ARTIKLER 66

ETTERHÅNDSSTRATIFISERING OG ESTIMERING INNEN DEL - BESTANDER

Av John Dagsvik

POST-STRATIFICATION AND ESTIMATION WITHIN SUBPOPULATIONS

OSLO 1974

STATISTISK SENTRALBYRÅ

ARTIKLER FRA STATISTISK SENTRALBYRÅ NR. 66

ETTERHÅNDSSTRATIFISERING OG ESTIMERING INNEN DEL - BESTANDER

Av John Dagsvik

POST-STRATIFICATION AND ESTIMATION WITHIN SUBPOPULATIONS

OSLO 1974

ISBN 82 - 537 - 0378 - 3

FORORD

Under bearbeidingen av resultatene fra de arbeidskraftundersøkelser Statistisk Sentralbyrå gjennomfører på utvalgsbasis, blir de som er intervjuet gruppert, etter kjønn og alder. En slik teknikk kalles etterhåndsstratifisering. Den blir nyttet dels i et forsøk på å redusere samplingsusikkerheten i de publiserte resultatene og dels for å skape samsvar mellom de tall som offentliggjøres fra arbeidskraftundersøkelsene og tall i annen befolkningsstatistikk.

Byrået har ønsket å få en vurdering av egenskapene ved tilgjengelige estimeringsmetoder som bygger på etterhåndsstratifisering. En slik vurdering blir lagt fram i denne artikkelen. I et forsøk på å gjøre de viktigste resultatene tilgjengelige også for folk som ikke har spesielle kunnskaper i matematisk statistikk, er framstillingen delt i to. De to første kapitlene gir en verbal beskrivelse av estimeringsmetodene og deres viktigste egenskaper. Den matematiske analyse er skilt ut og framstilt i kapittel 3 og i et matematisk appendiks.

Statistisk Sentralbyrå, Oslo, 20. juni 1974

Petter Jakob Bjerve

PREFACE

During the processing of the results from the Labour Force Surveys of the Central Bureau of Statistics of Norway, the respondents are classified according to sex and age. A technique like this is called post-stratification, and it is used in an attempt to reduce sampling error in the published results, and also to make these results consistent with figures from the usual population statistics.

The Bureau has wanted an evaluation of the statistical properties of current estimation methods based on post-stratification. Such an evaluation is presented in this article. In an attempt to make the main results accessible to those who do not have special knowledge of mathematical statistics, the presentation is made at two levels. The two first Chapters give a verbal description of the estimation methods and their main properties. The mathematical analysis has been separated from the rest and can be found in Chapter 3, and in a mathematical Appendix.

Central Bureau of Statistics, Oslo, 20 June 1974

Petter Jakob Bjerve

INNHOLD

CONTENTS

1. INNLEDNING OG HOVEDKONKLUSJON.

Ved mange utvalgsundersøkelser benyttes det en teknikk som kalles etterhåndsstratifisering. Denne teknikken går ut på at utvalget sorteres i klasser (etterhåndsstrata) etter spesielle kjennetegn, først når dataene foreligger. Metoden brukes i situasjoner der en på forhånd ikke vet hvilken av disse klassene den enkelte enhet tilhører. I Byrået benyttes etterhåndsstratifisering i forbindelse med arbeidskraftundersøkelsene (AKU), idet personene i utvalget inndeles etter kjønn og alder under databearbeidingen. Ett formål med denne framgangsmåten er den samme som for "vanlig" forhåndsstratifisering, nemlig ønsket om variansreduksjon. I Byråets AKU har en også en annen hensikt med den estimeringsmetoden som brukes. Fra det sentrale personregister kjenner en samlet antall personer i landet fordelt etter alder og kjønn. Det er derfor ønskelig at det samlede antall personer i en kjønns- og aldersgruppe som framkommer ved summering over AKU-estimatene for antallet i de enkelte sysselsettingsgrupper, skal samsvare med opplysningene fra personregisteret. Den metoden en bruker i dag, har denne egenskapen. En estimeringsmetode som har en egenskap av denne typen, vil vi si har samsvarsegenskapen.

I det prosjektet som beskrives her, har vi i hovedsak studert to måter til å utnytte etterhåndsstratifisering for estimering av samlet antall enheter med et bestemt kjennetegn i undersøkelsesbestanden. Den ene metoden brukes i Byråets arbeidskraftundersøkelser i dag, og vi vil betegne den a^*. Denne bygger på vår kunnskap om fordelingen på alder og kjønn i den samlede norske befolkning, og krever ikke at vi skal kjenne denne fordelingen i det enkelte utvalgsområde.

Den andre metoden vi studerer, krever kunnskap om fordelingen av befolkningen på alder og kjønn, også innen det enkelte utvalgsområde. Denne estimatoren vil vi kalle $\tilde{a}$. Den kan ikke brukes med dagens standard utvalgsplan, men den vil være aktuell når den nye standard utvalgsplan tas i bruk. Vi kjenner nemlig ikke befolkningens alders- og kjønnsfordeling innen de nåværende utvalgsområdene, mens disse data vil være tilgjengelig for kommunene, som blir utvalgsområder i den nye utvalgsplanen.

Vi studerer estimatorenes statistiske egenskaper, først og fremst forventning og varians. Det ser ut til at de tradisjonelle tilnærmingsformler for disse parametrene er for grove og kompliserte for vårt bruk. Ved å ta utgangspunkt i en tilnærmelse til den eksakte fordeling (Burr, 1973), har vi funnet finere tilnærmelsesformler. Vi har sammenliknet disse med de tilsvarende parametrene for den estimeringsmetode Byrået vanligvis bruker, der en ikke etterhåndsstratifiserer utvalget. Den tilsvarende estimatoren vil vi betegne med $\hat{a}$.

De estimatorene vi har studert, kan også brukes ved estimering av fordelinger osv. innen del-bestander. Vi kan f.eks. betrakte hvert etterhåndsstratum som en del-bestand. Det er nettopp det Byrået gjør når en er interessert i data for ett eller flere etterhåndsstrata. Totalen for den samlede bestand finnes så ved å summere estimatene for alle etterhåndsstrata.

I et forsøk på å gjøre de viktigste resultatene våre tilgjengelige også for folk som ikke er matematiske statistikere, har vi delt framstillingen i to hovedkapitler og et appendiks. I det første hovedkapitlet er det gitt en verbal beskrivelse av estimatorene og deres egenskaper. I det andre kapitlet tar vi utgangspunkt i den formelle beskrivelsen av en generell to-trinns utvalgsplan og gir matematiske bevis for de resultatene vi har funnet. Til slutt har vi et matematisk appendiks med bevis for noen hjelpesetninger som benyttes i det andre hovedkapitlet. Symboler og konvensjoner er stort sett i overensstemmelse med dem Hoem (1973) bruker. Han har også en oversikt over den eksisterende litteratur i avsnitt 7.6.

Hovedkonklusjon.

Vårt valg av estimeringsmetode vil være avhengig av den utvalgsplanen som ligger til grunn.

Med dagens standard utvalgsplan er de aktuelle metodene estimatorene a^* og $\hat{a}$. Vi har funnet at estimatoren a^*, som brukes i AKU i dag, har større varians enn estimatoren $\hat{a}$. Den er antakelig også noe forventningsskjev. Dette resultatet gjelder for alle mulige inndelinger i etterhåndsstrata, men forskjellen i varians vil avta når antall etterhåndsstrata øker så fremt oppdelingen i etterhåndsstrata ikke blir for "fin". Samtidig vil sannsynligvis skjevheten til a^* øke når antall etterhåndsstrata økes. Det eneste argumentet for å benytte a^* i stedet for $\hat{a}$, er at a^* har samsvarsegenskapen, mens $\hat{a}$ ikke har det.

I Byråets <u>nye</u> standard utvalgsplan vil også estimatoren $\tilde{a}$ være aktuell. Denne estimatoren er tilnærmet forventningsrett dersom antall etterhåndsstrata er under en viss grense. Denne grensen kan først bestemmes når den nye standard utvalgsplan er ferdig. Dersom etterhåndsstratifiseringen er slik at estimatoren $\tilde{a}$ er tilnærmet forventningsrett, vil $\tilde{a}$ som regel ha mindre varians enn $\hat{a}$. Estimatoren $\tilde{a}$ har samsvarsegenskapen, vel og merke innen de nye etterhåndsstrata som omtales ovenfor. Disse etterhåndsstrata blir større enn dem som brukes i dag (mai, 1974), og $\tilde{a}$ vil ikke ha samsvarsegenskapen innen de nåværende kjønns- og aldersklasser.

Jeg vil takke Olav Ljones og Ib Thomsen for nyttige diskusjoner under utarbeidelsen av manuskriptet, og Jan M. Hoem for god veiledning under manuskriptutformingen. Jeg vil rette en spesiell takk til Petter Laake som har lagt ned mye arbeid i gjennomgåelsen av artikkelmanuskriptet.

2. BESKRIVELSE AV ESTIMATORENE OG DERES EGENSKAPER.

2.1. Generelt om estimatorene.

I dette arbeidet studerer vi ulike metoder til å estimere en ukjent aggregatstørrelse a i undersøkelsesbestanden. Dette kan eksempelvis være samlet inntekt eller samlet forbruk i bestanden i en gitt periode. Felles for estimatorene er at en først estimerer a-verdien for hvert etterhåndsstratum. Totalstørrelsen a er summen av disse, og den estimeres ved summering av stratumestimatene. Dette kan en jo utnytte hvis en ønsker egen informasjon om ett eller flere etterhåndsstrata og ikke bare betrakter dem som et hjelpemiddel for estimeringen av a-verdien i den samlede bestand.

I det følgende vil vi for det meste konsentrere oss om estimeringen innen det enkelte etterhåndsstratum.

2.2. Estimatoren $\tilde{a}$.

Vi skal først se på den framgangsmåten som man antakelig ville velge automatisk hvis man ikke gav seg til å fundere nærmere over de praktiske og teoretiske problemene som foreligger. Etter denne metoden estimeres først a-verdien for hvert etterhåndsstratum innen det enkelte utvalgsområde ved å "blåse opp" a-verdien for utvalget på vanlig måte. Oppblåsningsfaktoren er da samlet antall trekke-enheter i etterhåndsstratumet dividert med antall enheter derfra i utvalget. Deretter blåses estimatene for hvert utvalgsområde opp tilsvarende, nemlig ved at en dividerer med sannsynligheten for at utvalgsområdet skal bli trukket ut. Ved å summere over utvalgsområder og forhåndsstrata innen det enkelte etterhåndsstratum får vi estimatoren $\tilde{a}(\ell)$ for a-verdien i ℓ-te etterhåndsstratum, som vi betegner med $a(\ell)$. Denne estimatoren er tidligere beskrevet av Hoem (1973, side 27).

Denne metoden krever kjennskap til bestandens fordeling på etterhåndsstrata innen hvert utvalgsområde. En har ikke slike data idag, så estimatoren er ikke aktuell med den eksisterende utvalgsplan. Derimot kjenner en befolkningens fordeling på alder og kjønn innen hver kommune, slik at $\tilde{a}$ kan være nyttig når en får den nye utvalgsplanen, der en tenker å bruke kommunene som utvalgsområder.

Denne estimatoren har imidlertid en svakhet som kan være betydningsfull for våre formål. For at den skal være forventningsrett, kreves det at sannsynligheten for å få tomme etterhåndsstrata er lik null innen hvert utvalgsområde. I Byråets nåværende standard utvalgsplan er denne sannsynligheten betydelig større enn null, og dette medfører at $\tilde{a}(\ell)$ ville underestimere $a(\ell)$ dersom den ble brukt. (Tilsvarende for estimatoren a.) Setning 3.3. (med eksempel) i neste hovedkapittel viser nemlig at dersom vi skal unngå en betydelig forventningsskjevhet i denne estimatoren, bør forholdet mellom utvalgsstørrelsen innen hvert primære utvalgsområde, og antall etterhåndsstrata gjennomsnittlig være større enn 6. Siden denne utvalgsstørrelsen ligger i nærheten av 30 i våre områder, betyr dette at vi bare kan bruke inntil ca. 5 etterhåndsstrata, og enda færre i mindre utvalgsområder. "Egentlig" burde vi altså bare etterhåndsstratifisere etter kjønn og et par grove aldersklasser.

I den nye utvalgsplanen vil antakelig utvalgsstørrelsen innen hver kommune bli vesentlig større enn 30, slik at antall etterhåndsstrata følgelig kan økes. Når dette skrives foreligger det ikke tilstrekkelige opplysninger om den nye utvalgsplanen til at vi kan vurdere dette nærmere. Når denne utvalgsplanen er fastlagt, vil vi ved hjelp av setning 3.3. kunne beregne maksimalt antall etterhåndsstrata.

Variansen til estimatoren $\tilde{a}(\ell)$ vil være mindre enn variansen til den vanlige estimeringsmetoden $\hat{a}(\ell)$ som brukes når utvalget ikke etterhåndsstratifiseres, dersom antall etterhåndsstrata ikke er "for stort" (Korollar 3.1.). Dette resultatet er forøvrig i samsvar med teorien for rent lotteriske utvalg.

