Interne notater

STATISTISK SENTRALBYRÅ

88/5 21. april 1988

REVISJON AV DEN GENERELLE UTVALGSPLAN

AV HÅKAN LÖVKVIST

Innhold

1. Innledning

Den utvalgsplan som brukes i forbindelse med Byråets husholdnings- og personundersøkelser ble lagt opp rundt 1975, og har stort sett vært uforandret siden. Av forskjellige grunner er det behov for å modifisere planen, kanskje mest fordi flere av kommunene som har vært med siden 1975 har blitt "utbrukt". Hensikten med dette notat er å beskrive hvordan utvalgsplanen er blitt revidert innen fylkene Nordland, Troms og Finnmark. Liknende revisjon av utvalgsplanen vil bli foretatt for resten av landet senere.

Under revisjonen har vi tatt hensyn til de svakheter i utvalgsplanen som er påvist i Haldorsen (1985), og derved har revisjonen antakeligvis ført til en mer effektiv utvalgsplan. Det er mulig å øke effektiviteten noe mer, men da måtte revisjonen gjøres som en total omlegging av utvalgsplanen, noe som lett kunne føre til brudd med tidligere resultater. Dessuten ville det være nødvendig å skifte ut et stort antall intervjuere, noe som ville medføre store administrative og økonomiske problemer.

I kapittel 2 og 3.1 blir de formål og ønsker som ligger til grunn for revisjonen behandlet. Vi beskriver de utgangspunkter og begreper som gjelde når den gamle utvalgsplanen ble utarbeidet og som til stor del også er aktuell for revisjonen. Vi begrunner dessuten de forandringer som vi har laget.

I kapittel 3.2 blir inndelingen i nye strata behandlet, mens kapittel 3.3 går løs på trekkingen av psu-er i første trinn.

Spørsmålet om hvordan vi kan forbedre metodene for revidering av utvalgsplanen (når dette blir aktuelt for resten av landet) begrunner vi i kapittel 3.4.

Dette notat beskriver en praktisk måte å bruke avansert statistisk metode på. Teoretiske begrunnelser blir mye sparsomt brukt. For den leser som er interessert av bevis av algoritmer etc., henviser vi til referansene.

2. Byråets utvalgsplan

Byråets utvalgsplan tar sikte på å oppfylle en rekke formål. En vil f.eks. ha mulighet for å utføre undersøkelser av varierende slag. I tillegg ønsker en å kunne variere antallet intervjuere med mellom omlag 100 og 350 for hele landet. Endelig er det behov for å kunne publisere tall både for hele landet og for enkelte geografiske områder (Thomsen, 1977).

Med utgangspunkt i disse formål fant en det formålstjenlig å lage en to-trinns utvalgsplan med stratifisering i første trinn. Stratum-grensene ble definert i tre trinn. Først ble landet delt inn i fem landsdeler. Med unntak av Oslo/Akershus ble landsdelene deretter delt inn i regioner med fylkesgrensene som utgangspunkt. Inndelingen ser ut på følgende måte:

1. Oslo/Akershus

2. Resten av Østlandet
   - 2.1 Østfold-Vestfold
   - 2.2 Hedmark-Oppland
   - 2.3 Buskerud/Telemark

3. Sørlandet/Vestlandet (- Møre)
   - 3.1 Agder-Rogaland
   - 3.2 Hordaland-Sogn og Fjordane

4. Møre/Trøndelag
   - 4.1 Møre og Romsdal
   - 4.2 Trøndelag

5. Nord-Norge
   - 5.1 Nordland
   - 5.2 Troms-Finnmark

I et tredje og siste trinn laget en en oppdeling i totalt 102 strata. Som grunnlag for stratifiseringen brukte en (foruten den ovennevnte regioninndelingen) blant annet geografisk beliggenhet og næringsstruktur.

Innenfor hvert stratum definierte en et antall utvalgsområder (psuer). Kommuner med mindre enn 3 000 innbyggere ble slått sammen med andre til ett psu, mens de øvrige kommunene utgjorde egne utvalgsområder.

TABELL 1.

OPPRINNELIG STRATUMINDELNING

STRATUM NR =88

	KOMMUNE	UTVOMR	ANT INNB
	1811 BINDAL	1	2167
	1825 GRANE	1	1766
	1826 HATTFJELLDAL	1	1759
	1812 SØMNA	2	2158
	1813 BRØNNØY	2	6826
	1822 LEIRFJORD	3	2357
	1827 DØNNA	3	1939
	1828 NESNA	3	1775
*	1832 HEMNES	4	4860
	1838 GILDESKÅL	5	2666
	1839 BEIARN	5	1546
	1842 SKJERSTAD	5	1269
	1848 STEIGEN	6	3452
			34540

STRATUM NR =89

	KOMMUNE	UTVOMR	ANT INNB
	1815 VEGA	1	1608
	1816 VEVELSTAD	1	730
	1818 HERØY	1	2102
	1834 LURØY	2	2410
	1835 TRÆNA	2	530
	1836 RØDØY	2	1807
	1856 RØST	3	704
	1857 VÆRØY	3	937
	1859 FLAKSTAD	3	1664
	1874 MOSKENES	3	1514
*	1860 VESTVÅGØY	4	10700
	1867 BØ	5	4053
	1868 ØKSNES	6	5021
			33780

STRATUM NR =90

	KOMMUNE	UTVOMR	ANT INNB
*	1915 BJARKØY	1	788
*	1917 IBESTAD	1	2395
	1927 TRANØY	2	2111
	1928 TORSKEN	2	1408
	1929 BERG	2	1288
	1936 KARLSØY	3	2882
	1938 LYNGEN	3	3677
	1941 SKJERVØY	4	3201
	1943 KVÆNANGEN	4	1695
	2014 LOPPA	5	1866
	2015 HASVIK	5	1462
	2016 SØRØYSUND	6	2342
	2018 MÅSØY	6	1977
	2022 LEBESBY	7	1797
	2023 GAMVIK	7	1482
	2024 BERLEVÅG	8	1467
	2028 BATSFJORD	8	2551
			34389

