Arbeidsnotater

STATISTISK SENTRALBYRÅ

IO 64/4 Oslo, 11. juni 1964

Varehandelsstatistikken

Vurdering av ny estimeringsmetode 1963

Av

Hans Olav Egede Larssen

I n n h o l d

HOEL/MF

1. Innledning

Varehandelsstatistikken for 1963 bygger på oppgaver fra et utvalg av bedrifter, trukket på grunnlag av det best mulig ajourførte bedriftsregister.

I registeret er bedriftene inndelt etter størrelse, som er definert ved totalomsetningen. Inndelingen er følgende:

Omsetningsgruppe:	Omsetning i 1 000 kr.	Utvalgsprosent: Engros	Detalj
0	0 - 9	1	1
1	10 - 19		
2	20 - 49	1	2
3	50 - 99	1	2,5
4	100 - 199	2	5
5	200 - 499	5	$16\frac{2}{3}$
6	500 - 999	10	25
7	1 000 - 1 999	$16\frac{2}{3}$	50
8	2 000 - 4 999	50	100
9	5 000 →	100	100
x	Uoppgitt	10	10

Omsetningsgruppe x, uoppgitt, faller i det følgende helt utenfor.

Det inndeles også etter næring, og dette skjer etter Standard for næringsgruppering. Man har 62-63, engros og 64-65, detalj. Detalj deles videre i 2 grupper. På denne måten inndeles i 3 grupper, som i det etterfølgende vil bli referert til som <u>hovedgrupper</u>:

62 - 63, Engros
641 - 642, Detalj, nærings- og nydelsesmidler
64 rest - 65, Annen detalj

Videre inndeling skjer i 3- og 4-sifrede næringsgrupper.

Geografisk inndeles i handelsfelt, handelsområde og handelsdistrikt.

For publikasjon har man to forskjellige inndelinger av hver hovedgruppe:

a) I 4-sifrede næringsgrupper eller kombinasjoner av disse.
b) En relativt grov inndeling etter næring kombinert med geografisk inndeling, hovedsakelig etter handelsområde.

De grupper som fremkommer ved den ene eller den andre av disse inndelinger, vil bli referert til som <u>undergrupper</u>.

2. Problemstillingen

Ved begge de nevnte undergruppe-inndelinger viser det seg at det totale antall strata blir stort og de enkelte strata ofte små. Særlig i de lavere omsetningsgrupper, hvor utvalget er helt ned til 1 o|o, vil dette kunne føre til at antall utvalgsenheter fra hvert stratum kan bli nesten - eller endog bokstavlig talt - forsvinnende lite. Selv om det innen undergruppen skal summeres over alle omsetningsgrupper, er dette et alvorlig problem.

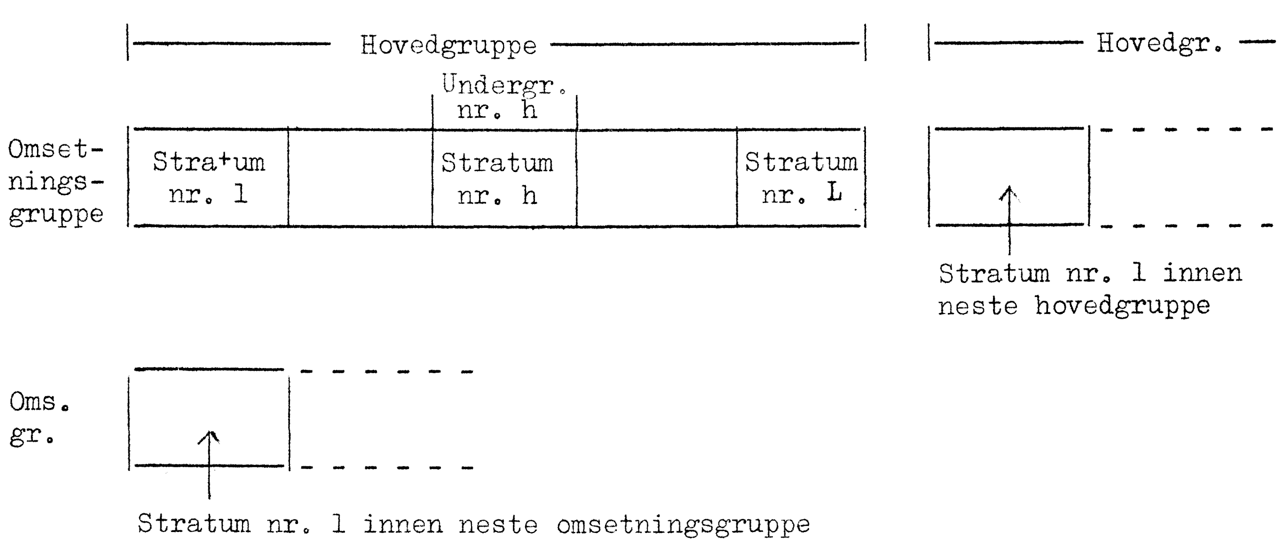
Problemet kan søkes løst etter 2 mulige retningslinjer.

1. Man oppgir å publisere tall overhodet for så små undergrupper og slår dem sammen i større grupper.

2. Man regner med at gjennomsnitt innen hver omsetningsgruppe for det kjennetegn som skal undersøkes, ikke varierer alt for mye fra undergruppe til undergruppe. Totaler for undergruppe kan i så fall fåes ved å multiplisere tilsvarende gjennomsnitt for den hovedgruppe enhetene tilhører med antall enheter totalt i undergruppen, samt summere over alle omsetningsgrupper.

Denne siste fremgangsmåte skal nå undersøkes nærmere. Den intuitive begrunnelse for at gjennomsnittsverdiene varierer lite fra undergruppe til undergruppe må være: Enhetene er allerede stratifisert etter totalomsetning. Stratifikasjonen bør da også være forholdsvis effektiv for kjennetegn som er korrelert med totalomsetning.

3. Estimatet $\bar{y}$; de ukorrigerte gjennomsnitts metode



Betrakt nå en omsetningsgruppe innen en hovedgruppe. Den herved definerte enhet (se fig.) videreinndeles etter undergrupper i L strata. Et vilkårlig av disse er stratum nr. h.

Det er nå hensiktsmessig å tenke seg en "bakenforliggende" uendelig populasjon. Enhetene i den faktiske, endelige populasjon tenkes å være utvalg fra denne hypotetiske populasjon, og de egentlige utvalg blir da å se på som underutvalg.

