

B O G A N M E L D E L S E R

Ernst Lykke Jensen: Repræsentative undersøgelsers teori og metode. II Restriktiv tilfældig udvælgelse. Universitetets statistiske Institut, København 1960, 212 sider.

I februar 1957 kom 1. bind af Lykke Jensens kompendium om Repræsentative undersøgelsers teori og metode; nemlig den del der vedrører »Simpel tilfældig udvælgelse«.

I december 1959 er anden del udgivet, som omfatter »Restriktiv tilfældig udvælgelse«. Bogen er inddelt i ialt 9 kapitler (kapitlerne 15-23). Jeg fremhævede i anmeldelsen af 1. bind, at dette bind var overlegent og godt skrevet. Det samme gælder 2. del af kompendiet. Når man læser 2. del med måske endnu større interesse end 1. del, er det bl. a. fordi Lykke Jensen ved gennemgangen af teorien for den restriktive tilfældige udvælgelse bestandig kommer ind på spørgsmål, som har direkte relevans for den praktiske statistiker.

Der er nogle gængse og ikke gængse spørgsmål, ved den restriktive udvælgelse som særlig interesserer, og som Lykke Jensen gennemgår: Hvorledes skal stikprøven fordeles på strata? Og hvor stor skal den være? Hvad betyder omkostningerne for fordelingen på strata? Og hvad betyder anvendelse af kvotientsskøn med hensyn til nedbringelse af usikkerheden? Hertil kommer et spørgsmål, som synes ret nyt og overmåde interessant: Hvorledes foretages den rette stratifikation?

I kapitel 15 omtales stratifikation af multipelt varierende analyseenheder.

Det karakteristiske for udvælgelsesproceduren ved stratifikation betones: 1) at samtlige strata skal være repræsenteret i stikprøven og 2) at stikprøveudtagningen skal foretages således, at stikprøverne fra de forskellige strata er stokastisk uafhængige.

1. Anmeldt af mig i Nationaløkonomisk Tidsskrift 1957, 95. bind, 3-4 hæfte.

Lykke Jensen fremhæver naturligvis det kendte, at man skal tilstræbe homogenitet af elementerne indenfor strata og heterogenitet mellem stratum-værdierne (gennemsnittene). Man opnår herigennem en betydelig nedsættelse af variansen, således at i jo højere grad det nævnte opfyldes, desto større bliver stratifikationsgevinsten.

Variansen er afhængig af stikprøvens størrelse og af spredningen i fordelingen af de enkelte elementer. Hvis spredningen inden for de enkelte strata er meget forskellig fra stratum til stratum, bør man benytte den såkaldte optimale fordeling af antal iagttagelser på de enkelte strata. Denne bestemmes ved at lade udvalgsprocenterne være proportionale med stratumspredningerne. Herved kan opnås en væsentlig større nedsættelse af usikkerheden end ved en proportional fordeling. Den optimale fordeling forudsætter dog, efter det anførte, kendskab til stratumspredningerne i populationen. Bl. a. af den grund konkluderer Lykke Jensen (s. 291) »at kun for en betragtelig forskel mellem stratumspredningerne er det besværet værd at erstatte den simple proportionale fordelingsprocedure med den optimale«.

Dette er rigtigt; men inden for den økonomiske statistik, som jo bliver mere og mere påkrævet, må man sikkert regne med, at spredningerne inden for strata kan være vidt forskellige, således at man måske på dette felt i mange tilfælde ved optimal fordeling kan opnå betydelige reduktioner i variansen. Det eksempel fra brugsforeningsstatistik, som Lykke Jensen omtaler, viser da også, at medens variansen ved proportional udvælgelse er 522,3 bliver den ved optimal udvælgelse kun 146,1.

For små lande betyder omkostningerne særdeles meget. Disse lande må derfor være særlig interesserede i spørgsmål om indpasningen af en omkostningsfunktion ved beregningerne. Lykke Jensen indfører en omkostningsfunktion af følgende form:

B O G A N M E L D E L S E R

Ernst Lykke Jensen: Repræsentative undersøgelsers teori og metode. II Restriktiv tilfældig udvælgelse. Universitetets statistiske Institut, København 1960, 212 sider.

