

Interne notater

STATISTISK SENTRALBYRÅ

85/4

20. februar 1985

BRUK AV UTVALG TIL VEIING AV TABELLER I FORBINDELSE MED TILLEGGS-
UNDERSØKELSEN TIL FOLKE- OG BOLIGTELLINGEN 1980

Av

Anders Rygh Swensen

INNHOLD

	Side
1. Innledning	2
2. Tallet på yrkesaktive	4
3. Fordeling på arbeidstid	9
4. Fordeling på næringer	11
5. Appendiks	14

FORORD

Dette notatet er utgitt som en del av prosjektet om kombinasjon av data fra registre og utvalg. Professor Emil Spjøtvoll, som er konsulent for prosjektet, har bidratt med nyttige kommentarer, og takkes herved. Det samme gjelder Sigrun Kaland, som har hjulpet til med en del av beregningsarbeidet.

1. INNLEDNING

Dette er et notat i forbindelse med et prosjekt om kombinasjon av opplysninger fra registre og utvalgsundersøkelser. Hensikten er å anvende noen enkle stratifiseringsmetoder på data for yrkesaktivitet og arbeidstid hentet fra Folke og Boligtellinga i 1980 (heretter FoB) og Tilleggsundersøkelsen til FoB (heretter TU). TU er et forsøk på å rekonstruere deler av resultatene i FoB med utgangspunkt i data fra ulike registre. Resultatene fra prosjektet er rapportert i Børke (1984).

Poenget med en undersøkelse av type TU er at en slik metode er adskillig billigere enn en klassisk folketelling med intervju eller utsending av skjemaer. Imidlertid er ofte data hentet fra ulike administrative registre ikke av en slik kvalitet at de uten videre kan danne grunnlaget for offisiell statistikk.

I forbindelse med TU ble det derfor laget en fil med opplysninger for et 10% utvalg av befolkningen. For hver person i dette utvalget er det mulig å sammenligne registerbaserte opplysninger med tilsvarende opplysninger fra FoB. Ved å lage tabeller basert på opplysninger fra de to delene av fila kan man derfor sammenligne de to mulighetene. Dette er nærmere beskrevet i Børke (1984), der også tilleggsundersøkelsesfila er dokumentert i detalj.

Vi skal i dette notatet presentere en enkel metode som foretar en veiing av tabeller basert på registeropplysninger. Veiingen blir foretatt med utgangspunkt i hvordan et sampel på ca. 10 000 personer hentet fra 10% utvalget er klassifisert i FoB og i registerdelen. De veide tabellene vil deretter bli sammenlignet med tilsvarende tabeller enten basert på FoB delen eller på registerdelen. På den måten kan man få et inntrykk om uoverensstemmelsene mellom de FoB baserte tabellene og de registerbaserte tabellene kan reduseres og hvor store reduksjoner en kan vente.

I den grad resultatet vurderes som vellykket kan man derfor tenke seg en framtidig folketelling som bestående av i hovedsak en registerdel, supplert med en FoB lignende undersøkelse for en del av befolkningen. Hensikten med den siste vil være å danne utgangspunkt for veiing av tabeller fra registerdelen og på den måten redusere eller eliminere feil slike tabeller er beheftet med uten noen form for etterbearbeiding.

Vi skal nå kort beskrive hovedideen bak veiingen, og illustrere hvordan den foretas for en situasjon der variabelen en er interessert i kan anta verdien A eller \bar{A} , f. eks. yrkesaktiv eller ikke

yrkesaktiv. Hensikten er å estimere hvor mange som har kjennetegn A i en befolkning på N personer. Fra registeret får en opplysning at N har kjennetegn A, men dette tallet er beheftet med feil. For et sampel på n personer undersøker en derfor nøyaktig hvor mange som har kjennetegnet A, og undersøker samtidig hvordan disse n personene er klassifisert i registeret. Resultatet for samplet kan sammenfattes i følgende 2x2 tabell:

		Register		
		\bar{A}	A	
Sample	\bar{A}	$n_{\bar{A}\bar{A}}$	$n_{\bar{A}A}$	$n_{\bar{A}}$
	A	$n_{A\bar{A}}$	n_{AA}	n_A
		$n_{\cdot\bar{A}}$	$n_{\cdot A}$	n

Resultatet fra tellingen i registeret kan sammenfattes i 1x2 tabellen:

			Register	
			\bar{A}	A
$N_{\bar{A}}$	N_A	N		

Siden de registerbaserte opplysningene er mangelfulle vil N_A være en dårlig estimator for det totale antall som har kjennetegn A. Fra 2x2 tabellen får man et inntrykk av hvordan skjevheten er og

$$\frac{N_{\bar{A}\bar{A}}}{n_{\cdot\bar{A}}} N_{\bar{A}} + \frac{n_{AA}}{n_{\cdot A}} N_A$$

vil være et bedre anslag for hvor mange som har kjennetegnet A.

