Arbeidsnotater

STATISTISK SENTRALBYRÅ

Dronningensgt. 16, Oslo-Dep., Oslo 1. Tlf. 41 38 20

IO 76/28 27. september 1976

METODEHEFTE NR. 19

NOTAT OM "BRUK AV SUPERPOPULASJONSMODELLER"

INNHOLD

FORORD

Metodehefter i serien Arbeidsnotater

I tilknytning til mange prosjekter i Statistisk Sentralbyrå utarbeides det mindre, upretensiøse notater for avklaring av spørsmål av metodisk interesse. Det kan dreie seg om utvalgsteknikk, alternative spørsmålsformuleringer, presentasjonsmetoder, begrepsavklaringer, diskusjon av "funn" i data, systemidéer eller andre temaer. Selv om mange slike notater bare har begrenset interesse i ettertid, vil det blant dem være noen som kunne fortjene å bli alminnelig tilgjengelig. Det kan også være nyttig å ha dem registrert sentralt slik at det blir lettere å få oversikt over det stoffet som foreligger, og lettere å referere tilbake til det. Byrået publiserer derfor leilighetsvis et passende antall notater av dette slaget samlet i metodehefter i serien Arbeidsnotater.

Kontorlederne bes holde øynene åpne for denne nye publiseringsmuligheten.

Forskningssjef Per Sevaldson er redaktør av metodeheftene. Førstekontorfullmektig Liv Hansen er redaksjonssekretær. Medarbeidere i Byrået som lager stoff som kan være aktuelt, bes sende dette til redaksjonen etter hvert som det blir ferdig. Retningslinjer for utformingen av inserater i metodeheftene finnes på side 46 til side 47 i Metodehefte nr. 9 (ANO IO 73/36).

BRUK AV SUPERPOPULASJONSMODELLER VED INNSAMLING OG ANALYSE AV DATA FRA UTVALGSUNDERSØKELSER

Av

Ib Thomsen

INNHOLD

1. INNLEDNING

I løpet av de siste årene er det publisert en rekke arbeider som behandler utvalgs-teoretiske problemer under den forutsetning at den endelige populasjon er tenkt generert av en stokastisk "mekanisme". Ideen er ikke ny, men interessen har tatt seg opp i de siste årene. Hensikten med dette notat er å gi en kort oversikt over arbeider som er skrevet innen dette feltet, og peke på områder hvor vi i Statistisk Sentralbyrå mener slike betraktningsmåter kan føre til fruktbare løsninger av viktige, praktiske problemer, som ikke har funnet noen tilfredsstillende løsning i den tradisjonelle utvalgsteori. Vi skal også gjøre oppmerksom på noen problemer, som kun i meget liten grad er blitt tatt opp innen teorien for statistiske utvalg, men som opptrer i forbindelse med analyse av data samlet inn ved slike utvalg.

Den "klassiske" utvalgsteori slik den er formulert i de fleste lærebøker [7], [17], er som følger: Vi tenker oss at vi har en bestemt endelig populasjon, f.eks. populasjonen av norske husholdninger. Til hvert element i populasjonen er det knyttet et (eller flere) kjennetegn som en ønsker å studere. F.eks. ønsker en å estimere de samlede utgifter til mat i den norske befolkning. For å gjøre dette, tas et utvalg fra populasjonen, og på grunnlag av resultatet i utvalget, estimerer vi det resultat vi ville ha fått dersom vi hadde undersøkt alle elementene i populasjonen. Utvalgsteorien er opptatt med å besvare følgende to spørsmål:

1. Hvordan skal en trekke utvalget?
2. Hvordan skal en på grunnlag av resultatet i utvalget kunne si noe om det en ville ha fått dersom alle elementer i populasjonen var blitt trukket?

Svaret er at hvis vi trekker ut de enheter som skal være med i utvalget på en slik måte at enhetene i populasjonen har kjente sannsynligheter for å komme med i utvalget, så kan det settes opp en stokastisk modell. De verdier vi observerer i utvalget er da stokastiske variable, og de data i populasjonen vi vil vite noe om, er parametre i sannsynlighetsfordelingen for disse variable. Ved å bruke metoder fra den teoretiske statistikken kan vi da estimere parametrene.