Følgende betegnelser innføres:

	Totalt			I stratum nr. h		
		Gjennomsnitt for			Gjennomsnitt for	
	Antall enheter	totalomsetning	kjennetegn som skal undersøkes	Antall enheter	totalomsetning	kjennetegn som skal undersøkes
Uendelig populasjon	∞	ξ	η	∞	ξ_h	η_h
Endelig populasjon	N	$\bar{A}$	$\bar{B}$	N_h	$\bar{A}_h$	$\bar{B}_h$
Utvalg	n	$\bar{x}$	$\bar{y}$	n_h	$\bar{x}_h$	$\bar{y}_h$
Varianser		σ^2	τ^2		σ_h^2	τ_h^2

For ikke å gjøre uttrykkene i dette og særlig senere avsnitt alt for kompliserte, antas at alle de L strata er like store. Dette er selvsagt ikke tilfelle innen varehandel, men resultatene har - om de bør brukes med noe forsiktighet - den fordel at de er oversiktlige.

Sammen med den helt realistiske forutsetning at utvalgsprosenten er like stor i alle strata:

$$\frac{n_h}{N_h} = \frac{n_k}{N_k} = \frac{n}{N} = \alpha \qquad \text{for alle h og k}$$

gir dette:

$$N_h = N_1 = \frac{N}{L}, \qquad n_h = n_1 = \frac{n}{L}$$

Egenskapene ved $\bar{y}$ som estimat for $\bar{B}_h$ skal nå undersøkes. Som mål for estimatets usikkerhet er det rimelig å velge

$$E = E(\bar{y} - \bar{B}_h)^2$$

dvs. forventningen av den kvadrerte avvikelse i hvert stratum, tatt over alle strata. Det som E uttrykker, er altså hvordan avvikelsene er <u>stort sett</u>.

Skulle det som interesserer være å vite hvor stor metodens usikkerhet er for ett stratum isolert, betraktes

$$E_h (\bar{y} - \bar{B}_h)^2 = E \left[(\bar{y} - \bar{B}_h)^2 \mid h \right]$$

som oppfattes som en betinget forventning for gitt h.

Betegnelsen E_h brukes heretter for å markere slike forventninger. For å finne E går man veien om disse og bruker så at:

$$E (\bar{y} - \bar{B}_h)^2 = E\, E_h (\bar{y} - \bar{B}_h)^2$$

Man har

$$E_h (\bar{y} - \bar{B}_h)^2 = E_h (\bar{y} - \eta)^2 + E_h (\bar{B}_h - \eta)^2 - 2E_h (\bar{y} - \eta)(\bar{B}_h - \eta)$$

$$= E_h (\bar{y} - \eta)^2 + E_h (\bar{B}_h - \eta_h)^2 + (\eta_h - \eta)^2 - 2E_h (\bar{y} - \eta)(\bar{B}_h - \eta_h)$$

(idet alle produktledd faller bort under forventningstegnet)

$$= \text{var}\, \bar{y} + \text{var}\, \bar{B}_h - 2 \,\text{cov}\, (\bar{y}, \bar{B}_h) + (\eta_h - \eta)^2$$

som under de foran gitte forutsetninger blir lik

$$= \frac{\tau^2}{n} + \frac{\tau_h^2}{N_1} - 2 \cdot \frac{1}{N_1} \cdot \frac{n_1}{n} \cdot \tau_h^2 + (\eta_h - \eta)^2$$

$$= \frac{\tau^2}{n} + \tau_h^2 \left(\frac{1}{N_1} - \frac{2}{N}\right) + (\eta_h - \eta)^2$$

Nå innføres $\tau_w^2 = E\, \tau_h^2$ og $\tau_b^2 = E\, (\eta_h - \eta)^2$,

$$\tau^2 = \tau_w^2 + \tau_b^2$$

Da fåes:

$$E = E\, E_h (\bar{y} - \bar{B}_h)^2 = \underline{\frac{\tau^2}{n} + \tau_w^2 \left(\frac{1}{N_1} - \frac{2}{N}\right) + \tau_b^2}$$

Ved innføring av utvalgsbrøken α og antall strata L kan dette omformes til

$$\underline{E = \frac{1}{N_1} \left[\frac{\tau^2}{L\alpha} + \tau_w^2 \left(1 - \frac{2}{L}\right) \right] + \tau_b^2}$$

For vanlig lineært estimat, utvalgsgjennomsnitt for hvert stratum, altså $\overline{y}_h$, har man

$$E_o = E\,(\overline{y}_h - \overline{B}_h)^2 = E\left[\tau_h^2\left(\frac{1}{n_1} - \frac{1}{N_1}\right)\right] = \underline{\tau_w^2\left(\frac{1}{n_1} - \frac{1}{N_1}\right)} = \underline{\frac{\tau_w^2}{N_1}\left(\frac{1}{\alpha} - 1\right)}$$

Som <u>"relativ effisiens"</u>av $\overline{y}$ til $\overline{y}_h$ betegnes <u>her</u>

$$e = \frac{E_o}{E}$$

4. Estimatet $\hat{\overline{B}}_h$, de korrigerte gjennomsnitts metode

4.1. Begrunnelse

Når det ble foreslått å bruke estimatet $\overline{y}$ for $\overline{B}_h$, var begrunnelsen at variasjonene i gjennomsnittlig totalomsetning fra stratum til stratum var små (alle $\overline{A}_h \approx \overline{A}$) i og med at kjennetegnet var brukt som grunnlag for stratifikasjon. Under forutsetning av tilstrekkelig korrelasjon mellom omsetning og tall for det kjennetegn man er interessert i, bør også denne andre størrelse variere så lite fra stratum til stratum (alle $\overline{B}_h \approx \overline{B}$) at estimatet for $\overline{B}$, nemlig $\overline{y}$, måtte kunne brukes som estimat for $\overline{B}_h$ i alle strata.

Men selv under slike forhold er det ikke gitt at $\overline{A}_h$ (og slett ikke $\overline{B}_h$) er <u>helt</u> like for alle strata. Populasjonsgjennomsnittet innen stratum kan fremdeles tenkes å variere noe, avhengig av enhetenes fordeling på størrelse innenfor hvert stratum. Er det så fremdeles samvariasjon mellom de to kjennetegn, nærmere presisert mellom $\overline{A}_h$ og $\overline{B}_h$, bør estimatet $\overline{y}$ med fordel kunne "korrigeres":

$$\hat{\overline{B}}'_h = \frac{\overline{A}_h}{\overline{A}} \cdot \overline{y} = K'_h \cdot \overline{y}$$

I det foreliggende tilfelle støter dette på den vanskelighet at alle $\overline{A}_h$, og dermed også $\overline{A}$, er ukjente størrelser. Bedriftsregisteret oppgir nemlig bare antall bedrifter i hver omsetningsgruppe, men sier ingenting om fordelingen <u>innen</u> gruppen.

Under omtalen av estimatet $\overline{y}$ ble enhetene i populasjonen betraktet som utvalg fra en hypotetisk uendelig populasjon. Det samme skal gjøres her.

