

Interne notater

STATISTISK SENTRALBYRÅ

81/35

2. desember 1981

EN GENERELL METODE FOR ENDRING AV EN TOTRINNS UTVALGSPLAN.

Anvendelse av metoden på undersøkelsen "Helse- og arbeidsmiljøgransking blant fiskere"

Av

Erling Siring

	Side
INNHold	
1. Innledning	1
2. Problemet ved å trekke et utvalg av fiskere	1
3. Metoden som ble brukt for å lage utvalgsplanen	2
3.1. Innledning	2
3.2. Keyfitz-metoden	3
4. Utvalgsplanen for undersøkelsen blant fiskere	4
5. En generell metode for endring av en totrinns utvalgsplan	9
5.1. En enkel prosedyre for sammenslåing av to strata	9
5.2. "Optimal" utnyttelse av tidligere uttrukne utvalgsområder etter en generell endring av strata og trekkesannsynligheter	10
Litteratur	13

1. INNLEDNING

Byråets standard-utvalgsplan er en totrinns utvalgsplan. I første trinn er det trukket områder som består av en eller flere kommuner. I annet trinn trekkes personer eller husholdninger fra disse områdene. I de uttrukne områdene, som kalles utvalgsområder, har en lokalt ansatte intervjuere.

Byråets standard-utvalgsplan er først og fremst konstruert med tanke på å produsere tall for hele befolkningen og spesielle geografiske områder. Hvis en ønsker tall for et mindre geografisk område eller for en spesiell gruppe av befolkningen, vil i mange tilfeller Byråets utvalgsplan være mindre egnet. Dette var også tilfellet i intervjuundersøkelsen "Helse- og arbeidsmiljøgranskning blant fiskere". Hvis en skulle ha trukket et utvalg av fiskere fra Byråets utvalgsområder, ville en fått et utvalg med dårlige egenskaper.

En løsning på dette problemet kunne ha vært å "glemme" standardutvalgsplanen, og lage en helt ny utvalgsplan. Dette ville imidlertid ha blitt en dyr løsning siden utnyttelsen av den fast ansatte intervjuerstaben ville ha blitt dårlig. Vi har søkt etter en løsning som utnytter intervjuerstaben på en tilfredstillende måte.

I kapitlene 2, 3 og 4 blir utvalgsplanen for undersøkelsen blant fiskere dokumentert, og i kapittel 5 blir det beskrevet en generell metode som kan brukes når vi ønsker følgende mål oppfylt: Å kunne trekke en størst mulig del av et utvalg innenfor Byråets utvalgsområder etter at vi har foretatt en "generell" endring av standard-utvalgsplanen. I undersøkelser der det er nødvendig med en endring av standard-utvalgsplanen, kan bruk av denne metoden bety en vesentlig reduksjon i reisekostnadene for intervjuerne.

Metoden som ble brukt for å endre Byråets standard-utvalgsplan til en passende utvalgsplan for undersøkelsen blant fiskere, beskrives i kapittel 3. Denne metoden er egentlig et spesialtilfelle av metoden i kapittel 5.

I kapittel 2 beskrives problemet ved å trekke utvalget av fiskere, og i kapittel 4 beskrives selve utvalgsplanen.

Lesere som bare er interessert i dokumentasjon av utvalgsplanen for undersøkelsen blant fiskere, trenger ikke å lese kapittel 5.

2. PROBLEMET VED Å TREKKE ET UTVALG AV FISKERE

Etter oppdrag fra Fiskeriteknologisk Institutt skulle vi trekke et utvalg av fiskere slik at alle fiskere i målpopulasjonen skulle ha samme sannsynlighet for å komme med i utvalget. Et slikt utvalg kalles selvveiende. Målpopulasjonen bestod i dette tilfellet av alle fiskere bosatt i 54 bestemte kommuner, heretter kalt fiskerikommuner.

En metodemessig sett enkel og god fremgangsmåte ville ha vært å trekke et enkelt tilfeldig utvalg fra disse kommunene. Men dette ville ha blitt en svært kostbar fremgangsmåte, hovedsakelig fordi reiseutgiftene for intervjuerne ville ha blitt store. I tillegg ville sikkert en feilkilde som frafall ha blitt ganske dominerende siden intervjuerne ikke ville hatt anledning til å gjøre mer enn ett forsøk på å komme i kontakt med mange av intervjuobjektene.

I Byrådet har en redusert reisekostnadene for intervjuerne ved å bruke en to-trinns utvalgsplan ved vanlige undersøkelser. I første trinn er det trukket utvalgsområder som består av en eller flere kommuner. I annet trinn trekkes personer eller husholdninger fra disse områdene. Trekkingen på første trinn foretas ikke for hver ny undersøkelse. En bruker nå utvalgsområder som ble trukket i 1974.

Før trekkingen på første trinn delte en landets ca. 440 kommuner inn i 102 strata (eller grupper) etter geografisk beliggenhet og næringsstruktur. Fra hvert av disse strataene trakk en så ett utvalgsområde som skulle representere sitt respektive stratum. Utvalgsplanen er nærmere beskrevet i Thomsen og Riderg (1974) og Thomsen (1977).

I utvalgsområdene har en lokalt ansatte intervjuere. Disse har hvert sitt faste og passende store område å intervju på. Ved en slik totrinns utvalgsplan oppnår en altså å få "samlet" intervjuobjektene innenfor mindre geografiske områder der en har fast ansatte intervjuere.

Når Byråets utvalgsplan i mange tilfeller er lite egnet for undersøkelser av mindre deler av befolkningen, har dette sammenheng med hvordan stratifiseringen av utvalgsområdene ble foretatt. Anta f.eks. at en skal trekke et utvalg av personer fra Finnmark. Hvis en ser nærmere på stratumindelingen, vil en oppdage at enkelte strata inneholder kommuner fra både Troms og Finnmark. Det betyr at noen av utvalgsområdene i Finnmark representerer kommuner i begge de nevnte fylkene. Dette medfører at en ikke får et selvveiende utvalg for Finnmark ved å trekke personer fra utvalgsområdene.

