationerna

U_1, U_{12} och U_2.

$$\hat{Y}_{LMVU} = \hat{Y}_2 + \hat{Y}_{12u} + \frac{(1-\Theta)}{1-\rho^2\Theta^2}\left[(\hat{Y}_{12m} - \hat{Y}_{12u}) + \rho(\hat{X}_{12} - \hat{X}_{12m})\right]$$

$$V(\hat{Y}_{LMVU}) = N_2^2\, S_2^2\left(\frac{1}{n_2} - \frac{1}{N_2}\right) + N_{12}^2\,\frac{S^2}{n_{12}}\left[\frac{1-\rho^2\Theta}{1-\rho^2\Theta^2} - \frac{n_{12}}{N_{12}}\right]$$

$$\Theta_{min} = \frac{1}{1+\sqrt{1-\rho^2}} \Rightarrow V_{opt}(\hat{Y}_{LMVU}) = N_2^2\, S_2^2\left(\frac{1}{n_2} - \frac{1}{N_2}\right) +$$

$$+ \quad N_{12}^2\,\frac{S^2}{n_{12}}\left[\frac{1+\sqrt{1-\rho^2}}{2} - \frac{n_{12}}{N_{12}}\right]$$

Motsvarande skattningar för $\bar{X}$ och X och dess varianser härleds analogt.

Skattningarna för differenserna $\bar{Y}-\bar{X}$ samt Y-X härleds också på liknande sätt och hjälpsats 2.2 kommer här till användning.

Observera att man ej kan få resultaten för $\hat{Y}-\hat{X}$ via resultaten för $\bar{Y}-\bar{X}$ och vice versa.

4.2 Resultat för LMVU-estimatorerna

LMVU-estimatorerna,dess varianser och optimala varianser ges i detta avsnitt för parametrarna X, $\bar{X}$, Y, $\bar{Y}$, Y-X och $\bar{Y}$-$\bar{X}$.

Följande beteckningar införs:

- e_{LMVU} = en LMVU-estimator med den i avsnitt 2 beskrivna urvalsplanen

- Optimala variansen för e_{LMVU} betecknas

  $V_{opt}(e_{LMVU}) = \min_{\theta} V(e_{LMVU})$ där $\theta = \frac{n_{12u}}{n_{12}}; 0 \leq \theta \leq 1.$

- θ_{min} = θ-värde som minimerar $V(e_{LMVU})$. Övriga beteckningar,se avsnitt 2.

Följande gäller samtliga nedanstående LMVU-estimatorer:

- $\rho = \pm 1$ och $\theta \to 1 \Rightarrow V(e_{LMVU}) \to V_{opt}(e_{LMVU})$

Skulle man i praktiken veta att ρ = 1 eller -1 använder man naturligtvis en annan strategi och utnyttjar det faktum att om n_{12m} = 2 så kan man exakt bestämma regressionssambandet mellan undersökningsvariablerna x och y i delpopulationen U_{12}.

- ρ = 0 medför att $V_{opt}(e_{LMVU})$ är oberoende av θ .

Pga detta särredovisas ej detta fall då uppdelningar görs för positiva och negativa korrelationer.

- Samtliga formler nedan är giltiga för $-1 < \rho < 1$.

a. Estimator av totalen X vid tidpunkt 1

$$\hat{X}_{LMVU} = \hat{X}_1 + \hat{X}_{12} - \frac{\rho(1-\Theta)\Theta}{1-\rho^2\Theta^2}\left[(\hat{Y}_{12m} - \hat{Y}_{12u}) + \rho(\hat{X}_{12} - \hat{X}_{12m})\right]$$

$$V(\hat{X}_{LMVU}) = N_1^2\, S_1^2\left(\frac{1}{n_1} - \frac{1}{N_1}\right) + N_{12}^2\,\frac{S^2}{n_{12}}\left[\frac{1-\rho^2\Theta}{1-\rho^2\Theta^2} - \frac{n_{12}}{N_{12}}\right]$$

$$\Theta_{min} = \frac{1}{1+\sqrt{1-\rho^2}} \Rightarrow V_{opt}(\hat{X}_{LMVU}) = N_1^2 S_1^2\left(\frac{1}{n_1} - \frac{1}{N_1}\right) + N_{12}^2\,\frac{S^2}{n_{12}}\left[\frac{(1+\sqrt{1-\rho^2})}{2} - \frac{n_{12}}{N_{12}}\right]$$

b. Estimator av medelvärdet $\bar{X}$ vid tidpunkt 1

$$\hat{\bar{X}}_{LMVU} = \hat{X}_{LMVU}/N$$

$$V(\hat{\bar{X}}_{LMVU}) = V(\hat{X}_{LMVU})/N^2;\quad V_{opt}(\hat{\bar{X}}_{LMVU}) = V_{opt}(\hat{X}_{LMVU})/N^2$$

c. Estimator av totalen Y vid tidpunkt 2

$$\hat{Y}_{LMVU} = \hat{Y}_2 + \hat{Y}_{12u} + \frac{(1-\Theta)}{1-\rho^2\Theta^2}\left[(\hat{Y}_{12m} - \hat{Y}_{12u}) + \rho(\hat{X}_{12} - \hat{X}_{12m})\right]$$

$$V(\hat{Y}_{LMVU}) = N_2^2\, S_2^2\left(\frac{1}{n_2} - \frac{1}{N_2}\right) + N_{12}^2\,\frac{S^2}{n_{12}}\left[\frac{1-\rho^2\Theta}{1-\rho^2\Theta^2} - \frac{n_{12}}{N_{12}}\right]$$

$$\Theta_{min} = \frac{1}{1+\sqrt{1-\rho^2}} \Rightarrow V_{opt}(\hat{Y}_{LMVU}) = N_2^2 S_2^2\left(\frac{1}{n_2} - \frac{1}{N_2}\right) + N_{12}^2\,\frac{S^2}{n_{12}}\left[\frac{(1+\sqrt{1-\rho^2})}{2} - \frac{n_{12}}{N_{12}}\right]$$

d. Estimator av medelvärdet $\bar{Y}$ vid tidpunkt 2

$$\hat{\bar{Y}}_{LMVU} = \hat{Y}_{LMVU}/N'$$

$$V(\hat{\bar{Y}}_{LMVU}) = V(\hat{Y}_{LMVU})/N'^2;\quad V_{opt}(\hat{\bar{Y}}_{LMVU}) = V_{opt}(\hat{Y}_{LMVU})/N'^2$$

e. Estimator av differansen Y-X

$$(\hat{Y}-\hat{X})_{LMVU} = \hat{Y}_{LMVU} - \hat{X}_{LMVU} = \hat{Y}_2 - \hat{X}_1 + \hat{Y}_{12u} - \hat{X}_{12} + \frac{(1-\Theta)}{1-\rho\Theta}\left[(\hat{Y}_{12m} - \hat{Y}_{12u}) + \rho(\hat{X}_{12} - \hat{X}_{12m})\right]$$