Men dette er nå i seg selv ikke nok for å komme videre.

De faktisk forekommende totalomsetninger for enkeltbedrifter betraktes som realisasjoner av en stokastisk variabel. Hvis dennes sannsynlighetsfordeling var kjent - og ikke for ubehagelig å hanskes med - kunne man av den finne sannsynlighet for omsetningsstørrelse innen bestemte intervaller, f.eks. de gitte omsetningsgrupper. For hver av disse kunne da også bestemmes lokal forventning og varians.

Ved å erstatte $\overline{A}_h$ og $\overline{A}$ med slike forventninger ξ_h og ξ ville man få et modifisert estimat for $\overline{B}_h$:

$$\hat{B}_h = \frac{\xi_h}{\xi} \circ \overline{y} = K_h \circ \overline{y}$$

I den faktiske situasjon er sannsynlighetsfordelingens form ukjent. Det samme er dens parametre.

Midlertidig innføres nå mere fullstendige betegnelser

		Undergruppe nr.						Hovedgruppen
		1	2		h		L	
Om-set-nings-gruppe nr.	0	ξ_{01}	ξ_{02}	- - - -	ξ_{0h}	- - -	ξ_{0L}	$\xi_{0\cdot}$
	1	ξ_{11}	ξ_{12}	- - - -	ξ_{1h}	- - -	ξ_{1L}	$\xi_{1\cdot}$
	g	ξ_{g1}	ξ_{g2}	- - - -	ξ_{gh}	- - -	ξ_{gL}	$\xi_{g\cdot}$
	9	ξ_{91}	ξ_{92}	- - - -	ξ_{9h}	- - -	ξ_{9L}	$\xi_{9\cdot}$
Alle oms. gr.		$\xi_{\cdot 1}$	$\xi_{\cdot 2}$	- - - -	$\xi_{\cdot h}$	- - -	$\xi_{\cdot L}$	$\xi_{\cdot\cdot}$

I denne terminologi blir estimatet

$$\hat{B}_{gh} = \frac{\xi_{gh}}{\xi_{g\cdot}} \circ \overline{y}_{g\cdot} = K_{gh} \circ \overline{y}_{g\cdot}$$

4.2. Den "bakenforliggende" fordeling

4.2.1. Fordelingens form

Bedriftsregisteret oppgir antall bedrifter i hver omsetningsgruppe, og dermed fraktiler i den faktiske fordeling av bedriftene. Disse må oppfattes som estimater for fraktilene i den "bakenforliggende" fordeling.

Hvis observerte fraktiler stort sett tilsvarer fraktiler i en fordeling av kjent form med ett eller annet verdisett for parametrene, antas nå at denne fordelingen med dette sett av parametre er den "bakenforliggende" sannsynlighetsfordeling.

Dristigheten ved å gjøre dette avhenger av hvordan tilnærmelsen til fordelingsfraktilene er. Men selv om denne er god, fins det ennu usikkerhetsfaktorer. Om den antatte "bakenforliggende" fordeling er "pen og jevn" (som de mest aktuelle fordelinger er), så behøver ikke dette å være realistisk. Men det er rimelig å gå ut fra det så lenge man ikke har direkte grunn til å tro at visse verdier av den variable innen omsetningsgruppe er særlig usannsynlige.

Av bedriftsregisteret fremgår det at fordeling for totalomsetning er meget skjev; fordelingen kan da (f.eks.) være logaritmisk-normal. Dette kan undersøkes nærmere ved hjelp av logaritmisk sannsynlighetspapir. Undersøkt ble hele materialet fra 1963 og dessuten et større materiale kjørt ut fra registeret i 1962. Punktene på papiret viste seg stort sett å ligge langs rette linjer. Tilnærmelsen til rett linje var best innen hovedgrupper og større undergrupper. I mindre undergrupper kunne avvikelsene være større, men noe i den retning måtte under enhver omstendighet ventes p.g.a. samplingfeil i relasjon til den "bakenforliggende" fordeling. Dette tatt i betraktning var tilnærmelsen til rett linje stort sett meget bra for hver hoved- og undergruppe.

Den "bakenforliggende" fordeling kan derfor forutsettes å være logaritmisk-normal.

4.2.2. Estimering av parametrene i fordelingen
(Se litt. nr. 1)

Når først fraktilene er avmerket på sannsynlighetspapiret og rette linjer er lagt som best mulig passer til punktene, er det lett å estimere fordelingsparametrene grafisk. Medianen μ, altså 50 o|o-fraktilen, estimeres ved abscissen til det punkt hvor den rette linje skjærer den horisontale 50 o|o- linje på papiret. Tilsvarende estimeres 16 o|o - og 84 o|o- fraktilene, μ_{16} og μ_{84}. Parameteren ε er identisk med variansen for logaritmen (med grunntall a) til de variable. Den estimeres ved:

$$\hat{\varepsilon} = \log_a \left[\frac{1}{2} \left(\frac{\hat{\mu}}{\hat{\mu}_{16}} + \frac{\hat{\mu}_{84}}{\hat{\mu}} \right) \right]$$

(Betegnelsene μ og ε kan i det foreliggende tilfelle referere til hovedgruppe. For undergruppe brukes betegnelsene μ_h og ε_h.)

Hovedgruppe-estimatene for μ og ε bygger på et stort antall observasjoner og må betraktes som temmelig presise. For undergrupper kan nok samplingfeil (relativt til den "bakenforliggende" fordeling) spille større rolle. Men også disse bygger - i motsetning til f.eks. ("under"-) utvalgsgjennomsnitt pr. stratum, $(\bar{y}_h)$, på observasjoner fra en hel undergruppe.

Det ville være av interesse å undersøke nærmere egenskapene ved estimatene $\hat{\mu}$ og $\hat{\varepsilon}$ $(\hat{\mu}_h$ og $\hat{\varepsilon}_h)$ og presisjonen av størrelser som nå skal beregnes på grunnlag av parameterverdiene.

4.2.3 Forventning (og varians) innen omsetningsgruppe
(Se litt. nr. 2)

Forventninger innen omsetningsgruppe g, $\xi_{.h}$ og ξ_{gh}, estimeres ved

$$\hat{\xi}_{g.} = \hat{\mu} \cdot e^{\frac{\hat{\varepsilon}^2}{2}} \cdot \frac{\phi(u_g - \hat{\varepsilon}) - \phi(u_{g-1} - \hat{\varepsilon})}{\phi(u_g) - (u_{g-1})}$$

$$\hat{\xi}_{gh} = \hat{\mu}_h \cdot e^{\frac{\hat{\varepsilon}_h^2}{2}} \cdot \frac{\phi(u_g - \hat{\varepsilon}_h) - \phi(u_{g-1} - \hat{\varepsilon}_h)}{\phi(u_g) - \phi(u_{g-1})}$$

u_{h-1} og u_h er henholdsvis nedre og øvre grense for omsetningsgruppe, og $\phi(u)$ er integralet opp til u over den normale fordeling (0,1).