Det samme problemet oppstår når en skal trekke et utvalg fra de 54 kommunene som her er kalt fiskerikommuner. De 54 kommunene utgjør ikke et eget stratum eller superstratum (dvs. et større stratum sammensatt av flere mindre strata). Ved bruk av standard-utvalgsplanen vil en derfor ikke få et selvveiende utvalg av fiskere.

Et annet problem er at bare 9 av de 54 fiskerikommunene hører med til utvalgsområdene. Det betyr at selv om en kunne få et selvveiende utvalg ved å trekke fra utvalgsområdene, ville tallene i undersøkelsen få uakseptabel høy varians (eller usikkerhet) med den planlagte utvalgsstørrelsen.

Hittil har vi drøftet to dårlige framgangsmåter for trekking av utvalget til undersøkelsen blant fiskerne. Våre krav var at utvalget skulle bli selvveiende, og at tallene i undersøkelsen skulle bli av rimelig høy sikkerhet. I tillegg ønsket vi at en størst mulig del av utvalget skulle bli trukket innenfor de nåværende utvalgsområdene for å få små reisekostnader.

For å få oppfylt våre krav og ønsker, gikk vi fram på følgende måte:

- i) Vi laget en passende stratifisering av fiskerikommunene, men passet samtidig på i størst mulig grad å bruke de samme stratumgrensene som i standard-utvalgsplanen.
- ii) Vi foretok så en "ny trekking" av utvalgsområder, med trekking av ett utvalgsområde fra hvert stratum. Vi brukte metoden som er beskrevet i neste avsnitt, til å trekke nye utvalgsområder. Metoden sikret at vi fikk "gjenvælt" flest mulig av de 9 fiskerikommunene som hører med til Byråets utvalgsområder. "Ny trekking" er satt i gåseøyne siden det egentlig ikke ble foretatt en helt ny trekking av utvalgsområder.
- iii) Det endelige utvalget av fiskere ble trukket fra de uttrukne fiskerikommunene. Utvalget ble trukket fra fiskermantallet i Garantikassen for fiskere.

3. METODEN SOM BLE BRUKT FOR Å LAGE UTVALGSPLANEN

3.1. Innledning

Metoden som beskrives i dette avsnittet, er en modifisert utgave av en metode som ble presentert av Keyfitz (1951). I sin opprinnelige form går metoden ut på å trekke to ustratifiserte utvalg på en slik måte at antall felles enheter i de to utvalgene blir størst mulig. Mer presist: Vi tenker oss at vi på et tidspunkt t_1 har trukket et utvalg med trekkesannsynligheter q_i , $i=1, \dots, N$, der N er antall enheter i populasjonen. Videre tenker vi oss at vi på tidspunktet t_2 ønsker oss et utvalg med nye trekkesannsynligheter p_i , $i=1, \dots, N$. Metoden går ut på å endre trekkesannsynlighetene fra q_i til p_i , $i=1, \dots, N$, på en slik måte at vi får beholde flest mulig av de opprinnelig uttrukne enhetene i det "nye" utvalget.

Vi skal kalle metoden for Keyfitz-metoden. Ved en totrinns utvalgsplan med primære trekkeenheter trukket fra strata, er den først og fremst velegnet når en ønsker å endre på trekkesannsynlighetene innen faste strata. Ved å anvende metoden på en spesiell måte har vi i dette tilfellet også brukt metoden til å endre på stratumgrensene. Metoden må for øvrig brukes med stor forsiktighet ved endring av stratumgrenser, og vi anbefaler at den generelle metoden i kapittel 5 brukes i slike situasjoner. Keyfitz-metoden er for øvrig en av byggestenene i denne metoden.

Når det gjelder direkte anvendelse av Keyfitz-metoden, er det i mange situasjoner behov for å endre på trekkesannsynligheter innen faste strata. F. eks. er det ofte et ønske at trekkesannsynlighetene på første trinn i totrinns utvalgsplaner skal være proporsjonale med antall innbyggere i de primære trekkeenheter. Når så folketallene endrer seg over tid, blir det behov for å endre på trekkesannsynlighetene. Tilsvarende kan det være behov for å endre på trekkesannsynlighetene på første

trinn i undersøkelser der målpopulasjonen er en annen enn i standard undersøkelser.

3.2. Keyfitz-metoden

Vi antar at vi fra et stratum har trukket ett utvalgsområde, og at trekkesannsynligheten til utvalgsområde nr. i var q_i , $i=1, 2, \dots, N$. N er antall primære trekkeenheter i stratomet. Vi antar videre at vi ønsker å endre trekkesannsynlighetene fra q_i til p_i , $i=1, 2, \dots, N$. Vi forutsetter at

$$\sum_i p_i = \sum_i q_i = 1. \text{ Følgende framgangsmåte maksimerer sannsynligheten for å beholde den tidligere uttrukne}$$

enheten (Keyfitz (1951)):

- i) En deler først enhetene inn i to grupper etter som de nye trekkesannsynlighetene er mindre eller større enn de gamle.
La $I = \{i: p_i \geq q_i\}$ og $D = \{i: p_i < q_i\}$
- ii) Hvis den opprinnelig uttrukne enheten hører med til mengden I , betraktes denne enheten som gjenvalgt.
- iii) Hvis den opprinnelig uttrukne enheten er enhet nr. d , der $d \in D$, foretas en trekning der en velger mellom å beholde enhet nr. d i utvalget eller å bytte ut enhet nr. d . Begivenheten å beholde enhet nr. d i utvalget gis sannsynligheten $\frac{p_d}{q_d}$.
- iv) Hvis den opprinnelig uttrukne enheten hører med i mengden D , og denne enheten ikke blir gjenvalgt, velges en ny enhet fra mengden I . Enhet nr. i i I gis betinget trekkesannsynlighet $\frac{p_i - q_i}{\sum_{i \in I} (p_i - q_i)}$

Vi skal nå gi et bevis for egenskapene til Keyfitz-metoden. Først skal vi se at hver enhet får riktig trekkesannsynlighet.