I februar 1957 kom 1. bind af Lykke Jensens kompendium om Repræsentative undersøgelsers teori og metode; nemlig den del der vedrører »Simpel tilfældig udvælgelse«.

I december 1959 er anden del udgivet, som omfatter »Restriktiv tilfældig udvælgelse«. Bogen er inddelt i ialt 9 kapitler (kapitlerne 15-23). Jeg fremhævede i anmeldelsen af 1. bind, at dette bind var overlegent og godt skrevet. Det samme gælder 2. del af kompendiet. Når man læser 2. del med måske endnu større interesse end 1. del, er det bl. a. fordi Lykke Jensen ved gennemgangen af teorien for den restriktive tilfældige udvælgelse bestandig kommer ind på spørgsmål, som har direkte relevans for den praktiske statistiker.

Der er nogle gængse og ikke gængse spørgsmål, ved den restriktive udvælgelse som særlig interesserer, og som Lykke Jensen gennemgår: Hvorledes skal stikprøven fordeles på strata? Og hvor stor skal den være? Hvad betyder omkostningerne for fordelingen på strata? Og hvad betyder anvendelse af kvotientsskøn med hensyn til nedbringelse af usikkerheden? Hertil kommer et spørgsmål, som synes ret nyt og overmåde interessant: Hvorledes foretages den rette stratifikation?

I kapitel 15 omtales stratifikation af multipelt varierende analyseenheder.

Det karakteristiske for udvælgelsesproceduren ved stratifikation betones: 1) at samtlige strata skal være repræsenteret i stikprøven og 2) at stikprøveudtagningen skal foretages således, at stikprøverne fra de forskellige strata er stokastisk uafhængige.

1. Anmeldt af mig i Nationaløkonomisk Tidsskrift 1957, 95. bind, 3-4 hæfte.

Lykke Jensen fremhæver naturligvis det kendte, at man skal tilstræbe homogenitet af elementerne indenfor strata og heterogenitet mellem stratum-værdierne (gennemsnittene). Man opnår herigennem en betydelig nedsættelse af variansen, således at i jo højere grad det nævnte opfyldes, desto større bliver stratifikationsgevinsten.

Variansen er afhængig af stikprøvens størrelse og af spredningen i fordelingen af de enkelte elementer. Hvis spredningen inden for de enkelte strata er meget forskellig fra stratum til stratum, bør man benytte den såkaldte optimale fordeling af antal iagttagelser på de enkelte strata. Denne bestemmes ved at lade udvalgsprocenterne være proportionale med stratumspredningerne. Herved kan opnås en væsentlig større nedsættelse af usikkerheden end ved en proportional fordeling. Den optimale fordeling forudsætter dog, efter det anførte, kendskab til stratumspredningerne i populationen. Bl. a. af den grund konkluderer Lykke Jensen (s. 291) »at kun for en betragtelig forskel mellem stratumspredningerne er det besværet værd at erstatte den simple proportionale fordelingsprocedure med den optimale«.

Dette er rigtigt; men inden for den økonomiske statistik, som jo bliver mere og mere påkrævet, må man sikkert regne med, at spredningerne inden for strata kan være vidt forskellige, således at man måske på dette felt i mange tilfælde ved optimal fordeling kan opnå betydelige reduktioner i variansen. Det eksempel fra brugsforeningsstatistik, som Lykke Jensen omtaler, viser da også, at medens variansen ved proportional udvælgelse er 522,3 bliver den ved optimal udvælgelse kun 146,1.

For små lande betyder omkostningerne særdeles meget. Disse lande må derfor være særlig interesserede i spørgsmål om indpasningen af en omkostningsfunktion ved beregningerne. Lykke Jensen indfører en omkostningsfunktion af følgende form:

$$C = c_0 + \sum_{h=1}^L c_h n_h$$

Denne udtrykker, at der findes visse faste omkostninger c_0 samt, at de variable må antages at være forskellige fra stratum til stratum.