Verdier for stratumvariansene benyttes ikke ved selve estimeringen av $\bar{B}_h$, men må tas i betraktning ved vurderingen av estimatene. De bestemmes som

$$\hat{\sigma}^2_{g.} = \hat{\mu}^2 \cdot e^{2\hat{\varepsilon}^2} \cdot \frac{\phi(u_g - 2\hat{\varepsilon}) - \phi(u_{g-1} - 2\hat{\varepsilon})}{\phi(u_g) - \phi(u_{g-1})} - \hat{\xi}^2_{g.}$$

$$\hat{\sigma}^2_{gh} = \hat{\mu}_h^2 \cdot e^{2\hat{\varepsilon}_h^2} \cdot \frac{\phi(u_g - 2\hat{\varepsilon}_h) - \phi(u_{g-1} - 2\hat{\varepsilon}_h)}{\phi(u_g) - \phi(u_{g-1})} - \hat{\xi}^2_{gh}$$

4.3. Estimatet og dets bruttovarians

$K_{gh} = \frac{\xi_{gh}}{\xi_{g.}}$ kan nå estimeres ved $\frac{\hat{\xi}_{gh}}{\hat{\xi}_{g.}}$. Det antas nå at feilen på estimatet er ubetydelig sammenlignet med feilen på estimater bygget på det egentlige utvalg (underutvalg i relasjon til "bakenforliggende" populasjon).

Etter at alle K_{gh} (og tilsvarende $K_{g.}$) er bestemt, er det bare nødvendig å betrakte en omsetningsgruppe om gangen. Da kan indeks for omsetningsgruppe (g) sløyfes og de opprinnelige betegnelser brukes.

Estimatet

$$\hat{\bar{B}}_h = \frac{\xi_h}{\xi} \cdot \bar{y} = K_h \cdot \bar{y}$$

skal undersøkes; idet forutsetninger og fremgangsmåte er analoge til dem i avsnitt 3.

$$E_h\,(\hat{\bar{B}}_h - \bar{B}_h)^2 = E_h\,(\frac{\xi_h}{\xi} \cdot \bar{y} - \bar{B}_h)^2$$

$$= (\frac{\xi_h}{\xi})^2 \cdot E_h(\bar{y} - \eta)^2 + E_h(\bar{B}_h - \frac{\xi_h}{\xi} \cdot \eta)^2 - 2\,\frac{\xi_h}{\xi}\,E_h(\bar{y} - \eta)\,(\bar{B}_h - \frac{\xi_h}{\xi} \cdot \eta)$$

$$= (\frac{\xi_h}{\xi})^2 \cdot E_h(\bar{y} - \eta)^2 + E_h(\bar{B}_h - \eta_h)^2 + (\frac{\xi_h}{\xi} \cdot \eta - \eta_h)^2 - 2\,\frac{\xi_h}{\xi}\,E_h(\bar{y} - \eta)\,(\bar{B}_h - \eta_h)$$

$$= (\frac{\xi_h}{\xi})^2 \cdot \text{var}\,\bar{y} + \text{var}\,\bar{B}_h - 2\,\frac{\xi_h}{\xi} \cdot \text{cov}\,(\bar{y}\,,\,\bar{B}_h) + (\frac{\xi_h}{\xi} \cdot \eta - \eta_h)^2$$

$$= (\frac{\xi_h}{\xi})^2 \cdot \frac{\tau^2}{n} + \frac{\tau_h^2}{N_1} - 2 \cdot \frac{\xi_h}{\xi} \cdot \frac{\tau_h^2}{N} + (\eta_h - \eta)^2 - 2\,\frac{\eta}{\xi}\,(\xi_h - \xi)\,(\eta_h - \eta)$$

$$+ (\frac{\eta}{\xi})^2 \cdot (\xi_h - \xi)^2$$

Som før innføres: $\tau_w^2 = E\tau_h^2$

$\tau_b^2 = E\,(\eta_h - \eta)^2$ og nå analogt $\sigma_b^2 = E\,(\xi_h - \xi)^2$, $\rho_b \sigma_b \tau_b = E(\xi_h - \xi)(\eta_h - \eta)$

I leddet av orden $(\frac{1}{N})$ forekommer produktet $\xi_h \tau_h^2$. Det skulle her være forsvarlig å sette

$$E(\xi_h \tau_h^2) \approx \xi \tau_w^2 , \quad \text{da blir } E(\frac{\xi_h}{\xi} \cdot \tau_h^2) \approx \tau_w^2$$

Videre er $E(\frac{\xi_h}{\xi}) = 1$ og $E(\frac{\xi_h}{\xi})^2 = 1 + \frac{\sigma_b^2}{\xi^2}$

Da fåes etter innsetting:

$$E_1 = E(\hat{\bar{B}}_h - \bar{B}_h)^2 = E\, E_h(\hat{\bar{B}}_h - \bar{B})^2$$

$$= \underline{(1 + \frac{\sigma_b^2}{\xi^2}) \cdot \frac{\tau^2}{n} + \tau_w^2 (\frac{1}{N_1} - \frac{2}{N}) + \tau_b^2 - 2\frac{\eta}{\xi} \rho_b \sigma_b \tau_b + (\frac{\eta}{\xi})^2 \cdot \sigma_b^2}$$

5. Sammenligning mellom estimatene $\bar{y}_h$, $\bar{y}$ og $\hat{\bar{B}}_h = \frac{\xi_h}{\xi} \cdot \bar{y}$

Sammenligning mellom ukorrigert gjennomsnitt og vanlig stratumgjennomsnitt kan gjøres direkte på grunnlag av de utledede formler. Når det gjelder korrigerte gjennomsnitts metode, er forholdene noe vanskeligere. De anvendte K_h-ene er i virkeligheten estimerte størrelser, mens formlene forutsetter at de er kjent eksakt som forholdet mellom forventning i undergruppe og forventning i hovedgruppe når observasjonene har de "bakenforliggende" fordelinger. I omsetningsgrupper med relativt høy utvalgsprosent, f.eks. $16\frac{2}{3}$ o|o eller 25 o|o, er det en viss risiko for at dette kan spille inn. Forholdet bør opplagt undersøkes nærmere. Inntil det er gjort bør sammenligningsresultatene for korrigerte gjennomsnitts metode brukes med en viss forsiktighet når det gjelder engros, oms. gr. 7, og detalj, oms. gr. 5 og 6.