STRATUM NR =93

	KOMMUNE	UTVOMR	ANT INNB
*	1805 NARVIK	1	18632
	1824 VEFSN	2	13156
			31788

STRATUM NR =94

	KOMMUNE	UTVOMR	ANT INNB
*	1820 ALSTAHAUG	1	7550
	1837 MELØY	2	6970
	1840 SALTDAL	3	5226
	1841 FAUSKE	4	10073
			29819

STRATUM NR =95

	KOMMUNE	UTVOMR	ANT INNB
*	1865 VÅGAN	1	9449
	1866 HADSEL	2	8733
	1870 SORTLAND	3	8192
			26374

STRATUM NR =96

	KOMMUNE	UTVOMR	ANT INNB
*	1901 HARSTAD	1	22101
	1931 LENVIK	2	11130
			33231

STRATUM NR =98

	KOMMUNE	UTVOMR	ANT INNB
	2001 HAMMERFEST	1	6969
*	2002 VARDØ	2	3136
	2003 VADSØ	3	5853
	2019 NORDKAPP	4	4222
			20180

STRATUM NR =99

	KOMMUNE	UTVOMR	ANT INNB
	2012 ALTA	1	14106
*	2030 SØRVARANGER	2	9730
			23836

STRATUM NR =100

	KOMMUNE	UTVOMR	ANT INNB
	1845 SØRFOLD	1	2973
	1849 HAMARØY	1	2371
*	1850 TYSFJORD	2	2704
*	1854 BALLANGEN	2	3188
	1851 LØDINGEN	3	2928
	1852 TJELDSUND	3	1853
	1853 EVENES	3	1797
	1871 ANDØY	4	6773
			24587

STRATUM NR =101

	KOMMUNE	UTVOMR	ANT INNB
*	1911 KVÆFJORD	1	3687
	1913 SKÅNLAND	2	3543
	1919 GRATANGEN	2	1555
	1920 LAVANGEN	3	1161
	1923 SALANGEN	3	2520
	1926 DYRØY	3	1682
	1922 BARDU	4	3971
	1924 MALSELV	5	7392
	1925 SØRREISA	5	3418
			28929

STRATUM NR =102

	KOMMUNE	UTVOMR	ANT INNB
	1933 BALSFJORD	1	6671
	1939 STORFJORD	2	1852
	1940 KÅFJORD	2	2911
	1942 NORDREISA	3	4699
	2011 KAUTOKEINO	4	2893
	2021 KARASJOK	4	2656
	2017 KVALSUND	5	1485
	2020 PORSANGER	5	4414
*	2025 TANA	6	3290
*	2027 NESSEBY	6	992
			31863
			353316

* Utvalgsområder trukket i 1975

Alle kommuner med flere enn 30 000 innbyggere ble definert som egne strata, og disse kommuner er derfor med i alle undersøkelser. Innen de øvrige strata ble ett psu trukket i første trinn med sannsynlighet proposjonal med antall innbyggere i 1974. De derved uttrukne kommunene har vært med i samtlige undersøkelser siden 1975.

Tabell 1 viser hvordan Byråets utvalgsplan for Nord-Norge så ut før revisjonen. Vi ser hvordan landsdelens kommuner (unntatt Rana, Bodø og Tromsø som er selvrepresentative) fordeler seg på de forskjellige strataene. Tabellen viser også trekkingen i første trinn før revisjonen.

3. Revisjon av utvalgsplanen

3.1 Hva vil vi forandre?

Den reviderte utvalgsplanen bygger på de samme grunnforutsetningene som den tidligere. Derfor er inndelingen i landsdeler, regioner og antall strata uforandret. Til tross for forandringer i befolkningsstrukturen i kommunene beholder vi også de tidligere utvalgsområdene. Dessuten lar vi fremdeles de største kommunene være selvrepresentative.

Det er to ting vi vil forandre. Den ene er stratumgrensene. Gjennom å bruke et program for multivariabel clusteranalyse vil vi finne frem til nye strata med så god homogenitet som er mulig. Den andre forandringen gjelder trekkingen i første trinn, hvor vi vil bruke sannsynligheter som er basert på ferske befolkningstall.*)

I forbindelse med disse forandringer er det to interesser som står i

motsetning til hverandre. På den ene siden vil vi fremdeles lage en selvveiende utvalgsplan med liten utvalgsvarians. På den andre siden er det blitt uttrykt ønsker fra Intervjukontoret om å få beholde i det minste noen av de primære utvalgsområdene som er i bruk. En av grunnene for dette er de administrative og økonomiske konsekvenser av å måtte skifte ut et flertall av intervjuerne. Ønsker om kontinuitet i mer eller mindre løpende undersøkelser drar også i retning av en "gradvis" utskiftning av utvalgsområder.

---

*) Vi har brukt innbyggerantall pr. 1.1.1987, foreløpige tall til revisjonen

3.2 Stratifisering

3.2.1 Metode

Før vi lager stratifiseringen er det to ting som vi må tenke på. For det første vil vi ha en lav stratumvarians. Vi definerer:

M_h = antallet psu-er i stratum h

H_{hj} = antallet innbyggere i stratum h, psu j.

$$N = \sum_{h=1}^{L} N_h = \sum_{h=1}^{L} \sum_{j=1}^{M_h} N_{hj}$$

hvor $h = \{1, \ldots, L\}$

og $j = \{1, \ldots, M_h\}$

n_{hj} = utvalgsstørrelsen i stratum h, psu j

$$n = \sum_{h=1}^{L} n_h = \sum_{h=1}^{L} \sum_{j=1}^{M_h} n_{hj}$$

Bare den n_{hj} som svarer til den uttrukne psu blir positiv

Y_{hij} = et observert verdi for individ i psu j i stratum h

hvor $i = \{1, \ldots, N_{hj}\}$

$$\bar{Y}_{hj} = \frac{1}{N_{hj}} \sum_{i=1}^{N_{hj}} Y_{hij}$$

$$\bar{Y}_h = \frac{1}{N_h} \sum_{j=1}^{M_h} N_{hj} \bar{Y}_{hj} = \frac{1}{N_h} \sum_{j=1}^{M_h} \sum_{i=1}^{N_{hj}} Y_{hij}$$