I motsetning til denne modellen kan en tenke seg at verdiene i den endelige populasjon er realisasjoner av stokastiske variable. Et utvalg kan da tenkes å være framkommet i to trinn: Først er den endelige populasjon trukket som et "stort" utvalg fra en uendelig populasjon, og deretter er det trukket et mindre utvalg fra den endelige populasjon. På grunnlag av observasjonene i utvalget kan vi gjøre generaliseringer til den endelige populasjon, eller til superpopulasjonen. I en gitt situasjon kan det være vanskelig å avgjøre om en bør gjøre det ene eller det andre. I [2] skriver Barnard:

> "Finally let me come to my last main point: This could be summed up by saying that we might do well, in sample surveys, to make more frequent use of models like those put forward by Prof. Sprott and Mr. Kalbfleisch - in which we regard our given sample, and the remainder of the finite population, as two samples from an underlying super population about which distributional form assumptions are in order."

I dette notatet skal vi se på nytten av en superpopulasjonsmodell både i tilfeller hvor interessen er knyttet til parametre i superpopulasjonen og i den endelige populasjonen. I avsnitt 3 skal vi gjøre rede for noen problemer som oppstår ved generalisering til superpopulasjonen. I avsnitt 4 skal vi se hvordan strategier fra den "vanlige" utvalgsteori kan vurderes og sammenliknes ved hjelp av superpopulasjonsmodeller. I avsnitt 5 skal gis noen eksempler på nye strategier, som er utviklet ved å bruke superpopulasjonsmodeller.

2. EN ENKEL MODELL

Mange av de problemer vi skal ta opp i dette notat kan belyses innen følgende enkle modell:

Vi skal tenke oss at den endelige populasjon består av N elementer, nummerert fra 1 til N. Til i-te element er knyttet to variabelverdier (Y_i, x_i), hvor x_i er kjente. Vi antar nå følgende superpopulasjonsmodell:

$$Y_i = \beta x_i + \alpha + U_i, \qquad (2.1)$$

hvor α og β er ukjente konstanter, og U_i er uavhengige, stokastiske variable med $E(U_i) = 0$ og $E(U_i^2) = \sigma_i^2$.

En utvalgsplan er definert som et sett, $\mathcal{S}$, av alle mulige utvalg. For ethvert $s \in \mathcal{S}$ er definert et tall $p(s)$, slik at $0 < p(s)$ og $\Sigma\, p(s) = 1$, hvor $p(s)$ er sannsynligheten for at utvalget s skal bli trukket.

3. ANALYTISKE UNDERSØKELSER

Med analytiske undersøkelser menes i dette notat undersøkelser hvor en ønsker å estimere parametre i superpopulasjonen. I modell (2.1) kan målet f.eks. være å lage en strategi (dvs. både en utvalgsplan og en estimator) for å estimere α og/eller β. Dette problemet er bl.a. behandlet i [26], hvor det er antatt at $\sigma_i^2 = \sigma$. En har tatt for seg følgende klasse av estimatorer:

$$\sum_{i \varepsilon s} a_i(s)\, Y_i \, , \qquad (3.1)$$

hvor $\sum_{i \varepsilon s}$ betyr summen over alle elementer i utvalget og $a_i(s)$ er konstanter som kan avhenge av utvalget og av x_i men som ikke avhenger av Y_i ($i = 1,2,\ldots,N$). Når en tar hensyn til repetisjoner både over superpopulasjonen og over mulige utvalg, vises det at det blant forventningsrette estimatorer for α innen klassen (3.1) ikke finnes noen med uniform minste varians. Hvis en derimot forlanger forventningsretthet over superpopulasjonen for hvert mulig utvalg, viser det seg at den vanlige minste kvadraters estimator har minst varians.

Det er verdt å merke seg at estimatorene er uavhengige av utvalgsplanen. En liknende konklusjon er framsatt i [28]. For mange statistikere med bakgrunn i utvalgsteori synes dette å være uakseptabelt, og en har derfor forsøkt å lage nye modeller, hvor estimatorene tilsynelatende avhenger av utvalgsmetoden [12], [19].

En litt annen, men kanskje beslektet problemstilling består i at en i stedet for β, velger å estimere β', hvor

$$\beta' = \sum_{i=1}^{N} Y_i \left(x_i - \frac{1}{N}\sum_{i=1}^{N} x_i\right) \Big/ \sum_{i=1}^{N} \left(x_i - \frac{1}{N}\sum_{i=1}^{N} x_i\right)^2 .$$

β' er minste kvadraters estimator for β hvis vi hadde hatt data for hele populasjonen [14], [15], [19].

Mens modeller av typen (3.1) har vært brukt i Statistisk Sentralbyrå til estimering av bl.a. inntektselastisiteter på grunnlag av forbruksundersøkelser, har jeg aldri sett noen grunn til å innføre parametre av typen β' [1].