Enhet nr. d , der $d \in D$, kan bli valgt ved at denne enheten er valgt opprinnelig, og at en så beholder denne enheten. Sannsynligheten for dette er:

$$q_d \cdot \frac{p_d}{q_d} = p_d$$

Enhet nr. i , der $i \in I$, kan bli valgt på to måter. Den ene er at enhet nr. i er valgt opprinnelig. Den andre er at enheten blir valgt som under pkt. iv. Sannsynligheten for at enhet nr. i skal bli valgt, blir da:

$$\begin{aligned} q_i + \sum_{d \in D} q_d \left(1 - \frac{p_d}{q_d}\right) \frac{p_i - q_i}{\sum_{i \in I} (p_i - q_i)} \\ = q_i + (p_i - q_i) \frac{\sum_{d \in D} (q_d - p_d)}{\sum_{i \in I} (p_i - q_i)} = p_i, \end{aligned}$$

$$\text{siden } \sum_{d \in D} (q_d - p_d) = \sum_{i \in I} (p_i - q_i)$$

Altså får hver enhet riktig sannsynlighet for å bli valgt. Under de gitte forutsetningene maksimerer metoden sannsynligheten for gjenvalg av den tidligere uttrukne enheten. Hvis den tidligere uttrukne enheten hører til gruppen I , blir enheten "gjenvalgt" med sannsynlighet 1. Hvis den tidligere uttrukne enheten hører med til gruppen D , blir enheten "gjenvalgt" med sannsynlighet $\frac{p_d}{q_d}$. Vi kan nå anta at vi kunne gi enhet nr. d en betinget sannsynlighet for "gjenvalg" lik γ_d , slik

$$\text{at } \gamma_d > \frac{p_d}{q_d}.$$

Den ubetingede sannsynligheten for valg av enhet nr. d blir da $q_d \cdot \gamma_d$, som er større enn p_d . Dette strider mot forutsetningen om at enhet nr. d skal ha sannsynlighet p_d for å bli valgt. Altså

gir metoden en maksimal sannsynlighet for "gjenvalg" av den tidligere uttrukne enheten under de gitte forutsetningene.

Som nevnt i innledningen i dette kapitlet, er Keyfitz-metoden først og fremst velegnet når en ønsker å endre på trekkesannsynlighetene innen faste strata. Men den kan også brukes når en ønsker å foreta mindre endringer av stratuminndelingen. Hvis en f.eks. ønsker å fjerne en primær trekkeenhet, kan en gi denne enheten ny trekkesannsynlighet $p_d = 0$. Som vi skal se i neste avsnitt, ble denne teknikken brukt for å fjerne kommuner som ikke var fiskerikommuner, fra utvalgsplanen.

Hvis en ønsker å ta med en ny primær trekkeenhet som ikke har vært med i utvalgsplanen før, kan en la $q_i = 0$ og $P_i > 0$.

I Siring og Thomsen (1981) er det vist hvordan en kan bruke Keyfitz-metoden til å endre på utvalgsplanen slik at en kan trekke et selvveiende utvalg fra Finnmark. Kommunene i Finnmark er inndelt i 4 strata. I to av disse strataene er det også med kommuner fra Troms. For å få fjernet disse kommunene fra de to strataene gir en dem "nye" trekkesannsynligheter lik 0.

Forutsetningen for at en skal kunne bruke Keyfitz-metoden ved endring av statuminndelingen, er at hvert av de nye strataene inneholder enheter fra bare ett av de opprinnelige strataene. Når en skal foreta endringer av trekkesannsynligheter og stratumgrenser, må en sørge for at en ikke lar seg påvirke av utfallet av den opprinnelige trekningen på første trinn. Ellers risikerer en at de nye trekkesannsynlighetene på første trinn blir feil.

4. UTVALGSPLANEN FOR UNDERSØKELSEN BLANT FISKERE

Problemet ved å trekke et utvalg av fiskere er beskrevet i kapittel 2. Konklusjonen i kapitlet er at det er dårlige framgangsmåter både å trekke et enkelt, tilfeldig utvalg fra de 54 fiskerikommunene og trekke et utvalg fra Byråets utvalgsområder. Det gjenstående alternativet er da å lage en tottrinns utvalgsplan med en noe annen stratuminndeling enn i Byråets standard utvalgsplan.

På det tidspunktet utvalgsplanen ble laget, hadde vi bare kjennskap til Keyfitz-metoden og ikke til metodene beskrevet i kapittel 5. Siden Keyfitz-metoden først og fremst er velegnet for endring av trekkesannsynligheter til primære trekkeenheter innen faste strata, var det naturlig å prøve å beholde Byråets stratuminndeling mest mulig intakt. Vi vurderte da å la stratumgrensene være uendret, og bare fjerne kommuner som ikke var fiskerikommuner fra utvalgsplanen. Dette kunne en gjøre ved å gi den sistnevnte typen kommuner nye trekkesannsynligheter lik 0.

Denne løsningen viste seg imidlertid heller ikke å være særlig god. En fikk for få strata, og i tillegg varierte antall fiskere i hvert stratum for mye. Ved en slik utvalgsplan ville usikkerheten til tallene blitt uakseptabel høy.