$$V(\hat{Y}-\hat{X})_{LMVU} = N_2^2\, S_2^2\left(\frac{1}{n_2} - \frac{1}{N_2}\right) + N_1^2 S_1^2\left(\frac{1}{n_1} - \frac{1}{N_1}\right) + \frac{2N_{12}^2 S^2(1-\rho)}{n_{12}}\left[\frac{1}{1-\rho\Theta} - \frac{n_{12}}{N_{12}}\right]$$

$$\underset{0<\rho<1}{\Theta_{min}} = 0 \Rightarrow V_{opt}(\hat{Y}-\hat{X})_{LMVU} = N_2^2\, S_2^2\left(\frac{1}{n_2} - \frac{1}{N_2}\right) + N_1^2\, S_1^2\left(\frac{1}{n_1} - \frac{1}{N_1}\right) + 2N_{12}^2 S^2(1-\rho)\left(\frac{1}{n_{12}} - \frac{1}{N_{12}}\right)$$

$$\underset{-1<\rho<0}{\Theta_{min}} = 1 \Rightarrow V_{opt}(\hat{Y}-\hat{X})_{LMVU} = N_2^2\, S_2^2\left(\frac{1}{n_2} - \frac{1}{N_2}\right) + N_1^2\, S_1^2\left(\frac{1}{n_1} - \frac{1}{N_1}\right) + 2N_{12}^2 \frac{S^2}{n_{12}}\left[1 - \frac{n_{12}}{N_{12}}(1-\rho)\right]$$

f. Estimator av differensen $\bar{Y}$-$\bar{X}$

$$(\hat{\bar{Y}}-\hat{\bar{X}})_{LMVU} = \hat{\bar{Y}}_{LMVU} - \hat{\bar{X}}_{LMVU} = (1\ Q_2)\bar{y}_2 - (1-Q_1)\bar{x}_1 + Q_2\,\bar{y}_{12u} - Q_1\,\bar{x}_{12} +$$

$$+ \quad \frac{(1-\Theta)(Q_2+Q_1\rho\Theta)}{1-\rho^2\Theta^2}\left[(\bar{y}_{12m} - \bar{y}_{12u}) + \rho(\bar{x}_{12} - \bar{x}_{12m})\right]$$

$$V(\hat{\bar{Y}}-\hat{\bar{X}})_{LMVU} = (1-Q_2)^2\, S_2^2\left(\frac{1}{n_2} - \frac{1}{N_2}\right) + (1-Q_1)^2\, S_1^2\left(\frac{1}{n_1} - \frac{1}{N_1}\right) +$$

$$+ \quad \frac{S^2}{n_{12}}\left[\frac{(Q_1^2 + Q_2^2)(1-\rho^2\Theta) - 2\,Q_1Q_2\,\rho\,(1-\Theta)}{1-\rho^2\Theta^2} - \frac{n_{12}}{N_{12}}\,(Q_1^2 + Q_2^2 - 2\,Q_1Q_2\,\rho)\right]$$

För beräkning av Θ_{min} betraktas två fall:

Fall 1: $0 < \rho < 1$

$$V(\Theta) = V(\hat{\hat{Y}}-\hat{\hat{X}})_{LMVU}\,;\ \frac{dV(\Theta)}{d\Theta} = 0 \text{ ger}$$

$$\theta = k \pm \sqrt{k^2 - \ell}\,;\ k = \frac{Q_1^2 + Q_2^2 - 2Q_1Q_2\,\rho}{(Q_1^2 + Q_2^2)\rho^2 - 2Q_1Q_2\,\rho} \quad \text{och} \quad \ell = \frac{1}{\rho^2} \tag{4.2.1}$$

Det kan bevisas att $V(\Theta) < V(1)$ då $0 \leq \Theta < 1$ och $0 < \rho < 1$.

Det följer då att

a. Om det existerar en lösning $\Theta \in [0,1[$ så är $\Theta = \Theta_{min}$, vilket ger $V_{opt}(\hat{\hat{Y}}-\hat{\hat{X}})_{LMVU}$.

b. Om det inte existerar en sådan lösning så är $\Theta_{min} = 0$.

$$\Theta_{min} = 0 \Rightarrow V_{opt}(\hat{\hat{Y}}-\hat{\hat{X}})_{LMVU} = (1-Q_2)^2\, s_2^2\left(\frac{1}{n_2} - \frac{1}{N_2}\right) + (1-Q_1)^2\, s_1^2\left(\frac{1}{n_1} - \frac{1}{N_1}\right) +$$

$$+ \; s^2(Q_1^2 + Q_2^2 - 2\,Q_1Q_2\rho)\left(\frac{1}{n_{12}} - \frac{1}{N_{12}}\right)$$

Fall 2: $-1 < \rho < 0$

Enligt (4.2.1) är $\Theta = k \pm \sqrt{k^2-\ell}$

Det kan bevisas att $V(0) > V(\Theta)$ då $0 < \Theta \leq 1$ och $-1 < \rho < 0$.

Av detta följer att

a. Om det existerar en lösning $\Theta \in]0,1]$ så är $\Theta = \Theta_{min}$, vilket ger $V_{opt}(\hat{\bar{Y}}-\hat{\bar{X}})_{LMVU}$.

b. Om det inte existerar en sådan lösning så är $\Theta_{min} = 1$.

$$\Theta_{min} = 1 \Rightarrow V_{opt}(\hat{\bar{Y}}-\hat{\bar{X}})_{LMVU} = (1-Q_2)^2\, S_2^2\left(\frac{1}{n_2} - \frac{1}{N_2}\right) + (1-Q_1)^2\, S_1^2\left(\frac{1}{n_1} - \frac{1}{N_1}\right) +$$

$$+ \frac{S^2}{n_{12}}\left[Q_1^2 + Q_2^2 - \frac{n_{12}}{N_{12}}\left(Q_1^2 + Q_2^2 - 2\, Q_1 Q_2 \rho\right)\right]$$

Följande specialfall fås då $Q = Q_1 = Q_2$:

$$\left[\begin{array}{l}\Theta_{min} = 0 \\ 0<\rho<1\end{array}\right. \Rightarrow V_{opt}(\hat{\bar{Y}}-\hat{\bar{X}})_{LMVU} = (1-Q)^2\left[S_2^2\left(\frac{1}{n_2} - \frac{1}{N_2}\right) + S_1^2\left(\frac{1}{n_1} - \frac{1}{N_1}\right)\right] +$$

$$+ 2\, Q^2\, S^2\, (1-\rho)\left(\frac{1}{n_{12}} - \frac{1}{N_{12}}\right)$$

$$\left[\begin{array}{l}\Theta_{min} = 1 \\ -1<\rho<0\end{array}\right. \Rightarrow V_{opt}(\hat{\bar{Y}}-\hat{\bar{X}})_{LMVU} = (1-Q)^2\left[S_2^2\left(\frac{1}{n_2} - \frac{1}{N_2}\right) + S_1^2\left(\frac{1}{n_1} - \frac{1}{N_1}\right)\right] +$$