$\hat{\bar{B}}_h$ skal først sammenlignes med $\bar{y}$. Man har

$$E_1 - E = \frac{\sigma_b^2}{\xi^2} \cdot \frac{\tau^2}{n} - 2 \cdot \frac{\eta}{\xi} \cdot \rho_b \sigma_b \tau_b + (\frac{\eta}{\xi})^2 \cdot \sigma_b^2$$

For at $\hat{\bar{B}}_h$ skal være best må $E_1 - E < 0$. Nå er nok n såvidt stor at første ledd innvirker lite på resultatet; dessuten er $\frac{\sigma_b^2}{\xi^2}$ neppe så svært stor. Første ledd sløyfes derfor. Betingelsen for at resten av uttrykket skal bli mindre enn 0 er at

$$\rho_b > \frac{1}{2} \cdot \frac{\frac{\sigma_b}{\xi}}{\frac{\tau_b}{\eta}} = \frac{C}{2}$$

(Størrelsen C er her definert også for senere bruk).

I spesialtilfellet $C = 1$ reduseres dette til

$$\rho_b > \frac{1}{2}$$

Konklusjon: Korrigerte gjennomsnitts metode er bedre enn ukorrigerte gjennomsnitts metode såfremt korrelasjonen mellom betingede forventninger gitt stratum for totalomsetning og for kjennetegn som skal undersøkes, overstiger størrelsen $\frac{C}{2}$.

$\bar{y}$ og $\bar{B}_h$ skal så sammenlignes med det rene stratumgjennomsnitt $\bar{y}_h$.

$\bar{y}$ er bedre enn $\bar{y}_h$ hvis $E - E_o < 0$.

$$E - E_o = \frac{\tau^2}{n} + \tau_w^2 \left(\frac{1}{N_1} - \frac{2}{N}\right) + \tau_b^2 - \tau_w^2 \left(\frac{1}{n_1} - \frac{1}{N_1}\right)$$

$$= \frac{\tau^2}{n} + \tau_b^2 + \tau_w^2 \left(\frac{2}{N_1} - \frac{1}{n_1} - \frac{2}{N}\right)$$

Videre, idet $\tau^2 = \tau_w^2 + \tau_b^2$

$$E - E_o = \tau_b^2 + \frac{\tau_b^2}{n} + \tau_w^2 \left[2 \left(\frac{1}{N_1} - \frac{1}{N}\right) - \left(\frac{1}{n_1} - \frac{1}{n}\right)\right]$$

Innføres antall strata, L, og utvalgsbrøken, α, fåes:

$$E - E_o = \tau_b^2 + \frac{\tau_b^2}{n} + 2 \cdot \frac{\tau_w^2}{n} (L - 1) \left(\alpha - \frac{1}{2}\right)$$

Her er de to første ledd positive. Skal det hele bli negativt, må siste ledd desto mer være negativt. Men det kan det bli bare hvis

$$\alpha < \frac{1}{2}$$

$\hat{B}_h$ er bedre enn $\bar{y}_h$ hvis $E_1 - E_o < 0$.

$$E_1 - E_o = (E_1 - E) + (E - E_o)$$

$$= \frac{\sigma_b^2}{\xi^2} \cdot \frac{\tau^2}{n} - 2 \cdot \frac{\eta}{\xi} \cdot \rho_b \sigma_b \tau_b + \frac{\eta^2}{\xi^2} \cdot \sigma_b^2$$

$$+ \tau_b^2 + \frac{\tau_b^2}{n} + 2 \cdot \frac{\tau_w^2}{n} (L - 1)(\alpha - \frac{1}{2})$$

Innsetting av $C = \dfrac{\frac{\sigma_b}{\xi}}{\frac{\tau_b}{\eta}}$ eller $\sigma_b = C \cdot \frac{\xi}{\eta} \cdot \tau_b$

gir nå:

$$E_1 - E_o = \frac{\sigma_b^2}{\xi^2} \cdot \frac{\tau^2}{n} + \tau_b^2 \left[(1 - \rho_b^2) + (C - \rho_b)^2 \right]$$

$$+ \frac{\tau_b^2}{n} + 2 \cdot \frac{\tau_w^2}{n} (L - 1)(\alpha - \frac{1}{2})$$

Her er det igjen bare siste ledd som har mulighet for å bli negativt, og forutsetningen for $E_1 - E_o < 0$ blir også

$$\underline{\alpha < \frac{1}{2}}$$

Konklusjon: <u>For at ukorrigerte eller korrigerte gjennomsnitts metode skal være fordelaktig må under enhver omstendighet utvalget være på mindre enn 50 o|o. Er utvalget 50 o|o eller større, bør vanlig lineært estimat foretrekkes - dvs. gjennomsnitt pr. stratum.</u>

Forholdet mellom de tre estimeringsmetoder skal nå undersøkes nærmere ved at relative effisienser mellom $\bar{y}_h$, $\bar{y}$ og $\hat{B}_h$ bestemmes. Man har:

$$E_o = \tau_w^2 \left(\frac{1}{n_1} - \frac{1}{N_1}\right) = \underline{\frac{\tau_w^2}{n} \cdot L(1 - \alpha)}$$

og, etter innsetting av $\tau^2 = \tau_b^2 + \tau_w^2$:

$$E = \tau_w^2 \left(\frac{1}{n} + \frac{1}{N_1} - \frac{2}{N}\right) + \frac{\tau_b^2}{n} + \tau_b^2$$

$$= \frac{\tau_w^2}{n} \left[1 + (L - 2)\,\alpha\right] + \frac{\tau_b^2}{n} + \tau_b^2$$

Etter innsetting av C blir:

$$E_1 = \left(1 + C^2 \cdot \frac{\tau_b^2}{\eta^2}\right) \cdot \frac{\tau^2}{n} + \tau_w^2 \left(\frac{1}{N_1} - \frac{2}{N}\right) + \tau_b^2 \left[(1 - \rho_b^2) + (C - \rho_b)^2\right]$$

Leddet $\left(1 + C^2 \cdot \frac{\tau_b^2}{\eta^2}\right)$ settes nå lik 1, da $C^2 \cdot \frac{\tau_b^2}{\eta^2} = \frac{\sigma^2}{\xi^2}$ ut fra materialet anslås til å være av størrelsesorden 0,001. Men da er:

$$E_1 \approx \frac{\tau_w^2}{n} \left[1 + (L - 2)\,\alpha\right] + \frac{\tau_b^2}{n} + \tau_b^2 \left[(1 - \rho_b^2) + (C - \rho_b)^2\right]$$