Vi har mer spesielt studert effekten av etterhåndsstratifisering på variansen til estimatoren $\tilde{a}$ for den samlede a-verdi. I setning 3.5. og korollar 3.2. nedenfor er det etablert et kriterium for når etterhåndsstratifisering har henholdsvis "positiv" og "negativ" effekt på variansen. Denne effekten viser seg å være knyttet til fordelingen av forholdet mellom den relative a-verdi i det enkelte etterhåndsstratum innen hvert utvalgsområde på den ene siden og relativ a-verdi for alle etterhåndsstrata i hvert område på den annen. Dersom dette forholdet varierer lite omkring 1, kan en risikere at estimatoren $\tilde{a}$ har større varians enn estimatoren $\hat{a}$ som brukes når utvalget <u>ikke</u> etterhåndsstratifiseres. Dersom vår estimand er liten, må dette forholdet variere mye for at estimatoren $\tilde{a}$ ikke skal være dårligere enn estimatoren $\hat{a}$. (Se eksempel 3.)

2.3. Estimatoren a^*.

Dersom fordelingen på etterhåndsstrata er kjent for den samlede bestand, mens fordelingen på etterhåndsstrata ikke er kjent innen de enkelte utvalgsområder, kan estimeringen foregå på følgende måte:

En finner samlet a-verdi for ℓ-te etterhåndsstratum for hele utvalget under ett og blåser så opp dette med en faktor som er lik antall enheter i etterhåndsstratumet i totalbestanden dividert med samlet antall enheter i etterhåndsstratumet i utvalget. Metoden brukes i dag i tilknytning til Byråets arbeidskraftsundersøkelse.

Ved denne metoden benyttes ikke den informasjonen en måtte ha om forhåndsstratifiseringen, slik at det er naturlig å regne med at denne estimatoren kan ha større varians enn estimatoren $\check{a}$. (Som vi tidligere har nevnt, har imidlertid estimatoren a^* samsvarsegenskapen.)

I kapittel 3 har vi funnet et uttrykk for forventning og varians til a^*. De viser at denne estimatoren er noe forventningsskjev og at den har større varians enn estimatoren $\hat{a}$. Forventningsskjevheten har nær sammenheng med størrelsen $\hat{N}(\underset{\sim}{J},\ell)$, som er summen av de oppblåste antall enheter fra ℓ-te etterhåndsstratum i hvert av de uttrukne utvalgsområdene. Dersom denne størrelsen har liten varians, vil estimatoren a^* ha liten forventningsskjevhet. I motsatt fall kan vi risikere en betydelig skjevhet. Variansen til $\hat{N}(\underset{\sim}{J},\ell)$ er igjen avhengig av antall utvalgsområder i utvalget, i den forstand at variansen avtar dersom antall uttrukne utvalgsområder øker. For å kunne studere skjevheten til estimatoren a^* nærmere, må vi imidlertid ha kjennskap til bestandens fordeling på etterhåndsstrata innen hvert utvalgsområde. Korollar 3.3., side 31, viser at variansen til a^{**}, og dermed variansen til a^*, avtar med økende antall etterhåndsstrata innenfor en viss grense. Samtidig ser det ut som om skjevheten øker med økende antall etterhåndsstrata. Dersom estimatoren a^* benyttes, er antakelig det antall etterhåndsstrata vi har i dag, det maksimale vi kan bruke hvis forutsetningene for teorien skal holde.

2.4. Estimatoren a^{**}.

Dersom bestandens fordeling på etterhåndsstrata er kjent innen hvert utvalgsområde, kan vi redusere skjevheten til estimatoren a^* ved å erstatte det kjente samlede antall enheter i ℓ-te etterhåndsstratum med størrelsen $\hat{N}(\underset{\sim}{J},\ell)$, som estimerer dette antallet. Imidlertid har denne modifiserte estimatoren a^{**}, ikke lenger den ønskede samsvarsegenskapen. Denne nye estimatoren har en varians av samme størrelsesorden som variansen til estimatoren a^*, og det er derfor bedre å benytte estimatoren $\hat{a}$ istedenfor estimatoren a^{**}.

3. MATEMATISK ANALYSE AV ESTIMATORENES STATISTISKE EGENSKAPER.

3.1. Formell beskrivelse av utvalgsplanen.

For å lette lesingen skal vi repetere noen definisjoner fra Hoem (1973) og innføre noen nye.

Befolkningen er inndelt i et antall forhåndsstrata. Forhåndsstratum nr. i har M_i primære utvalgsområder, (p.s.u.). P.s.u. nr. j i stratum i inneholder $N_i(j)$ trekkeenheter. Til trekkeenhet nr. k er det tilknyttet en verdi $a_i(j,k)$ som vi kan måle.
Vi lar

$$N_i = \sum_j N_i(j)$$

være antall trekkeenheter i stratum i og

$$N = \sum_i N_i$$

antall trekkeenheter totalt. Videre lar vi

$$a_i(j) = \sum_k a_i(j,k)$$

være a-verdien i p.s.u. nr. j i stratum nr. i.

Fra de M_i utvalgsområdene i stratum i trekkes det ut m_i områder. La $\pi_i(j)$ være sannsynligheten for at p.s.u. nr. j i stratum i skal bli med i utvalget, og la $\pi_i(j,q)$ være sannsynligheten for at både p.s.u. nr. j og p.s.u. nr. q skal bli med i utvalget. Numrene på de p.s.u. som trekkes ut, betegnes $J_{i1}, J_{i2}, \ldots, J_{im_i}$. Videre lar vi $\underset{\sim}{J}_i$ og $\underset{\sim}{J}$ betegne henholdsvis settet av numrene på de uttrukne p.s.u. i stratum i, og for hele landet. Fra det r-te uttrukne p.s.u. trekkes det $n_{ir}(\underset{\sim}{J})$ trekkeenheter rent lotterisk. Numrene på trekkeenhetene som velges ut fra område nr. J_{ir}, betegnes $K_{ir1}, K_{ir2}, \ldots$. La

$$X_{irs} = a_i(J_{ir}, K_{irs}),$$

$$\bar{X}_{ir} = \sum_s X_{irs}/n_{ir}(J),$$

og

$$X = \sum_{irs}\sum\sum X_{irs}.$$

Vi antar at utvalget av trekkeenheter inndeles i v etterhåndsstrata på grunnlag av informasjon som foreligger i dataene. Vi innfører indeksvariabelen $I_i(j,k,\ell)$. Denne antar verdien 1 dersom trekkeenhet nr. (i,j,k) tilhører etterhåndsstratum nr. ℓ, og 0 ellers.

$$N_i(j,\ell) = \sum_k I_i(j,k,\ell)$$

blir dermed antall trekkeenheter i etterhåndsstratum nr. ℓ i p.s.u. nr. (i,j). La

$$a_i^{(\ell)}(j) = \sum_k a_i(j,k)\ I_i(j,k,\ell)$$

være totalen innen etterhåndsstratum ℓ, p.s.u. nr. j og stratum i og la

$$\bar{a}_i^{(\ell)}(j) = a_i^{(\ell)}(j)/N_i(j,\ell)$$

være det tilsvarende gjennomsnitt. La videre

$$\sigma_i^2(j) = \sum_k \{a_i(j,k) - \bar{a}_i(j)\}^2/\{N_i(j) - 1\},$$

$$\sigma_i^2(j,\ell) = \sum_k I_i(j,k,\ell)\ \{a_i(j,k) - \bar{a}_i^{(\ell)}(j)\}^2/\{N_i(j,\ell) - 1\}$$

og

$$N(\ell) = \sum_{i,j} N_i(j,\ell).$$

En ny tilfeldig variabel $X_{irs}(\ell)$ defineres ved

$$X_{irs}(\ell) = I_i(J_{ir}, K_{irs}, \ell).$$

Analogt med tidligere symboler innføres

$$X_{ir}(\ell) = \sum_s X_{irs}(\ell),$$

$$X(\ell) = \sum_{i,r} X_{ir}(\ell),$$

og

$$b_{ir}(\underset{\sim}{J}) = n_{ir}(\underset{\sim}{J})/N_i(J_{ir}).$$

Vi noterer at $X_{ir}(\ell)$ er antallet i utvalget i etterhåndsstratum ℓ, forhåndsstratum i og p.s.u. J_{ir}.

Vi tar utgangspunkt i estimatorene

$$(3.1.1)\qquad \hat{a}(\ell) = \sum_{i,r} \P_i^{-1}(J_{ir})\, b_{ir}^{-1}(\underset{\sim}{J}) \sum_s X_{irs} X_{irs}(\ell)$$

og

$$(3.1.2)\qquad \hat{a} = \sum_{\ell} \hat{a}(\ell).$$

Dersom

$$b(\underset{\sim}{J}) = \P_i(J_{ir})\, b_{ir}(\underset{\sim}{J})$$

er uavhengig av i og r, blir utvalget "selvveiende", og estimatorene ovenfor får formen

$$\hat{a}(\ell) = b^{-1}(\underset{\sim}{J}) \sum_{i,r,s} X_{irs} X_{irs}(\ell)$$

og

$$\hat{a} = b^{-1}(\underset{\sim}{J})\, X$$

siden

$$\sum_{\ell} X_{irs}(\ell) = 1.$$

Estimatorene (3.1.1) og (3.1.2) har følgende egenskaper:

<u>Setning 3.1</u>: La

$$b_i^{-1}(j) = E\{b_{ir}^{-1}(\underset{\sim}{J}) \mid J_{ir} = j\}.$$

Da gjelder følgende:

(i) $E\{\hat{a}(\ell)\} = a(\ell)$.

(ii) $$\mathrm{Var}\, E\{\hat{a}(\ell) \mid \underset{\sim}{J}\} = \sum_i \sum_{p<q} \{\pi_i(p)\, \pi_i(q) - \pi_i(p,\, q)\} \cdot \left\{\frac{a_i^{(\ell)}(p)}{\pi_i(p)} - \frac{a_i^{(\ell)}(q)}{\pi_i(q)}\right\}^2 .$$

(iii) $$\mathrm{Var}\, E\{\hat{a} \mid \underset{\sim}{J}\} = \sum_i \sum_{p<q} \{\pi_i(p)\, \pi_i(q) - \pi_i(p,\, q)\} \cdot \left\{\frac{a_i\,(p)}{\pi_i(p)} - \frac{a_i(q)}{\pi_i(q)}\right\}^2 .$$

(iv) $$E\, \mathrm{Var}\, \{\hat{a}(\ell) \mid \underset{\sim}{J}\} = \sum_{i,j} \pi_i^{-1}(j)\, \{b_i^{-1}(j) -1\}\, N_i(j)/\{N_i(j) -1\}$$

$$\cdot\{[\, N_i(j,\, \ell) -1\,]\, \sigma_i^2(j,\ell) + [\, \bar{a}_i^{(\ell)}(j)\,]^2\, N_i(j,\ell)\, [1-N_i(j,\ell)/N_i(j)\,]\},$$

(v) $$E\, \mathrm{Var}\, \{\hat{a} | \underset{\sim}{J}\} = \sum_{i,j} \pi_i^{-1}(j)\, \{b_i^{-1}(j) -1\}\, N_i(j)\, \sigma_i^2(j).$$

Bevis:

Punktene (i), (ii), (iii) og (v) er vist hos Des Raj (1968, side 118.)

Bevis for punkt (iv): La

$$\tau_i^2(j,\ell) = \sum_k \{I_i(j,k,\ell)\, a_i(j,k) -a_i^{(\ell)}(j)/N_i(j)\}^2/\{N_i(j) -1\}.$$

Det følger av resultatene hos Des Raj at

$$E\, \mathrm{Var}\{\hat{a}(\ell) \mid \underset{\sim}{J}\} = \sum_{i,j} \pi_i^{-1}(j)\{b_i^{-1}(j) -1\}\, N_i(j)\, \tau_i^2(j,\ell).$$

Ved en enkel omforming finner vi at

$$\{N_i(j) -1\}\, \tau_i^2(j,\ell) = \{N_i(j,\ell) -1\}\, \sigma_i^2(j,\ell) +$$

$$[\, \bar{a}_i^{(\ell)}(j)\,]^2\, N_i(j,\ell)\, \{1 - N_i(j,\ell)/N_i(j)\}.$$

Q.E.D.

3.2. Estimatoren $\tilde{a}$.

Vi skal her se på den "klassiske" måten å utnytte etterhåndsstratifiseringen på under estimeringen. Denne er blant annet beskrevet hos Hoem (1973, side 27), og forutsetter at bestandens fordeling på etterhåndsstrata er kjente for de uttrukne utvalgsområdene. Estimatorene er definert ved

$$(3.2.3)\quad \tilde{a}(\ell) = \sum_{i,r} \pi_i^{-1}(J_{ir})\, N_i(J_{ir},\ell) \sum_s X_{irs}\, X_{irs}(\ell)/X_{ir}(\ell),$$

$$(3.2.4)\quad \tilde{a} = \sum_{\ell} \tilde{a}(\ell).$$

I dagens standard utvalgsplan er $\pi_i(J_{ir}) = m_i/M_i$. Altså er $\pi_i(J_{ir})$ bare avhengig av forhåndsstrataene og ikke av de enkelte p.s.u. innen hvert forhåndsstratum.