$$+ 2\, Q^2\, \frac{S^2}{n_{12}}\left(1 - \frac{n_{12}}{N_{12}}(1-\rho)\right)$$

4.3 Variansjämförelser vid två olika strategier

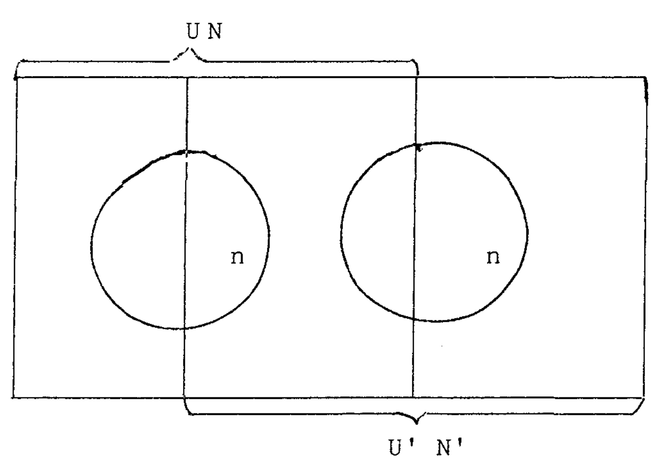
I detta avsnitt skall vi jämföra varianserna för LMVU-estimatorerna för den tidigare beskrivna strategin (här kallad D1) med varianserna för en enklare strategi D0. Syftet med detta är att undersöka hur stora variansreduktioner man kan få om man använder den mer komplicerade strategin D1.

Strategi D0 kan beskrivas på följande sätt:

Vid tidpunkt 1 dras ett OSU omfattande n objekt från de N objekten i U.

Vid tidpunkt 2 dras ett nytt OSU omfattande n objekt från de N' objekten i U', oberoende av stickprovet vid tidpunkt 1. Observera att dessa urval slumpmässigt kan överlappa varandra.

Figur över urvalsplanen i D0:



Följande förenklande antaganden görs vid jämförelserna av varianserna vid de båda strategierna:

a. $Q_1 = Q_2 = Q \Rightarrow N = N'$ och $N_1 = N_2$

b. $0 \leq \rho < 1$

c. $n_1 = n_2$

d. $\frac{n_2}{n} = \frac{N_2}{N} = 1-Q$ och $\frac{n_{12}}{n} = \frac{N_{12}}{N} = Q$

e. Populationsvarianserna för U, U', U_1, U_{12} och U_2 är lika:
$S_U^2 = S_{U'}^2 = S_1^2 = S_{12x}^2 = S_{12y}^2 = S_2^2 = S^2$.

f. N, N', N_1, N_{12} och N_2 är stora och stickprovsfraktionerna

$\frac{n}{N} = \frac{n}{N'}$, $\frac{n_1}{N_1} = \frac{n_2}{N_2}$, $\frac{n_{12u}}{N_{12}}$, $\frac{n_{12m}}{N_{12}}$ och $\frac{n_{12}}{N_{12}}$ är ignorerbara.

Då är estimatorerna och deras varianser i strategin DO som följer

$$\hat{Y}_{DO} = \frac{N\sum^{n} y_i}{n}, \quad \hat{\hat{Y}}_{DO} = \hat{Y}_{DO}/N \quad \text{vid tidpunkt 2}$$

$$\hat{X}_{DO} = \frac{N\sum^{n} x_i}{n}, \quad \hat{\hat{X}}_{DO} = \hat{X}_{DO}/N \quad \text{vid tidpunkt 1}$$

$$V(\hat{Y}_{DO}) = V(\hat{X}_{DO}) = N^2 \frac{S^2}{n}$$

$$V(\hat{\hat{Y}}_{DO}) = V(\hat{\hat{X}}_{DO}) = \frac{S^2}{n}$$

$$V(\hat{Y}_{DO} - \hat{X}_{DO}) = 2N^2 \frac{S^2}{n}$$

$$V(\hat{\hat{Y}}_{DO} - \hat{\hat{X}}_{DO}) = 2 \frac{S^2}{n}$$

Vi kan då uttrycka jämförelser av varianser genom följande varianskvoter:

$$\frac{V(\hat{Y}_{LMVU})}{V(\hat{Y}_{DO})} = \frac{V(\hat{\hat{Y}}_{LMVU})}{V(\hat{\hat{Y}}_{DO})} = \frac{V(\hat{X}_{LMVU})}{V(\hat{X}_{DO})} = \frac{V(\hat{\hat{X}}_{LMVU})}{V(\hat{\hat{X}}_{DO})} = (1-Q) + Q\frac{1-\rho^2\Theta}{1-\rho^2\Theta^2} \tag{4.3.1}$$

$$\frac{V(\hat{Y}-\hat{X})_{LMVU}}{V(\hat{Y}_{DO}-\hat{X}_{DO})} = \frac{V(\hat{\hat{Y}}-\hat{\hat{X}})_{LMVU}}{V(\hat{\hat{Y}}_{DO}-\hat{\hat{X}}_{DO})} = (1-Q) + Q\frac{(1-\rho)}{1-\rho\Theta} \tag{4.3.2}$$

Om vi sätter in optimala värden för Θ, dvs $\Theta = \dfrac{1}{1+\sqrt{1-\rho^2}}$

i (4.3.1) och $\Theta = 0$ i (4.3.2), kan dessa kvoter skrivas

$$\frac{V_{opt}(\hat{Y}_{LMVU})}{V(\hat{Y}_{DO})} = \frac{V_{opt}(\hat{\bar{Y}}_{LMVU})}{V(\hat{\bar{Y}}_{DO})} = \frac{V_{opt}(\hat{X}_{LMVU})}{V(\hat{X}_{DO})} = \frac{V_{opt}(\hat{\bar{X}}_{LMVU})}{V(\hat{\bar{X}})} = (1-Q) + Q\frac{(1+\sqrt{1-\rho^2})}{2} \qquad (4.3.3)$$

och

$$\frac{V_{opt}(\hat{Y}-\hat{X})_{LMVU}}{V(\hat{Y}_{DO}-\hat{X}_{DO})} = \frac{V_{opt}(\hat{\bar{Y}}-\hat{\bar{X}})_{LMVU}}{V(\hat{\bar{Y}}_{DO}-\hat{\bar{X}}_{DO})} = 1 - Q\rho \qquad (4.3.4)$$

Tabell 1: *Variansreduktion genom användning av strategin D1 i stället för D0 vid estimation av parametrarna* Y, $\bar{Y}$, X, $\bar{X}$. Jämförelsen uttrycks enligt (4.3.3).

Q \ ρ	0.99	0.80	0.60	0.40	0.20
1	0.57	0.80	0.90	0.96	0.99
0.99	0.57	0.80	0.90	0.96	0.99
0.90	0.61	0.82	0.91	0.96	0.99
0.80	0.66	0.84	0.92	0.97	0.99
0.70	0.70	0.86	0.93	0.97	0.99
0.60	0.74	0.88	0.94	0.97	0.99
0.50	0.79	0.90	0.95	0.98	0.99

Tabell 2: Variansreduktioner genom användning av strategin D1 i stället för D0 vid estimation av parametrarna Y-X och $\bar{Y}$-$\bar{X}$. Jämförelsen uttrycks enligt (4.3.4).