Det en gjør her er å estimere andelen med kjennetegn A som klassifiseres uten kjennetegn A i registeret som $\frac{n_{\bar{A}\bar{A}}}{n_{\cdot\bar{A}}} N_A$. Tilsvarende blir antallet som klassifiseres med kjennetegn A i registeret som virkelig har det, estimert som $\frac{n_{AA}}{n_{\cdot A}} N_A$.

I tillegg til estimatet for det totale antall med kjennetegn A er det relativt enkelt å bestemme standardavviket til estimatoren ovenfor og på den måten få et inntrykk av de tilfeldige variasjonene når en

varierer samplet. Dette er nærmere beskrevet i appendikset til notatet.

Vi skal nå beskrive datagrunnlaget noe mer i detalj. Hvordan de registerbaserte opplysningene i 10% utvalget av befolkningen på TU fila er framkommet, er beskrevet i Børke (1984). Samplet som brukes ved veiingen er nettoutvalget, d. v. s. de som har svart i en arbeidskraftsundersøkelse (AKU) fra 1980. Det eneste som er hentet herfra er imidlertid utvalgsmetoden. Ingen av AKU opplysningene er benyttet. Derimot vil valget av et slikt sample medføre at bare personer mellom 16 og 74 år vil bli inkludert.

Dessverre ser det ut til at resultatet er blitt et sample som er skeivt i forhold til 10% utvalget på TU fila. Hvis man for eksempel betrakter antallet som er klassifisert som ikke yrkesaktiv i registerdelen, er det henholdsvis 29% og 33% for henholdsvis samplet og hele 10% utvalget. Dette er en meget sterk skjevhet, som nok forklarer mange av uoverensstemmelsene mellom FoB baserte og veide tabeller i det etterfølgende. Når resultatet likevel er en forbedring i forhold til rene registerbaserte tabeller, må det tas som et uttrykk for at det kan være adskillig større gevinst med et bedre sample.

2. TALLET PÅ YRKESAKTIVE

Som yrkesaktive anses personer som har et eller flere arbeidsforhold som til sammen varer minst 100 timer i året. For registerdelen er opplysningene hentet fra felt 37 på TU fila, for FoB delen fra felt 72 og 77.

Veiingen er foretatt for hele landet, for hele landet oppdelt etter kjønn, for hvert fylke, innen hvert fylke oppdelt etter kjønn, for hver kommunetype og innen hver kommunetype oppdelt etter kjønn.

Resultatene er gjengitt i tabell 2 som kan sammenlignes med tabell 7 i Børke (1984). Kolonne 1 er hentet fra Børke (1984) og er identisk med de publiserte FoB tallene, jfr NOS B299 1980. Kolonne 2 er de tilsvarende registerbaserte tallene for 16-74 gruppa i 10% utvalget på TU fila, d.v.s antallet yrkesaktive mellom 16 og 74 år fra registerdelen av TU fila multiplisert med 10 og dividert med folkemengden i de respektive kategoriene. Vi legger merke til at tallene i kolonne 2 gjennomgående er lavere enn i kolonne 1, noe som skyldes at personer under 16 og over 74 år er sortert ut.

I kolonne 3 er det estimerte antall yrkesaktive antall i hver kategori multiplisert med 10 og dividert med folkemengden. For å avgjøre

kvaliteten på veiingsmetoden er den viktige størrelsen differensen mellom kolonne 2 og 3. Den er angitt i kolonne 4. Kolonne 5 er hentet fra Børke (1984) og angir differensen mellom antallet yrkesaktive fra registerdelen av TU fila multiplisert med 10 og dividert med folkemengden. For å eliminere virkningen av at de registerbaserte tallene er basert på et 10% utvalg hadde det trolig vært mer dekkende å angi differensen mellom tallene for FoB og registerdelen av TU fila. Slik de nå er presentert skyldes avvikene i kolonne 5 både uoverensstemmelser mellom klassifisering i FoB og registerdelen og tilfeldige variasjoner som skyldes TU fila er basert på et 10% utvalg.