<u>Setning 3.2</u>: La

$$U_i(j,\ell) = N_i(j,\ell)\, E\{X_{ir}^{-1}(\ell) \mid J_{ir} = j\} - b_i^{-1}(j).$$

Dersom

$$\Pr\{X_{ir}(\ell) > 0 \mid \underset{\sim}{J}\} \simeq 1$$

for alle i, r og ℓ, gjelder følgende:

(i) $E\{\tilde{a}(\ell)\} \simeq a(\ell)$.

(ii) $\operatorname{Var} E\{\tilde{a}(\ell) \mid \underset{\sim}{J}\} \simeq \operatorname{Var} E\{\hat{a}(\ell) \mid \underset{\sim}{J}\}$.

(iii) $\operatorname{Var} E\{\tilde{a} \mid \underset{\sim}{J}\} \simeq \operatorname{Var} E\{\hat{a} \mid \underset{\sim}{J}\}$.

(iv) $E \operatorname{Var}\{\tilde{a}(\ell) \mid \underset{\sim}{J}\} =$

$$\sum_{i,j} \pi_i^{-1}(j)\{N_i(j,\ell)\left[b_i^{-1}(j) - 1 + U_i(j,\ell)\right] \sigma_i^2(j,\ell)\}.$$

(v) $E \operatorname{Var}\{\tilde{a} \mid \underset{\sim}{J}\} = \sum_{\ell} E \operatorname{Var}\{\tilde{a}(\ell) \mid \underset{\sim}{J}\}$.

<u>Bevis</u>: Det er hensiktsmessig å benytte notasjonene

$$\underset{\sim}{X} = \{X_{ir}(\ell) : i = 1,2, \ldots; r = 1, 2, \ldots; \ell = 1, 2, \ldots\}.$$

og

$$\tilde{X}_{ir}(\ell) = \sum_s X_{irs}\, X_{irs}(\ell)/X_{ir}(\ell).$$

Fra Hoem (1973, side 28) vet vi at når $X_{ir}(\ell) > 0$, er

$$(3.2.5)\quad E\{\tilde{X}_{ir}(\ell) \mid \underset{\sim}{J} = \underset{\sim}{j}, \underset{\sim}{X} = \underset{\sim}{x}\} = \bar{a}_i^{(\ell)}(j_{ir}).$$

Dersom $\Pr\{X_{ir}(\ell) > 0 \mid \underset{\sim}{J} = \underset{\sim}{j}\} \simeq 1$, blir

derfor $E\{\tilde{a}(\ell)\} \simeq a(\ell)$, og vi har vist (i).

Under forutsetningene ovenfor får vi videre at

$$E\{\tilde{a}(\ell) \mid \underset{\sim}{J} = \underset{\sim}{j}\} = \sum_{i,r} \P_i^{-1}(j_{ir})\, a_i^{(\ell)}(j_{ir})$$

$$= E\{\hat{a}(\ell) \mid \underset{\sim}{J} = \underset{\sim}{j}\},$$

slik at (ii) og (iii) følger umiddelbart.

Det er videre klart at

$$\operatorname{cov}\{\tilde{X}_{ir}(\ell), \tilde{X}_{ir}(\ell') \mid \underset{\sim}{J} = \underset{\sim}{j}, \underset{\sim}{X} = \underset{\sim}{x}\} = 0$$

når $\ell \neq \ell'$, slik at

$$\operatorname{cov}\{\tilde{X}_{ir}(\ell), \tilde{X}_{ir}(\ell') \mid \underset{\sim}{J} = \underset{\sim}{j}\} =$$

$$E\operatorname{cov}\{\tilde{X}_{ir}(\ell), \tilde{X}_{ir}(\ell') \mid \underset{\sim}{J} = \underset{\sim}{j}, \underset{\sim}{X}\} +$$

$$\operatorname{cov}\{\bar{a}_i^{(\ell)}(j_{ir}), \bar{a}_i^{(\ell')}(j_{ir}) \mid \underset{\sim}{J} = \underset{\sim}{j}\} = 0.$$

Altså er $\tilde{X}_{ir}(\ell)$ og $\tilde{X}_{ir}(\ell')$ ukorrelerte når $\underset{\sim}{J}$ er gitt. Av resultater hos Hoem (1973, side 29-30) følger det da at

$$E\ \mathrm{Var}\{N_i(J_{ir},\ell)\ \tilde{X}_{ir}(\ell)\ |\ \underset{\sim}{J}\} =$$

$$E\{\sigma_i^2(J_{ir},\ell)\ [N_i(J_{ir},\ell)\ E\{X_{ir}^{-1}(\ell)\ |\ \underset{\sim}{J}\} -1]\ N_i(J_{ir},\ell)\}$$

Q.E.D.

<u>Korollar 3.1</u>: Dersom

$$b_i^{-1}(j) -1 + U_i(j) \approx b_i^{-1}(j) -1$$

og

$$N_i(j,\ell)/\{N_i(j,\ell) -1\} \approx 1,$$

er

$$\mathrm{Var}\{\tilde{a}(\ell)\} \leq \mathrm{Var}\{\hat{a}(\ell)\}.$$

<u>Bevis</u>: Det er nok å vise at

$$E\ \mathrm{Var}\{\tilde{a}(\ell)\ |\ \underset{\sim}{J}\} \leq E\ \mathrm{Var}\{\hat{a}(\ell)\ |\ \underset{\sim}{J}\}.$$

Fra setning 3.1 og 3.2 følger det direkte at

$$E\ \mathrm{Var}\{\hat{a}(\ell)\ |\ \underset{\sim}{J}\} - E\ \mathrm{Var}\{\tilde{a}(\ell)\ |\ \underset{\sim}{J}\}$$

$$= \sum_{i,j} \pi_i^{-1}(j)\ (b_i^{-1}(j) -1)\ \{\sum_k a_i^2(j,k)\ I_i(j,k,\ell) - [a_i^{(\ell)}(j)]^2/N_i(j)$$

$$-[\sum_k a_i^2(j,k)\ I_i(j,k,\ell) - [a_i^{(\ell)}(j)]^2\ /N_i(j,\ell)]\ N_i(j,\ell)/(N_i(j,\ell)-1)\}$$

$$\geq 0$$

Q.E.D.

Setning 3.3: La

$$\lambda_{ir}(\ell) = N_i(j_{ir},\ell)\; n_{ir}(\underset{\sim}{j})/N_i(j_{ir}) = b_{ir}(\underset{\sim}{j})\; N_i(j_{ir},\ell).$$

Vi har da:

$$\Pr\{X_{ir}(\ell) = 0 \mid \underset{\sim}{J} = \underset{\sim}{j}\} \simeq e^{-\lambda_{ir}(\ell)}.$$

Bevis: Når $\underset{\sim}{J} = \underset{\sim}{j}$, er $X_{ir}(\ell)$ hypergeometrisk fordelt med parametre $n_{ir}(\underset{\sim}{j})$, $N_i(j_{ir},\ell)$ og $N_i(j_{ir})$. Ved å bruke tilnærmingsformelen i lemma 1 i appendikset får vi at

$$\Pr\{X_{ir}(\ell) = 0 \mid \underset{\sim}{J} = \underset{\sim}{j}\} \simeq e^{-\lambda_{ir}(\ell)}\,(1 - c_{ir}(\ell)).$$

Nå vil $c_{ir}(\ell)$ være liten sammenliknet med 1, slik at vi kan sette

$$e^{-\lambda_{ir}(\ell)}\,(1 - c_{ir}(\ell)) \simeq e^{-\lambda_{ir}(\ell)}.$$

Q.E.D.

Eksempel: I Byråets AKU er

$$n_{ir}(\underset{\sim}{j})/N_i(j_{ir}) = b_{ir}(\underset{\sim}{j}) \simeq 1/50.$$

Dersom vi betrakter $e^{-6} = 1,5\cdot 10^{-3}$ som neglisjerbart, vil $N_i(j_{ir},\ell) \geq 300$ medføre at

$$\Pr\{X_{ir}(\ell) = 0 \mid \underset{\sim}{J} = \underset{\sim}{j}\} \simeq 0.$$

Anta at $N_i(j_{ir}) \simeq 2\,000$. Dersom vi bruker ettårige aldersklasser, blir antall etterhåndsstrata lik 120, slik at den gjennomsnittlige verdi av $N_i(j_{ir},\ell)$ blir mindre enn 20, og

$$e^{-\lambda_{ir}(\ell)} \geq e^{-1}.$$

Dette viser at estimatoren $\tilde{a}$ gir en betydelig forventningsskjevhet. Hoem (1973, side 28) viser forøvrig at $\tilde{a}$ vil underestimere a.

Med v etterhåndsstrata blir $N_i(j,\ell)$ gjennomsnittlig lik $N_i(j)/v$ slik at

$$\lambda_{ir}(\ell) \simeq n_{ir}(\underset{\sim}{j})/v.$$

Setning 3.4: La forutsetningene være de samme som i setning 3.2 og anta dessuten at $N_i(j)/\{N_i(j) -1\} \approx 1$. Da er

$$\mathrm{Var}\ \hat{a} - \mathrm{Var}\ \tilde{a} \approx$$

$$\sum_{i,j,\ell} \pi_i^{-1}(j) \{(b_i^{-1}(j) -1) N_i(j,\ell) (\bar{a}_i^{(\ell)}(j) - \bar{a}_i(j))^2$$

$$- (N_i(j,\ell) U_i(j,\ell) + b_i^{-1}(j) -1) \sigma_i^2(j,\ell)\}.$$

Bevis: Ved en enkel omforming får vi

$$\sum_{\ell} (N_i(j,\ell) -1) \sigma_i^2(j,\ell)$$

$$= \sum_k a_i^2(j,k) - \sum_{\ell} N_i(j,\ell) \left[\bar{a}_i^{(\ell)}(j)\right]^2$$

$$= \sum_k (a_i(j,k) - \bar{a}_i(j))^2 - \sum_{\ell} N_i(j,\ell) (\bar{a}_i^{(\ell)}(j) - \bar{a}_i(j))^2$$

$$= (N_i(j) -1) \sigma_i^2(j) - \sum_{\ell} N_i(j,\ell) (\bar{a}_i^{(\ell)}(j) - \bar{a}_i(j))^2.$$

Setningen følger nå ved at en bruker setning 3.1 og 3.2. Q.E.D.

Setning 3.4 gir altså et uttrykk for "gevinsten" ved etterhånds-stratifisering. Som vi ser, inneholder den både et positivt og et negativt ledd. Vi skal nå se nærmere på differansen $\mathrm{Var}\ \hat{a} - \mathrm{Var}\ \tilde{a}$. Vi vil finne at parameteren

$$h_i(j,\ell) = \bar{a}_i^{(\ell)}(j)/\bar{a}_i(j)$$

spiller en sentral rolle i den forbindelse. Vi nøyer oss med å betrakte situasjonen der $a_i(j,k)$ er en binær variabel. Da blir

$$\sigma_i^2(j,\ell) = \bar{a}_i^{(\ell)}(j) (1 - \bar{a}_i^{(\ell)}(j)) N_i(j,\ell)/(N_i(j,\ell) -1).$$

Setning 3.5: Anta at $a_i(j,k)$ er en binær variabel og la

$$\eta'(x,t) = \frac{-x\,|t^{-1}-2|}{2(x+1)} + \{(\frac{x|t^{-1}-2|}{2(x+1)})^2 + \frac{x(t^{-1}-1)}{x+1}\}^{\frac{1}{2}},$$

$$\eta''(x,t) = \eta'(x,t) + \frac{x\,|t^{-1}-2|}{x+1},$$

$$\lambda = \lambda_i(j,\ell) = N_i(j,\ell)\; b_i(j),$$

$$2\,c = 2\,c_i(j,\ell) = N_i^{-1}(j,\ell) + N_i^{-1}(j)\; b_i^{-1}(j),$$

$$\alpha_i(j,\ell) = \{N_i^{-1}(j,\ell) + (\lambda\phi(\lambda)\; -1)\; (1 - b_i(j))^{-1}\}\; N_i(j,\ell)/\; \{N_i(j,\ell)-1\}$$

og

$$\beta_i(j,\ell) = \lambda\phi(\lambda)\; -1 - 3c,$$

der $\phi\;(\cdot)$ er definert like før lemma 4 i appendikset. La α og β være slik at $\alpha \geq \alpha_i(j,\ell)$ og $\beta \leq \beta_i(j,\ell)$ for alle i, j, ℓ.