3.1 Tidsserie / tverrsnittsanalyse av data fra gjentatte undersøkelser

I [3] er det skissert noen problemer i forbindelse med analytisk bruk av gjentatte undersøkelser. En oppstiller her følgende strukturrelasjon:

$$y_{it} = a_0 + \sum_{j=1}^{J} a_j q_{ji} + \sum_{n=1}^{N} b_n x_{nt} + \sum_{k=1}^{K} c_k z_{kit} + u_i + v_t + w_{it}$$

hvor a-ene, b-ene og c-ene er ukjente konstanter,

y_{it} = Verdi av den endogene variable for husholdning nr. i, i år nr. t

q_{ji} = Verdi av husholdningsspesifikk eksogen variabel nr. j for husholdning nr. i.
Med husholdningsspesifikk variabel menes en variabel som bare varierer over husholdning, ikke over tid (eksempler kan være hovedpersonens fødselsår og kjønn).

x_{nt} = Verdi av tidsspesifikk eksogen variabel nr. n ($n = 1,2,\ldots,N$) i år nr. t.
Med tidsspesifikk variabel menes en variabel som bare varierer over tid, ikke over husholdning (eksempler er konsumpriser, konjunktursituasjonen, skattesatser m.m.).

z_{kit} = Verdi av kombinert eksogen variabel nr. k for husholdning nr. i, i år t.
Med kombinert variabel menes en variabel som varierer både over husholdning og over tid (eksempler er husholdningens løpende inntekt og formue, antall husholdningsmedlemmer etc.).

u_i = Restleddskomponent for husholdning nr. i.

v_t = Tidsspesifikk restleddskomponent i år nr. t.

w_{it} = Kombinert restleddskomponent for husholdning nr. i, i år nr. t.

Slike relasjoner er mye brukt ved analyse av et komplett tidsserie-/tverrsnittsmateriale. I situasjoner hvor vi ikke har et komplett sett med data, hvilket er tilfelle ved roterende utvalg, finnes det nesten ingen litteratur.

Viktige problemstillinger er:

(i) Finnes det flere anvendelsesområder for roterende utvalg enn for utvalg trukket uavhengig av hverandre?

(ii) Kan det sies noe om forskjeller i estimeringspresisjon ved bruk av de to innsamlingsmetoder?

(iii) Er det mulig på grunnlag av et ukomplett tidsserie-/tverrsnittsmateriale som fremkommer ved bruk av roterende utvalg, kombinert med visse forutsetninger, å konstruere et fullstendig materiale? F.eks. ved bruk av en "missing observations" teknikk.

En av årsakene til at slike problemer ikke er blitt tatt opp kan være at de forutsetter en ganske stor mengde data, samlet inn over et lengre tidsrom. Modellen er aktuell i Statistisk Sentralbyrå når det foreligger flere data fra de løpende forbruksundersøkelser.

4. BRUK AV SUPERPOPULASJONSMODELLER TIL VURDERING AV TIDLIGERE KJENTE STRATEGIER

Teknikken med å bruke superpopulasjonsmodeller for å vurdere gitte strategier er ikke ny. I [7] har Cochran gjort bruk av en superpopulasjonsmodell for å sammenlikne rateestimering kombinert med enkel lotterisk utvalgstrekking, og Horvitz - Thompson estimering. Senere er liknende modeller brukt av Hanurav (1967), Brewer (1963). I disse arbeidene har en tatt for seg en modell av typen (2.1). Hensikten med å innføre en superpopulasjonsmodell er nå ikke å estimere α og/eller β. En ønsker nå å estimere ΣY_i, og innfører superpopulasjonsmodellen for å vurdere forskjellige strategier når modellen er riktig. La $\hat{Y}_s$ betegne en estimator med tilhørende utvalgsplan. Det kriteriet en ofte bruker for sammenlikning er

$$M(\hat{Y}_s) = E\{\sum_s p(s)(\hat{Y}_s - Y)^2\}. \tag{4.1}$$

Vi skal se på et enkelt eksempel, hentet fra [11].

La

$$\hat{Y}'_s = \frac{\sum_{i\varepsilon s} Y_i}{\sum_{i\varepsilon s} x_i} \sum_{i=1}^{N} x_i ,$$ kombinert med enkel lotterisk trekking.