Det foreliggende materiale gir ikke opplysninger om η eller τ_b for noe av de kjennetegn som skal undersøkes. Derfor er det heller ikke kjent om forholdet $\frac{\tau_b}{\eta}$ (variasjonskoeffisient for stratumforventninger (betingede forventninger)) varierer fra kjennetegn til kjennetegn. Men som grunnlag for de følgende effisiensberegninger er det kanskje ikke urimelig å forutsette at variasjonskoeffisienten er konstant, dvs. alle $\frac{\tau_b}{\eta} = \frac{\sigma_b}{\xi}$, eller:

$$C = \frac{\frac{\sigma_b}{\xi}}{\frac{\tau_b}{\eta}} = 1$$

I så fall er

$$E_1 = \frac{\tau_w^2}{n} \left[1 + (L - 2)\,\alpha\right] + \frac{\tau_b^2}{n} + 2\,\tau_b^2\,(1 - \rho_b)$$

Relative effisienser blir etter dette :

For ukorrigert gjennomsnitt, $\bar{y}$, relativt til stratumgjennomsnitt, $\bar{y}_h$:

$$e = \frac{E_o}{E} = \frac{L(1-\alpha)}{1 + (L-2)\alpha + \frac{\tau_b^2}{\tau_w^2}\left[1 + n\right]}$$

For korrigert gjennomsnitt $\hat{\bar{B}}_h$ relativt til stratumgjennomsnitt, $\bar{y}_h$ (når $C = 1$, $1 + \frac{\sigma_b^2}{\xi^2} = 1$) :

$$e_1 = \frac{E_o}{E_1} = \frac{L(1-\alpha)}{1 + (L-2)\alpha + \frac{\tau_b^2}{\tau_w^2}\left[1 + 2n\,(1 - \rho_b)\right]}$$

Effisiens for korrigert gjennomsnitt, $\hat{\bar{B}}_h$, relativt til ukorrigert gjennomsnitt, $\bar{y}$, bestemmes som :

$$\frac{e_1}{e} = \frac{E}{E_1}$$

I tabell 1 A - C er gitt de relative effisienser e, e_1 og $\frac{e_1}{e}$. Resultatene er for $C = 1$ og $L = 10$ strata. Utvalgsbrøker er $\alpha = \frac{1}{5}$ og $\alpha = \frac{1}{20}$.

I tabellene er gruppert etter verdier av ρ_b og N_1. ρ_b, korrelasjonen mellom stratumforventning for totalomsetning og for undersøkt kjennetegn, er ikke kjent fra materialet. (Det bør gjøres oppmerksom på at totalomsetning her refererer til trekningstidspunkt, mens undersøkt kjennetegn refererer til tellingstidspunkt). Tabellene gir resultater for 3 (eller 4) av de mulige verdier : (0,3), 0,5, 0,7, 0,9. For $\rho_b = 0{,}5$ er resultatene for $\bar{y}$ og $\hat{\bar{B}}_h$ identiske, $e = e_1$.

I tabellene er angitt resultater for forskjellige verdier av N_1. Når N_1 her er valgt og ikke n, som inngår i formlene, er dette for å se hvordan det går når stratumstørrelsen varieres, noe som er av langt større interesse i praksis.

Forholdet mellom variasjoner mellom strata og variasjoner innen strata, uttrykt ved $\frac{\tau_b^2}{\tau_w^2}$, er av stor betydning for metodenes relative effisiens. Beregninger er foretatt for 3 verdier av denne størrelse. I tabell 1 A forutsettes en verdi som tilsvarer $\frac{\sigma_b^2}{\sigma_w^2}$ i engros, omsetningsgruppe 2. I tabell 1 B er verdien noe større, hentet fra detalj (totalt), omsetningsgruppe 6. Nå kan det være et spørsmål om $\frac{\tau_b^2}{\tau_w^2}$ i virkeligheten er lik $\frac{\sigma_b^2}{\sigma_w^2}$ og kan erstattes slik av denne. For å unngå feilaktig å konkludere med at (u-) korrelerte gjennomsnitt bør brukes, er i tabell 1 C gitt relative effisienser for det tilfelle at $\frac{\tau_b^2}{\tau_w^2}$ har en verdi noe større enn begge $\frac{\sigma_b^2}{\sigma_w^2}$ - verdiene, slik at en viss sikkerhetsmargin oppnås.

Tabell 2 gir verdier av e_1 for samme verdisett av C og $\frac{\tau_b^2}{\tau_w^2}$ som i tabell 1 A. Mens ρ_b holdes fast på 0,7, er det gitt effisienser for flere utvalgsbrøker α. For antall strata har man som før L = 10 og nå dessuten L = 30.

(NB! Mange kombinasjoner av N_1, L og α vil gi et utvalg pr. stratum, n_1, som ikke er noe helt tall. I praksis vil dette innebære at med den gitte totale utvalgsprosent vil utvalgsprosenten innen strata måtte variere noe, og helt ned til 0 i enkelte strata. At det er slik kan vanskelig tale mot (u-) korrigerte gjennomsnitts metode).

<u>Konklusjon: Med de gitte verdier av L, α, ρ_b og $\frac{\tau_b}{\tau_w}$ (og C = 1) vil ukorrigerte eller korrigerte gjennomsnitts metode være fordelaktig fremfor rent stratumgjennomsnitt bare strataene ikke er alt for store. I små strata er gevinsten stor. Er korrelasjonen ρ_b relativt høy, vil korrigerte gjennomsnitt være å foretrekke fremfor ukorrigerte. Gevinsten er her liten når strataene er små, men blir betydelig etter hvert som stratumstørrelsen øker.</u>

Når det gjelder strata med relativt høy utvalgsprosent (som her $\alpha = \frac{1}{5}$, dvs. 20 o|o) må resultatene for korrigert gjennomsnitt gis med visse forbehold. Dette skyldes at "korreksjonen" $\frac{\xi_h}{\xi}$ i virkeligheten må anslås på grunnlag av materialet; og feilen kan tenkes å ha betydning ved sammenligningen med ukorrigert gjennomsnitt.

Uten å kjenne korrelasjonen ρ_b mellom stratumforventninger for totalomsetning på trekningstidspunkt og for det aktuelle kjennetegn på tellingstidspunkt er det ikke mulig å trekke noen endelig praktisk konklusjon av den foretatte undersøkelse. Men inntil videre er det vel nærliggende å tro at korrelasjonen er såpass høy at korrelerte gjennomsnitts metode er fordelaktig. Det er imidlertid av avgjørende betydning at dette undersøkes nærmere.

Stratumstørrelsen innen varehandel er varierende. Men det er vanskelig å tenke seg at dette kan endre konklusjonene; de aller fleste strata er mindre enn de kritiske grenser (i tabellene markert med streker), i hvert fall når $\rho_b > 0{,}5$.

De få strata som er større enn de kritiske grenser, krever særskilt behandling. Dette gjelder i enkelte undergrupper innen næring- og nydelse. Her bør brukes rent stratumgjennomsnitt (lineært estimat).