(i) Dersom

$$|h_i(j,\ell)\; -1| > \eta'\{\alpha,\; \bar{a}_i(j)\}$$

for alle i, j, ℓ, er

$$\text{Var}\; \hat{a} \geq \text{Var}\; \tilde{a}.$$

(ii) Dersom

$$|h_i(j,\ell)\; -1| \leq \eta''(\beta,\; \bar{a}_i(j))$$

for alle i, j, ℓ, er

$$\text{Var}\; \hat{a} \leq \text{Var}\; \tilde{a}.$$

<u>Bevis</u>: (i): La $|h_i(j,\ell)\; -1| \geq \eta$.

Siden

$$h_i(j,\ell)\; \{1 - h_i(j,\ell)\; \bar{a}_i(j)\}$$

er voksende i $h_i(j,\ell)$ når

$$\bar{a}_i^{(\ell)}\; (j) < \tfrac{1}{2},$$

og avtagende når

$$\bar{a}_i^{(\ell)}\; (j) > \tfrac{1}{2},$$

er

$$h_i(j,\ell)\ \{1 - h_i(j,\ell)\ \bar{a}_i(j)\}$$

$$\leq \begin{cases} (1+\eta)\ \{1 - [1+\eta]\ \bar{a}_i(j)\} & \text{når } \bar{a}_i^{(\ell)}(j) > \tfrac{1}{2}, \\ (1-\eta)\ \{1 - [1-\eta]\ \bar{a}_i(j)\} & \text{når } \bar{a}_i^{(\ell)}(j) \leq \tfrac{1}{2}. \end{cases}$$

La videre

$$x \geq \{N_i^{-1}(j,\ell) + U_i(j,\ell)\ (b_i^{-1}(j) - 1\}\ N_i(j,\ell)/(N_i(j,\ell) - 1).$$

Vi finner da at uttrykket

$$\Delta_i(j) = \sum_\ell \{N_i(j,\ell)\ (\bar{a}_i^{(\ell)}(j) - \bar{a}_i(j))^2$$

$$- \sigma_i^2(j,\ell)\ [U_i(j,\ell) N_i(j,\ell)\ b_i(j)/(1 - b_i(j)) + 1]\}$$

$$\geq \sum_\ell \{N_i(j,\ell)\ (\bar{a}_i^{(\ell)}(j) - \bar{a}_i(j))^2$$

$$- \sigma_i^2(j,\ell)\ (N_i(j,\ell) - 1)\ x\}.$$

Det følger herav at

$$\Delta_i(j) \geq \bar{a}_i(j) \sum_\ell \{N_i(j,\ell)\ \bar{a}_i(j)\ (h_i(j,\ell) - 1)^2$$

$$- x N_i(j,\ell)\ h_i(j,\ell)\ [1 - \bar{a}_i(j)\ h_i(j,\ell)]\}$$

$$\geq a_i(j)\ \{\bar{a}_i(j)\ (1+x)\eta^2 + x\ |\ 1 - 2\bar{a}_i(j)\ |\ \eta$$

$$- x\ (1 - \bar{a}_i(j))\}$$

$$(3.2.6)\quad = a_i(j)\ (\eta - \eta')\ (\eta + \eta' - \tfrac{x}{x+1}\ |\ 1/\bar{a}_i(j) - 2|).$$

Når $\eta \geq \eta\{x, \bar{a}_i(j)\}$, er uttrykket i (3.2.6) positivt og (i) følger umiddelbart ved å bruke setning 3.4. Vi skal deretter vise at vi kan velge $x = \alpha$. Fra setning 1(ii) i appendikset følger det at

$$E\{\varepsilon(X)/X\} \leq \phi(\lambda),$$

noe som medfører at

$$U_i(j,\ell)\ b_i(j) \leq \lambda\phi(\lambda) - 1.$$

Dermed er (i) vist.

(ii): Resonementet er helt analogt med det under punkt (i), og vi skal derfor nøye oss med å vise at vi kan velge $x = \beta$. Fra setning 1(ii) i appendikset følger at

$$E\{\varepsilon(X)/X\} \geq \phi(\lambda) - 2c\lambda^{-1}(1 + 3\lambda^{-1})$$

$$\geq \phi(\lambda) - 3c\lambda^{-1},$$

når $\lambda > 6$. Derav:

$$U_i(j,\ell)\ b_i(j) \geq \lambda\ \phi(\lambda) - 1 - 3c.$$

Ulikheten ovenfor medfører derfor at

$$N_i^{-1}(j,\ell) + U_i(j,\ell)\ (b_i^{-1}(j) - 1) \geq \beta_i(j,\ell).$$

Q.E.D.

<u>Korollar 3.2</u>: Anta at $\bar{a}_i(j) \leq k$ for alle i og j. Dersom

$$|U_i(j,\ell) - 1| \leq \eta''(\beta,k)$$

for alle i, j, ℓ, er

$$\mathrm{Var}\ \tilde{a} \geq \mathrm{Var}\ \hat{a}.$$

Bevis: Det er nok å vise at $\eta''(\beta,t)$ er en avtagende funksjon i t. Dette kan gjøres ved å studere den deriverte av funksjonen $\eta''(\beta,t)$. Beviset er nokså teknisk og er ikke tatt med her.

Eksempel 1: Setter vi k = 0,5 i korollaret foran, blir

$$\eta''(\beta,1/2) = \{\frac{\lambda\phi(\lambda) -1 -3c}{\lambda\phi(\lambda) -3c}\}^{\frac{1}{2}} .$$

La $\lambda \simeq 7$ og $c \simeq 5\%$. Av tabellen på side 242 hos Stegun og Abramowitz (1954) og av lemma 4 i appendikset finner vi at $7\phi(7) \simeq 1{,}22$, slik at

$$\eta''(\beta,1/2) \simeq 0{,}25.$$

Dersom $h_i(j,\ell)$ varierer mindre enn 25% omkring 1, vil altså $\tilde{a}$ være dårligere enn â.

Eksempel 2: La λ og c være som i eksempel 1 og anta at k = 1/4. Da blir

$$\eta''(\beta,1/4) \simeq 0{,}5.$$

Eksempel 3: For samme λ og c som ovenfor og k = 1/6 blir

$$\eta''(\beta,1/6) \simeq 0{,}72.$$

Disse eksemplene viser at dersom vår estimand er liten, må $h_i(j,\ell)$ variere mye for at etterhåndsstratifisering skal lønne seg.

3.3. Estimatoren a^{**}.

I det foregående avsnitt har vi sett at estimatoren $\tilde{a}$ vil underestimere a når antallet trekke-enheter i etterhåndsstrataene innen hvert p.s.u. er lite. Dette problemet kan reduseres vesentlig ved å benytte andre "oppblåsningsfaktorer" enn de som inngår i estimatoren $\tilde{a}$. La

$$\hat{N}(\underset{\sim}{J},\ell) = \sum_{i,r} \pi_i^{-1}(J_{ir})\, N_i(J_{ir},\ell).$$

Vi merker oss at

$$E\,\{\hat{N}(\underset{\sim}{J},\ell)\} = N(\ell),$$

slik at $\hat{N}(\underset{\sim}{J},\ell)$ kan betraktes som en estimator for samlet antall enheter i ℓ-te etterhåndsstratum. Estimatorene $a^{**}(\ell)$ og a^{**} defineres ved

$$(3.3.1) \quad a^{**}(\ell) = \hat{N}(\underset{\sim}{J},\ell) \sum_{i,r,s} x_{irs} X_{irs}(\ell)/X(\ell)$$

og

$$(3.3.2) \quad a^{**} = \sum_{\ell} a^{**}(\ell).$$

Spørsmålet er nå om denne estimeringsmetoden har gunstige egenskaper når det gjelder forventning og varians sammenliknet med estimatoren $\hat{a}$.

<u>Setning 3.6</u>: La

$$2c_{ir}(\underset{\sim}{J},\ell) = N_i^{-1}(J_{ir},\ell) + n_{ir}^{-1}(\underset{\sim}{J}),$$

$$\rho_{ir}(\underset{\sim}{J},\ell) = N_i(J_{ir},\ell)/\P_i(J_{ir})\ \hat{N}(\underset{\sim}{J},\ell),$$

$$\Theta(\underset{\sim}{J},\ell) = \sum_{i,r} c_{ir}(\underset{\sim}{J},\ell)\ \rho_{ir}^2(\underset{\sim}{J},\ell),$$

og

$$\lambda_{ir}(\underset{\sim}{J},\ell) = b_{ir}(\underset{\sim}{J})\ N_i(J_{ir},\ell).$$

Dersom

$$1 + c_{ir}(\underset{\sim}{j},\ell)\ \lambda_{ir}^2(\underset{\sim}{j},\ell) \simeq 1$$

for alle i, r, ℓ og $\underset{\sim}{j}$, er

$$\text{(i)} \quad E\{a^{**}(\ell)\} - a(\ell) \simeq 2 \sum_{i,r} E\{\P_i^{-1}(J_{ir}) \left[\rho_{ir}(\underset{\sim}{J},\ell)\ c_{ir}(\underset{\sim}{J},\ell) - \Theta(\underset{\sim}{J},\ell)\right] a_i^{(\ell)}(J_{ir})\}.$$

Dersom vi i tillegg kan anta at

$$\rho_{ir}(\underset{\sim}{j},\ell)\; c_{ir}(\underset{\sim}{j},\ell) - \Theta(\underset{\sim}{j},\ell) \simeq 0$$

for alle i, r, ℓ og $\underset{\sim}{j}$, blir

(ii) $\quad E\{a^{**}(\ell)\} \simeq a(\ell).$

<u>Bevis:</u> Av likning (3.2.5) følger at

$$E\{\sum_s X_{irs}\, X_{irs}(\ell)/X(\ell) \mid \underset{\sim}{J} = \underset{\sim}{j},\ \underset{\sim}{X} = \underset{\sim}{x}\} = \bar{a}_i^{(\ell)}(j_{ir})\; x_{ir}(\ell)/x(\ell),$$

hvor vi tolker $x_{ir}(\ell)/x(\ell)$ som 0 når $x(\ell) = 0$. Av setning 2(ii) i appendikset får vi at

$$E\{X_{ir}(\ell)/X(\ell) \mid \underset{\sim}{J} = \underset{\sim}{j}\} \simeq \left[1 + 2c_{ir}(\underset{\sim}{j},\ell)\,\rho_{ir}(\underset{\sim}{j},\ell) - 2\Theta(j,\ell)\right] \rho_{ir}(\underset{\sim}{j},\ell),$$

noe som medfører at

$$E\{\sum_s X_{irs}\, X_{irs}(\ell)/X(\ell) \mid \underset{\sim}{J} = \underset{\sim}{j}\} = \bar{a}_i^{(\ell)}(j_{ir}) \left[1 + 2c_{ir}(\underset{\sim}{j},\ell)\,\rho_{ir}(\underset{\sim}{j},\ell) - 2\Theta(\underset{\sim}{j},\ell)\right] \rho_{ir}(\underset{\sim}{j},\ell).$$

Herav:

$$(3.3.3)\quad E\{a^{**}(\ell) \mid \underset{\sim}{J} = \underset{\sim}{j}\} = \sum_{i,r} \P_i^{-1}(j_{ir})\; a_i^{(\ell)}(j_{ir}) \left[1 + 2c_{ir}(\underset{\sim}{j},\ell)\,\rho_{ir}(j,\ell) - 2\Theta(\underset{\sim}{j},\ell)\right],$$

som gir (i) umiddelbart. Punkt (ii) følger direkte fra punkt (i). Q.E.D.

<u>Merknad:</u> Vi observerer at $c_{ir}(\underset{\sim}{J},\ell)$ og $\rho_{ir}(\underset{\sim}{J},\ell)$ er <u>kjente</u> størrelser.

Eksempel: I Byråets AKU er de gjennomsnittlige verdier av n_{ir} og $N_i(j,\ell)$ lik henholdsvis 30 og 20, dvs. $c_{ir}(\underset{\sim}{J},\ell)$ er gjennomsnittlig lik 5%. Antar vi videre at $\P_i(j)$ er tilnærmet konstant blir $\rho_{ir} \simeq 1/300$, siden det er 300 utvalgsområder i utvalget. Dermed blir

$$\Theta = 5 \cdot 10^{-2} \cdot \frac{1}{3} \cdot 10^{-2} \simeq 2 \cdot 10^{-4}$$

og

$$\Theta^2 = 4 \cdot 10^{-8} \simeq 0.$$

Dette viser at antagelsen $\Theta^2 \simeq 0$ er en rimelig antagelse i våre anvendelser.