La dessuten

$$\hat{Y}''_s = \frac{1}{n} \sum_{i\varepsilon s} \frac{Y_i}{\pi_i} ,$$ hvor $\pi_i = x_i / \sum_{i=1}^{N} x_i$, n er utvalgsstørrelsen. Utvalget trekkes ved at en foretar n trekninger med tilbakelegging. I hver trekking har element i sannsynligheten $x_i / \sum_{i=1}^{N} x_i$ for å bli trukket.

Superpopulasjonsmodellen er modellen (2.1) med $\sigma_i = x_i^{2\gamma} \sigma^2$. For store verdier av N viser det seg da at $\hat{y}''_s$ er bedre enn $\hat{y}'_s$ hvis

$$\gamma > \tfrac{1}{2} .$$

Også innen et annet område av utvalgsteorien har en brukt en liknende framgangsmåte for å vurdere forskjellige strategier mot hverandre. F.eks. for å vurdere forskjellige måter å stratifisere på har en brukt modeller av typen (2.1) [9], [27].

I tillegg til at en her har vurdert gitte strategier, bestemmes også deler av strategien ved å minimalisere forventet kvadratavvik. Dette gjelder f.eks. konstruksjonen av strataene samt bestemmelse av optimalt antall strata.

5. BRUK AV SUPERPOPULASJONSMODELLER FOR Å LAGE NYE STRATEGIER

I de siste årene er det skrevet en del arbeider som har til hensikt å finne optimale strategier under forskjellige superpopulasjonsmodeller. Ericson bruker i [10] et bayesiansk argument med β som stokastisk variabel. Kalbfleisch og Sprott anvender et "fidusial" argument. Her skal vi særlig trekke fram en rekke arbeider av Royall [20], [21], [22], [23], samt noen arbeider av Scott og Smith [24], [25].

5.1 Optimal strategi når superpopulasjonsmodellen er en lineær, heteroscedastisk regresjonsmodell

La situasjonen være den samme som ovenfor. I [20] finner Royall den strategi som minimerer forventet utvalgsvarians. Under modell (2.1) viser den beste estimator seg å være

$$\hat{y}^{**} = \sum_{i \in s} Y_i + \{N - n(s)\}\, \hat{\alpha}^{**} + \hat{\beta}^{**} \sum_{i \notin s} x_i \, , \qquad (5.1)$$

hvor $\hat{\alpha}^{**}$ og $\hat{\beta}^{**}$ er de vanlige veide minste kvadraters estimatorer for α og β. Denne estimator viser seg å være "best" for enhver utvalgsplan. Den beste utvalgsplan viser seg å bestå av et utvalg, nemlig det utvalg som minimerer uttrykket

$$\left(\frac{1}{N} \sum_{i=1}^{N} x_i - \frac{1}{N} \sum_{i \in s} x_i\right)^2 \Big/ \sum_{i \in s} \left(x_i - \frac{1}{N} \sum_{j \in s} x_j\right)^2 . \qquad (5.2)$$

I tilfellet med enkel tilfeldig utvalg og $\sigma_i^2 = \sigma^2$ er estimatoren (5.1) identisk med den vanlige regresjonsestimator. Estimatorer av typen (5.1) er brukt gjennom flere år i Statistisk Sentralbyrå ved estimering av gjennomsnitt innen små delklasser [1].

Derimot har en aldri tatt sikte på å lage en optimal utvalgsplan, definert ved (4.2). Begrunnelsen har vært at et selvveiende utvalg vil være mer robust overfor svakheter i modellen. Andre teoretiske arbeider av Royall og Herson bekrefter dette synet [22], [23] .

5.2 Optimal strategi ved gjentatte undersøkelser når de endelige populasjoner tenkes generert som en tidsserie

Ved Statistisk Sentralbyrås løpende undersøkelser er samplingfeilen ved de publiserte estimater av betydelig størrelse. Det er ofte av spesiell interesse å studere kvartalsvise og årlige endringer i tallene, og nettopp for estimater av endringstall er samplingfeilen svært stor. De metodene som er studert for å øke presisjonen av estimatene, har til nå vært hentet fra den tradisjonelle utvalgsteori. Disse metodene har det til felles at en ikke utnytter noe om sammenhengen mellom påfølgende verdier av de størrelser en ønsker å estimere. Ofte er det imidlertid kjent at det er en sammenheng mellom f.eks. forbruket i Norge ett år og de følgende. I noen tilfelle kan sammenhengen være av typen beskrevet i pkt. 3.1, mens en i andre situasjoner kan bruke modeller fra tidsserieanalyse. I [4], [25] har en sett på konsekvensene av å forutsette at de endelige populasjonstotaler er realisasjoner av en stokastisk prosess. Følgende enkle modell illustrerer teknikken, [25]:

La Θ_t betegne realisasjonen av en stasjonær stokastisk prosess, av følgende enkle type:

$$\Theta_t = \lambda \Theta_{t-1} + \varepsilon_t, \quad 0 \leq \lambda \leq 1, \tag{5.2.1}$$

hvor ε_t er ukorrelerte stokastiske variable, med $E(\varepsilon_t) = 0$ og $E(\varepsilon_t^2) = \sigma^2$. Vi antar nå at y_t er en estimator basert på et utvalg for Θ_t, og at

$$y_t = \Theta_t + e_t, \tag{5.2.2}$$

hvor e_t er ukorrelerte stokastiske variable, med $E(e_t) = 0$ og $E(e_t^2) = s^2$. Vi forutsetter altså uavhengighet mellom trekningene ved forskjellige tidspunkter. Modellen (5.2.1) induserer nå følgende prosess for y_t:

$$y_t = \lambda y_{t-1} + w_t - \beta w_{t-1},$$

hvor $\{w_t\}$ er ukorrelerte, med $E(w_t) = 0$, $E(w_t^2) = \nu^2$. Da $cov(y_t, y_{t-k}) = cov(\Theta_t, \Theta_{t-k}) + cov(e_t, e_{t-k})$ er

$$\beta \nu^2 = \lambda s^2$$

$$(1+\beta^2)\nu^2 = (1+\lambda^2) s^2 + \sigma^2.$$

Fra resultater fra tidsrekkeanalysen [5; pp. 152-4], følger nå at beste estimator for y_t gitt $y_{t-1}, y_{t-2}, \ldots,$ er

$$\hat{y}_t = \{ (\lambda - \beta) / \beta\} \sum_{j=1}^{\infty} \beta^j y_{t-j}.$$

I [5] er vist at beste, lineære, forventningsrette estimator for Θ_t gitt $y_t, y_{t-1}, \ldots,$ da er

$$\Theta_t = (1-\pi) \sum_{j=0} \beta^j y_{t-j} ,$$

hvor $\pi = s^2/\nu^2 = \beta/\lambda$. Det er også vist at reduksjonen i varians, sammenliknet med å bruke y_t er $(1-\pi)$. Dvs. at reduksjonen avhenger av forholdet mellom samplingvariansen og variansen på "noisen" i den induserte modellen for y_t.

Når en skal bruke slike modeller i praksis, har det vist seg at det oppstår en del problemer:

(a) Modellene i praksis er sjelden av en så enkel type som (5.2.1). Vanligvis er det sesongvariasjoner, som kan føre til at det blir matematisk komplisert å finne beste, lineære estimator.

(b) For et gitt sett av observasjoner, $\{y_t\}$, skal en bestemme modellen for $\{\Theta_t\}$. Ofte fører dette til flere mulige valg, som alle er konsistente med $\{y_t\}$. Kva er konsekvensene av feil valg av modell for $\{\Theta_t\}$?

(c) I metoden ovenfor forutsettes parametrene i prosessen $\{\Theta_t\}$ å være kjente. Hva er effekten av at en bruker estimerte verdier?

I Statistisk Sentralbyrå arbeider vi med å bruke metoder fra tidsserieanalyse til estimering i arbeidskraftundersøkelsene [8]. I løpet av et års tid regner vi med å vite mer om effektiviteten til slike metoder i praksis.

6. SLUTTMERKNADER

Ikke alle muligheter ved bruk av superpopulasjonsmodeller er blitt tatt opp her, bl.a. fordi vi beveger oss i delvis ukjent terreng. Spesielt vil jeg nevne problemene skissert i avsnitt 3 og avsnitt 5.2. Dessuten vil jeg nevne at vi ved Statistisk Sentralbyrå arbeider med andre superpopulasjonsmodeller, hvor både Y-ene og x-ene antar verdiene 0 og 1. Estimatorene vi har funnet vil bli brukt i en undersøkelse som har til formål å oppdatere resultatene fra Folke- og boligtellingen 1970. Endelig ser vi på mulighetene for ved hjelp av superpopulasjonsmodeller å estimere tall for delklasser, som er for små til at utvalgene tillater estimering ved vanlige metoder.