Q \ ρ	0.99	0.80	0.60	0.40	0.20
1	0.01	0.20	0.40	0.60	0.80
0.99	0.02	0.21	0.41	0.60	0.80
0.90	0.11	0.28	0.46	0.64	0.82
0.80	0.21	0.36	0.52	0.68	0.84
0.70	0.31	0.44	0.58	0.72	0.86
0.60	0.41	0.52	0.64	0.76	0.88
0.50	0.50	0.60	0.70	0.80	0.90

Slutsatserna vid ovan gjorda jämförelser med optimala θ-värden är att:

a. vid estimation av nivåer är D1 överlägsen D0 om Q och ρ är stora ($\rho \geq 0.80$ och $Q \geq 0.80$). (Samma resultat gäller för negativa ρ.)

b. vid estimation av differenser kan stora variansreduktioner erhållas då $\rho \geq 0.40$ och $Q \geq 0.60$. (För negativa värden på ρ är varianserna lika stora för de båda strategierna.)

c. om ρ är positiv är variansreduktionerna för estimatorerna av differenserna större än de för nivåerna.

Diagrammen 1 och 2 visar funktionerna (4.3.1) och (4.3.2) då $0 \leq \theta \leq 1$ och för olika värden på Q och ρ. För givna θ-värden kan precisionsvinsten vid användandet av D1 enkelt utläsas. Diagrammen illustrerar också problemet att välja lämpligt θ-värde då effektiva estimatorer för nivåer och differenser önskas samtidigt. Sådana diagram är lämpliga att använda då man gör avvägningen mellan nivå- och förändringsskattningar.

Diagram 1:

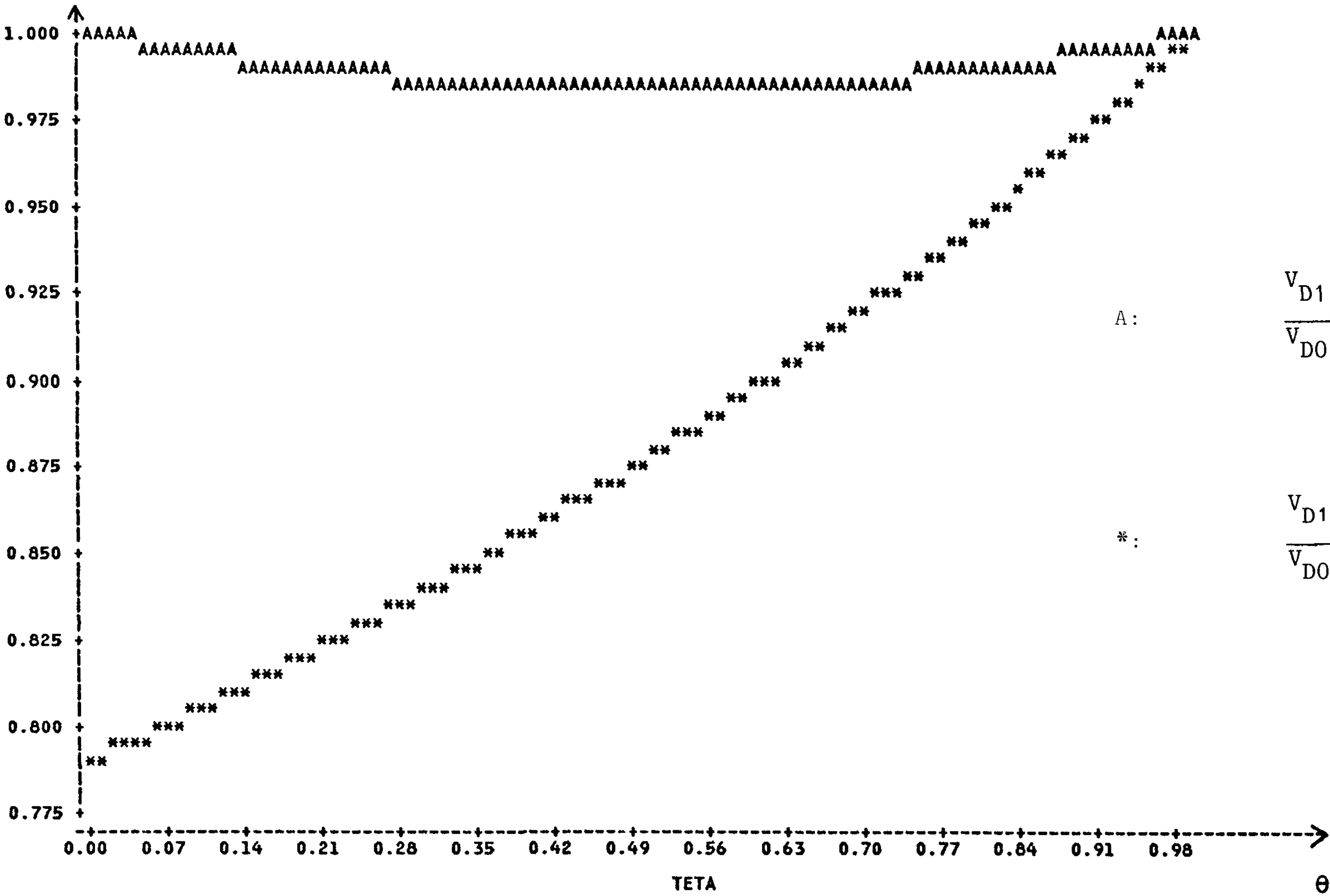
S T A T I S T I C A L   A N A L Y S I S   S Y S T E M   1