I den utvalgsplanen vi til slutt valgte, delte vi opp noen av de opprinnelige strataene i flere deler. I et par tilfeller slo vi også sammen opprinnelige strata. På forhånd hadde vi bestemt oss for å lage 16-17 nye strata med omtrent like mange fiskere i hvert stratum.

Som i Byråets standard-utvalgsplan bestemte vi oss for å trekke én primær trekkeenhet fra hvert stratum. De primære trekkeenhetene lot vi være de samme som i standard-utvalgsplanen, der en primær trekkeenhet består av en større eller flere mindre kommuner. Primære trekkeenheter som ikke var fiskerikommuner, ble gitt nye trekkesannsynligheter lik 0.

Vi ønsket så at trekkesannsynlighetene på første trinn skulle være proporsjonale med antall fiskere i de primære trekkeenhetene. Nå var det imidlertid slik at vi ikke visste nøyaktig hvor mange fiskere som bodde i de enkelte fiskerikommunene. Vi måtte bruke tall på fiskere fra folketellingen i 1970 til beregning av trekkesannsynlighetene. Om ikke tallene på fiskere i fiskerikommunene er de samme i dag som i 1970, er trekkesannsynlighetene på 1. og 2. trinn justert slik i forhold til hverandre at utvalget av fiskere likevel er selvveiende. Det at vi ikke har proporsjonal allokering kan imidlertid føre til at variansene eller standardavvikene blir noe større enn ved proporsjonal allokering. Avvikene fra proporsjonalitet er i dette tilfelle antakeligvis så små at eventuelle presisjonstap blir ubetydelige.

Under følger en oversikt over stratuminndelingen av de primære trekkeenheter, antall fiskere i 1970, opprinnelige trekkesannsynligheter og nye eller ønskede trekkesannsynligheter. At kommuner er slått sammen til en primær trekkeenheter er markert enten ved "&" mellom kommunene eller at det er klammaparenteser rundt dem. "Andre" primære trekkeenheter betegner andre trekkeenheter i det opprinnelige stratomet. Det kan være både kommuner som ikke er fiskerikommuner og fiskerikommuner som er kommet i et av de andre "nye" strataene. "v" markerer opprinnelig trukket enhet.

Navn på de primære trekkeenheter	Antall fiskere 1. nov. 1970	Opprinnelig trekkesanns. q_i .	Ny ønsket trekkesanns. p_i .
<u>Stratum 1</u>			
I alt	421	Uinteressant	1,0000
Hvaler	167	p.g.a.	0,3967
Kvitsøy & Utsira	110	helt ny	0,2613
Sund	144	trekking	0,3420
<u>Stratum 2</u>			
I alt	956	1,000	1,0000
Austevoll	432	0,207	0,4519
v { Øygården Fedje Solund }	524	0,247	0,5481
Andre	-	0,546	-
<u>Stratum 3</u>			
I alt	585	1,0000	1,0000
Vågsøy & Selje	585	0,5460	1,0000
Andre	-	0,4540	-
<u>Stratum 4</u>			
I alt	958	1,0000	1,0000
Frøya	608	0,1697	0,6347
v Osen	123	0,2165	0,1284
Vikna	227	0,1547	0,2370
Andre	-	0,4591	-
<u>Stratum 5</u>			
I alt	779	1,0000	1,0000
v Sande & Herøy	779	0,3552	1,0000
Andre	-	0,6448	-
<u>Stratum 6</u>			
I alt	770	1,0000	1,0000
Giske	423	0,1690	0,5494
Midsund & Sandøy	347	0,2109	0,4506
Andre	-	0,6201	-

Navn på de primære trekkeenheter	Antall fiskere 1. nov. 1970	Opprinnelig trekkesanns. q_i .	Ny (ønsket) trekkesanns. p_i .							
<u>Stratum 7</u>										
I alt	725	1,0000	1,0000							
Averøy	320	0,1633	0,4414							
Smøla	405	0,1014	0,5586							
Andre	-	0,7353	-							
<u>Stratum 8</u>										
I alt	540	1,0000	1,0000							
Gildeskål	180	0,1772	0,3333							
Steigen	215	0,1080	0,3981							
Dønna	145	0,1708	0,2685							
Andre	-	0,5440	-							
<u>Stratum 9</u>										
I alt	1 051	1,0000	1,0000							
Vega & Herøy	460	0,1298	0,4377							
<table border="0" style="display: inline-table; vertical-align: middle;"> <tr> <td style="font-size: 2em; vertical-align: middle;">{</td> <td style="padding: 0 5px;">Lurøy</td> <td rowspan="3" style="vertical-align: middle;">}</td> </tr> <tr> <td></td> <td>Træna</td> </tr> <tr> <td></td> <td>Rødøy</td> </tr> </table>	{	Lurøy	}		Træna		Rødøy	591	0,1442	0,5623
{	Lurøy	}								
	Træna									
	Rødøy									
Andre	-	0,7260	-							
<u>Stratum 10</u>										
I alt	1 420	1,0000	1,0000							
<table border="0" style="display: inline-table; vertical-align: middle;"> <tr> <td style="font-size: 2em; vertical-align: middle;">{</td> <td style="padding: 0 5px;">Røst</td> <td rowspan="3" style="vertical-align: middle;">}</td> </tr> <tr> <td></td> <td>Værøy</td> </tr> <tr> <td></td> <td>Moskenes</td> </tr> </table>	{	Røst	}		Værøy		Moskenes	800	0,1537	0,5634
{	Røst	}								
	Værøy									
	Moskenes									
v Vestvågøy	620	0,2989	0,4366							
Andre	-	0,5474	-							
<u>Stratum 11</u>										
I alt	976	1,0000	1,0000							
Bø	419	0,1290	0,4293							
Øksnes	557	0,1445	0,5707							
Andre	-	0,7265	-							
<u>Stratum 12</u>										
I alt	813	1,0000	1,0000							
v Bjarkøy & Ibestad	293	0,0962	0,3604							
<table border="0" style="display: inline-table; vertical-align: middle;"> <tr> <td style="font-size: 2em; vertical-align: middle;">{</td> <td style="padding: 0 5px;">Tranøy, Torsken</td> <td rowspan="2" style="vertical-align: middle;">}</td> </tr> <tr> <td></td> <td>& Berg</td> </tr> </table>	{	Tranøy, Torsken	}		& Berg	520	0,1359	0,6396		
{	Tranøy, Torsken	}								
	& Berg									
Andre	-	0,7679	-							
<u>Stratum 13</u>										
I alt	1 181	1,0000	1,0000							
Karlsøy & Lyngen	687	0,1809	0,5817							
Skjervøy & Kvænangen	494	0,1412	0,4183							
Andre	-	0,6779	-							