$$S_{hj}^2 = \frac{\sum_{i=1}^{N_{hj}} (Y_{hij} - \bar{Y}_{hj})^2}{N_{hj} - 1}$$

$$\hat{\bar{Y}}_h = \sum_{j=1}^{M_h} I_{hj} \quad \bar{Y}_{hj}$$

hvor $I_{hj} = \begin{cases} 1 \text{ hvis psu h trekkes i første trinn} \\ 0 \text{ ellers} \end{cases}$

det vil si, I_{hj} er en stokatisk variabel for trekking i første trinn, slik at $P\ (I_{hj} = 1) = \frac{N_{hj}}{N_h}$ = trekksannsynligheten

$$\hat{\bar{Y}}_{hj} = \frac{1}{n_{hj}} \sum_{i=1}^{n_{hj}} Y_{hij}$$

Variansen for estimatoren $\hat{\bar{Y}}$ blir

$$V\ (\hat{\bar{Y}}_h) = \frac{1}{N_h} \sum_{j=1}^{M_h} N_{hj}\ (\bar{Y}_{hj} - \bar{Y}_h)^2 + \frac{1}{N_h} \sum_{j=1}^{M_h} (N_{hj} - n_{hj})\ \frac{1}{n_{hj}}\ S^2_{hj}$$

$$= V_h^{\text{mellom psu-er}} + V_h^{\text{innenfor psu-er}}$$

således få vi den totale variansen

$$V\ (\hat{\bar{Y}}) = \frac{1}{N^2} \sum_{h=1}^{L} N_h^2\ V\ (\hat{\bar{Y}}_h)$$

Ettersom psu-ene, og derved også $V_h^{\text{innenfor psu-er}}$, er blitt gitt på forhånd, bortser vi fra disse og definerer

$$V^*\ (\hat{\bar{Y}}) = \frac{1}{N^2} \sum_{h=1}^{L} N_h^2\ V_h^{\text{mellom psu-er}} = \frac{1}{N^2} \sum_{h=1}^{L} N_h\ \Sigma\ (\hat{\bar{Y}}_{hj} - \bar{Y}_h)^2$$

Dette uttrykk vil vi altså minimere.

Det andre vi må tenke på er at vi bør trekke omtrent like mange observasjoner fra hvert stratum i annet trinn. Dette vil gi oss store økonomiske og administrative fordeler når vi skal fordele intervjuerne over de uttrukne utvalgsområdene.

Totalt sett er det mest formåletjenlig å definere strataene slik at vi får en situasjon hvor utvalgstrekkingen i annet trinn med like allokering blir identisk med tilsvarende trekking med proposjonal allokering. Vi ønsker altså en situasjon hvor

$$n_h = n \frac{N_h}{N} = \frac{n}{L}$$

Dette oppnår vi gjennom å la N_h være like over alle strata. Derved har vi definert to restriksjoner som er relevante for stratifiseringen: *)

1. Strataene må være like i størrelse.
2. Strataene må være mest mulig homogene.

---

*) Det bør sies at disse to restriksjoner ikke kan bli tilfredsstilt fullt ut. På den andre siden mener vi å ha funnet en stratifiseringsplan som gjør det rimelig å bruke proporsjonal allokering med hensyn til ovennevnte kriterier.

For å få fram homogene strata har vi valgt å stratifisere i to trinn. Vi har først brukt et program for multivariabel klusteranalyse for å få en kvantitativ gruppering av utvalgsområdene, hvoretter vi har "justert" stratifiseringen på en mere kvalitativ eller "erfaringsmessig" grunn.

Kvantitativ metode

I første trinn har vi brukt klusteranalyse med Average-Linkage-metoden. Denne metode gir oss en rekke av utvalgsområder ordnet etter noen variable som er blitt definert på forhånd. Metodens algoritme blir presentert i appendiks A.

For å oppfylle formålet med stratifiseringen er det ønskelig først å finne frem til et antall variable som forventes å ha høy korrelasjon med de relevante spørsmålene i kommende utvalgsundersøkelser. Ettersom den nye generelle utvalgsplanen skal brukes til varierende undersøkelser av samfunnsvitenskaplig natur kan det være rimelig å se på næringsstruktur og liknende i utvalgsområdene.

Vi har valgt å bruke andelen yrkesaktive i fem forskjellige næringsgrupper fra folke- og boligtellingen 1980. Disse næringsgruppene er:

1. Jordbruk og skogbruk
2. Fiske og fangst
3. Industri
4. Bank- og finansieringsvirksomhet, forsikringsvirksomhet o.s.v.
5. Offentlig og privat tjenesteyting

Dessuten har vi tatt med graden av sentralitet, som er blitt definert som en dikotom variabel på kommunenivå med verdien 1 hvis kommunen hører til tettsteder på nivå 0 eller høyre eller hvis reisetiden til en slik kommune er mindre enn 45 minutter, og verdien 0 ellers. (Statistisk Sentralbyrå, 1985).

For Nord-Norges del er vi interessert i å lage 12 nye strata, seks for Nordland og seks for Troms og Finnmark. Utvalgsområdene Rana, Bodø og Tromsø, som fremdeles vil være selvrepresentative, er ikke tatt med.

Gjennom å lage et trediagram som beskrives klusteranalysens "klynge-struktur" på forskjellige nivåer får vi et godt overblikk over hvilke utvalgsområder som er omlag like med hensyn til stratifiserings-variablene. Figur 1 viser et slikt diagram. Klusteranalyseresultatene for Nord-Norges to regioner er her blitt presentert for nivåene 3, 6, 9, 12 og 18 klynger.

Figur 1 viser også hvordan vi har laget en "foreløpig" stratuminndel-ing. Vi har her tatt hensyn til at stratene må være like med hensyn til størrelse (tallene innenfor parentes gir antallet innbyggere). Samtidig har vi i størst mulig utstrekning stratifisert med hensyn til de klyngene som er mest like. Derved har vi funnet fram til de strata som blir presentert i tabell 2.