10:00 THURSDAY, JANUARY 7, 1982

Q=0.7   RO=0.3

PLOT OF F*TETA   LEGEND: A = 1 OBS, B = 2 OBS, ETC.
PLOT OF G*TETA   SYMBOL USED IS *

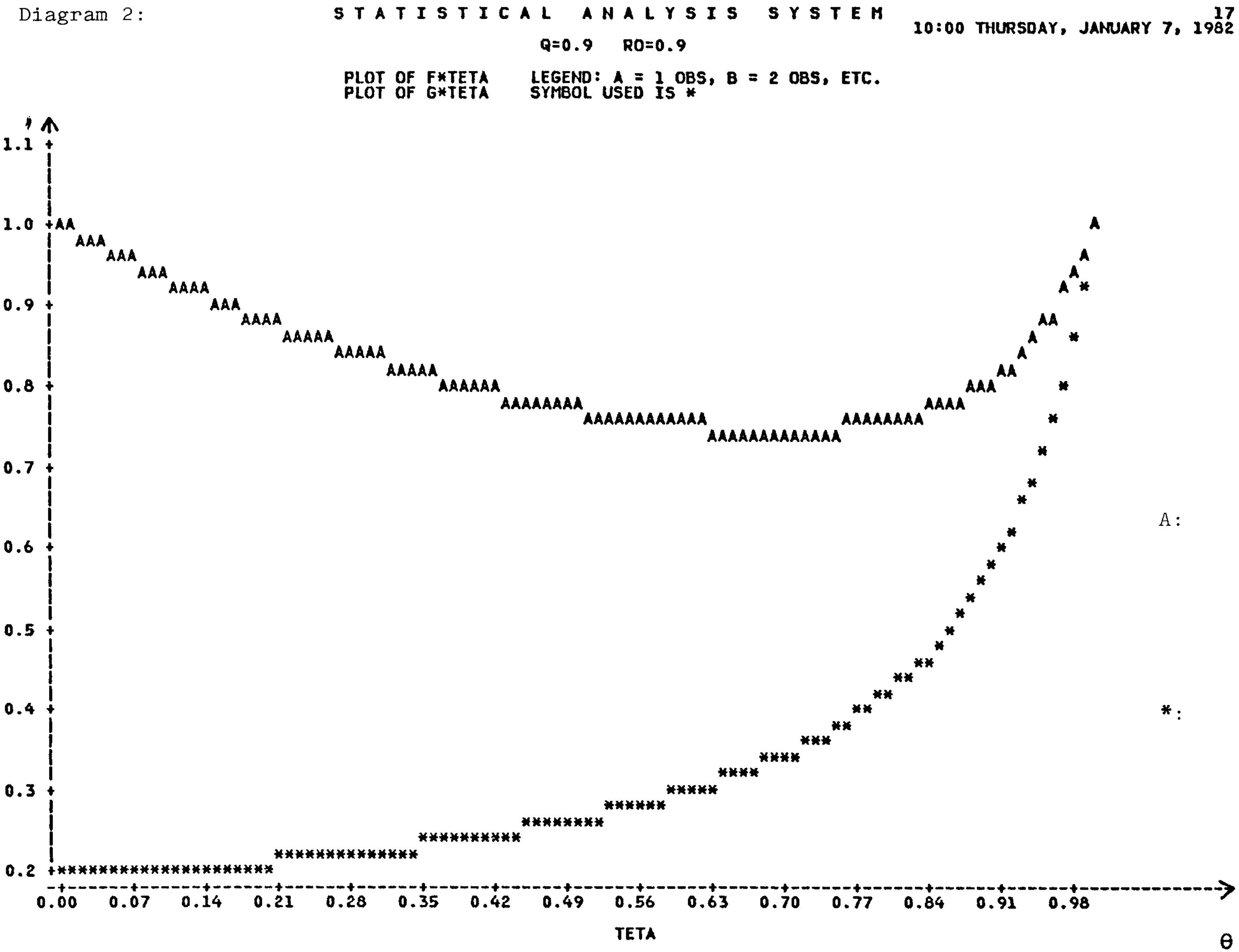
$Q = 0.7$
$\rho = 0.3$



A: $\dfrac{V_{D1}(\text{LMVU-est. för nivå})}{V_{D0}(\text{est. för nivå})}$ enligt funktion (4.3.1)

*: $\dfrac{V_{D1}(\text{LMVU-est. för differens})}{V_{D0}(\text{estimator för differens})}$ enligt funktion (4.3.2)

Diagram 2:



Q = 0.9

ρ = 0.9

A: $\dfrac{V_{D1}(\text{LMVU-est. för nivå})}{V_{D0}(\text{estimator för nivå})}$ enligt funktion (4.3.1)

*: $\dfrac{V_{D1}(\text{LMVU-est. för differens})}{V_{D0}(\text{estimator för differens})}$ enligt funktion (4.3.2)

5 KVOTESTIMATORN, DESS VARIANS SAMT OPTIMALA VARIANS

Förändringar uttrycks oftare som kvoter än som differenser. Betrakta kvoten $R = \frac{Y}{X}$ och dess estimator

$$\hat{R} = \frac{\hat{Y}}{\hat{X}} = \frac{a(\hat{X}_{12m} - \hat{X}_{12}) + b\,\hat{Y}_{12m} + (1-b)\,\hat{Y}_{12u} + \hat{Y}_2}{\hat{X}_1 + (1-c)\,\hat{X}_{12} + c\,\hat{X}_{12m} + d\,(\hat{Y}_{12m} - \hat{Y}_{12u})} \tag{5.1}$$

där $E(\hat{Y}) = Y, E(\hat{X}) = X$; a, b, c och d är godtyckliga konstanter.

Väntevärdet för $\hat{R}$ ges av $E(\hat{R}) = R - \frac{1}{X}\,Cov(\hat{R},\hat{X})$. $\hat{R}$ är alltså inte en väntevärdesriktig skattning. Eftersom $\frac{|bias\ i\ \hat{R}|}{\sqrt{V(\hat{R})}} \leq \frac{\sqrt{V(\hat{X})}}{X} = CV(\hat{X})$ kan biasen ignoreras om CV för $\hat{X}$ är liten, t ex mindre än 0.1. (Se Cochran (1977) sid 162.) Detta antages gälla i detta avsnitt. Eftersom $\hat{R}$ är en kvotestimator behandlas ej fallet med negativt ρ.

Variansen för $\hat{R}$ är

$$V(\hat{R}) \doteq V\left(\frac{\hat{Y}-R\hat{X}}{X}\right) = \frac{1}{X^2}\left[V(\hat{Y}) + R^2V(\hat{X}) - 2RCov(\hat{Y},\hat{X})\right] \tag{5.2}$$

$\hat{R}$ har alltså approximativt samma varians som $\left(\frac{\hat{Y}-R\hat{X}}{X}\right)$. Eftersom $\hat{Y}$ och $\hat{X}$ är väntevärdesriktiga estimatorer kan vi använda hjälpsats 2.2 (med $\alpha = \frac{1}{X}$ och $\beta = \frac{R}{X}$) vilket leder till $\left(\frac{\hat{Y}-R\hat{X}}{X}\right)_{LMVU} = \left(\frac{\hat{Y}_{LMVU}-R\hat{X}_{LMVU}}{X}\right)$.

Med beteckningen

$$\hat{R}_{MV} = \frac{\hat{Y}_{LMVU}}{\hat{X}_{LMVU}} \tag{5.3}$$

(MV = minimum variance)

får vi

$$V(\hat{R}_{MV}) \doteq V\left(\frac{\hat{Y}_{LMVU}-R\hat{X}_{LMVU}}{X}\right) = \frac{1}{X^2}\left[V(\hat{Y}_{LMVU}) + R^2V(\hat{X}_{LMVU}) - 2RCov(\hat{Y}_{LMVU},\hat{X}_{LMVU})\right] \quad (5.4)$$

I avsnitt 4.2 har vi beräknat

$\hat{Y}_{LMVU}$, $\hat{X}_{LMVU}$, $V(\hat{Y}_{LMVU})$, $V(\hat{X}_{LMVU})$ samt $V(\hat{Y}-\hat{X})_{LMVU} = V(\hat{Y}_{LMVU}-\hat{X}_{LMVU})$.