Navn på de primære trekkeenheter	Antall fiskere 1. nov. 1970	Opprinnelig trekkesanns. q_i .	Ny (ønsket) trekkesanns. p_i .
<u>Stratum 14</u>			
I alt	1 094	1,0000	1,0000
Loppa & Hasvik	393	0,0987	0,3592
Sørøysund & Måsøy	701	0,1271	0,6408
Andre	-	0,7742	-
<u>Stratum 15</u>			
I alt	725	1,0000	1,0000
Lebesby & Gamvik	368	0,0997	0,5076
Berlevåg & Båtsfjord	357	0,1203	0,4924
Andre	-	0,7800	-
<u>Stratum 16</u>			
I alt	683	Uinteressant	1,0000
Vardø	250	p.g.a.	0,3660
Nordkapp	321	helt ny	0,4700
Kvalsund	112	trekning	0,1640

De tre sørligste fiskerikommunene var enslige og små fiskerikommuner i hver sine opprinnelige strata, og ble derfor slått sammen til ett nytt stratum, stratum 1. Tilsvarende består stratum 16 av fiskerikommuner fra to forskjellige opprinnelige strata. Hvert av de andre strataene består av enheter fra bare ett av de opprinnelige strataene.

For å få trukket ett primært utvalgsområde fra hvert stratum, og samtidig få utnyttet Byråets utvalgsområder best mulig, anvendte vi Keyfitz-metoden innen hvert av de nye strataene. Metoden er imidlertid ikke anvendelig i strata som består av enheter fra mer enn ett opprinnelig stratum, og den kunne derfor ikke anvendes innen stratum 1 og 16. Da vi på det tidspunktet utvalgsplanen ble laget, ikke hadde kjennskap til metodene i kapittel 5, måtte vi foreta en helt ny trekning i stratum 1 og 16. Det var forøvrig ingen tidligere uttrukne områder i noen av disse strataene, så i dette tilfellet hadde vi derfor ikke tjent noe på å anvende en av metodene i kapittel 5.

Vi skal nå vise hvordan vi anvendte Keyfitz-metoden i hvert av strataene 2 - 15.

Stratum 2 inneholdt et av Byråets utvalgsområder. Den nye trekkesannsynligheten til dette utvalgsområdet var større enn den opprinnelige trekkesannsynligheten. Etter Keyfitz-metodens pkt. (ii) beholdt en da dette området som representant for stratum 2.

På tilsvarende måte ble tidligere uttrukne utvalgsområder "gjenvolgt" i strataene 5, 10 og 12.

I stratum 4 hadde vi fiskerikommunen Osen, som sammen med 2 ikke-fiskerikommuner utgjør et av Byråets utvalgsområder. I vår nye utvalgsplan var trekkesannsynligheten til denne trekkeenheten blitt sterkt redusert. Den viktigste årsaken til dette var at de to ikke-fiskerikommunene var blitt fjernet.

Etter Keyfitz-metodens pkt. (iii) skulle en da foreta et valg mellom å beholde dette utvalgsområdet eller å bytte det ut. Begivenheten å beholde området ble gitt sannsynligheten $\frac{p_d}{q_d} = 0,593$, der

p_d var ny trekkesannsynlighet og q_d opprinnelig trekkesannsynlighet. Utfallet av trekningen ble at vi skulle beholde Osen som utvalgsområde.

Stratum 3 og 5 inneholdt bare én trekkeenheter hver, og disse enhetene ble selvsagt valgt.

I resten av strataene, strataene 6, 7, 8, 9, 11, 13, 14 og 15 hørte den opprinnelige uttrukne enheten med til gruppen "andre" trekkeenheter med nye trekkesannsynligheter lik 0. Etter Keyfitz-metodens pkt. (iii) ble da de opprinnelig uttrukne områdene byttet ut med sannsynlighet $1 - \frac{p_d}{q_d} = 1$.

Nye områder ble så trukket blant enheter som hadde fått økt sine trekkesannsynligheter i den nye utvalgsplanen. Etter Keyfitz-metodens pkt. (iv) valgte en da blant disse enhetene slik at enhet nr. i ble gitt betinget trekkesannsynlighet lik

$$\frac{p_i - q_i}{\sum_{i \in I} (p_i - q_i)}$$

I stratum 6 f.eks. hørte alle trekkeenhetene unntatt "andre" med i gruppen I av områder som hadde fått økt sine trekkesannsynligheter. En valgte da mellom de to aktuelle områdene i gruppe I med sannsynligheter beregnet etter formelen over. Områdene Midsund & Sandøy og Giske fikk da betingede trekkesannsynligheter lik henholdsvis 0,3866 og 0,6134. Midsund & Sandøy ble valgt.