Stiller man kravet om minimalisering af variansen til et *givet* pengebeløb fører dette til, at de strata, hvor det er dyrt at indsamle observationer, vil få en mindre repræsentation i stikproven end strata, hvor det er billigt at fremskaffe observationer. Ved at tage hensyn til omkostningerne vil man følgelig blive tvunget til at afvige fra den foran omtalte optimale fordeling.

Lykke Jensen har i kapitel 15 en omtale af udvælgelse af kun en enhed fra hvert stratum. Denne metode har været anvendt ved de danske arbejdsstyrketællinger. For at kunne beregne variansen er det nødvendigt at samle (L) strata eksempelvis i $\frac{L}{2}$ grupper á 2 strata. Den varians, man kommer til, er ikke et centralt skøn, men vil gennemsnitlig være en overvurdering af den sande varians.

I dette afsnit fremhæver Lykke Jensen også, at grunden til at anvende denne fremgangsmåde kan være den, at man ønsker en stærk stratificering. Vore erfaringer med de danske arbejdsstyrketællinger synes at vise, at den store effekt ved stratificeringen opnås ved hovedinddelingen i få strata (hovedstaden, hovedstadens forstæder, provinsbyerne med forstæder, hymæssige bebyggelser og rene landkommuner). Den følgende underinddeling efter erhverv m.v. betød derimod ikke så meget. Dette synspunkt understreger Lykke Jensen meget stærkt, således som det vil fremgå af det følgende.

Ved omtalen af kvotientsskøn i kapitel 16 nævner Lykke Jensen to hovedberegningssmåder, nemlig:

$$Y = \beta_{oh} X_{h\mu} + P_h \mu$$

og

$$Y_{h\mu} = \beta_0 X_{h\mu} + P_h \mu$$

I første tilfælde beregnes kvotientsskøn inden for hvert stratum (β_{oh}), som herefter sammenevjes. I det andet tilfælde regnes med en fælles kvotient (β_0) for samtlige iagttagelser. Ingen af de to metoder fører til centrale skøn. Ved små stikprøver og et stort antal strata kan det første kvotientsskøn medføre en betydelig skævhed i beregningerne.

Ved gennemgangen af de spørgsmål, som hidtil er omtalt, kan Lykke Jensen i vidt omfang henholde sig til de »klassiske« fremstillinger af stikproveteori. Det gælder derimod ikke problemerne i kapitel 19, hvor spørgsmålene omkring optimal stratifikation vel hører til de nyeste landvindinger. Lykke Jensen konkluderer i dette kapitel, at medens en grov stratifikation, eksempelvis en opdeling på 5 strata, ofte vil betyde meget for en nedsættelse af variansen, vil en yderligere opdeling almindeligvis ikke føre til en væsentlig yderligere nedsættelse.

Lykke Jensen behandler *optimal stratifikation* under forudsætning af *proportional fordeling* og *optimal fordeling*. Han nævner, at betingelsen for optimal stratifikation (minimalisering af variansen) ved proportional fordeling er, at grænsen (d) mellem to vilkårlige strata er et simpelt gennemsnit af middelværdierne i de pågældende strata. Delpunkterne kan bestemmes gennem en iterationsproces.

Når man, som nævnt, kun behøver få strata for eventuelt at opnå en ønsket effekt med hensyn til en nedsættelse af skønnets varians, skyldes dette, at der er en tendens til, at skønnets varians – ved proportional fordeling – aftager omvendt proportionalt med kvadratet på antallet af strata.