Setning 3.7: La forutsetningene være de samme som i setning 3.6(ii). Da er

(i) $$\text{Var } E\{a^{**}(\ell) \mid \underset{\sim}{J}\} \simeq \text{Var } E\{\hat{a}(\ell) \mid \underset{\sim}{J}\}$$

og

(ii) $$\text{Var } E\{a^{**} \mid \underset{\sim}{J}\} \simeq \text{Var } E\{\hat{a} \mid \underset{\sim}{J}\}.$$

Dersom $b^{-1}(\underset{\sim}{j})\ \hat{N}^{-1}(\underset{\sim}{j},\ell) + 1 \simeq 1$ for alle $\underset{\sim}{j}$ og ℓ, gjelder følgende:

(iii) $$E\ \text{Var}\{a^{**}(\ell) \mid \underset{\sim}{J}\} \simeq$$

$$\sum_{i,j} \P_i^{-1}(j)\ \{N_i(j,\ell)\ [\bar{a}_i^{(\ell)}(j)]^2\ [1 + b_i^{-1}(j)\ (1 + N_i(j,\ell)\ N_i^{-1}(j))]$$

$$+ [N_i(j,\ell) - 1]\ \sigma_i^2(j,\ell)\ [b_i^{-1}(j) - N_i^{-1}(j,\ell)]\}.$$

(iv) $$E\ \text{Var}\{a^{**} \mid \underset{\sim}{J}\} \geq E\ \text{Var}\{\hat{a} \mid \underset{\sim}{J}\}$$

$$+ \sum_{i,j} \P_i^{-1}(j)\ \{N_i(j)\ \sigma_i^2(j) + 2N_i^{-1}(j)\ b_i^{-1}(j) \sum_{\ell} [a_i^{(\ell)}(j)]^2\}.$$

<u>Bevis:</u> (i) og (ii): Av likning (3.3.3) har vi

$$E\{a^{**}(\ell) \mid \underset{\sim}{J} = \underset{\sim}{j}\} \simeq \sum_{i,r} \pi_i^{-1}(j_{ir})\, a_i^{(\ell)}(j_{ir}),$$

noe som gir

$$E\{a^{**} \mid \underset{\sim}{J} = \underset{\sim}{j}\} \simeq \sum_{i,r} \pi_i^{-1}(j_{ir})\, a_i(j_{ir})$$

$$= E\{\hat{a} \mid \underset{\sim}{J} = \underset{\sim}{j}\}.$$

Vi ser herav at vi kan få en formel for Var $E\{a^{**}(\ell) \mid \underset{\sim}{J}\}$ ved å erstatte $a_i(j)$ med $a_i^{(\ell)}(j)$ i formelen for Var $E\{\hat{a} \mid \underset{\sim}{J}\}$. Det tilsvarende gjelder for Var $E\{a^{**} \mid \underset{\sim}{J}\}$. Formlene (i) og (ii) er dermed bevist.

(iii): Hoem (1973, side 29-30) har vist at

$$\mathrm{Var}\{\sum_s X_{irs}\, X_{irs}(\ell)/X(\ell) \mid \underset{\sim}{J} = \underset{\sim}{j},\ \underset{\sim}{X} = \underset{\sim}{x}\}$$

$$= [x_{ir}(\ell)/x^2(\ell)]\,(1 - x_{ir}(\ell)/N_i(j_{ir},\ell))\; \sigma_i^2(j_{ir},\ell).$$

Av korollaret til setning 3 i appendikset får vi når vi benytter tilnærmelsen $(N_i(j)-1)/N_i(j) \simeq 1$, at

$$E\{[X_{ir}(\ell)/X^2(\ell)]\,(1 - X_{ir}(\ell)/N_i(j_{ir},\ell)) \mid \underset{\sim}{J} = \underset{\sim}{j}\} \simeq$$

$$\rho_{ir}(\underset{\sim}{j},\ell)\, b^{-1}(\underset{\sim}{j})\, \hat{N}^{-1}(\underset{\sim}{j},\ell)\, \left[1 - (1 + b_{ir}(\underset{\sim}{j}))/N_i(\underset{\sim}{j},\ell)\right],$$

og at

$$\mathrm{Var}\{X_{ir}(\ell)/X(\ell) \mid \underset{\sim}{J} = \underset{\sim}{j}\} \simeq$$

$$\rho_{ir}(\underset{\sim}{j},\ell)\, b^{-1}(\underset{\sim}{j})\, \hat{N}^{-1}(\underset{\sim}{j},\ell)\, \left[1 + N_i(j_{ir},\ell)/N_i(j_{ir}) + b_{ir}(j)\right].$$

Herav:

$$E\ \mathrm{Var}\{a^{**}(\ell)\mid \underset{\sim}{J}=\underset{\sim}{j},\ \underset{\sim}{X}\}\simeq$$

$$\sum_{i,r}\pi_i^{-2}(j_{ir})\,\sigma_i^{2}(j_{ir},\ell)\,(N_i(j_{ir},\ell)-1)\,(b_{ir}^{-1}(\underset{\sim}{j})-N_i^{-1}(j_{ir},\ell))$$

og

$$\mathrm{Var}\ E\{a^{**}(\ell)\mid \underset{\sim}{J}=\underset{\sim}{j},\ \underset{\sim}{X}\}$$

$$\sum_{i,r}\pi_i^{-2}(j_{ir})\,b_{ir}^{-1}(\underset{\sim}{j})\,N_i(j_{ir},\ell)\,[\bar{a}_i^{(\ell)}(j_{ir})]^2\{1+$$

$$N_i(j_{ir},\ell)/N_i(j_{ir})+b_{ir}(\underset{\sim}{j})\}.$$

Ved å bruke identiteten

$$\mathrm{Var}\{a^{**}(\ell)\mid \underset{\sim}{J}=\underset{\sim}{j}\}=E\ \mathrm{Var}\{a^{**}(\ell)\mid \underset{\sim}{J}=\underset{\sim}{j},\ \underset{\sim}{X}\}$$

$$+\ \mathrm{Var}\ E\{a^{**}(\ell)\mid \underset{\sim}{J}=\underset{\sim}{j},\ \underset{\sim}{X}\},$$

får vi (iii).

(iv): Av setning 4 i appendikset følger det at

$$\sum_{\ell\neq\ell'}\mathrm{cov}\,\{E(a^{**}(\ell)\mid \underset{\sim}{J}=\underset{\sim}{j},\underset{\sim}{X}),\ E(a^{**}(\ell')\mid \underset{\sim}{J}=\underset{\sim}{j},\underset{\sim}{X})\}$$

$$\geq-\sum_{\ell\neq\ell'}\ \sum_{i,r}a_i^{(\ell)}(j_{ir})\,a_i^{(\ell')}(j_{ir})\,\pi_i^{2}(j_{ir})/n_{ir}(\underset{\sim}{j})$$

$$=\sum_{i,r}\pi_i^{-2}(j_{ir})\{-N_i(j_{ir})\,\bar{a}_i^{\,2}(j_{ir})\,b_{ir}^{-1}(\underset{\sim}{j})$$

$$+\sum_{\ell}[a_i^{(\ell)}(j_{ir})]^2/n_{ir}(\underset{\sim}{j})\}.$$

Siden

$$\mathrm{cov}\{a^{**}(\ell),\ a^{**}(\ell') \mid \underset{\sim}{J} = \underset{\sim}{j}\} =$$

$$E\ \mathrm{cov}\{a^{**}(\ell),\ a^{**}(\ell') \mid \underset{\sim}{J} = \underset{\sim}{j},\ \underset{\sim}{X}\} +$$

$$\mathrm{cov}\{E(a^{**}(\ell) \mid \underset{\sim}{J} = \underset{\sim}{j},\ \underset{\sim}{X}),\ E(a^{**}(\ell') \mid \underset{\sim}{J} = \underset{\sim}{j},\ \underset{\sim}{X})\}$$

$$= \mathrm{cov}\{E(a^{**}(\ell) \mid \underset{\sim}{J} = \underset{\sim}{j},\ \underset{\sim}{X}),\ E(a^{**}(\ell') \mid \underset{\sim}{J} = \underset{\sim}{j},\ \underset{\sim}{X})\},$$

får vi ved å gjennomføre samme regningen som i beviset for setning 3.4, at

$$(3.3.3)\quad E\ \mathrm{Var}\{a^{**} \mid \underset{\sim}{J}\} \geq$$

$$E\Big[\sum_{i,r}\{\P_i^{-2}(J_{ir})\ b_{ir}^{-1}(\underset{\sim}{J})\ N_i(J_{ir})\ \sigma_i^{2}(J_{ir}) +$$

$$\sum_{\ell}\ \P_i^{-2}(J_{ir})([a_i^{(\ell)}(J_{ir})]^2\ (2n_{ir}^{-1}(\underset{\sim}{J}) + N_i^{-1}(J_{ir},\ell))$$

$$- (1 - N_i^{-1}(J_{ir},\ \ell))\ \sigma_i^{2}(J_{ir},\ \ell))\}\Big].$$

Nå er

$$\{1 - N_i^{-1}(j,\ \ell)\}\ \sigma_i^{2}(j,\ \ell) = N_i^{-1}(j,\ \ell)\ \sum_k[a_i^{(\ell)}(j,\ k)]^2 - [\bar{a}_i^{(\ell)}(j)]^2$$

$$\leq N_i^{-1}(j,\ \ell)\ [a_i^{(\ell)}(j)]^2\ (1 - N_i^{-2}(j,\ \ell))$$

$$\leq N_i^{-1}(j,\ \ell)\ [a_i^{(\ell)}(j)]^2,$$

slik at

$$E\ \mathrm{Var}\{a^{**} \mid \underset{\sim}{J}\} \geq$$

$$E\{\sum_{i,r}\P_i^{-2}(J_{ir})\ [(b_{ir}^{-1}(\underset{\sim}{J}) - 1)\ N_i(J_{ir})\ \sigma_i^{2}(J_{ir})$$

$$+ N_i(J_{ir})\ \sigma_i^{2}(J_{ir}) + 2\sum_{\ell}\ [a_i^{(\ell)}(J_{ir})]^2/n_{ir}(\underset{\sim}{J})]\},$$

som er identisk med (iv). Q.E.D.

Korollar 3.3: Høyre side av ulikhet (iv) i setning 3.7. øker når to og to etterhåndsstrata slås sammen.

Bevis: Anta at vi slår sammen etterhåndsstratum nr. ℓ og etterhåndsstratum nr. $\ell + 1$. Det er nok å vise at

$$\sum_{\ell} \left[a_i^{(2\ell)}(j) + a_i^{(2\ell-1)}(j)\right]^2 >$$

$$\sum_{\ell} \left[a_i^{(2\ell)}(j)\right]^2 + \sum_{\ell} \left[a_i^{(2\ell-1)}(j)\right]^2.$$

Men denne ulikheten følger av Cauchy - Schwartz ulikhet, og korollaret er dermed bevist. Q.E.D.

Ulikheten (iv) i setning 3.7 er ganske fin slik at vi ikke ville gjøre så stor feil om vi erstattet ulikhetstegnet med tilnærmet likhetstegn. Dette kan vi slutte av de tilnærmelsene som er gjort under utledningen, der ulikhetstegnet bare opptrer i forbindelse med kovariansleddene og under omformingen av ulikhet 3.3.

Resultatet ovenfor tyder derfor på at variansen innen utvalgsområdene for estimatoren a^{**} vil avta når antall etterhåndsstrata øker.

3.4. Estimatoren a^*.

Vi skal til slutt se på den estimatoren som brukes i Byråets arbeidskraftundersøkelse i dag. Denne estimatoren framkommer ved å erstatte $\hat{N}(\underset{\sim}{J}, \ell)$ med sin forventningsverdi $N(\ell)$ i estimatoren a^{**}. Skrevet med symboler kan altså denne estimatoren uttrykkes ved

$$(3.4.1)\quad a^*(\ell) = N(\ell)\, a^{**}(\ell)/\hat{N}(\underset{\sim}{J}, \ell)$$

og

$$(3.4.2)\quad a^* = \sum_{\ell} a^*(\ell)$$

Estimatoren er beskrevet nærmere av Hoem (1973, side 31). Følgende resultater følger umiddelbart:

Setning 3.8: Dersom Var $\hat{N}(\underset{\sim}{J}, \ell)$ er liten, har estimatorene $a^{*}(\ell)$ og a^{*} tilnærmet samme egenskaper som estimatorene $a^{**}(\ell)$ og a^{**}.

Bevis: Hvis Var $\hat{N}(\underset{\sim}{J}, \ell)$ er liten, er

$$\Pr\{N(\ell)/\hat{N}(\underset{\sim}{J}, \ell) \simeq 1\} \simeq 1,$$

slik at

$$E\{a^{*}(\ell)\} \simeq E\{a^{**}(\ell)\}$$

og

$$\mathrm{Var}\{a^{*}(\ell)\} \simeq \mathrm{Var}\{a^{**}(\ell)\}$$

Q.E.D.

Setning 3.9: La forutsetningene være de samme som i setning 3.6 (ii). Da er

$$E\{a^{*}(\ell)\} \simeq E\{N(\ell)\, \hat{N}^{-1}(\underset{\sim}{J}, \ell) \sum_{i,r} \P_i^{-1}(J_{ir})\, a_i^{(\ell)}(J_{ir})\}.$$

Beviset følger direkte av likning (3.4.1) og (3.3.3).