Kovarianstermen kan då beräknas eftersom

$$Cov\ (\hat{Y}_{LMVU}, \hat{X}_{LMVU}) = \frac{1}{2}\left[V(\hat{Y}_{LMVU}) + V(\hat{X}_{LMVU}) - V(\hat{Y}-\hat{X})_{LMVU}\right].$$

Genom insättning av dessa uttryck i (5.3) och (5.4) får vi för $0 \leq \rho < 1$:

$$\hat{R}_{MV} = \frac{\hat{Y}_2 + \hat{Y}_{12u} + \frac{(1-\Theta)}{(1-\rho^2\Theta^2)}\left[(\hat{Y}_{12m} - \hat{Y}_{12u}) + \rho(\hat{X}_{12} - \hat{X}_{12m})\right]}{\hat{X}_1 + \hat{X}_{12} - \frac{\rho\Theta(1-\Theta)}{(1-\rho^2\Theta^2)}\left[\hat{Y}_{12m} - \hat{Y}_{12u}) + \rho(\hat{X}_{12} - \hat{X}_{12m})\right]} = \frac{\hat{Y}_{LMVU}}{\hat{X}_{LMVU}} \quad (5.5)$$

$$V(\hat{R}_{MV}) \doteq \frac{1}{X^2}\left[N_2^2\ S_2^2\ (\frac{1}{n_2} - \frac{1}{N_2}) + R^2\ N_1^2\ S_1^2\ (\frac{1}{n_1} - \frac{1}{N_1}) - N_{12}^2\ \frac{S^2}{n_{12}}\left[\frac{n_{12}}{N_{12}}\ (1+R^2-2\ R\rho)\right] + \right.$$

$$\left. + N_{12}^2\ \frac{S^2}{n_{12}}\left[\frac{(1+R^2)(1-\rho^2\Theta)-2R\rho(1-\Theta)}{(1-\rho^2\Theta^2)}\right]\right] \quad (5.6)$$

Minimumvariansen för $\hat{R}_{MV}$ erhålles genom att minimera $V(\hat{R}_{MV})$ med avseende på Θ. Detta leder till en andragradsekvation i Θ.

Låt $R > 0, 0 < \rho < 1$ och $V(\Theta) = V(\hat{R}_{MV})$.

Då leder $\frac{dV(\Theta)}{d\Theta} = 0$ till

$$\Theta = k \pm \sqrt{k^2 - \ell}; \quad k = \frac{(1+R^2-2R\rho)}{\rho(\rho+\rho R^2-2R)} \quad \text{och } \ell = \frac{1}{\rho^2}$$

Det kan visas att $V(\Theta) < V(1)$ när $0 \leq \Theta < 1$, $R>0$ och $0<\rho<1$. Pga detta kan vi beräkna optimala Θ-värden.

a. Om det existerar en lösning $\Theta \in [0,1[$ så är $\Theta = \Theta_{min}$, vilket ger $V_{opt}(\hat{R}_{MV})$.

b. Om det inte existerar en sådan lösning är $\Theta_{min} = 0$.

$$\Theta_{min} = 0 \Rightarrow$$

$$\hat{R}_{MV} = \frac{\hat{Y}_2 + \hat{Y}_{12}}{\hat{X}_1 + \hat{X}_{12}}$$

$$V_{opt}(\hat{R}_{MV}) \doteq \frac{1}{X^2}\left[N_2^2 S_2^2 \left(\frac{1}{n_2} - \frac{1}{N_2}\right) + R^2 N_1^2 S_1^2 \left(\frac{1}{n_1} - \frac{1}{N_1}\right) + N_{12}^2 S^2 \left(\frac{1}{n_{12}} - \frac{1}{N_{12}}\right)(1 + R^2 - 2R\rho)\right]$$

$\rho = 0 \Rightarrow V_{opt}(\hat{R}_{MV})$ är oberoende av Θ.

$\rho = 1$ och $\Theta \to 1 \Rightarrow V(\hat{R}_{MV}) \to V_{opt}(\hat{R}_{MV})$.

Som exempel visas optimala Θ-värden i tabellen nedan för olika värden på R och ρ.

Tabell: Optimala Θ-värden (gäller även då populationen är konstant). $\Theta_{min} \Rightarrow V_{opt}(\hat{R}_{MV})$

R \ ρ	0,01	0,10	0,25	0,40	0,50	0,75	0,90	0,99
0,1	0	0	0,11	0,28	0,35	0,49	0,62	0,85
0,5	0	0	0	0	0	0	0,21	0,66
1	0	0	0	0	0	0	0	0
1,5	0	0	0	0	0	0	0	0,48
2	0	0	0	0	0	0	0,21	0,66
5	0	0	0	0,02	0,14	0,37	0,54	0,82
10	0	0	0,11	0,28	0,35	0,49	0,62	0,85

6 SAMMANFATTNING OCH KOMMENTARER

I SCBs löpande undersökningar förekommer oftast att man för successiva mättidpunkter önskar estimera nivåer och förändringar. I detta sammanhang har man sällan beaktat problematiken med föränderliga populationer samt konflikten att effektivt skatta både nivåer och förändringar med samma urvalsplan.

Rapporten har behandlat dessa problem under följande förutsättningar:

a Föränderlig population

b Ändlig populationsstorlek

c Två successiva tidpunkter

d Inga ram-, mät- och bortfallsproblem förekommer.

Strategier har angivits för att på effektivt sätt estimera nivåer och förändringar. Dessa strategier bygger på matchade urvalsförfaranden mellan tidpunkter och leder för nivåer till sammansatta estimatorer bestående av två OSU-estimatorer och en två-fas regressionsestimator. Förändringsskattningarna består av differanser eller kvoter av sådana estimatorer.

Vid en jämförelse mellan två strategier med matchade respektive oberoende urval får man att stora variansreduktioner är möjliga med den matchade strategin trots att populationen är föränderlig. Speciellt gäller detta förändringsskattningarna. För nivåerna är variansreduktionen måttligare. Strategin med matchade urval ger också möjlighet att i efterhand förbättra tidigare skattningar.

Då man för samma urvalsplan önskar effektivt skatta förändringar och nivåer uppstår en konflikt och man måste då kompromissa angående strategin enligt de prioriteringar man har mellan de två skattningstyperna. Det är angivet i rapporten hur man kan förfara i denna situation.

Prioriteringsproblem kan också uppstå mellan olika förändringsskattningar samt naturligtvis då man har flera undersökningsvariabler vid en given tidpunkt att ta hänsyn till (massvariabelproblemet).

För användning av strategier med matchade urval krävs att man på förhand åtminstone har ungefärlig ännedom om t ex korrelationskoefficienter mellan undersökningsvariabler vid olika tidpunkter samt om in- och utflöde av populationsobjekt över tiden.

Sådan information ackumuleras vid successiva undersökningar och kan utnyttjas vid fastställandet av strategin.

Det är en fördel om en viss reguljäritet finns angående den nämnda hjälpinformationen. Man behöver då inte råka ut för att förutsättningarna för effektiv estimation snabbt förändras.