I vurderingen af optimal stratifikation ved *optimal fordeling* fremhæver Lykke Jensen, at ved normale fordelinger vil optimal stratifikation med optimal fordeling ikke være væsentlig bedre end optimal stratifikation med proportional fordeling. For udpræget skæve fordelinger er den optimale stratifikation med optimal fordeling derimod den optimale stratifikation med pro-

portional fordeling meget overlegen, idet den ensidige hale (skævhed) får indflydelse på effekten af de to stratifikationsmetoder. Dette forhold må få stor betydning for den økonomiske statistik, hvor man jo ofte har skæve fordelinger; men man må se i øjnene, at dette kan besværliggøre forarbejderne ved samplingens tilrettelæggelse.

I forbindelse med disse betragtninger kunne det have været interessant at have fået betydningen af indførelse af en kostfunktion behandlet. Ønskes en minimalisering af variansen, må omkostningernes indflydelse på *fordelingen* (allokeringen), der ikke længere bliver optimal, øve indflydelse på den ønskelige stratifikation.

Lykke Jensen har en indgående behandling af to-trins udvælgelsen. Han betoner, det kendte, at hvert trin i stikprøveudtagningen yder sit bidrag til variansen. Han omtaler (s. 446), at det ved indsamling af statistiske oplysninger kan være praktisk, at udvalgsbrøken på andet trin er konstant, og at man da kan foretage en systematisk udvælgelse af sekundærenhederne. Ved de danske arbejdsstyrkeundersøgelser valgte man ikke denne fremgangsmåde. Antallet af sekundærenheder bestemtes under hensyn til, hvad man mente, interviewerne kunne overkomme, hvorved udvalgsbrøken blev forskellig fra stratum til stratum. Det synes, at denne fremgangsmåde fra et økonomisk synspunkt var relevant.

I de afsluttende bemærkninger diskuterer Lykke Jensen spørgsmålet om bibeholdelse af enhederne, når man foretager sammenligninger mellem forskellige tidspunkter.

Ønsker man således at sige noget om forskydninger fra et tidspunkt til et andet, opnås størst nøjagtighed, når man bibeholder enhederne, idet variansen

$$V(y) = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma$$

hvor ρ almindeligvis er positiv. Ønsker man derimod størst mulig præcision på et gennemsnit over to (eller flere) tidspunkter, kommer variansen på en sum på tale:

$$V(y) = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2$$

Da ρ almindeligvis er positiv, hvis man anvender de samme enheder, bør man sørge for, at skønnene på de to tidspunkter er stokastisk uafhængige, hvorved $\rho = 0$. Dette opnås ved at udvælge en tilfældig stikprøve på hvert undersøgelsestidspunkt.

Ved de danske arbejdsstyrkeundersøgelser, hvor man var interesseret i forskydningerne, valgte man den første fremgangsmåde; men da man ikke stadig kunne ulejlig de samme lejlighedsindehavere, var man opmærksom på, at man (desværre) blev nødt til en succesiv udskiftning af enheder.

Selv med de vanskeligheder in mente, som gælder for små lande med hensyn til stikprøveundersøgelser, skal der ikke megen spådomskunst til for at sige, at stikprøverne på mangfoldige områder er fremtidens statistiske redskab. Det er da af meget stor betydning, at der nu foreligger et dansk værk, som på alle områder indenfor de givne rammer opfylder de krav, man som »konsument« kan stille til det. Værket er et betydeligt arbejde.

*Kjeld Bjerke**

* Kontorchef i Det statistiske Departement. Lektor på Handelshøjskolen.

Karen Gredal: Moderne forbrugeres motiver og adfærd. Einar Harcks forlag, København 1959, 144 s., Pris kr. 18,00

Karen Gredal, der underviser på Handelshøjskolen i København, er blevet opfordret til at udbygge konsumentteorien, og det foreliggende værk er svaret på denne opfordring. Dispositionen i bogen er i så henseende også ganske klar, idet forfatteren — en for en — behandler en række faktorer, der har betydning for konsumenternes adfærd.

Forfatteren tager udgangspunkt i grænse-nyttelæren og valghandlingsteorien, idet hun i kapitel II gennemgår disse to teorier. Kapitlet slutter derefter med at hævde det ofte fremførte synspunkt, at begge teorier er temmeligt virkelighedsfjerne.