Et krav til utvalget av fiskere var at utvalget skulle være selvveiende, dvs. at alle fiskerne i de 54 fiskerikommunene skulle ha samme sannsynlighet for å bli trukket ut. Som vi har sett, varierte trekkesannsynlighetene på første trinn mellom de primære trekkeenhetene. For at utvalget skulle bli selvveiende, måtte vi også la trekkesannsynlighetene på annet trinn variere mellom de primære trekkeenhetene. Betingelsen for at et utvalg skal være selvveiende, er at produktet av trekkesannsynlighetene på første og annet trinn er det samme for alle de primære trekkeenhetene.

Under følger en oversikt over de uttrukne primære trekkeenhetene og trekkesannsynlighetene som ble brukt på annet trinn. Forventet antall som skulle trekkes fra hvert primært utvalgsområde, er også tatt med. Disse tallene er beregnet ut fra folketellingsdata fra 1970, og avviker derfor litt fra antall fiskere som virkelig ble trukket.

Uttrukne fiskerikommuner	Trekkesannsynlighet på 2. trinn	Forventet antall som skulle trekkes
I alt		1 441
Hvaler	0,2654	44
Øygard & Fedje & Solund	0,1921	101
Vågsøy & Selje	0,1053	62
Osen	0,8201	101
Sande & Herøy	0,1053	82
Midsund & Sandøy	0,2337	81
Smøla	0,1885	76
Dønna	0,3922	57
Vega & Herøy	0,2406	111
Vestvågøy	0,2412	150
Øksnes	0,1845	103
Bjarkøy & Ibestad	0,2922	86
Karlsøy & Lyngen	0,1810	124
Sørøysund & Måsøy	0,1643	115
Lebesby & Gamvik	0,2074	76
Vardø	0,2877	72

Trekningen på 2. trinn ble foretatt fra Fiskermantallet i Garantikassen for fiskere.

Alle de 9 fiskerikommunene som hører med til Byråets utvalgsområder, ble "gjenvælt" som nye utvalgsområder. Ut fra folketellingstallene fra 1970 var forventet andel av utvalget som skulle trekkes innenfor Byråets utvalgsområder, 36 prosent. Hvis en skulle ha trukket et enkelt tilfeldig utvalg fra de 54 fiskerikommunene, ville den samme prosentandelen ha blitt ca. 17. Metoden som ble anvendt, ga derfor en anelig kostnadmessig gevinst ved at en sparte reiseutgifter for intervjuerne.

5. EN GENERELL METODE FOR ENDRING AV EN TOTRINNS UTVALGSPLAN.

Tidligere i dette notatet har vi sett at det i mange situasjoner kan være behov for å endre en totrinns utvalgsplan. I kap. 3 ble det presentert en metode til endring av trekkesannsynligheter innen faste strata, slik at en får beholde et maksimalt antall tidligere uttrukne enheter. Metoden kan også brukes når en skal foreta enkle endringer i stratumgrenser, slik som f.eks. å dele et stratum i to deler.

I Kish and Scott (1971) er det presentert metoder som er mer generelle. Etter en helt generell endring av stratumgrenser og trekkesannsynligheter sikrer metodene at en får beholde det bortimot maksimale antall tidligere uttrukne primære trekkeenheter. I dette kapitlet skal vi presentere en forenklet versjon av en av disse metodene. Hensikten med kapitlet er å gi et innblikk i tankegangen i metodene til Kish og Scott, og forsøke å gjøre disse nyttige metodene bedre kjent i Byrået. Først skal vi imidlertid presentere en enkel prosedyre for sammenslåing av to strata.

5.1. En enkel prosedyre for sammenslåing av to strata

Vi antar at vi ønsker å slå sammen 2 strata A og B til ett stratum, og at det tidligere er blitt trukket ett utvalgsområde fra hvert av strataene A og B med trekkesannsynligheter proporsjonale med folketallene i utvalgsområdene. Vi antar videre at vi ønsker bare ett utvalgsområde trukket fra det nye stratomet, og at trekkesannsynlighetene fortsatt skal være proporsjonale med folketallene i utvalgsområdene. Vi antar også at alle de primære trekkeenheterene i de to strataene A og B skal være med i det nye stratomet.

La N_i være antall innbyggere i utvalgsområde nr. i og N_A og N_B antall innbyggere totalt i henholdsvis stratum A og B. Vi har at den opprinnelige trekkesannsynligheten til utvalgsområde nr. i i stratum A var $\frac{N_i}{N_A}$, og at den opprinnelige trekkesannsynligheten til utvalgsområde nr. j i stratum B var $\frac{N_j}{N_B}$. Vi ønsker nå at disse trekkesannsynlighetene endres til henholdsvis $\frac{N_i}{N_A + N_B}$ og $\frac{N_j}{N_A + N_B}$.

I det nye stratomet har vi altså to tidligere uttrukne enheter. For å få trukket ett utvalgsområde fra det nye stratomet med ønsket trekkesannsynlighet og samtidig få gjenvalgt et av de tidligere uttrukne utvalgsområdene, kan en bruke følgende framgangsmåte:

Metode I

Velg mellom stratum A og B med sannsynligheter lik henholdsvis $\frac{N_A}{N_A + N_B}$ og $\frac{N_B}{N_A + N_B}$, og

betrakt den opprinnelige uttrukne enheten i det uttrukne stratomet som trukket fra det nye stratomet.

Bevis for at metoden gir riktige trekkesannsynligheter:

Anta at trekkeenheter nr. i er en av de opprinnelige trekkeenheterene i stratum A. At denne trekkeenheteren skal bli trukket fra det nye stratomet kan skje ved at stratum A velges, og at trekkeenheter nr. i er den opprinnelige uttrukne enheten i stratum A. Sannsynligheten for dette er:

$$\frac{N_i}{N_A} \cdot \frac{N_A}{N_A + N_B} = \frac{N_i}{N_A + N_B}.$$

Tilsvarende kan en vise at en trekkeenheter i stratum B får riktig trekkesannsynlighet.