En ikke uvesentlig fordel ved å bruke superpopulasjonsmodeller består av at en minsker den avstand som ofte finnes mellom utvalgsstatistikeren og de personer som analyserer data. Superpopulasjonsmodellene blir nå et felles utgangspunkt, og resultater fra analysene av undersøkelsene kan nå også bli brukt ved valg av strategi i en senere undersøkelse.

7. REFERANSER

[1] Amundsen, A. (1960): Metoder i analysen av forbruksdata. Artikler nr. 6. Statistisk Sentralbyrå.

[2] Barnard, G.A. (1969): Summary Remarks. New Developments in Survey Sampling. Wiley - Inter Science, N.Y.

[3] Biørn, E. (1974): Løpende forbruksundersøkelser: Noen momenter for å vurdere hensiktsmessigheten fra et økonometrisk synspunkt av å anvende roterende utvalg. Stensil KEB/WTD 29.8.74. Statistisk Sentralbyrå.

[4] Blight, B.J.N. and Scott, A.J. (1973): A stochastic model for Repeated surveys. J. R. Statist. Soc., B, 35, 61-66.

[5] Box, G.E.P. and Jenkins, G.M. (1970): Time series analysis, forecasting and control. San Francisco: Holden-Day, Inc.

[6] Brewer, K.R.W. (1963): Ratio estimation and finite populations: Some results deducible from the assumption of an underlying stochastic process. Aust. J. Statist. 5, 93-105.

[7] Cochran, W.G. (1953): Sampling techniques, 1st edn 122-24, Wiley, N.Y.

[8] Dagsvik, J. (1975): Prosjektskisse til prosjektet "Estimering i AKU ved hjelp av metoder fra tidsserieanalysen. Stensil JDa/MFo, 9/7-75. Statistisk Sentralbyrå.

[9] Dalenius, T. (1957): Sampling in Sweden. Almqvist and Wiksell, Stockholm.

[10] Ericson, W.A. (1969): Subjective bayesian models in sampling finite populations. (With discussion.) J. R. Statist. Soc. B, 31, 195-224.

[11] Foreman, E.K. and Brewer, K.R. (1971): The efficient use of supplementary information in standard sampling procedures. J. R. Statist., Soc. B, 33, 291-400.

[12] Fuller, W. A. (1973): Regression analysis for sample surveys. Presented at the 39th session of the ISI, Vienna.

[13] Hanurav, T.V. (1967): Optimum utilization of auxiliary information: π p s sampling of two units from a stratum. J. R. Statist., Soc. B, 29, 374-91.

[14] Hartley, H.O. and Sielken, Jr. (1975): A "Superpopulation viewpoint" for finite population sampling". Biometrics 31. (2), 411-422.

[15] Jønrup, H. (1975): A note on the theory of regression analysis in sample surveys. Contributed paper at the 40th session of the ISI, Warszawa.

[16] Kalbfleisch, J.D. and Sprott, D.A. (1969): Applications of likelihood and fiducial probability to sampling finite populations. New development in survey sampling. pp. 358-89.

[17] Kish, L. (1965): Survey Sampling. Wiley, N.Y.

[18] Kish, L. and Frankel, M. (1974): Inference from complex samples. J. Roy. Statist. Soc., Ser. B.

[19] Konijn, H. S. (1962): Regression analysis in sample surveys. J.A.S.A. 57.

[20] Royall, R.M. (1970): On finite population sampling theory under certain linear regression models. Biometrika, 57, 377-87.

[21] Royall, R.M. (1971): Linear regression models in finite population sampling theory. Foundations of Statistical Inference. Holt, Rinehart and Winston of Canada, Ltd.

[22] Royall, R.M. and Herson, H.J. (1973): Robust estimation in finite populations. I. J.A.S.A. 68, 880-889.

[23] Royall, R.M. and Herson, H.J. (1973): Robust estimation in finite populations. II. J.A.S.A. 68, 890-893.

[24] Scott, A. and Smith, T.M.F. (1969): A note on estimating secondary characteristics in multivariate surveys. Sankhyā, A, 497-498.

[25] Scott, A. and Smith, T.M.F. (1974): Analysis of repeated surveys using time series methods. J. A. S. A. 69, 674-678.

[26] Thomsen, I. (1974): Design and estimation problems when estimating a regression coefficient from survey data. Publiseres i Metrika.

[27] Thomsen, I. (1975): On the effect of stratification when two stratifying variables are used. Publiseres i J. A. S. A.

[28] Brewer, K. R. W. and Mellor, R. W. (1973): The effect of sample structure on analytical surveys. Austral. J. Statist. 15, 3.