Til slutt bør det presiseres at både ved ukorrigerte og ved korrigerte gjennomsnitts metode må estimatorene ventes å være forventningsskjeve. Og skjevheten vil ikke forsvinne om antall utvalgsenheter økes; estimatorene er m.a.o. neppe konsistente.

Dette innebærer også at de gitte "effisienser" ikke uten videre er vanlige effisienser i den forstand at de skulle uttrykke forholdet mellom antall enheter man måtte ha i utvalget under bruk av forskjellige metoder for å oppnå samme presisjon ved hver metode. Bare for det lineære estimat (rent gjennomsnitt) $\bar{y}$ vil variansen kunne ventes å gå mot 0 når antall utvalgsenheter vokser; bruttovariansene for de andre estimater går isteden mot en størrelse som avhenger av de "sanne" verdiers variasjon fra næring til næring innen omsetningsgruppe. (Poenget ved korrigerte gjennomsnitts metode er at denne størrelse reduseres ved utnyttelse av eventuell korrelasjon mellom totalomsetning og det kjennetegn som skal undersøkes).

Hvis tabellene "Effisiens av $\hat{\bar{B}}_h$ relativt til $\bar{y}_h$" gir en verdi, f.eks. $e_1 = \frac{E_o}{E_1} = 5$, så innebærer dette at tilsvarende gode resultater som ved korrigerte gjennomsnitts metode kunne oppnås ved rent gjennomsnitt av observasjonene fra et 5 ganger så stort utvalg. Men om tabellene "Effisiens av

$\hat{\bar{B}}_h$ relativt til $\bar{y}$" gir f.eks. $\frac{e_1}{e} = \frac{E}{E_1} = 1{,}2$, så forteller dette tallet bare hvordan disse estimaters bruttovarianser - tatt i "gjennomsnitt over alle strata" - forholder seg til hverandre.

Tabell 1 A

$\frac{\tau_b^2}{\tau_w^2} = 0{,}010749$ L = 10 C = 1

Effisiens av $\hat{\bar{B}}_h$ relativt til $\bar{y}_h$, dvs. $e_1 = \frac{E_o}{E_1}$

(herunder tall for effisiens av $\bar{y}$ relativt til $\bar{y}_h$, dvs. $e = \frac{E_o}{E}$)

ρ_b / N_1	$\alpha = \frac{1}{5}$: 0,3	0,5 (= e)	0,7	0,9	$\alpha = \frac{1}{20}$: 0,3	0,5 (= e)	0,7	0,9
1	3,029	3,040	3,049	3,059	6,698	6,708	6,719	6,729
10	2,748	2,831	2,920	3,015	6,393	6,487	6,583	6,683
20	2,490	2,631	2,788	2,966	6,085	6,257	6,439	6,633
50	1,944	2,171	2,457	2,832	5,317	5,656	6,043	6,487
100	1,424	1,680	2,050	2,631	4,391	4,875	5,479	6,256
200	0,927	1,158	1,541	2,307	3,259	3,822	4,621	5,846
500	0,453	0,598	0,882	1,679	1,837	2,318	3,140	4,874
1 000	0,245	0,332	0,516	1,160	1,064	1,401	2,049	3,827

Effisiens av $\hat{\bar{B}}_h$ relativt til $\bar{y}$, dvs. $\frac{e_1}{e} = \frac{E}{E_1}$

ρ_b / N_1	$\alpha = \frac{1}{5}$: 0,3	0,5	0,7	0,9	$\alpha = \frac{1}{20}$: 0,3	0,5	0,7	0,9
1	0,997	1,000	1,003	1,007	0,998	1,000	1,002	1,003
10	0,971	1,000	1,031	1,065	0,986	1,000	1,015	1,030
20	0,947	1,000	1,060	1,128	0,972	1,000	1,029	1,060
50	0,896	1,000	1,132	1,305	0,940	1,000	1,068	1,147
100	0,847	1,000	1,220	1,567	0,901	1,000	1,124	1,283
200	0,801	1,000	1,331	1,992	0,853	1,000	1,209	1,529
500	0,757	1,000	1,473	2,806	0,792	1,000	1,355	2,103
1 000	0,737	1,000	1,554	3,494	0,760	1,000	1,463	2,732

Tabell 1 B

$\frac{\tau_b^2}{\tau_w^2} = 0{,}028496$ L = 10 C = 1

Effisiens av $\hat{\bar{B}}_h$ relativt til $\bar{y}_h$, dvs. $e_1 = \frac{E_o}{E_1}$

(herunder tall for effisiens av $\bar{y}$ relativt til $\bar{y}_h$, dvs. $e = \frac{E_o}{E}$)

N_1 \ ρ_b	$\alpha = \frac{1}{5}$			$\alpha = \frac{1}{20}$		
	0,5 (= e)	0,7	0,9	0,5 (= e)	0,7	0,9
1	2,979	3,004	3,030	6,585	6,611	6,637
10	2,501	2,693	2,917	6,046	6,274	6,520
20	2,123	2,415	2,801	5,544	5,939	6,396
50	1,460	1,844	2,502	4,437	5,118	6,047
100	0,961	1,326	2,125	3,329	4,160	5,545
?00	0,569	0,843	1,626	2,220	3,027	4,754
500	0,257	0,406	0,962	1,110	1,665	3,330
1 000	0,133	0,216	0,567	0,605	0,950	2,213

Effisiens av $\hat{\bar{B}}_h$ relativt til $\bar{y}$, dvs. $\frac{e_1}{e} = \frac{E}{E_1}$

N_1 \ ρ_b	$\alpha = \frac{1}{5}$			$\alpha = \frac{1}{20}$		
	0,5	0,7	0,9	0,5	0,7	0,9
1	1,000	1,008	1,017	1,000	1,004	1,008
10	1,000	1,077	1,166	1,000	1,038	1,078
20	1,000	1,138	1,319	1,000	1,071	1,154
50	1,000	1,263	1,714	1,000	1,153	1,363
100	1,000	1,380	2,211	1,000	1,250	1,666
200	1,000	1,482	2,858	1,000	1,364	2,141
500	1,000	1,580	3,743	1,000	1,500	3,000
1 000	1,000	1,624	4,263	1,000	1,570	3,658

Tabell 1 C

$\frac{\tau_b^2}{\tau_w^2} = 0,04$ $\qquad$ $L = 10$ $\qquad$ $C = 1$

Effisiens av $\hat{\hat{B}}_h$ relativt til $\bar{y}_h$, dvs. $e_1 = \frac{E_o}{E_1}$

(herunder tall for effisiens av $\bar{y}$ relativt til $\bar{y}_h$, dvs. $e = \frac{E_o}{E}$)