Sammenligner man differensene i kolonne 4 og 5 ser man at tallene i kolonne 4 gjennomgående er mindre. Dette er mest markert for fylkestallene, men også for kommunetyper er det en viss bedring. Videre er det en jevnere fordeling av fortegnene i kolonne 4 enn i kolonne 5 slik at det gjennomsnittlige avvik vil bli mindre. Det er også verdt å legge merke til at så store avvik som de størst i kolonne 5 ikke forekommer i kolonne 4. Dette er avvik som ikke skyldes tilfeldige variasjoner. Tabell 1 nedenfor, der gjennomsnittet og gjennomsnittet av absoluttverdiene er beregnet for tallene i kolonne 4 og 5, gir et inntrykk av forbedringen.

Tabell 1. Sammenfatning av differensen mellom kol. 4 og kol. 5 i tabell 2

	Begge kjønn		Menn		Kvinner	
	Fylke	Kommune- type	Fylke	Kommune- type	Fylke	Kommune- type
Gj. avvik kol. 4 .	-0,05	-0,3	0,13	0,60	-0,79	-0,71
Gj. avvik kol. 5 .	1,65	2,1	1,61	1,98	1,76	2,22
Gj. abs. avvik kol. 4	1,06	1,15	1,38	1,44	1,60	0,93
Gj. abs. avvik kol. 5	1,65	2,11	1,61	1,98	1,83	2,22

Tar en med i vurderingen at avvikene i kolonne 4 trolig skyldes det skeive utvalget som er brukt i veiingen, er det grunn til å tro at en ved å bruke et mer representativt utvalg kan oppnå ganske vesentlige gevinster ved å veie registerbaserte tabeller av den typen vi har med å gjøre her.

Tabell 2. Andel yrkesaktive estimert ved etterstratifisering sammenlignet med tall fra FoB.
 Fylker og kommunetyper

Begge kjønn Fylke	FoB	FoButv.	Est.	Diff. Kol. 2- Kol. 3	Diff. FoB - Reg.
ALLE	49,9	49,4	49,8	-0,4	1,4
Østfold	48,4	48,6	47,8	0,8	0,2
Akershus	54,8	53,7	53,8	-0,1	0,5
Oslo	54,4	53,7	54,9	-1,2	0,9
Hedmark	49,5	48,8	51,0	-2,2	1,0
Oppland	49,8	48,7	48,3	0,4	2,4
Buskerud	51,1	49,4	48,6	0,8	0,9
Vestfold	48,9	47,9	50,2	-2,3	0,9
Telemark	47,4	46,5	47,6	-1,1	0,0
Aust-Agder	45,5	43,9	41,5	2,4	2,0
Vest-Agder	47,5	47,3	46,5	0,8	1,4
Rogaland	48,5	48,6	47,8	0,8	1,0
Bergen/Hordaland	48,9	49,0	49,8	-0,8	0,8
Sogn og Fjordane	49,9	48,7	48,2	0,5	4,5
Møre og Romsdal	48,5	48,4	49,3	-0,9	3,0
Sør-Trøndelag	49,8	49,7	51,3	-1,6	1,9
Nord-Trøndelag	48,0	47,5	46,9	0,6	2,9
Nordland	48,7	47,2	46,3	0,9	3,1
Troms	48,7	48,8	50,5	-1,7	2,6
Finmark	48,9	49,1	49,9	-0,6	0,4

Begge kjønn Kommunetype	FoB	FoButv.	Est.	Diff. Kol. 2- Kol. 3	Diff. FoB - Reg.
Landbruk	47,9	47,1	47,8	-0,7	3,6
Mindre sentr. landbruk/ind.	46,8	45,3	48,3	-3,0	3,4
Sentralt landbruk/ind.	49,1	48,6	50,3	-1,7	0,8
Fiskeri	45,7	44,4	42,8	1,6	4,6
Mindre sent. ind.	47,9	47,3	47,8	-0,5	1,8
Sentral ind.	48,6	48,4	47,6	0,8	0,6
Særlig sentr. ind./tj.yt.	53,0	52,5	53,1	-0,3	0,7
Øvrig ind./tj.yt.	49,0	49,1	49,4	-0,3	1,2
Andre	48,7	48,4	46,9	1,5	2,3