"Kvalitativ modifisering"

Informasjonen som stratifiseringvariablene gir oss er ikke helt full-stendig. Det finnes kunnskap om befolkningens struktur, f.eks. lokale tradisjoner, seder og bruk o.s.v. som det ikke foreligger data for, eller som er svært vanskelig å kvantifisere.

Her må vi bruke informasjon på kvalitativ nivå, hentet inn fra "eksperter" med omfattende kjennskap og erfaring om de berørte kommunene/utvalgsområdene. *)

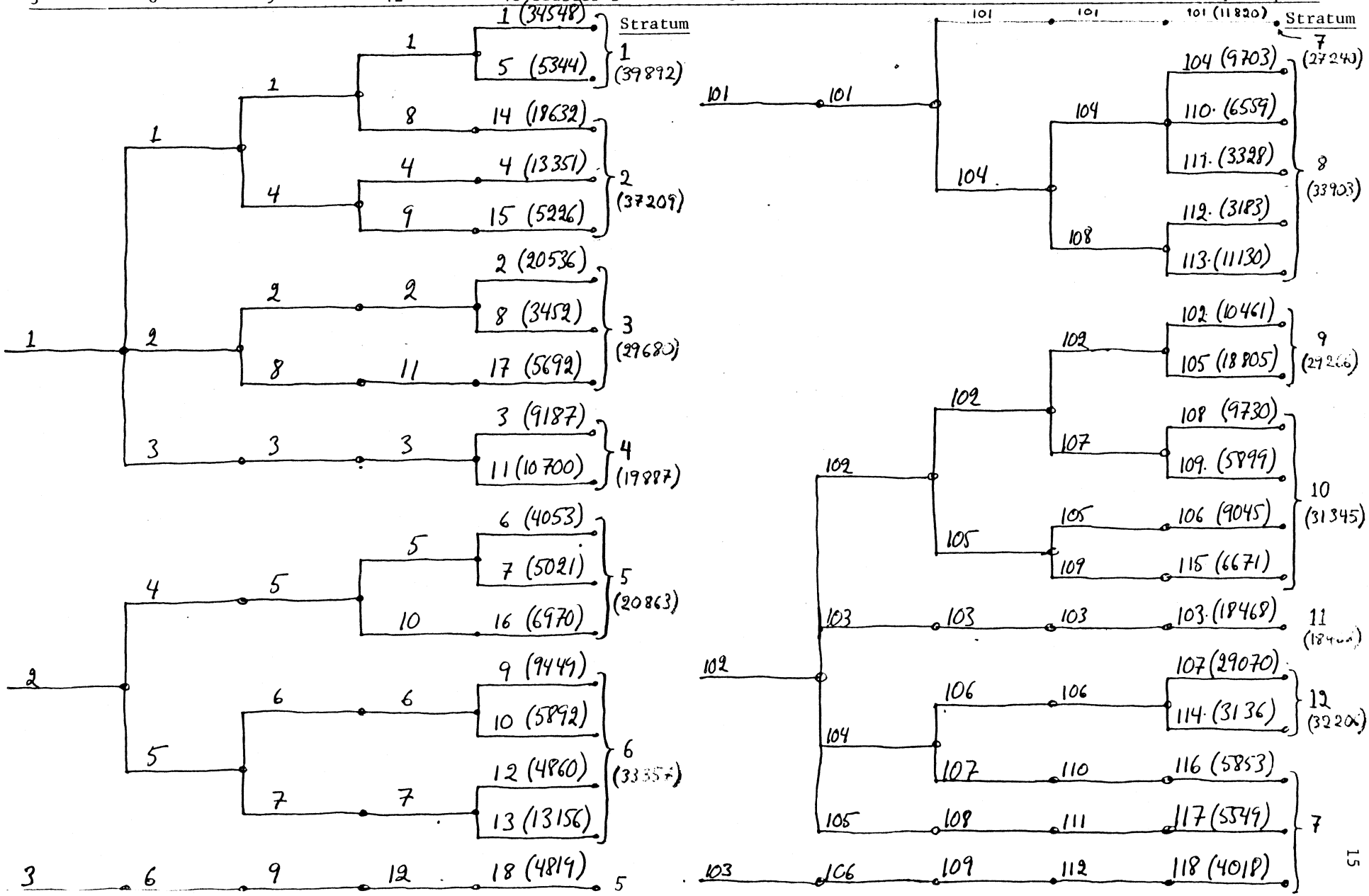
Tabell 3 viser hvordan Nord-Norges kommuner (unntatt Rana, Bodø og Tromsø) fordeler seg på strataene etter den "geografiske modifiseringen". Hvis vi sammenligner med tabell 2, som viser tilsvarende fordeling før denne "finjustering", ser vi at følgende endringer er blitt gjort:

---

*) Lars Østby, Seksjon for sosiodemografi, har vært behjelpelig i dette sammenheng og takkes herved.

1820 Alstahaug	er blitt flyttet fra stratum				1	til stratum			6
1845 Sørfold/									
1849 Hamarøy	"	"	"	"	"	1	"	"	3
1822 Leirfjord/									
1827 Dønna/									
1828 Nesna	"	"	"	"	"	3	"	"	4
1860 Vestvågøy	"	"	"	"	"	4	"	"	5
1837 Meløy	"	"	"	"	"	5	"	"	4
1865 Vågan	"	"	"	"	"	6	"	"	1
2011 Kautokeino/									
2021 Karasjok	"	"	"	"	"	7	"	"	10
2003 Vadsø	"	"	"	"	"	7	"	"	12
2014 Loppa/									
2015 Hasvik	"	"	"	"	"	8	"	"	7
1913 Skånland/									
1919 Gratangen	"	"	"	"	"	9	"	"	11

FIGUR 1.