Då det gäller att använda strategier med matchade urval i SCBs undersökningar måste stratifierad urvalsplan, fler än två successiva mättidpunkter samt ram-, mät- och bortfallsproblem inkorporeras i strategin. Man kan även tänka sig att använda hjälpinformation från ramen i estimatorerna samtidigt som man har en matchad urvalsplan.

Sett i detta perspektiv är rapporten att uppfatta som ett första led i arbetet att konstruera praktiskt användbara strategier.

Syftet med att använda matchade strategier i SCBs löpande statistikgrenar är att reducera kostnaderna med oförändrade precisionskrav, samtidigt som man beaktar alla avvägningar som måste göras mellan nivå- och förändringsskattningar samt mellan olika undersökningsvariabler.

Denna rapport finns även i en kortare engelsk version, se Forsman och Garås (1982).

7 LITTERATURFÖRTECKNING

Nedan följer en förteckning över litteratur angående estimation vid successiva tidpunkter och därmed närbesläktade problemområden.

Abraham, T.P., Khosla, R.K. and Kathuria, O.P. (1969): Some investigations on the use of successive sampling in pest and disease surveys. Journal of the Indian Society of Agricultural Statistics, 21, p. 43-57.

Anderson, H. (1978): Estimation in repeated sampling from a population with varying composition. Department of Mathematical Statistics, University of Lund.

Arnab, R. (1979): On strategies of sampling finite populations on successive occations with varying probabilities. Sankhyā ser. C, 41, p. 141-155.

Arnab, R. (1980): Two-stage sampling over two occasions. Australian Journal of Statistics, 22, p. 349-357.

Arnab, R. (1981): Sampling on two occations with varying probabilities. Australian Journal of Statistics, 23, p. 360-364.

Bailar, B.A. (1975): The Effects of Rotation Group Bias on estimates from panel surveys, Journal of the American Statistical Association, 70, No 349, p. 23-30.

Bailar, B.A. (1980): Rotation sample biases and their effects on estimates of change. U.S. Bureau of the Census.

Bershad, M.A. (1952): Best linear estimate - sampling for a time series. Unpublished manuscript, U.S. Bureau of the Census, Washington, DC.

Bershad, M.A. and Nisselson, H. (1962): Sampling for weekly time series. The American Statistical Association, Proceedings of the Social Statistics Section 1962, p. 207-213.

Blight, B.J.N. and Scott, A.J. (1973): A stochastic model for repeated surveys. Journal of the Royal Statistical Society B, vol. 35, p. 61-66.

Bäcklund, S. (1972): Tillämpning av Jales metod vid stratifierat urval. Skattning av en förändringsparameter. Bestämning av variansen i denna skattning, när enheter byter stratum mellan två tidpunkter. PM, SCB.

Cassel, C. (1978): The limiting variances and covariances for composite estimation of parameters of domains of study in repeated sampling from a changing population. National Central Bureau of Statistics, Sweden.

Cassel, C. (1979): Varianser för tre urvalstyper vid föränderliga populationer. PM, SCB.

Coburn, T.C. and Warde, W.D. (1980): A sample survey design to estimate the means of two overlapping subpopulations. American Statistical Association. Proceedings of the section on survey research methods. p 188-193.

Cochran, W.G. (1977): Sampling Techniques. 3rd ed, Wiley. New York.

Cohen, S.E. (1969): A device for sampling from finite populations on sucessive occasions. The American Statistician, vol. 23 No. 4, p. 26-27.

Dahmström, P. (1971): Sampling and estimation for surveys on several occasions - a review. Statistical Review 1971:1, p. 25-47.

Dahmström, P. och Malmberg, S. (1971): Utredningsarbetet rörande sammansatta skattningar i AKU. PM, SCB.

Eckler, A.R. (1955): Rotation sampling. The Annals of Mathematical Statistics, 26, p. 664-685.

Fellegi, I.P. (1963): Sampling with varying probabilities without replacement: rotating and non-rotating samples. Journal of the American Statistical Association, 58, p. 183-201.

Forsman, G. and Garås, T. (1982): Optimal estimation of change in sample surveys. Research report. Statistics Sweden.

Fuller, W.A. and Burmeister, L.F. (1972): Estimators for samples selected from two overlapping frames. The American Statistical Association, Proceedings of the Social Statistics Section, p. 245-249.

Garås, T. (1981): Förändringsskattningar. PM, SCB.

Gleason, C.P. (1980): An estimator of change from an updated frame. American Statistical Association. Proceedings of the section on survey research methods, p. 242-244.

Goldstein, H. (1979): The Design and Analysis of Longitudinal Studies. Academic Press. London.

Graham, J.E. (1973): Composite estimation in two cycle rotation sampling designs. Communications in Statistics, 1, p. 419-431.

Gurney, M. and Daly, J.F. (1965): A multivariate approach to estimation in periodic sample surveys. American Statistical Association, Proceedings of the Social Statistical Section, p. 242-257.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953): Sample Survey Methods and Theory. Vol. I and II. New York.

Hansen, M.H., Hurwitz, W.N., Nisselson, H. and Steinberg, J. (1955): The redesign of the census current population survey. Journal of the American Statistical Association, 50, p. 701-719.

Hanson, R.H. (1978): The Current Population Survey - Design and Methodology. Technical paper 40, U.S. Bureau of the Census, Washington, DC.

Hartley, H.O. (1962): Multiple frame surveys. The American Statistical Assoiation, Proceedings of the Social Statistics Section, p. 203-206.

Hartley, H.O. (1974): Multiple frame methodology and selected applications. Sankhyā C, vol. 36, p. 99-118.

Huang, E.T. and Ernst, L.R. (1981): Comparison of an alternative estimator to the current composite estimator in CPS. U.S. Bureau of the Census.

Isaki, C.T., Wolter, K.M., Sturdevant, T.R., Monsour, N.J. and Trager, M.L. (1976): Sample redesign of the Census Bureaus Monthly Business Surveys. American Statistical Association, Proceedings of the Business and Economic Statistics Section, p. 90-98.

Jessen, R.J. (1942): Statistical investigation of a sample survey for obtaining farm facts. Iowa State College of Agriculture and Mechanic Arts, Research Bulletin 304, June 1942, Ames, Iowa.

Kish, L. (1965): Survey Sampling, Wiley, New York.

Konijn, H.S. (1973): Statistical Theory of Sample Survey Design and Analysis. North-Holland, Amsterdam.

Kulldorff, G. (1954): Estimation in repeated sample surveys and optimum allocation of the samplew. Fil. lic. thesis. University of Lund.

Kulldorff, G. (1963): Some problems of optimum allocation for sampling on two occasions. Review of the International Statistical Inst., 31, p. 24-57.

Kulldorff, G. (1979): Optimum allocation for sampling on many occasions. Invited paper 42nd session of ISI, Manila, 1979.

Lindström, H., Thorslund, M. och Wärneryd, B. (1981): Förändringsskattningar i ULF. SCB.

Loverre, J.M. (1979): Sampling for change. American Statistical Association. Proceedings of the section on survey research methods, p. 343-346.