Tabell 2 (forts.). Andel yrkesaktive estimert ved etterstratifisering sammenlignet med tall fra FoB.
 Fylker og kommunetyper

Menn Fylke	FoB	FoButv.	Est.	Diff. Kol. 2- Kol. 3	Diff. FoB - Reg.
ALLE	59,1	58,4	58,5	-0,1	1,4
Østfold	58,3	55,4	56,4	-1,0	0,1
Akershus	62,7	60,6	60,7	-0,1	0,2
Oslo	62,7	62,2	64,5	-2,3	0,8
Hedmark	58,7	57,3	60,2	-2,9	0,8
Oppland	58,2	57,5	55,2	2,3	1,8
Buskerud	60,2	58,2	57,7	0,5	0,6
Vestfold	59,0	56,6	58,1	-1,5	2,2
Telemark	58,0	56,0	56,6	-0,6	0,1
Aust-Agder	55,7	54,8	53,3	1,5	2,4
Vest-Agder	58,0	58,4	57,6	0,8	0,8
Rogaland	58,4	59,2	59,8	-0,6	0,9
Bergen/Hordaland	58,2	58,0	59,6	-1,6	1,2
Sogn og Fjordane	58,5	56,4	57,7	-1,3	3,8
Møre og Romsdal	58,8	57,8	58,2	-0,4	3,7
Sør-Trøndelag	58,8	58,8	59,5	-0,7	1,4
Nord-Trøndelag	56,8	56,5	55,6	0,9	2,3
Nordland	56,8	56,5	54,2	2,3	2,6
Troms	57,0	56,9	57,1	-0,2	2,7
Finmark	56,2	55,2	52,7	2,5	0,8

Menn Kommunetype	FoB	FoButv.	Est.	Diff. Kol. 2- Kol. 3	Diff. FoB - Reg.
Landbruk	57,0	56,0	54,7	1,3	2,7
Mindre sentr. landbruk/ind.	56,9	55,7	57,3	-1,6	2,6
Sentralt landbruk/ind.	58,1	57,7	58,5	-0,8	0,0
Fiskeri	56,3	53,7	51,1	2,6	5,3
Mindre sent. ind.	58,5	57,4	57,0	0,4	1,4
Sentral ind.	58,7	58,3	56,7	1,6	0,5
Særlig sentr. ind./tj.yt.	61,4	61,1	62,5	-1,4	0,5
Øvrig ind./tj.yt.	58,3	57,9	57,9	0,0	1,9
Andre	57,1	55,6	52,3	3,3	2,9

Tabell 2 (forts.). Andel yrkesaktive estimert ved etterstratifisering sammenlignet med tall fra FoB.
 Fylker og kommunetyper

Kvinner Fylke	FoB	FoButv.	Est.	Diff. Kol. 2- Kol. 3	Diff. FoB - Reg.
ALLE	40,9	40,6	41,4	-0,8	1,5
Østfold	38,7	39,0	39,3	-0,3	0,2
Akershus	46,9	46,3	47,2	-0,9	0,6
Oslo	47,2	46,4	46,4	0,0	1,1
Hedmark	40,3	40,1	41,4	-1,3	1,3
Oppland	41,3	39,8	41,5	-1,7	3,1
Buskerud	42,1	40,7	39,6	1,1	1,1
Vestfold	38,5	38,9	42,7	-3,8	-0,4
Telemark	36,9	36,9	39,1	-2,3	-0,2
Aust-Agder	35,6	33,0	30,2	2,8	2,1
Vest-Agder	37,1	36,6	36,1	0,5	2,1
Rogaland	38,7	38,0	36,5	1,5	1,2
Bergen/Hordaland	39,8	40,3	40,0	0,3	0,5
Sogn og Fjordane	40,9	40,8	39,8	1,0	5,4
Møre og Romsdal	38,1	38,8	41,0	-2,2	2,2
Sør-Trøndelag	40,9	40,9	43,3	-2,4	2,4
Nord-Trøndelag	37,6	38,3	38,5	-0,2	3,4
Nordland	37,6	37,7	38,4	-0,7	3,4
Troms	40,2	40,5	43,8	-3,3	2,4
Finmark	41,1	42,6	45,8	-3,2	0,1