Appendiks: Matematiske hjelperesultater.

I dette appendikset har vi samlet en del generelle matematiske setninger som er benyttet i kapitlet foran. Disse resultatene er nye tilnærmelser til forventning, varians og kovarians til brøker hvor teller er en hypergeometrisk fordelt variabel og nevner er en sum av slike variable. I utledningene har vi tatt utgangspunkt i en artikkel av Burr (1973), hvor han blant annet gir tilnærmete uttrykk for den hypergeometriske fordeling.

Lemma 1: La X være hypergeometrisk fordelt med punktsannsynlighet

$$P(x, n, k, N) = \binom{k}{x}\binom{N-k}{n-x} \Big/ \binom{N}{n} \quad \text{for } x = 0,1, \ldots, k.$$

Følgende formler gjelder:

(i) $P(x, n, k, N) = (e^{-\lambda}\lambda^x/x!)\left[1 + \delta + c(x - (x-\lambda)^2)\right],$

(ii) $P(x, n, k, N) = \binom{n}{x}(\lambda/n)^x(1-\lambda/n)^{n-x}$

$\cdot \left[1 + \eta + (1/2k)(x-(x-\lambda)^2)\right],$

med tilhørende genererende funksjoner;

(iii) $E\{s^X\} = e^{\lambda(s-1)}\left[1 + \delta - c\lambda^2(1-s)^2\right]$

og

(iv) $E\{s^X\} = (1 + \lambda/n(s-1))^n\left[1 + \eta - \lambda^2(1-s)^2/2k\right],$

der

$$c = 1/2n + 1/2k, \quad \lambda = nk/N = E\{X\},$$

$$\delta = 0(1/n^2) + 0(1/k^2) \quad \text{og} \quad \eta = 0(1/n^2).$$

Formlene (i) og (ii) finnes i Burr (1973) og (iii) (iv) følger direkte av (i) og (ii).

At en funksjon $f(x) = 0\{g(x)\}$, betyr at det finnes en konstant M slik at

$$|f(x)| \leq M\ |g(x)|.$$

Lemma 2: La X og Y være simultant hypergeometrisk fordelte med punktsannsynlighet

$$P(x, y, n, k_1, k_2, N) = \binom{k_1}{x}\binom{k_2}{y}\binom{N-k_1-k_2}{n-x-y} \Big/ \binom{N}{n}.$$

Den genererende funksjon for (X, Y) er

$$E\{s^X z^Y\} \approx [1 - \lambda_1^2(1 - s)^2/2k_1]\,[1 - \lambda_2^2(1 - s)^2/2k_2]$$

$$\cdot[1 + \lambda_1(s - 1)/n + \lambda_2(z - 1)/n]^n,$$

hvor vi har neglisjert ledd av typen

$$O(1/n^2),\ O(1/k_i^2) \text{ og } O(1/nk_i)\ ,i = 1,\ 2.$$

<u>Bevis:</u> La

$$\lambda_3 = \lambda_1/(1 - k_2)N) = \lambda_1/(1 - \lambda_2/n).$$

Ved å benytte lemma 1 (iv) finner vi at

$$E\{s^X \mid Y = y\} = [1 + O(1/(n - y)^2) - \lambda_3^2(s - 1)^2/2k_1]\,[1 + \lambda_3(s - 1)/n]^{n-y}.$$

$$\approx [1 - \lambda_1^2(s - 1)^2/2k_1]\,[1 + \lambda_3(s - 1)/n]^{n-y}.$$

Av lemma 1(ii) følger videre

$$E\{s^X z^Y\} = E\{s^Y E\{s^X \mid Y\}\} \approx$$

$$(1 - \lambda_1^2(s - 1)^2)\sum_y \binom{n}{y}[1 + \lambda_3(s - 1)/n]^{n-y}(1 - \lambda_2/n)^{n-y}$$

$$\cdot(z\lambda_2/n)^y[1 + (1/2k_2)(y - (y - \lambda_2)^2)],$$

som etter litt regning reduserer seg til

$$[1 - \lambda_1^2(s - 1)^2/2k_1]\,[1 - \lambda_2^2(z - 1)^2/2k_2]$$

$$\cdot[\lambda_2 z/n + 1 - \lambda_2/n + \lambda_3(1 - \lambda_2/n)(s - 1)/n]^n$$

$$\approx [1 - \lambda_1^2(s - 1)^2/2k_1]\,[1 - \lambda_2^2(z - 1)^2/2k_2]$$

$$\cdot[1 + \lambda_1(s - 1)/n + \lambda_2(z - 1)/n]^n.$$

Q.E.D.

Lemma 3: La X_j, $j = 1, 2, \dots$ være uavhengige tilfeldige diskrete variable med punktsannsynligheter $f_j(\cdot)$, $j = 1, 2, \dots$, og genererende funksjoner $\hat{f}_j(\cdot)$. Definer $0/0 = 0$. Da er

$$\text{(i)} \qquad E\{X_i/\sum_j X_j\} = \int_0^1 \hat{f}_i{}'(u) \prod_{i \neq j} \hat{f}_j(u)\, du$$

og

$$\text{(ii)} \qquad E\{(X_i/\sum_j X_j)^2\} = \int_0^1 v^{-1}\{\int_0^v \left[\hat{f}_i{}'(u) + u\hat{f}_i{}''(u)\right] \prod_{i \neq j} \hat{f}_j(u)\, du\}\, dv.$$

Bevis: La $Z = \sum X_j$ og $g_i(\cdot) = \prod_{j \neq i} f_j(\cdot)$.

Variabelparet (X_i, Z) har sannsynlighetstetthet

$$h_i(x, z) = f_i(x)\, g_i(z - x)$$

og genererende funksjon

$$\hat{h}_i(u, v) = \sum_{x,z} f_i(x)\, g_i(z - x)\, (uv)^x v^{z - x}$$

$$= \hat{f}_i(uv)\, \hat{g}_i(v).$$

Betrakt den genererende funksjonen

$$\hat{h}_i(u, v) = \sum_{x,z} u^x v^z h_i(x, z).$$

Ved å derivere partielt m.h.p. u får vi

$$v^{-1} \frac{\partial \hat{h}_i}{\partial u}(u, v) = \sum_{x,z>0} x u^x v^{z-1} h_i(x, z).$$

Integrasjon m.h.p. v gir deretter

$$\sum_{x,z>0} x h_i(x, z)/z = \int_0^1 v^{-1} \frac{\partial \hat{h}_i}{\partial u}(1, v)\, dv$$

$$= \int_0^1 \hat{f}_i{}'(v)\, \hat{g}_i(v)\, dv,$$

hvilket beviser (i). Punkt (ii) bevises analogt ved å observere at

$$\sum_{x,z>o} x^2 h_i(x,z)/z^2 = \int_o^1 v^{-1}\{\int_o^v u^{-1} \left[\frac{\partial^2 \hat{h}_i(1,u)}{\partial u^2} + \frac{\partial \hat{h}_i(1,u)}{\partial u}\right] du\} dv.$$

Q.E.D.

<u>Definisjon:</u> Funksjonen $\phi(\cdot)$ er definert ved

$$\phi(x) = e^{-x}\{\int_o^1 (e^{tx}-1)\, dt/t\} \text{ for } x > o.$$

<u>Lemma 4:</u> Funksjonen $\phi(\cdot)$ kan uttrykkes ved den eksponensielle integralfunksjonen Ei (x):

(i) $\phi(x) = e^{-x}(\mathrm{Ei}(x) - \ln x - C),$

hvor C er Eulers konstant.

(ii) Når $x>>1$, har $\phi(\cdot)$ følgende asymptotiske rekkeutvikling

$$\phi(x) \sim x^{-1}(1 + x^{-1} + 2!\, x^{-2} + 3!\, x^{-3} + ..).$$

Den eksponensielle integralfunksjon er definert blant annet hos Jahnke, Emde og Løsch (1960) og hos Stegun og Abramowitz (1954), hvor den også er tabellert. Ved å rekkeutvikle $\phi(x)$ får vi (i). Siden $e^{-x}(\ln x + C) \approx 0$ for store x får vi at

$$\phi(x) \simeq \mathrm{Ei}(x)e^{-x} \sim x^{-1}(1 + x^{-1} + 2!\, x^{-2} + ..).$$

(Se Jahnke, Emde og Løsch, 1960, side 18.)

Q.E.D.

Setning 1: La X være hypergeometrisk fordelt med forventning λ.

La

$$\varepsilon(x) = \begin{cases} 1 & \text{når } x > o, \\ 0 & \text{ellers.} \end{cases}$$

Vi har

(i) $$E\{\varepsilon(X)/X\} = (1 + \delta - c\lambda^2)\,\phi(\lambda) + c(\lambda + 1 - (1 + 2\lambda)\,e^{-\lambda}),$$

der c og δ er definert i lemma 1.

(ii) Når $\lambda > 6$ gjelder tilnærmet

$$\phi(\lambda) - 2\,c\lambda^{-1}(1 + 3\lambda^{-1}) \leq E\{\varepsilon(X)/X\} \leq \phi(\lambda).$$

Bevis: (i): Ved å integrere potensrekken $t^{-1}\hat{f}(t)$ ledd for ledd finner vi at

$$\sum_{x>o} f(x)/x = \int_o^1 \{\hat{f}(t) - \hat{f}(0)\}\, t^{-1} dt.$$

Lemma 1(iii) gir oss

$$\int_o^1 \{\hat{f}(t) - \hat{f}(0)\} t^{-1} dt = \int_o^1 (1 + \delta - c\lambda^2)\, e^{-\lambda}(e^{\lambda t} - 1)\, t^{-1} dt$$

$$+ \int_o^1 c\lambda^2 (2 - t) e^{(t-1)\lambda} dt = (1 + \delta - c\lambda^2)\,\phi(\lambda)$$

$$+ c(1 + \lambda - (1 + 2\lambda) e^{-\lambda}).$$

Dermed er (i) bevist.

(ii): Fra lemma 4(ii) kan vi slutte at

$$\lambda\phi(\lambda) \geq 1 + \lambda^{-1} + 2\lambda^{-2},$$

noe som medfører at

$$c(\lambda + 1 - \lambda^2\phi(\lambda)) \leq -\,2c\lambda^{-1}.$$

Herav:

$$E\{\varepsilon(X)/X\} \leq \phi(\lambda) + \delta\phi(\lambda) - 2c\lambda^{-1}.$$

Siden δ er av størrelsesorden $n^{-2} + k^{-2}$, er høyre side mindre enn $\phi(\lambda)$.
Når $\lambda > 6$ er $(1 + 2\lambda)e^{-\lambda} \simeq 0$ slik at

$$E\{\varepsilon(X)/X\} \geq \phi(\lambda) - c(\lambda^2\phi(\lambda) -1 -\lambda).$$

Bruk av lemma 4(ii) gir

$$\lambda^2\phi(\lambda) - 1 - \lambda \simeq 2\lambda^{-1}(1 + 3\lambda^{-1}).$$

Q.E.D.

<u>Setning 2</u>: La X_j, $j = 1, 2, \ldots$ være uavhengige hypergeometrisk fordelte variable med forventninger λ_j. La

$$\Lambda = \sum_j \lambda_j, \quad \rho_i = \lambda_i/\Lambda,$$

$$\Theta' = \sum_j \rho_j^2 c_j/(1 - c_j \lambda_j^2) \text{ og } \Theta = \sum_j c_j\rho_j^2.$$

Anta at $\Lambda >> 1$ og $(\Theta')^2 \simeq c_i\rho_i\Theta' \simeq o$.