N_1 \ ρ_b	$\alpha = \frac{1}{5}$			$\alpha = \frac{1}{20}$		
	0,5 (= e)	0,7	0,9	0,5 (= e)	0,7	0,9
1	2,941	2,974	3,008	6,507	6,552	6,552
10	2,326	2,564	2,857	5,793	6,090	6,419
20	1,887	2,222	2,703	5,163	5,655	6,250
50	1,205	1,587	2,326	3,893	4,657	5,793
100	0,752	1,075	1,887	2,762	3,598	5,163
200	0,429	0,654	1,370	1,744	2,474	4,241
500	0,188	0,300	0,752	0,830	1,277	2,762
1 000	0,097	0,158	0,429	0,443	0,707	1,746

Effisiens av $\hat{\hat{B}}_h$ relativt til $\bar{y}$, dvs. $\frac{e_1}{e} = \frac{E}{E_1}$

N_1 \ ρ_b	$\alpha = \frac{1}{5}$			$\alpha = \frac{1}{20}$		
	0,5	0,7	0,9	0,5	0,7	0,9
1	1,000	1,011	1,023	1,000	1,007	1,007
10	1,000	1,102	1,228	1,000	1,051	1,108
20	1,000	1,178	1,432	1,000	1,095	1,211
50	1,000	1,317	1,930	1,000	1,196	1,488
100	1,000	1,430	2,509	1,000	1,303	1,869
200	1,000	1,524	3,193	1,000	1,419	2,432
500	1,000	1,596	4,000	1,000	1,539	3,328
1 000	1,000	1,629	4,423	1,000	1,596	3,941

Tabell 2

Effisiens av $\hat{\bar{B}}_h$ relativt til $\bar{y}_h$, dvs. $e_1 = \frac{E_o}{E_1}$,

når $\frac{\tau_b^2}{\tau_w^2} = 0{,}010749$, $\rho_b = 0{,}7$ og $C = 1$

L = 10

N_1 \ α	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{1}{10}$	$\frac{1}{20}$
1	1,419	3,053	4,957	6,721
2	1,410	3,038	4,940	6,705
5	1,385	2,994	4,887	6,660
10	1,345	2,924	4,803	6,585
20	1,271	2,792	4,643	6,441
50	1,092	2,459	4,222	6,044
100	0,884	2,053	3,666	5,481
200	0,640	1,541	2,903	4,621
500	0,351	0,883	1,786	3,143
1 000	0,200	0,516	1,089	2,049

L = 30

N_1 \ α	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{1}{10}$	$\frac{1}{20}$
1	1,466	3,611	7,052	11,777
2	1,457	3,590	7,017	11,730
5	1,430	3,529	6,912	11,591
10	1,387	3,431	6,745	11,367
20	1,309	3,251	6,434	10,945
50	1,120	2,809	5,652	9,846
100	0,902	2,290	4,699	8,437
200	0,650	1,672	3,515	6,555
500	0,354	0,924	2,001	3,928
1 000	0,201	0,529	1,165	2,356

6. Oversikt over estimeringsfremgangsmåten for varehandelsstatistikk 1963

a) Fra bedriftsregisteret ble kjørt ut oppgaver over bedriftenes fordeling på størrelsesgrupper etter totalomsetning (omsetningsgrupper). Dette ble gjort separat for:

A. Hovedgrupper:
- I. Engros
- II. Detalj, nærings- og nydelsesmidler
- III. Detalj, annet.

B. Hver enkelt undergruppe som det skulle publiseres tall for (eller eventuelt for mindre grupper, som da ble slått sammen før beregninger og publisering).

b) På dette grunnlag beregnet Gruppen for metodespørsmål "korreksjonsfaktorer" for gjennomsnittene i de enkelte omsetningsgrupper innen hver undergruppe.

c) Det ble beregnet lineære estimater på totalene for:
1) Antall bedrifter
2) Omsetning
3) Innkjøp
osv.

for hver omsetningsgruppe innen I, II og III (hver for seg). Oppblåsning med faktor for hvert stratum ble foretatt. (Betegnelsen "stratum" er her, i motsetning til ellers, brukt i den vanlige betydning innen varehandelsstatistikk).

Gjennomsnitt pr. bedrift for
2) Omsetning
3) Innkjøp
osv.

ble beregnet.

d) For hver undergruppe ble beregnet gjennomsnitt for hver omsetningsgruppe på grunnlag av gjennomsnitt under c) og "korreksjonsfaktorene".

e) Totaler for 2), 3), osv. under c) ble beregnet ved å multiplisere de <u>"korrigerte gjennomsnitt"</u> fra d) med antall bedrifter pr. omsetningsgruppe (innen undergruppe) <u>fra registeret</u>. Det ble summert over alle omsetningsgrupper.

\- - -

I 1963 ble den beskrevne <u>korrigerte gjennomsnitts metode</u> brukt i alle omsetningsgrupper, unntatt:

a) Engros, omsetn.gr. 9 og detalj, omsetn.gr. 8 og 9. Her ble foretatt fullstendig telling, og vanlige lineære estimater ble brukt.

b) Omsetn.gr. "uoppgitt".

7. Konklusjoner og forslag

1. Fremgangsmåten fra 1963 foreslås beholdt, men med følgende forandringer:

a) I omsetningsgrupper hvor utvalget er 50 o|o, bør anvendes vanlige lineære estimater. Dette gjelder engros, omsetn.gr. 8 og detalj, omsetn.gr. 7.

b) Rikstall for næringsgruppene 6411, 6412 og 6424 (kolonial, landhandel og tobakk- sjokolade- frukt) bør estimeres ved vanlige lineære estimater også i omsetningsgruppene 4, 5 og 6. En mulighet er å skille næringsgruppene beregningsteknisk fra nærings- og nydelsesmidler og betrakte dem som tre egne hovedgrupper.

2. Så snart som mulig bør korrelasjon mellom stratumgjennomsnitt for totalomsetning på trekningstidspunkt og stratumgjennomsnitt for hvert enkelt kjennetegn på tellingstidspunkt undersøkes. Hvis denne er vesentlig lavere enn forutsatt, dvs. omkring 0,5 eller lavere, vil ukorrigerte gjennomsnitts metode eller kanskje rent gjennomsnitt være aktuelle.

3. Med henblikk på strata hvor utvalgsprosenten er 16 $\frac{2}{3}$ eller 25 (engros, omsetn.gr. 7 og detalj, omsetn.gr. 5 og 6) vil videre undersøkelse av korrigerte gjennomsnitts metode være av interesse.

Litteratur

1. Aitchison & Brown: The Lognormal Distribution. Cambridge, University Press, 1957.

2. E. Lykke Jensen: Repræsentative undersøgelsers teori og metode, bind II. Universitetets statistiske Institut, København 1960.