Kvinner Fylke	FoB	FoButv.	Est.	Diff. Kol. 2- Kol. 3	Diff. FoB - Reg.
Landbruk	38,5	37,8	39,2	-1,3	4,6
Mindre sentr. landbruk/ind.	36,4	34,6	33,6	1,0	4,3
Sentralt landbruk/ind.	39,8	39,3	42,0	-2,7	1,6
Fiskeri	34,4	34,4	34,5	-0,1	3,6
Mindre sent. ind.	37,0	36,9	38,6	-1,7	2,1
Sentral ind.	38,5	38,5	38,5	0,0	0,7
Særlig sentr. ind./tj.yt.	45,1	44,5	44,5	0,0	0,8
Øvrig ind./tj.yt.	40,0	40,4	41,2	-0,8	0,7
Andre	40,0	41,0	41,6	-0,8	1,6

Standardavviket kan ventes å være av størrelsesorden 0.1% for hele befolkningen og for hvert kjønn på landsbasis, 0.8% for de minste fylkene, 1.0% for hvert kjønn i de minste fylkene. Forutsetningen er at fordelingen av antallet yrkesaktive er av noenlunde samme karakter som i tabell 2, at registeropplysninger samles for hele befolkningen og at et sampel på 100 000 benyttes ved veiingen.

3. FORDELING PÅ ARBEIDSTID

Tabell 3 angir de resultatene en får ved veiing når yrkesaktive er klassifisert i fem kategorier. Estimeringen er foretatt for totalbefolkningen mellom 16 og 74 år og for samme gruppe splittet opp etter kjønn.

I først rad er de offisielle folketellingstallene gjengitt. Neste rad er tallene for personer mellom 16 og 74 år fra FoB delen av TU fila multiplisert med 10. Rad 3 er tallene fra registerdelen av TU fila blåst opp til landsnivå. Tallene er hentet fra Børke (1984), tabell 8. Rad 4 gir de tilsvarende tallene som framkommer ved å veie de registerbaserte tallene. Avviket mellom rad 2 og 4 angir den viktigste uoverensstemmelsen, og den er tabulert i rad 5. I rad 6 er tallene i rad 4 gitt i prosent av tallene i rad 2. Disse tallene er de mest relevante for sammenligning med forholdet mellom FoB tall og registerbaserte tall. Disse forholdene er også hentet fra Børke (1984) og er tabulert i rad 7. Her er antallet yrkesaktive i FoB med uoppgitt timetall fordelt proporsjonalt på de fire andre kategoriene yrkesaktive. Disse tallene er så sammenlignet med tilsvarende registerbaserte tall.

Ved sammenligning av rad 6 og 7 kan en trekke noe av samme konklusjon som for tallene for yrkesaktive, selv om forbedringen ser ut til å være noe klarere. Veiingen fjerner de største skjevhetene, og de resterende kan trolig føres tilbake til at samplet som er brukt ved veiingen er skjeivt.

Tabell 3. Tall for arbeidstid estimert ved etterstratifisering sammenlignet med tall fra FoB

	Ikke yrkesaktive		Yrkesaktive	100-499t	500-999t	1 000-1 299t	13 001-t	Uopp-gitt	% av yrkesaktive
Begge kjønn									
FoB	4 091 132	2 049 490	2 041 642	301 701	239 348	167 287	1 211 502	121 804	49,9
FoB utv.	2 906 620	884 130	2 022 490	298 460	236 950	165 820	1 201 620	119 640	49,4
Reg.	4 069 910	2 096 990	1 972 920	189 850	238 800	172 750	1 352 930	-	48,5
Est	2 905 950	876 870	2 029 080	350 490	245 110	167 490	1 159 950	106 040	49,6
Diff: rad 2 - rad 3	670	7 260	-6 590	-50 030	-8 160	-1 670	41 670	13 600	-0,2
Est i % av FoB utv.	100,00	99,2	100,4	117,4	103,4	101,0	96,5	88,6	-
Reg i % av FoB just	99,5	102,3	96,6	59,2	93,8	97,1	105,0	-	-
Menn									
FoB	2 027 083	829 647	1 197 436	108 995	72 099	61 320	883 598	71 425	59,1
FoB utv.	1 451 330	267 100	1 184 230	107 770	70 150	59 620	877 420	69 270	58,4
Reg.	2 016 010	852 720	1 163 280	76 310	62 510	39 620	973 480	-	57,7
Est	1 451 160	272 740	1 178 410	134 130	76 530	58 890	846 550	62 310	58,1
Diff: rad 2 - rad 3	170	-5 640	5 820	-26 360	-6 380	730	-30 870	-6 960	0,3
Est i % av FoB utv.	100,0	102,1	99,5	124,5	109,1	98,8	96,5	90,0	-
Reg i % av FoB just	99,5	102,8	97,1	65,8	81,6	60,8	103,6	-	-
Kvinner									
FoB	2 064 049	1 219 843	844 206	192 706	167 249	105 967	327 904	50 380	40,9
FoB utv.	1 455 290	617 030	838 260	190 690	166 800	106 200	324 200	50 370	40,6
Reg.	2 053 900	1 244 260	809 640	113 540	176 260	133 130	379 450	-	39,4
Est	1 454 790	601 130	853 650	217 670	167 630	108 370	316 020	43 940	41,3
Diff: rad 2 - rad 3	500	15 900	-15 390	-26 980	-830	-2 190	-8 180	-6 430	-0,7
Est i % av FoB utv.	100,1	97,4	98,2	114,1	100,5	102,1	97,5	87,2	-
Reg i % av FoB just	99,5	102,0	95,9	55,4	99,1	118,1	108,8	-	-