NORDLAND
3
6
9
12
18 cluster's
Stratum
1 (34548)
5 (5344)
14 (18632)
4 (13351)
15 (5296)
2 (20536)
8 (3459)
17 (5692)
3 (9187)
11 (10700)
6 (4053)
7 (5021)
16 (6970)
9 (9449)
10 (5892)
12 (4860)
13 (13156)
18 (4819)
1 (39892)
2 (37209)
3 (29680)
4 (19887)
5 (20863)
6 (33354)
5
TROMS og FINNMARK
101 (11820)
104 (9703)
110 (6559)
111 (3328)
112 (3183)
113 (11130)
102 (10461)
105 (18805)
108 (9730)
109 (5899)
106 (9045)
115 (6671)
103 (18468)
107 (29070)
114 (3136)
116 (5853)
117 (5349)
118 (4018)
7 (27240)
8 (33903)
9 (27266)
10 (31345)
11 (18468)
12 (32206)
7

TABELL 2.

STRATUMINDELNING FØR GEOGRAFISK MODIFISEING

STRATUM NR =1

KOMMUNE	UTVOMR	ANT INNB
1820 ALSTAHAUG	1	7550
1845 SØRFOLD	1A	2973
1849 HAMARØY	1A	2371
1866 HADSEL	2	8733
1870 SORTLAND	3	8192
1841 FAUSKE	4	10073
		39892

STRATUM NR =2

KOMMUNE	UTVOMR	ANT INNB
1805 NARVIK	1	18632
1840 SALTDAL	3	5226
1851 LØDINGEN	3A	2928
1852 TJELDSUND	3A	1853
1853 EVENES	3A	1797
1871 ANDØY	4	6773
		37209

STRATUM NR =3

KOMMUNE	UTVOMR	ANT INNB
1811 BINDAL	1	2167
1825 GRANE	1	1766
1826 HATTFJELLDAL	1	1759
1812 SØMNA	2	2158
1813 BRØNNØY	2	6826
1822 LEIRFJORD	3	2357
1827 DØNNA	3	1939
1828 NESNA	3	1775
1838 GILDESKÅL	5	2666
1839 BEIARN	5	1546
1842 SKJERSTAD	5	1269
1848 STEIGEN	6	3452
		29680

STRATUM NR =4

KOMMUNE	UTVOMR	ANT INNB
1815 VEGA	1	1608
1816 VEVELSTAD	1	730
1818 HERØY	1	2102
1834 LURØY	2	2410
1835 TRÆNA	2	530
1836 RØDØY	2	1807
1860 VESTVÅGØY	4	10700
		19887

STRATUM NR =5

KOMMUNE	UTVOMR	ANT INNB
1837 MELØY	2	6970
1856 RØST	3	704
1857 VÆRØY	3	937
1859 FLAKSTAD	3	1664
1874 MOSKENES	3	1514
1867 BØ	5	4053
1868 ØKSNES	6	5021
		20863

STRATUM NR =6

KOMMUNE	UTVOMR	ANT INNB
1865 VÅGAN	1	9449
1824 VEFSN	2	13156
1850 TYSFJORD	2A	2704
1854 BALLANGEN	2A	3188
1832 HEMNES	4	4860
		33357

STRATUM NR =7

KOMMUNE	UTVOMR	ANT INNB
2003 VADSØ	3	5853
2019 NORDKAPP	4	4222
2011 KAUTOKEINO	4A	2893
2021 KARASJOK	4A	2656
2016 SØRØYSUND	6	2342
2018 MASØY	6	1977
2022 LEBESBY	7	1797
2023 GAMVIK	7	1482
2024 BERLEVÅG	8	1467
2028 BÅTSFJORD	8	2551
		27240

STRATUM NR =8

KOMMUNE	UTVOMR	ANT INNB
1915 BJARKØY	1	788
1917 IBESTAD	1	2395
1927 TRANØY	2	2111
1928 TORSKEN	2	1408
1929 BERG	2	1288
1931 LENVIK	2A	11130
1936 KARLSØY	3	2882
1938 LYNGEN	3	3677
1941 SKJERVØY	4	3201
1943 KVÆNANGEN	4	1695
2014 LOPPA	5	1866
2015 HASVIK	5	1462
		33903

STRATUM NR =9

KOMMUNE	UTVOMR	ANT INNB
2012 ALTA	1	14106
1913 SKÅNLAND	2	3543
1919 GRATANGEN	2	1555
1920 LAVANGEN	3	1161
1923 SALANGEN	3	2520
1926 DYRØY	3	1682
1942 NORDREISA	3A	4699
		29266

STRATUM NR =10

KOMMUNE	UTVOMR	ANT INNB
1933 BALSFJORD	1	6671
2030 SØRVARANGER	2	9730
1939 STORFJORD	2A	1852
1940 KÅFJORD	2A	2911
2017 KVALSUND	5	1485
2020 PORSANGER	5	4414
2025 TANA	6	3290
2027 NESSEBY	6	992
		31345

STRATUM NR =11

KOMMUNE	UTVOMR	ANT INNB
1911 KVÆFJORD	1	3687
1922 BARDU	4	3971
1924 MÅLSELV	5	7392
1925 SØRREISA	5	3418
		18468

STRATUM NR =12

KOMMUNE	UTVOMR	ANT INNB
1901 HARSTAD	1	22101
2001 HAMMERFEST	1A	6969
2002 VARDØ	2	3136
		32206
		353316

TABELL 3.

STRATUMINDELNING ETTER GEOGRAFISK MODIFISEING

STRATUM NR =1

KOMMUNE	UTVOMR	ANT INNB
*1865 VÅGAN	1	9449
1866 HADSEL	2	8733
1870 SORTLAND	3	8192
1841 FAUSKE	4	10073
		36447

STRATUM NR =2

KOMMUNE	UTVOMR	ANT INNB
*1805 NARVIK	1	18632
1840 SALTDAL	3	5226
1851 LØDINGEN	3	2928
1852 TJELDSUND	3	1853
1853 EVENES	3	1797
1871 ANDØY	4	6773
		37209

STRATUM NR =3

KOMMUNE	UTVOMR	ANT INNB
1845 SØRFOLD	1	2973
1849 HAMARØY	1	2371
1811 BINDAL	1A	2167
1825 GRANE	1A	1766
1826 HATTFJELLDAL	1A	1759
*1812 SØMNA	2	2158
*1813 BRØNNØY	2	6826
1838 GILDESKÅL	5	2666
1839 BEIARN	5	1546
1842 SKJERSTAD	5	1269
1848 STEIGEN	6	3452
		28953