Hvis det er slik at innbyggertallene i de primære trekkeenheterene har endret seg fra N_i til M_i siden den opprinnelige trekningen, og vi nå ønsker at de nye trekkesannsynlighetene skal være

$\frac{M_i}{M_A + M_B}$, kan en bruke metode I til å velge en enhet med sannsynlighet $\frac{N_i}{N_A + N_B}$, og siden bruke Keyfitz-metoden til å endre trekkesannsynlighetene til $\frac{M_i}{M_A + M_B}$.

Hvis endringene i trekkesannsynlighetene er små, er den ovenfor beskrevne framgangsmåten nær optimal, dvs. sannsynligheten for å velge en helt ny enhet fra det nye stratomet vil være svært liten. Metoden er imidlertid ikke særlig generell ved at den forutsetter at alle de primære trekkeenheterene i de opprinnelige strataene A og B er med i det nye stratomet.

5.2. "Optimal" utnyttelse av tidligere uttrukne utvalgsområder etter en generell endring av strata og trekkesannsynligheter.

Metoden som skal presenteres her, kan i sin mest generelle versjon brukes til trekking av primære trekkeenheter etter en helt vilkårlig endring av stratumgrenser og trekkesannsynligheter. Metoden er nær ved å være optimal med hensyn til "gjenvalg" av flest mulig enheter trukket fra opprinnelige strata.

Siden metoden er detaljert beskrevet i Kish and Scott (71), skal vi her nøye oss med å presentere den i en forenklet form for å gi et innblikk i tankegangen i den. Vår innskrenking i generalitet består i at vi skal anta at "nye" strata består av deler av høyst to opprinnelige strata.

Før vi ser nærmere på metoden, skal vi komme med en advarsel i forbindelse med bruken av den: Når en foretar endringer i trekkesannsynligheter og stratumgrenser, må en sørge for at en ikke lar seg påvirke av utfallet av den opprinnelige trekningen på første trinn. Ellers risikerer en at de nye trekkesannsynlighetene på 1. trinn blir feil.

For å unngå feil av denne typen kan en la en person som ikke kjenner til utfallet av den opprinnelige trekningen på første trinn eller metodene beskrevet i dette notatet, foreta den nye stratumindelingen og fastsettingen av de nye trekkesannsynlighetene.

Vi skal nå anta at vi etter en omlegging av stratumgrensene har fått et nytt stratum som består av deler av to opprinnelige strata. De to delene kaller vi A og B. Som tidligere i dette notatet skal vi anta at det ble trukket ett utvalgsområde fra hvert av de opprinnelige strataene, og at det skal trekkes et utvalgsområde fra hvert av de nye strataene.

Vi antar at den opprinnelige trekkesannsynligheten til primær trekkeenheter nr. i var q_i , og at vi ønsker denne endret til P_i . Vi lar P_A og P_B betegne summene av alle trekkesannsynlighetene i henholdsvis A og B.

$$P_A = \sum_{i \in A} P_i \quad \text{og} \quad P_B = \sum_{i \in B} P_i$$

Metoden går ut på følgende:

Metode II

a) En foretar en ordning av de to stratumdelen A og B etter størrelsen av P_A og P_B . Vi kaller nå de to stratumdelen for A_1 og A_2 slik at $P_{A_1} > P_{A_2}$.

b) Hvis A_1 inneholder en tidligere trukket enhet, som vi kan kalle enhet nr. k, går en fram på følgende måte:

i) Hvis $P_k \geq q_k$, betraktes enhet nr. k som valgt fra det nye stratumet.

ii) Hvis $P_k < q_k$, velger en enhet nr. k som utvalgsområde fra det nye stratumet med betinget trekkesannsynlighet $\frac{P_k}{q_k}$.

Vi ser at i) og ii) over faller sammen med pkt. ii) og iii) i Keyfitz-metoden.

c) Hvis det ikke er blitt valgt noen enhet under b), ser en på A_2 . En går nå fram som under b), men bytter ut den opprinnelige trekkesannsynligheten q_k i (i) og (ii) med

$$r_{k2} = q_k \left(1 - \sum_{I_1} q_i - \sum_{D_1} P_d \right) = q_k \cdot Q_1,$$

der $I_1 = \{i \in A_1 : P_i \geq q_i\}$, $D_1 = \{d \in A_1 : P_d < q_d\}$

og Q_1 er sannsynligheten for at det ikke er blitt valgt noe utvalgsområde under b).

d) Hvis det ikke er blitt valgt et utvalgsområde verken under b) eller c), trekkes et utvalgsområde fra mengden I av trekkeenheter, der $I = \{i : P_i \geq r_i\}$ med r_i definert som følger:

$$r_i = \begin{cases} q_i & \text{hvis } i \in A_1 \\ r_{i2} & \text{hvis } i \in A_2 \end{cases}$$

Enhet nr. $i \in I$ gis følgende betingede trekkesannsynlighet:

$$\frac{P_i - r_i}{\sum_I (P_i - r_i)}$$

Bevis for at hver enhet får den riktige trekkesannsynlighet P_k , $k = 1, 2, \dots$:

Vi definerer først følgende:

$$I_2 = \{i \in A_2 : P_i \geq r_i\}$$

$$D_2 = \{d \in A_2 : P_i < r_i\}$$

$$D = D_1 \cup D_2$$

Anta at enhet nr. i er en enhet i I_1 . Sannsynligheten for at enhet nr. i skal bli valgt blir:

$\Pr(\text{enhet nr. } i \text{ velges}) = \Pr(\text{enhet nr. } i \text{ er en opprinnelig valgt enhet}) + \Pr(\text{ikke noe valg under b) og c}) \cdot \Pr(\text{enhet nr. } i \text{ velges under d))$.

$$= q_i + Q_1 \left(1 - \sum_{I_2} q_i - \sum_{D_2} \frac{P_d}{Q_1}\right) \frac{P_i - r_i}{\sum_I (P_i - r_i)}$$

$$(*) = q_i + \left(1 - \sum_I r_i - \sum_D P_d\right) \frac{P_i - r_i}{\sum_I (P_i - r_i)}$$

Ved å bruke $\sum_D P_d + \sum_I P_i = 1$ i (*)

får en: $\Pr(\text{enhet nr. } i \text{ velges})$

$$= q_i + \sum_I (P_i - r_i) \frac{P_i - r_i}{\sum_I (P_i - r_i)} = P_i,$$

dvs. enhet nr. i har fått riktig trekkesannsynlighet.