Narain, R.D. (1953): On the recurrence formula in sampling on successive occasions. Journal of the Indian Society of Agricultural Statistics, 5, p. 96-99.

Narain, R.D. (1954): The general theory of sampling on successive occasions. Bulletin of the International Statistical Institute, 34, p. 87-89.

Narain, R.D. (1954): On the recurrence formula in sampling on successive occasions. Bulletin of the International Statistical Institute, 34, p. 201-202.

Nilsson, P. (1969): Om användningen av roterande urvalsmönster successivt återkommande statistiska undersökningar. Handelshögskolan, Göteborg.

Ohlsson, E.B. (1966): Some problems of estimation when sampling for time series (In Swedish). Statistisk tidskrift (Statistical Review), 4, p. 104-114.

Patterson, H.D. (1950): Sampling on successive occasions with partial replacement of units. Journal of Royal Statistical Society ser. B, 12, p. 241-255.

Prabhu-Ajgaonkar, S.G. and Tikkiwal, B.D. (1961): On the theory of univariate successive sampling. Abstract. The Annals of Mathematical Statistics, 32, p. 340.

Prabhu-Ajgaonkar, S.G. (1968): The theory of univariate sampling on successive occasions under the general correlation pattern. The Australian Journal of Statistics, 10, p. 56-63.

Raj, D. (1965): On sampling over two occasions with probability proportionate to size. The Annals of Mathematical Statistics, 36, p. 327-330.

Raj, D. (1968): Sampling Theory. McGraw-Hill, New York.

Rao, C.R. (1952): Some Theorems on Minimum Variance Unbiased Estimation. Sankhyā, 12.

Rao, G.D. (1966): On sampling schemes to estimate difference in prices on two occasions. Sankhyā ser. B, 28, p. 219-228.

Rao, J.N.K. and Graham, J.E. (1964): Rotation designs for sampling on repeated occasions. Journal of the American Statistical Association, 59, p. 492-509.

Rao, J.N.K. and Shimizugawa, H. (1968): A note on sampling over two occasions. Journal of the Indian Society of Agricultural Statistics, 20, p. 15-21.

Råbäck, G. (1980): Varianser för tre urvalstyper vid föränderliga populationer. PM. SCB.

Saito, K. (1957): Some results in the theory of sampling on successive occasions with partial replacement of units. Reports of Statistical Application Research, 4, p. 125-131.

Scott, A.J. and Smith, T.M.F. (1974): Analysis of repeated surveys using time series methods. Journal of the American Statistical Association, 69, p. 674-678.

Scott, A.J., Smith, T.M.F. and Jones, R.G. (1977): The application of time series methods to the analysis of repeated surveys. Int. Stat. Rev., 45, p. 13-28.

Sen, A.R. (1972): Successive sampling with p ($p \geq 1$) auxiliary variables. Annals of Mathematical Statistics, 43, p. 2031-2034.

Sen, A.R. (1973 a): Theory and application of sampling on repeated occasions with several auxiliary variables. Biometrics, 29, p. 381-385.

Sen, A.R. (1973 b): Some theory of sampling on successive occasions. Australian Journal of Statistics, 15, p. 105-110.

Singh, D. (1968): Estimates in successive sampling using a multi-stage design. Journal of the American Statistical Association, 63, p. 99-112.

Singh, D. and Kathuria, O.P. (1969): On two-stage successive sampling. The Australian Journal of Statistics, 11, p. 59-66.

Singh, D. and Singh, B.D. (1965): Double sampling for stratification on successive occasions. Journal of the American Statistical Association, 60, p 784-792.

Technical Paper No. 7 (1963): U.S. Bureau of the Census, The Current Population Survey - A Report on Methodology. Technical Paper No. 7. U.S. Government Printing Office, Wash., DC, 1963.

Thorburn, D. (1981): Framskrivningar med hjälp av longitudinella studier och övergångsmatriser. Statistisk Tidskrift 1981:3, sid. 189-201.

Tikkiwal, B.D. (1953): Optimum allocation in successive sampling. Journal of the Indian Society of Agricultural Statistics, 5, p. 100-102.

Tikkiwal, B.D. (1954): Theory of successive multiphase sampling. (Preliminary report.) Abstract. The Annals of Mathematical Statistics, 25, p. 174.

Tikkiwal, B.D. (1955): On the efficiency of estimates in successive multiphase sampling. (Preliminary report.) Abstract. The Annals of Mathematical Statistics, 26, p. 158.

Tikkiwal, B.D. (1956): A further contribution to the theory of univariate sampling on successive occasions. Journal of the Indian Society of Agricultural Statistics, 8, p. 84-90.

Tikkiwal, B.D. (1958): An examination of the effect of matched sampling on the efficiency of estimators in the theory of successive sampling. Journal of the Indian Society of Agricultural Statistics, 10, p. 16-22.

Tikkiwal, B.D. (1958): Theory of successive two-stage sampling. (Preliminary report.) Abstract. The Annals of Mathematical Statistics, 29, p. 1291.

Tikkiwal, B.D. (1964): A note on two-stage sampling on successive occasions. Sankhyā ser. A, 26, p. 97-100.

Tikkiwal, B.D. (1967): Theory of multiphase sampling from a finite or an infinite population on successive occasions. Review of the International Statistical Institute, 35, p. 247-263.

Wahlström, S och Malmberg, S. (1971): De svenska arbetskrafts-undersökningarna III. Variansformler för genomsnitts- och förändringsskattningar. Stat. Tidskrift 1971:4, p. 316-320.

Waite, P.J. and Cole, S.J. (1980): Selection of a new sample panel for the annual survey of manufactures. U.S. Bureau of the Census.

Waksberg, J. and Pearl, R.B. (1964): The Current Population Survey: A case history in panel operations. The American Statistical Association, Proceedings of the Social Statistics Section 1964, p. 217-228.

Wolter, K.M., Isaki, C.T., Sturdevant, T., Monsour, N. and Mayes, F. (1976): Sample selection and estimation aspects of the Census Bureaus Monthly Business Surveys. American Statistical Association, Proceedings of the Business and Economic Statistics Section, p. 99-109.

Wolter, K.M. (1979): Composite estimation in finite populations. Journal of the American Statistical Association, 74, p. 604-613.

Woodruff, R.S. (1959): The use of rotating samples in the census bureau's monthly surveys. The American Statistical Association, Proceedings of the Social Statistics Section 1959, p. 130-138.

Woodruff, R.S. (1963): The use of rotating samples in the census bureau's monthly surveys. Journal of the American Statistical Association, 58, p. 454-467.

Woodruff, S.M. (1980): A class of ratio composite estimators. American Statistical Association. Proceedings of the section on survey research methods, p. 200-204.

Wretman, J. (1973): Variansformler för en differens mellan medelvärden i två överlappande stickprov. SCB. Metodinformation 73:1.

de Vries, M. (1961): Practical estimators for sampling on successive occasions. Statistica Nederlandica, 15, p. 123-133.

Yates, F. (1949): Sampling Methods for Censuses and Surveys. 1st ed. 1949, 2nd ed. 1953 and 3rd ed. 1960. London.