STRATUM NR =4

KOMMUNE	UTVOMR	ANT INNB
*1815 VEGA	1	1608
*1816 VEVELSTAD	1	730
*1818 HERØY	1	2102
1834 LURØY	2	2410
1835 TRÆNA	2	530
1836 RØDØY	2	1807
1837 MELØY	2A	6970
1822 LEIRFJORD	3	2357
1827 DØNNA	3	1939
1828 NESNA	3	1775
		22228

STRATUM NR =5

KOMMUNE	UTVOMR	ANT INNB
1856 RØST	3	704
1857 VÆRØY	3	937
1859 FLAKSTAD	3	1664
1874 MOSKENES	3	1514
*1860 VESTVÅGØY	4	10700
1867 BØ	5	4053
1868 ØKSNES	6	5021
		24593

STRATUM NR =6

KOMMUNE	UTVOMR	ANT INNB
1820 ALSTAHAUG	1	7550
1824 VEFSN	2	13156
1850 TYSFJORD	2A	2704
1854 BALLANGEN	2A	3188
*1832 HEMNES	4	4860
		31458

STRATUM NR =7

KOMMUNE	UTVOMR	ANT INNB
*2019 NORDKAPP	4	4222
2014 LOPPA	5	1866
2015 HASVIK	5	1462
2016 SØRØYSUND	6	2342
2018 MÅSØY	6	1977
2022 LEBESBY	7	1797
2023 GAMVIK	7	1482
2024 BERLEVÅG	8	1467
2028 BATSFJORD	8	2551
		19166

STRATUM NR =8

KOMMUNE	UTVOMR	ANT INNB
1915 BJARKØY	1	788
1917 IBESTAD	1	2395
1927 TRANØY	2	2111
1928 TORSKEN	2	1408
1929 BERG	2	1288
1931 LENVIK	2A	11130
*1936 KARLSØY	3	2882
*1938 LYNGEN	3	3677
1941 SKJERVØY	4	3201
1943 KVÆNANGEN	4	1695
		30575

STRATUM NR =9

KOMMUNE	UTVOMR	ANT INNB
*2012 ALTA	1	14106
1920 LAVANGEN	3	1161
1923 SALANGEN	3	2520
1926 DYRØY	3	1682
1942 NORDREISA	3A	4699
		24168

STRATUM NR =10

KOMMUNE	UTVOMR	ANT INNB
1933 BALSFJORD	1	6671
*2030 SØRVARANGER	2	9730
1939 STORFJORD	2A	1852
1940 KÅFJORD	2A	2911
2011 KAUTOKEINO	4	2893
2021 KARASJOK	4	2656
2017 KVALSUND	5	1485
2020 PORSANGER	5	4414
2025 TANA	6	3290
2027 NESSEBY	6	992
		36894

STRATUM NR =11

KOMMUNE	UTVOMR	ANT INNB
1911 KVÆFJORD	1	3687
*1913 SKÅNLAND	2	3543
*1919 GRATANGEN	2	1555
1922 BARDU	4	3971
1924 MÅLSELV	5	7392
1925 SØRREISA	5	3418
		23566

STRATUM NR =12

KOMMUNE	UTVOMR	ANT INNB
*1901 HARSTAD	1	22101
2001 HAMMERFEST	1A	6969
2002 VARDØ	2	3136
2003 VADSØ	3	5853
		38059
		353316

*Utvalgsområder trukket etter revisjonen

3.2.2. En vurdering av den nye stratuminndelingen.

For å vurdere "kvaliteten" til den nye stratuminndelingen har vi laget noen beregninger på grunnlag av variansen mellom psu-er. Vi har brukt uttrykket

$$V^*(\bar{\hat{Y}}) = \frac{1}{N^2} \sum_{h=1}^{L} N_h \sum (\bar{Y}_{hj} - \bar{Y}_h)^2$$

Dette variansmål ble presentert i avsnitt 3.2.1.

Resultatene av våre variansberegninger blir presentert i tabell 4. Vi har brukt de fem kontinuerlige variablene for andel ansatt innen respektive landbruk, fiske, industri, finans og privat eller offentlig tjenesteyting. Variansen for landbruk, fiske og finans har økt noe, mens industri og tjenesteyting oppviser en kraftig reduksjon. Dette impliserer en forbedring av stratuminndelingen.

En sammenligning mellom stratuminndelingen før og etter den "geografiske modifiseringen" viser forholdsvis små forskjeller. Varianstallene for landbruk og industri har etter modifiseringen gått opp noe, mens tilsvarende tall for fiske, finans og tjenesteyting er blitt lavere. Dette er interessant å notere seg ettersom vi har forventet oss en større økning av stratumvariansene på grunn av modifiseringen.

3.3 Trekking i første trinn.

I forbindelse med trekkingen av utvalgsområder i første trinn ønsker vi at så mange av de gamle psu-ene som er mulig blir "gjenvalgt", men nå med sannsynligheter proposjonalt med antall innbyggere i 1987. Leslie Kish og Alastair Scott har beskrevet en generell metode for

TABELL 4.

EN SAMMENLIKNING AV STRATUMVARIANSER FOR DE KVANTITATIVE CLUSTERANALYSE-VARIABLENE.

	Gammel stratifisering	Ny strat. før den geogr. mod.	Ny strat. etter den geogr. mod.
VAR(landbruk)	1,667	1,941	1,986
VAR(fiske)	1,190	1,398	1,215
VAR(industri)	3,167	1,724	1,902
VAR(finans)	0,031	0,055	0,038
VAR(tjeneste)	2,039	1,414	1,401

trekking i første trinn, hvor en går ut fra den betingede trekksannsynligheten Pr (trekksannsynligheten etter revisjonen|trekksannsynlighet for revisjonen). Metoden maksimerer antallet gjenvalg med hensyn til alle tenklige situasjoner som kan oppstå når såvel trekksannsynligheter som stratumgrenser er blitt endret. Algoritmen for Kish & Scott's metode blir presentert i appendiks B (Kish & Scott, 1971).