Anta at enhet nr. d hører med til D_1 . $\Pr(\text{enhet nr. } d \text{ velges}) = \Pr(\text{enhet nr. } d \text{ er en opprinnelig valgt enhet}) \cdot \Pr(\text{enhet nr. } d \text{ gjenvelges} \mid \text{enhet nr. } d \text{ valgt tidligere}) = q_d \cdot \frac{P_d}{q_d} = P_d$.

Tilsvarende som vi har vist at enheter i I_1 og D_1 får riktige trekkesannsynligheter, kan det vises at enheter i I_2 og D_2 får riktige trekkesannsynligheter.

Når et nytt stratum består av enheter fra flere enn to opprinnelige strata, blir metode II atskillig mer komplisert enn beskrevet her.

Når et nytt stratum består av enheter fra bare ett opprinnelig stratum, tas pkt. c) vekk. Metoden vil da falle sammen med Keyfitz-metoden i kap. 3. Det betyr at hvis en hadde anvendt metode II ved trekkingen av fiskerikommuner, ville framgangsmåten for trekking av fiskerikommuner i stratum 2 - 15 (jfr. kap. 4) ha blitt den samme som den som ble brukt.

Metode II er ikke optimal med hensyn til det å "gjenvelge" flest mulig tidligere trukne enheter, men i de fleste tilfeller vil den være svært nær ved å være optimal. At ikke metoden er optimal skal vi se av følgende:

Anta at A_1 inneholder en tidligere trukken enhet (enhet nr. k), og at A_2 ikke gjør det, og at $p_k < q_k$. Sannsynligheten for å velge en helt ny enhet blir $1 - \frac{p_k}{q_k}$. Anta nå at A_1 og A_2 hadde vært byttet om. Det betyr at q_k erstattes med r_{k2} (jfr. pkt. c)). Vi har at $r_{k2} < q_k$. Hvis nå $r_{k2} < p_k$, vil sannsynligheten for å gjenvelge en helt ny enhet være 0. Hvis $p_k < r_{k2}$, vil sannsynligheten for å velge en helt ny enhet være $1 - \frac{p_k}{r_{k2}}$. Vi ser altså at sannsynligheten for å velge en helt ny enhet endrer seg med ordningen av stratumdelene A_1 og A_2 . Det betyr at metode II ikke er optimal.

Selv om en finner fram til et kriterium for en optimal ordning av stratumdelene, får en likevel ikke en optimal metode. Skal en få minimert sannsynligheten for å velge en helt ny trekkeenhet, må en anvende en prosedyre tilsvarende metode II direkte på de enkelte trekkeenhetene.

En optimal metode vil kreve at en ordner trekkeenhetene slik at en først behandler enheter som har større trekkesannsynlighet i det "nye" stratomet enn det de hadde opprinnelig. Når en plasserer en enhet i rekkefølgen, må imidlertid de opprinnelige trekkesannsynlighetene til de gjenværende enhetene erstattes med modifiserte sannsynligheter slik som q_k blir erstattet med r_{k2} under c). I praktiske situasjoner kan en slik framgangsmåte bli svært arbeidskrevende. Dette sammen med at gevinsten blir liten i forhold til å bruke metode II, gjør at det i få tilfeller vil være bryet verdt å anvende en slik metode. Når det gjelder metode II og ordningen av stratumdeler, anbefaler Kish og Scott en ordning slik at $p_{A1} > p_{A2} > \dots$. Grunnen til at de anbefaler dette ordningskriteriet er at det er det som medfører minst regnearbeid i praktiske situasjoner. Med de tekniske hjelpemidler som vi er i besittelse av i dag, betyr imidlertid mengden av regnearbeid svært lite. Hvis metode II skal anvendes i Byrådet, bør en vurdere andre ordningskriterier enn det som Kish og Scott anbefaler.

$$\text{La } D_i = \{d \in A_i : p_d < q_d\}$$

F.eks. kan en ordning som er slik at $\sum_{D_1} (q_d - p_d) < \sum_{D_2} (q_d - p_d) < \dots$, bety en forbedring av metode II.

Uansett hvilket ordningskriterium en velger, må en sørge for at ordningen blir uavhengig av utfallet av den opprinnelige trekningen.

LITTERATUR

- Keyfitz, N. (1951): *Sampling with probabilities proportional to size: Adjustment for changes in the probabilities.* Am.stat. Ass. Journal, March - 51.
- Kish, L. and Scott, A. (1971): *Retaining Units after Changing Strata and Probabilities.* Journal of the Am.stat. Ass, Sept. 1971, Vol. 66, No. 335.
- Siring, E. og Thomsen, I. (1981): *Metoder for estimering av tall for fylker v.h.a. utvalgsundersøkelser.* Statistisk Sentralbyrå. Rapporter 81/6.
- Thomsen, I. og Rideng, A. (1974): *Oversikt over arbeidet med ny utvalgsplan.* Stensil ITh/ARi/GHu, 21/5 -74, Statistisk Sentralbyrå.
- Thomsen, I. (1977): *Prinsipper og metoder for Statistisk Sentralbyrås utvalgsundersøkelser.* Samfunnsøk. studier nr. 33, Statistisk Sentralbyrå -77.