4. FORDELING PÅ NÆRINGER

Næringsopplysningene er hentet fra felt 74 på TU fila for FoB tallene og fra felt 49 for registeropplysningene.

Ved fordeling av yrkesaktive på næringer er det en kompliserende faktor at næringsopplysningene kan være forskjellig for FoB delen og registerdelen.

Det er derfor rimelig å benytte en estimeringsmetode som tar hensyn til dette. La n_{ijkl} være antallet i samplet i kategori $ijkl$ der i og j betegner yrkesaktiv/ikke yrkesaktiv i henholdsvis FoB og registerdelen, og k og l betegner næringsklassifiseringen i henholdsvis FoB og registerdelen. Marginalfordelingene betegnes med: f.eks. $n_{i.k.} = \sum_{jl} n_{ijkl}$, og N_{jl} angir antallet i registeret i kategori jl .

Da er en rimelig estimator for antallet yrkesaktive (heretter estimator I) i næringskategori k

$$\sum_{jl} \frac{n_{ijkl}}{n_{.j.l.}} N_{jl}$$

Denne estimatoren er beregnet på grunnlag av tall for hele befolkningen og for hvert kjønn.

I tillegg er det regnet ut en estimator (heretter estimator II) ved å foreta veiingen innen hver næringskategori slik denne er oppgitt i registeret

$$\sum_j \frac{n_{ij.l.}}{n_{.j.l.}} N_{jl}$$

Det er grunn til å tro at denne estimatoren gjennomgående vil være noe for lav.

Resultatene er sammenfattet i tabell 4 og 5, som kan sammenlignes med tabell 9 og 10 i Børke (1984).

Første rad i tabell 4 er basert på FoB tallene fra TU fila for personer mellom 16 og 74 år. Neste rad er tilsvarende tall fra registerdelen. EstI og EstII er resultatet av veiing med metode I og II henholdsvis. Prosenttallene i rad 4 og 6 er tallene i rad 3 og 5 i prosent av tallene i rad 1.

Tallene for sekundærnæringer, tertiærnæringer og totaltallene fra iiii-kommer ved å summere over næringer.

Ved veiingen er det for tallene fra FoB i tillegg til de ni næringsgruppene benyttet en kategori for ikke yrkesaktive og en for uoppgitte. For de registerbaserte opplysningene er det benyttet en

Tabell 4. Antall yrkesaktive estimert ved etterstratifisering sammenlignet med tall fra FoB.
 Kjønn og næring