Resultatene av trekkingen i første trinn etter revisjonen blir presentert i tabell 5. Som fremgår av tabellen er seks utvalgsområder blitt "gjenvalgt". Tabellen presenterer også de nye trekksannsynlighetene og de proposjonale allokeringsvektene med hele landet som basis (hensyn er tatt til selvveiende utvalg).

Allokeringsvektene skal brukes slik, at vi multipliserer dem med utvalgsstørrelsen. Hvis vi f.eks. ønsker å trekke 5 000 individer i hele landet, skal 5 000 * 0,008729 = 44 individer av disse blir trukket fra Vågan.

3.4 Konklusjon

Gjennom å revidere den generelle utvalgsplanen regner vi med å ha redusert kostnadene for kommende intervjuundersøkelser. Først og fremst har restratifiseringen gitt oss et antall nye utvalgsområder som vi mener skal gi oss resultater med høy presisjon, selv om utvalgene er små. Vi har også fått en ny plan for "strategisk" fordeling av intervjuere, slik at intervjukostnadene blir redusert.

Imidlertid finnes det rom for en del forbedringer. Når vi f.eks. trekker i første trinn får vi en betinget skjevhet ved trekking i annet trinn. Goodman & Kish presenterte i 1950 en metode som kan

TABELL 5.

UTVALGSOMRÅDER SOM ER BLITT TRUKKET FRA DEN NYE STRATIFISERINGSPLANEN FOR NORDNORGE.

Nytt stratum-nr	Kommune	Ant innb i utv.omr	Ant innb i strat.	Ubetinget trekke-sanns.	Allo-kerings-vekt[x]	
1	1865 Vågan	9 449	36 447	0,259	0,008729	(gjenvalg)
2	1805 Narvik	18 632	37 209	0,501	0,008912	(gjenvalg)
3	1812 Sømna 1813 Brønnøy	8 984	28 953	0,310	0,006935	
4	1815 Vega 1816 Vevelstad 1818 Herøy	4 440	22 228	0,120	0,005324	
5	1860 Vestvågøy	10 700	24 593	0,435	0,005890	(gjenvalg)
6	1832 Hemnes	4 860	31 458	0,154	0,007535	(gjenvalg)
7	2019 Nordkapp	4 222	19 166	0,220	0,004590	
8	1936 Karlsøy 1938 Lyngen	6 559	30 575	0,215	0,007323	
9	2012 Alta	14 106	24 168	0,584	0,005789	
10	2030 Sør-Varanger	9 730	36 894	0,264	0,008837	(gjenvalg)
11	1913 Skånland 1919 Gratangen	5 089	23 566	0,216	0,005644	
12	1901 Harstad	22 101	38 059	0,581	0,009116	(gjenvalg)
13	1833 Rana[xx]	25 136	25 136	1	0,006020	
14	1804 Bodø	35 024	35 024	1	0,008389	
15	1902 Tromsø	48 857	48 857	1	0,011702	

[x] Proporsjonal allokering, med 4 175 171 innb for hele landet som basis.

[xx] Kommunene Rana, Bodø og Tromsø er, som tidligere, selvrepresentative med trekkesannsynlighet 1.

brukes for å redusere denne skjevhet, s.k. "controlled selection". Metoden bygger i sin enkleste form på at en først grupperes to strata, A_1 og A_2, som er omlag like med hensyn til en bestemt stratifiseringvariabel (som kan være kontinuerlig, ordinal eller dikotom). Deretter rangordner en utvalgsområdene i A_1 i oppgående ledd med hensyn til stratifiseringsvariabelen, mens en gjør tilsvarende med A_2, men i nedgående ledd. Gjennom å beregne de kumulative sannsynlighetene for de rangordnede utvalgsområdene i A_1 og A_2, og deretter trekke en psu fra hvert og et av disse samtidlig, får en så en balansering av utvalgsdesignen (Hess, Riedel & Fitzpatrick, 1961).

Gjennom "controlled selection" beholder vi trekksannsynlighetene i første trinn slik vi ønsker dem, samtidig som altså den betingete skjevheten blir redusert.

4. Litteratur

Cochran, W. G.: "Sampling Techniques", 3rd ed, Wiley 1977.

Dahmström, P., Hagnell, M.: "The Formation of Strata Using Cluster Analysis", Statistisk Tidskrift 1974:6, s. 477-486.

Green; P. E.: "Analyzing Multivariate Data", The Dryden Press 1978.

Haldorsen, T.: "Statistiske egenskaper ved Byråets standard utvalgsplan", Rapporter, 85/34, Statistisk Sentralbyrå 1985.

Hess, I., Riedel, D. C. and Fitzpatrick, T. B.: "Probability Sampling of Hospitals and Patients", The University of Michigan 1961.

Kish, L. and Scott, A.: "Retaining Units after Changing Strata and Probabilities", Journal of the Am. stat. Ass, Sept. 1971, Vol. 66, No. 335.

SAS: "SAS User's Guide: Statistics", version 5 edition, SAS Institute Inc. 1985.

Siring, E.: "En generell metode for endring av en totrinns utvalgsplan", Interne Notater 81/35, 2/12 1981, Statistisk Sentralbyrå 1981.

Statistisk Sentralbyrå: "Standarder for kommuneklassifisering", SNS, nr. 4, Statistisk Sentralbyrå 1985.

Thomsen, I.: "Prinsipper og metoder for Statistisk Sentralbyrås utvalgsundersøkelser", SØS nr. 33, Statistisk Sentralbyrå 1977.

Thomsen, I. og Rideng, A: "Oversikt over arbeidet med ny utvalgsplan", Stensil ITh/ARi/GHu, 21/5-74, Statistisk Sentralbyrå 1974

Appendiks A: Klusteranalysens algoritme.

Klusteranalyse med den s.k. average-linkage-metoden er en teknikk for inndeling av en rekke observasjoner i homogene grupper med hensyn til en eller flere variabler. Algoritmen kan bli beskrevet på følgende måte:

Definere et mål for avstanden d_{KL} mellom to klynger, C_k og C_L.

$$d_{KL} = \frac{1}{N_K N_L} \sum_{i=1}^{N_K} \sum_{j=1}^{N_L} d\,(Y_i, Y_j)$$

N_K = antallet observasjoner i klynge C_k

N_L = antallet observasjoner i klynge C_L

og $d\,(Y_i, Y_j)$ = et mål på avstanden mellom to observasjoner. Ved arbeidet med den nye generelle utvalgsplanen har vi brukt definisjonen

$$d\,(Y_i\ Y_j) = \sum_{p=1}^{P} (Y_{ip} - Y_{jp})^2$$

hvor P = antallet variabler i clusteranalysen.