Da er

(i) $\quad \rho_i(1 + 2c_i\rho_i - 2\Theta') \leq E\{X_i/\sum_j X_j\} \leq \rho_i(1 + 2c_i\rho_i - 2\Theta).$

(ii) $\quad$ Dersom vi i tillegg har $\Theta \simeq \Theta'$, er

$$E\{X_i/\sum_j X_j\} \simeq \rho_i(1 + 2c_i\rho_i - 2\Theta).$$

Bevis: La $V_i = X_i / \sum_j X_j$. Av lemma 1(iii) og lemma 3(i), følger det at

$$E\{V_i\} = \lambda_i \int_o^1 \left[1 + \delta - c_i\lambda_i^2(u-1)^2 + 2c_i\lambda_i(1-u)\right] e^{\Lambda(u-1)}$$

$$\cdot \prod_{i\neq j} \left[1 + \delta - c_j\lambda_j^2(u-1)^2\right] du,$$

som etter variabelskiftet $1 - u = \Lambda t$ blir lik

$$\int_o^\Lambda \rho_i(1 + \delta - c_i\rho_i^2 t^2 + 2c_i\rho_i t)e^{-t} \prod_{i\neq j} (1-c_j\rho_j^2 t^2)dt.$$

Ved å bruke middelverdisetningen får vi

$$\log (1 - c_j\rho_j^2 t^2) \geq -c_j\rho_j^2 t^2/(1 - c_j\lambda_j^2)$$

slik at

$$\prod_j (1 - c_j\rho_j^2 t^2) \geq e^{-\Theta' t^2} \geq 1 - \Theta' t^2.$$

Dermed blir

$$E\{V_i\} \geqq \int_o^\Lambda \rho_i(1 + \delta - \Theta' t^2 + 2c_i\rho_i t(1 - (\Theta' - c_i\rho_i^2)t^2))e^{-t}dt.$$

Når Λ er stor og n liten, er

$$\int_o^\Lambda t^n e^{-t}dt \simeq \Gamma(n + 1) = n!.$$

Siden $c_i\rho_i\Theta' \simeq (\Theta')^2 \simeq o$, blir

$$E\{V_i\} \geqq \int_o^\Lambda \rho_i(1 - \Theta' t^2 + 2c_i\rho_i t)e^{-t}dt$$

$$\simeq \rho_i(1 - 2\Theta' + 2c_i\rho_i).$$

Videre finner vi at

$$\prod_j (1 - c_j\rho_j^2 t^2) \leq e^{-\Theta t^2} \leq 1 - \Theta t^2 + \Theta^2 t^4/2!$$

$$\simeq 1 - \Theta t^2,$$

slik at

$$E\{V_i\} \leqq \int_o^{\Lambda} \rho_i(1 + \delta - \Theta t^2 + 2c_i\rho_i t)e^{-t}dt$$

$$\simeq \rho_i(1 - 2\Theta + 2c_i\rho_i).$$

Q.E.D.

Lemma 5: Anta at $\Lambda >> 1$ og $n \, \varepsilon\{0, 1, 2, 3\}$. Da er

$$\int_o^1 v^{-1} \int_{\Lambda(1-v)}^{\Lambda} t^n e^{-t} dt dv \sim (n + 1)! \, \Lambda^{-1}(1 + \Lambda^{-1}).$$

Bevis: La I_n betegne venstre side her. Vi finner at $I_o = \phi(\Lambda)$ og $I_1 = \Lambda\phi(\Lambda) - 1 + \phi(\Lambda)$. Når Λ er stor, er

$$\Lambda\phi(\Lambda) - 1 \sim \Lambda^{-1} + \Lambda^{-2} + \ldots \sim \phi(\Lambda).$$

(Symbolet $\sim$ betyr som vanlig "asymptotisk lik".) Dermed blir

$$I_1 \sim 2\Lambda^{-1}(1 + \Lambda^{-1}).$$

Beviset for $n = 2$ og $n = 3$ går helt analogt.

Q.E.D.

Setning 3: Anta at forutsetningene i setning 2(ii) er oppfylt og at $1 + \Lambda^{-1} \simeq 1$.

Vi har da

(i) $\quad \mathrm{Var}\{X_i/\sum_j X_j\} \approx \Lambda^{-1}\rho_i\left[1 + 2c_i\lambda_i - \Theta(6 - 4\lambda_i)\right]$

og

(ii) $\quad E\{X_i(k_i - X_i)/(\sum_j X_j)^2\} \approx \Lambda^{-1}\rho_i(k_i + 4k_ic_i\rho_i$

$+ 6\theta - 6\Theta k_i - 1 - 2c_i\lambda_i).$

<u>Bevis</u>: (i): Ved derivasjon av de genererende funksjoner i lemma 1 får vi

$$\hat{f}_i'(u) + u\hat{f}_i''(u) \approx \lambda_i\left[1 - c_i\lambda_i^2(u-1)^2 + 2c_i\lambda_i(1-u) + \right.$$

$$\left.\lambda_i u(1 - c_i\lambda_i^2(u-1)^2 + 2c_i\lambda_i(1-u)) + 2c_i\lambda_i u(\lambda_i(1-u) - 1)\right]e^{\lambda_i(u-1)}.$$

Etter variabelskiftet $\Lambda(1-u) = t$ blir

$$\int_o^1 v^{-1}\int_o^v(\hat{f}_i(u) + u\hat{f}_i'(u))\,\hat{g}_i(u)\,du\,dv \approx$$

$$\rho_i\int_o^1 v^{-1}\int_{\Lambda(1-v)}^{\Lambda}\left[(1 - c_i\rho_i^2t^2 + 2c_i\rho_it)\,(1 + \rho_i(\Lambda - t))\right.$$

$$\left. + 2c_i\rho_i(\Lambda - t)\,(\rho_it - 1)\right]\,(1 - (\Theta - c_i\rho_i^2)t^2)e^{-t}dt\,dv.$$

Siden $c_i\rho_i^2 \approx c_i\rho_i\Theta \approx \Theta^2 \approx o$, er integralet ovenfor tilnærmet lik

$$\rho_i\int_o^1 v^{-1}\int_{\Lambda(1-v)}^{\Lambda}\left[1 + 4c_i\rho_it + 2c_i\rho_i\Lambda(1 + \rho_it)\right]\,(1 - \Theta t^2)e^{-t}dt\,dv.$$

Lemma 5 og lemma 3(ii) gir

$$E\{V_i^2\} = \rho_i\Lambda^{-1}(1 + 8c_i\rho_i + 2c_i\rho_i\Lambda - 6\Theta - 12c_i\rho_i\Lambda\Theta)$$

$$\approx \rho_i\Lambda^{-1}(1 + 2\Lambda c_i\rho_i - 6\Theta),$$

noe som medfører at

$$\mathrm{Var}\{V_i^2\} = E\{V_i^2\} - (E\{V_i\})^2$$

$$\simeq \Lambda^{-1}\rho_i(1 + 2\Lambda c_i\rho_i - 6\Theta)$$

$$- \rho_i^2(1 - 4\Theta + 4c_i\rho_i)$$

$$= \Lambda^{-1}\rho_i(1 + 2\Lambda c_i\rho_i - 6\Theta - \Lambda\rho_i + 4\Theta\Lambda\rho_i - 4c_i\rho_i^2\Lambda)$$

$$\simeq \Lambda^{-1}\rho_i(1 + 2c_i\lambda_i - 6\Theta + 4\Theta\lambda_i).$$

Punkt (i) er dermed bevist.

(ii): Vi merker oss først at analogt til lemma 3 har vi

$$E\{X_i/(\Sigma X_j)^2\} = \int_o^1 v^{-1} \int_o^v \hat{f}_i'(u) \prod_{i\neq j} \hat{f}_j(u)\, du\, dv.$$

Dette medfører at

$$E\{X_i/(\Sigma X_j)^2\} \simeq \rho_i \int_o^1 v^{-1} \int_{\Lambda(1-v)}^{\Lambda} (1 - c_i\rho_i^2 t^2 + 2c_i\rho_i t)$$

$$\cdot e^{-t} \prod_{i\neq j} (1 - c_j\rho_j t^2)\, dt \simeq$$

$$\rho_i \int_o^1 v^{-1} \int_{\Lambda(1-v)}^{\Lambda} (1 - \Theta t^2 + 2c_i\rho_i t)e^{-t} dt.$$

Ved bruk av lemma 5 blir det siste integralet tilnærmet lik

$$\Lambda^{-1}\rho_i(1 - 6\Theta + 4c_i\rho_i).$$

Herav

$$E\{k_i X_i/(\Sigma X_j)^2\} - E\{V_i^2\} \simeq$$

$$\Lambda^{-1}\rho_i(k_i - 6k_i\Theta + 4k_i c_i\rho_i + 6\Theta - 1 - 2c_i\lambda_i).$$

Q.E.D.

Av setning 3 får vi umiddelbart følgende resultat:

Korollar 1: Dersom θ og $c_i\rho_i$ er små, er

(i) $\mathrm{Var}\{X_i/\sum_j X_j\} \simeq \Lambda^{-1}\rho_i(1 + k_i/N_i + n_i/N_i)$,

og

(ii) $E\{X_i(k_i - X_i)/(\sum_j X_j)^2\} \simeq \Lambda^{-1}\rho_i[k_i(1 - N_i^{-1}) - 1 - n_i/N_i]$.

Lemma 6: La (X_j, Y_j), $j = 1, 2, ..,$ være parvis uavhengige tilfeldige diskrete variable med punktsannsynligheter $f_j(\cdot,\cdot)$ og genererende funksjoner $\hat{f}_j(\cdot,\cdot)$. Da er

$$E\{X_iY_i/(\sum_j X_j)(\sum_j Y_j)\} = \int_0^1\int_0^1 \frac{\partial^2}{\partial u\partial v}\hat{f}_i(u,v)\prod_{i\neq j}\hat{f}_j(u,v)\,du\,dv.$$

Bevis: La $Z_1 = \sum_j X_j$ og $Z_2 = \sum_j Y_j$. Analogt til beviset for lemma 3 er det lett å innse at (X_i, Y_i, Z_1, Z_2) har genererende funksjon lik

$$\hat{f}_i(u_1v_1, u_2v_2)\prod_{i\neq j}\hat{f}_j(v_1, v_2).$$

Ved å derivere og integrere på tilsvarende måte som i beviset av lemma 3, får vi resultatet ovenfor.

Q.E.D.

Setning 4: La (X_i, Y_i), $i = 1, 2, ..,$ være parvis uavhengige simultant hypergeometrisk fordelte variable med forventninger

$(\lambda_{i1}, \lambda_{i2})$, $i = 1, 2, \ldots .$ La

$$\Lambda_k = \sum_j \lambda_{jk} \text{ og } \rho_{ik} = \lambda_{ik}/\Lambda_k \text{ for } k = 1,2.$$

Dersom $\Lambda_k >> 1$ for $k = 1, 2$, er

(i) $E\{X_iY_i/(\sum_j X_j)(\sum_j Y_j)\} \leq \rho_{i1}\rho_{i2}(1 - n_i^{-1})$.

(ii) Hvis vi i tillegg har at $\Theta_k - c_{ik}\rho_{ik} \simeq 0$ for $k = 1, 2$, er

$$\mathrm{Cov}\{X_i/\sum_j X_j,\ Y_i/\sum_j Y_j\} \geq -\rho_{i1}\rho_{i2}/n_i.$$

Bevis: Vi benytter resultatet i lemma 2 og får ved derivasjon at

$$\frac{\partial^2}{\partial s\partial z}\hat{f}_i(s, z) = \lambda_{i1}\lambda_{i2}(1 - n_i^{-1}) \left[1 - \lambda_{i1}^2(s - 1)^2/2k_{i1} + \lambda_{i1}(1 - s)/k_{i1}\right] \left[1 - \lambda_{i2}^2(z - 1)^2/2k_{i2} + \lambda_{i2}(1 - z)/k_{i2}\right] (1 + \lambda_{i1}(s - 1)/n_i + \lambda_{i2}(z - 1)/n_i)^{n_i-2}.$$

Siden

$$(1 + \lambda_{i1}(s-1)/n_i + \lambda_{i2}(z - 1)/n_i)^{n_i-2} \leq e^{-\lambda_{i1}(1 - s) - \lambda_{i2}(1 - z)},$$

får vi etter variabelskiftet $u = \Lambda_1(1 - s)$ og $v = \Lambda_2(1 - z)$ at

$$\int_0^1\int_0^1 \frac{\partial^2}{\partial s\partial z}\hat{f}_i(s, z) \prod_{i\neq j}\hat{f}_j(s, z)\, ds\, dz \leq \rho_{i1}\rho_{i2}(1 - n_i^{-1}) \int_0^{\Lambda_1}\int_0^{\Lambda_2}(1 + u\rho_{i1}/k_{i1})\ (1 + v\rho_{i1}/k_{i2})e^{-u-v}du\, dv.$$

Det siste integralet er tilnærmet lik

$$(1 + \rho_{i1}/k_{i1})\ (1 + \rho_{i2}/k_{i2}) \simeq (1 + (\Lambda_1^{-1} + \Lambda_2^{-1})\ n_i/N_i) \simeq 1.$$

Dermed er (i) bevist. Punkt (ii) følger umiddelbart ved å bruke setning 2.

Q.E.D.

REFERANSER

[1] Burr, Irving W. (1973): Some approximate relations between terms of the hypergeometric, binomial and Poisson distributions. Communications in Statistics, 1: 297-301.

[2] Des Raj (1968): Sampling theory. McCraw-Hill, London.

[3] Hoem, Jan M. (1973): Statistisk Sentralbyrås utvalgsundersøkelser: Elementer av det matematiske grunnlaget. Statistisk Sentralbyrå, Artikkel 58.

[4] Jahnke, Emde og Lösch (1960): Tables of higher functions. B.G. Teubner Verlagsgesellschaft, Stuttgart.

[5] Stegun, I.A. og Abramowitz, M. (1954): Handbook of Matematical functions. U.S. National Bureau of Standards.

Utkommet i serien ART

Issued in the series Artikler fra Statistisk Sentralbyrå (ART)

* utsolgt *out of sale*

Nr. 1* Odd Aukrust: Investeringenes effekt på nasjonalproduktet *The Effects of Capital Formation on the National Product* 1957 28 s.

" 2* Arne Amundsen: Vekst og sammenhenger i den norske økonomi 1920 - 1955 *Growth and Interdependence in Norwegian Economy* 1957 40 s.

" 3* Statistisk Sentralbyrås forskningsavdeling: Skattlegging av personlige skattytere i årene 1947 - 1956 *Taxation of Personal Tax Payers* 1957 8 s.

" 4 Odd Aukrust og Juul Bjerke: Realkapital og økonomisk vekst 1900 - 1956 *Real Capital and Economic Growth* 1958 32 s. kr. 3,50

" 5 Paul Barca: Utviklingen av den norske jordbruksstatistikk *Development of the Norwegian Agricultural Statistics* 1958 23 s. kr. 2,00

" 6 Arne Amundsen: Metoder i analysen av forbruksdata *Methods in Family Budget Analyses* 1960 24 s. kr. 5,00

" 7 Arne Amundsen: Konsumelastisiteter og konsumprognoser bygd på nasjonalregnskapet *Consumer Demand Elasticities and Consumer Expenditure Projections Based on National Accounts Data* 1963 44 s. kr. 5,00

" 8 Arne Øien og Hallvard Borgenvik: Utviklingen i personlige inntektsskatter 1952 - 1964 *The Development of Personal Income Taxes* 1964 30 s. kr. 5,00

" 9 Hallvard Borgenvik: Personlige inntektsskatter i sju vest-europeiske land *Personal Income Taxes in Seven Countries in Western Europe* 1964 16 s. kr. 5,00

" 10 Gerd Skoe Lettenstrøm og Gisle Skancke: De yrkesaktive i Norge 1875 - 1960 og prognoser for utviklingen fram til 1970 *The Economically Active Population in Norway and Forecasts up to 1970* 1964 56 s. kr. 6,00

" 11* Hallvard Borgenvik: Aktuelle skattetall 1965 *Current Tax Data* 1965 38 s. kr. 6,00

" 12 Idar Møglestue: Kriminalitet, årskull og økonomisk vekst *Crimes, Generations and Economic Growth* 1956 63 s. kr. 7,00

" 13 Svein Nordbotten: Desisjonstabeller og generering av maskinprogrammer for granskning av statistisk primærmateriale *Decision Tables and Generation of Computer Programs for Editing of Statistical Data* 1965 11 s. kr. 4,00

" 14 Gerd Skoe Lettenstrøm: Ekteskap og barnetall - En analyse av fruktbarhetsutviklingen i Norge *Marriages and Number of Children - An Analysis of Fertility Trend in Norway* 1965 29 s. kr. 6,00

" 15* Odd Aukrust: Tjue års økonomisk politikk i Norge: Suksesser og mistak *Twenty Years of Norwegian Economic Policy: An Appraisal* 1965 38 s. kr. 6,00

" 16 Svein Nordbotten: Long-Range Planning, Progress- and Cost-Reporting in the Central Bureau of Statistics of Norway *Langtidsprogrammering, framdrifts- og kostnadsrapportering i Statistisk Sentralbyrå* 1966 17 s. kr. 4,00

Nr. 17* Olav Bjerkholt: Økonomiske konsekvenser av nedrustning i Norge *Economic Consequences of Disarmament in Norway* 1966 25 s. kr. 4,00

" 18 Petter Jakob Bjerve: Teknisk revolusjon i økonomisk analyse og politikk? *Technical Revolution in Economic Analysis and Policy?* 1966 23 s. kr. 4,00

" 19 Harold W. Watts: An Analysis of the Effects of Transitory Income on Expenditure of Norwegian Households 1968 28 s. kr. 5,00

" 20 Thomas Schiøtz: The Use of Computers in the National Accounts of Norway *Bruk av elektronregnemaskiner i nasjonalregnskapsarbeidet i Norge* 1968 28 s. kr. 5,00

" 21 Petter Jakob Bjerve: Trends in Quantitative Economic Planning in Norway *Utviklingstendensar i den kvantitative økonomiske planlegginga i Norge* 1968 29 s. kr. 5,00

" 22 Kari Karlsen og Helge Skaug: Statistisk Sentralbyrås sentrale registre *Registers in the Central Bureau of Statistics* 1968 24 s. kr. 3,50

" 23 Per Sevaldson: MODIS II A Macro-Economic Model for Short-Term Analysis and Planning *MODIS II En makroøkonomisk modell for korttidsanalyse og planlegging* 1968 40 s. kr. 4,50

" 24* Olav Bjerkholt: A Precise Description of the System of Equations of the Economic Model MODIS III *Likningssystemet i den økonomiske modell MODIS III* 1968 30 s. kr. 4,50

" 25 Eivind Hoffmann: Prinsipielt om måling av samfunnets utdanningskapital og et forsøk på å måle utdanningskapitalen i Norge i 1960 *On the Measurement of the Stock of Educational Capital and an Attempt to Measure Norway's Stock of Educational Capital in 1960* 1968 60 s. kr. 5,00

" 26 Hallvard Borgenvik: Aktuelle skattetall 1968 *Current Tax Data* 1969 40 s. kr. 7,00

" 27 Hallvard Borgenvik: Inntekts- og formuesskattlegging av norske kapitalplasseringer i utlandet *Income and Net Wealth Taxes of Norwegian Investment in Foreign Countries* 1969 40 s. kr. 7,00

" 28 Petter Jakob Bjerve og Svein Nordbotten: Automasjon i statistikkproduksjonen *Automation of the Production of Statistics* 1969 30 s. kr. 7,00

" 29 Tormod Andreassen: En analyse av industriens investeringsplaner *An Analysis of the Industries Investment Plans* 1969 26 s. kr. 5,00

" 30 Bela Balassa og Odd Aukrust: To artikler om norsk industri *Two Articles on Norwegian Manufacturing Industries* 1969 40 s. kr. 5,00

" 31* Hallvard Borgenvik og Hallvard Flø: Virkninger av skattereformen av 1969 *Effects of the Taxation Reform of 1969* 1969 35 s. kr. 7,00

" 32 Per Sevaldson: The Stability of Input-Output Coefficients *Stabilitet i kryssløpskoeffisienter* 1969 40 s. kr. 7,00

Nr. 33 Odd Aukrust og Hallvard Borgenvik: Inntektsfordelingsvirkninger av skattereformen av 1969 *Income Distribution Effects of the Taxation Reform of 1969* 1969 29 s. kr. 7,00

" 34 Odd Aukrust og Svein Nordbotten: Dataregistrering, dataarkiver og samfunnsforskning *Data Registration, Data Banks and Social Research* 1970 43 s. kr. 7,00

" 35 Odd Aukrust: PRIM I A Model of the Price and Income Distribution Mechanism of an Open Economy *PRIM I En modell av pris- og inntektsfordelingsmekanismen i en åpen økonomi* 1970 61 s. kr. 7,00

" 36 Arne Amundsen: Konsumets og sparingens langsiktige utvikling *Consumption and Saving in the Process of Long-Term Growth* 1970 18 s. kr. 5,00

" 37 Steinar Tamsfoss: Om bruk av stikkprøver ved kontoret for intervjuundersøkelser, Statistisk Sentralbyrå *On the Use of Sampling Surveys by the Central Bureau of Statistics, Norway* 1970 46 s. kr. 7,00

" 38 Svein Nordbotten: Personmodeller, personregnskapssystemer og persondataarkiver *Population Models, Population Accounting Systems and Individual Data Banks* 1970 28 s. kr. 7,00

" 39 Julie E. Backer: Variasjoner i utviklingen hos nyfødte barn *Variations in the Maturity Level of New Born Infants* 1970 36 s. kr. 7,00

" 40 Svein Nordbotten: Two Articles on Statistical Data Files and Their Utilization in Socio-Demographic Model Building *To artikler om statistiske dataarkiver og deres bruk i sosio-demografisk modellbygging* 1971 30 s. kr. 7,00

" 41 Per Sevaldson: Data Sources and User Operations of MODIS, a Macro-Economic Model for Short Term Planning *Datagrunnlag og brukermedvirkning ved MODIS, en makroøkonomisk modell for planlegging på kort sikt* 1971 31 s. kr. 7,00

" 42 Erik Biørn: Fordelingsvirkninger av indirekte skatter og subsidier *Distributive Effects of Indirect Taxes and Subsidies* 1971 42 s. kr. 5,00

" 43 Hallvard Borgenvik og Inger Gabrielsen: Aktuelle skattetall 1970 *Current Tax Data* 1971 53 s. kr. 7,00

" 44 Vidar Ringstad: PRIM II En revidert versjon av pris- og inntektsmodellen *PRIM II A Revised Version of the Price and Income Model* 1972 43 s. kr. 7,00

" 45 Jan M. Hoem: Purged and Partial Markov Chains *Lutrede og partielle Markovkjeder* 1972 16 s. kr. 5,00

" 46 Jan M. Hoem: Two Articles on the Interpretation of Vital Rates *To artikler om tolking av befolkningsrater* 1972 33 s. kr. 7,00

" 47 Inger Gabrielsen: Aktuelle skattetall 1972 *Current Tax Data* 1972 58 s. kr. 8,00

" 48 Vidar Ringstad: Om estimering av økonomiske relasjoner fra tverrsnitts-, tidsrekke- og kombinert tverrsnitts tidrekke-data *On the Estimation of Economic Relations Using Cross Section-, Time Series- and Combined Cross Section- Time Series- Data* 1972 26 s. kr. 7,00

Nr. 49 Jan M. Hoem: On the Statistical Theory of Analytic Graduation *Statistisk teori for analytisk glatting* 1972 41 s. kr. 7,00

" 50 Henry M. Peskin: National Accounting and the Environment *Nasjonalregnskap og miljøverdier* 1972 60 s. kr. 8,00

" 51 Eivind Gilje: Analytic Graduation of Age-Specific Fertility Rates *Analytisk glatting av aldersspesifikke fødselsrater* 1972 49 s. kr. 8,00

" 52 Jan M. Hoem og Arne Rideng: Kommentarer til Statistisk Sentralbyrås framskriving av folkemengden i kommunene 1972-2000 *Comments to the Regional Population Projections for Norway* 1972 29 s. kr. 7,00

" 53 Juul Bjerke: Estimering av konsumfunksjoner på grunnlag av nasjonalregnskapsdata 1865-1968 *Estimating Consumption Functions from National Accounts Data* 1972 60 s. kr. 8,00

" 54 Jan M. Hoem: Usikkerhet ved befolkningsprognoser *Inaccuracy of Population Projections* 1973 63 s. kr. 8,00

" 55 Erik Biørn: Prognoser for de langsiktige endringer i sammensetningen av det private konsum *Long Term Forecasts for the Changes in the Composition of the Private Consumption* 1973 71 s. kr. 8,00

" 56 Jan M. Hoem: Inhomogeneous Semi-Markov Processes, Select Actuarial Tables, and Duration-Dependence in Demography *Inhomogene semi-markovprosesser, selekte aktuartabeller og varighetsavhengighet i demografi* 1973 54 s. kr. 8,00

" 57 Svein Brenna: Revisjon av indeksene for utenrikshandelen *Revision of Indices for Foreign Trade* 1973 47 s. kr. 7,00

" 58 Jan M. Hoem: Statistisk Sentralbyrås utvalgsundersøkelser: Elementer av det matematiske grunnlaget *The Sample Surveys of the Central Bureau of Statistics of Norway: Basic Mathematical Elements* 1973 59 s. kr. 8,00

" 59 Inger Gabrielsen: Aktuelle skattetall 1973 *Current Tax Data* 1973 63 s. kr. 8,00

" 60 Per Sevaldson: Om oppstilling og bruk av regionalt nasjonalregnskap *Construction and Use of Regional National Accounts* 1973 74 s. kr. 7,00

" 61 Jan M. Hoem: Levels of Error in Population Forecasts *Usikkerhetsnivåer ved befolkningsprognoser* 1973 46 s. kr. 8,00

" 62 Arne Rideng og Bjørn Lied Tønnesen: Statistisk Sentralbyrås regionale befolkningsframskrivinger Nåværende opplegg og utviklingsplaner 1974 *The Regional Population Projections of the Central Bureau of Statistics of Norway Current Procedure and Plans for the Future* 1974 25 s. kr. 7,00

" 63 Erik Biørn: Estimering av makro-konsumfunksjoner for etterkrigstiden: metodespørsmål og empiriske resultater *Estimating Aggregate Consumption Functions for the Post-War Period: Methodological Problems and Empirical Results* 1974 84 s. kr. 8,00

" 64 Terje Assum: Hvem har nytte av forbrukerservice? *To Whose Benefit is the Consumer Service?* 1974 22 s. kr. 5,00

Publikasjonen utgis i kommisjon hos
H. Aschehoug & Co., Oslo, og er til salgs hos alle bokhandlere
Pris kr. 7,00

Omslag trykt hos Grøndahl & Søn, Oslo

ISBN 82-537-0378-3