Begynn å betrakte hver observasjon som en klynge. Se på avstanden d_{KL} (som er lik $d(Y_i, Y_j)$, ettersom N_k og N_L er lik 1) for alle tenklige parvise kombinasjoner av alle de N klyngene (observasjonene) i materialet. Slå sammen de to klynger som er mest like. Sammenlikne d_{KL}-verdiene mellom alle tenklige parvise kombinasjoner for de N-1 "nye" klyngene og slå sammen det "klynge-par" som har lavest verdi i denne rekken. Fortsett med proseduren til alle observasjoner er blitt grupperte i en stor "super-klynge".

Det finnes en rekke alternativer til average-linkage-metoden. De to viktigste er complete-linkage og single-linkage. I den førstnevnte defineres d_{KL} som det maksimale avstandet mellom to enkelte observasjoner i klyngene C_K og C_L, mens den andre på tilsvarende måte måler det minimale avstanden mellom disse to cluster's.

Grunnen til at average-linkage-metoden er mest brukbar i dette sammenheng er at den lager klynger med forholdsvis lav varians.

Appendiks B: Algoritme for trekking i første trinn med Kish & Scott's metode.

En generell metode for trekking i første trinn, hvor en i størst mulig utstrekning ønsker å beholde de opprinnelig utttrukne utvalgsområdene, finnes beskrevet av Kish & Scott i 1971. Metoden går ut fra at det nye stratumet inneholder en rekke gamle stratumdeler, med eller uten enheter (=utvalgsområder) som tidligere er blitt trukket. Før en går inn på selve algoritmen, sett

$\{A_1, A_2, \ldots\ldots, A_k, \ldots\ldots, A_n\}$ = de gamle stratumdelene i det nye stratumet

π_j = den opprinnelige trekksannsynligheten for enhet j, hvor

$j \in A_k$

P_j = den nye trekksannsynligheten for samme enhet

$P_K = \sum_{j \in A_k} P_j$

og $\sum_{k \in n} P_K = 1$

Vi ordner stratumdelene etter størrelse på P_K, slik at

$$P_1 < P_2 < \ldots < P_K < \ldots < P_n$$

Dessuten definierer vi indeksene (for enheter)

$i \varepsilon I_k$

hvor I_k = den delmengde i stratumdel A_K

hvor $\pi_j \leq P_j$

og $d \varepsilon D_k$

hvor, på tilsvarende måte,

D_k = den delmengde i stratumdel A_k

hvor $\pi_j > P_j$

hvorved gjelder, at

$I_k \cup D_k = A_k$ og $I_k \cap D_k = \emptyset$

Bermerk at indeksene i, j og d betegner en enhet, mens k betegner en stratumdel!

Algoritmen for trekking ifølge Kish & Scott's metode kan beskrives i fire trinn:

Trinn 1:

Se om stratumdel A_1 inneholder en enhet som opprinnelig er blitt trukket. Hvis denne enhet finnes i delmengde I_1, så er den blitt gjenvalgt. Ellers, hvis enheten finnes i D_1, så trekk med sannsynligheten P_d/π_d for "gjenvalg".

Trinn 2:

Hvis ikke noen enhet er blitt gjenvalgt fra A_1, eller hvis denne stratumdel ikke inneholder noen opprinnelig trukket enhet, se om stratumdel A_2 inneholder en slik enhet. La

$$r_{2_j} = \pi_j Q_1$$

hvor $Q_1 = (1 - \sum_{i \varepsilon I_1} \pi_i - \sum_{d \varepsilon D_1} P_d)$

= sannsynligheten for at ikke noen enhet i stratumdel A_1 ble trukket

Lage en ny definisjon av delmengdene I_k og D_k. La

I_2 = den delmengde i stratumdel A_2, hvor $r_{2j} \leq P_j$

og D_2 = den delmengde i stratumdel A_2, hvor

$$r_{2j} > P_j$$

Hvis enheten finnes i delmengde I_2, så er den blitt gjenvalgt. Ellers hvis enheten finnes i D_2, så trekk med sannsynligheten P_d/r_{2d} for "gjenvalg".

Trinn 3:

Hvis ikke noen enhet er blitt gjenvalgt fra $\{A_1, \ldots, A_{k-1}\}$, eller hvis disse stratumdeler ikke inneholder noen opprinnelig trukket enhet, se om stratumdel A_k inneholder en slik enhet. La

$$r_{kj} = \pi_j \, Q_{k-1}$$

hvor $Q_{k-1} = (Q_{k-2} - \sum_{i \in I_{k-1}} r_{(k-1)i} - \sum_{d \in D_{k-1}} P_d) =$

= sannsynligheten for at ikke noen enhet i $\{A_1, \ldots, A_{k-1}\}$ ble trukket

Lage en ny definisjon av delmengdene I_k og D_k etter samme prinsipp som i trinn 2, og bruk samme prosedyre for "gjenvalg" (la sannsynligheten for gjenvalg, gitt at enheten finnes i D_k, være P_d/r_{kd}).

Trinn 4:

Hvis trinn 1 - trinn 3 ikke har resultert i noen trekking fra stratumdelene $\{A_1, A_2, \ldots, A_n\}$, så beregne

$$P(k,i \mid \text{den gamle enheten "ikke gjenvalgt"}) = \frac{P_i - r_{ki}}{\sum_{k \in n} \sum_{i \in I_k} (P_i - r_{ki})}$$

hvor $r_{ki} = \pi_i$ hvis $k=1$

og $\{I_1, I_2, \ldots, I_k, I_n\}$ er de delmengder som er blitt definert i trinn 1 - trinn 3.

Trekk deretter en enhet med denne betingede trekksannsynlighet.