		Begge kjønn	Menn	Kvinner
Næring				
Alle med kjent næring	FoB utv.	2 006 390	1 117 640	830 750
	Reg.	1 852 780	1 088 680	764 100
	Est I	2 029 850	1 184 050	844 416
		101,2	105,9	101,6
	Est II	1 783 950	1 062 260	719 550
	88,9	95,0	86,6	
Primær næring	FoB utv.	160 990	110 130	50 860
	Reg.	134 530	103 070	31 460
	Est I	175 290	117 670	55 150
		108,9	106,8	108,4
	Est II	136 250	107 580	26 100
	84,6	97,7	51,3	
Sekundærnæring	FoB utv.	594 630	477 900	116 730
	Reg.	562 670	449 510	113 160
	Est I	593 160	487 730	107 900
		99,8	102,1	92,4
	Est II	547 290	441 970	106 080
	92,0	92,5	90,9	
Olje/Bergverk	FoB utv.	10 900	9 570	1 330
	Reg.	8 820	7 460	1 360
	Est I	9 680	7 920	1 760
		88,8	82,8	132,3
	Est II	8 390	7 010	1 360
	76,9	73,2	102,3	
Industri	FoB utv.	409 250	307 710	101 540
	Reg.	404 850	304 910	99 940
	Est I	414 170	321 260	94 370
		101,2	104,4	93,0
	Est II	392 030	299 850	93 130
	95,8	97,4	91,7	
Kraft/Vannforsyning	FoB utv.	18 280	15 560	2 720
	Reg.	18 060	15 440	2 620
	Est I	19 440	17 020	2 910
		106,3	109,4	107,0
	Est II	17 840	15 210	2 620
	97,5	97,8	96,3	
Bygg/Anlegg	FoB utv.	156 200	145 060	11 140
	Reg.	130 940	121 700	9 240
	Est I	149 870	141 530	8 860
		95,9	97,6	77,7
	Est II	129 050	119 900	8 970
	82,6	82,7	80,5	
Tertærnæring	FoB utv.	1 250 770	587 610	663 160
	Reg.	1 155 580	536 100	619 480
	Est I	1 261 400	578 650	681 360
		100,3	98,5	102,7
	Est II	1 100 410	512 710	587 370
	88,0	87,2	88,6	
Varehandel	FoB utv.	355 400	163 370	192 030
	Reg.	325 260	151 810	173 450
	Est I	371 080	167 970	202 060
		104,4	102,8	105,2
	Est II	307 450	145 360	161 820
	86,5	89,0	84,3	

Tabell 4 (forts.). Antall yrkesaktive estimert ved etterstratifisering sammenlignet med tall fra FoB. Kjønn og næring

		Begge kjønn	Menn	Kvinner
Næring				
Transport	FoB utv.	173 670	129 380	44 290
	Reg.	144 170	105 140	39 030
	Est I	172 040	126 480	46 110
		99,0	97,8	104,1
	Est II	140 660	103 170	37 420
	81,0	80,0	84,5	
Bank	FoB utv.	103 460	55 680	47 780
	Reg.	98 980	53 130	45 850
	Est I	105 690	57 170	48 680
		102,1	102,7	101,9
	Est II	95 860	52 220	43 740
	92,7	93,8	91,5	
Tj. yting	FoB utv.	618 240	239 180	379 060
	Reg.	587 170	226 020	361 150
	Est I	612 590	227 030	384 510
		99,0	94,9	101,4
	Est II	556 440	211 960	344 390
	90,0	88,6	90,7	
Annet	FoB utv.	16 140	8 610	7 530
	Reg.	18960/69270	12450/41320	6510/27950
	Est I	5360/16250	1770/8640	3570/7540
	Est II	208050/59420	97490/34710	111330/24640

Tabell 5. Andel yrkesaktive estimert ved etterstratifisering sammenlignet med tall fra FoB. Kjønn og næring

Næring	Begge kjønn				Menn				Kvinner			
	FoB utv.	Reg.	Est I	Est II	FoB utv.	Reg.	Est I	Est II	FoB utv.	Reg.	Est I	Est II
Alle	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Primærnæring	8,0	7,3	8,6	7,6	9,4	9,5	10,0	10,1	6,1	4,1	6,5	3,6
Sekundærnæring	29,6	30,3	29,2	30,7	40,6	41,3	41,2	41,6	14,1	14,8	12,8	14,7
Olje/Bergverk	0,5	0,4	0,5	0,5	0,8	0,7	0,7	0,7	0,2	0,2	0,2	0,2
Industri ..	20,4	21,9	20,4	22,0	26,2	28,0	27,1	28,2	12,2	13,1	11,2	12,9
Kraft/Vannforsyning	0,9	0,8	1,0	1,0	1,3	1,4	1,4	1,4	0,3	0,3	0,3	0,3
Bygg/anlegg	7,8	7,1	7,4	7,2	12,3	11,2	12,0	11,3	1,3	1,2	1,0	1,2
Tertiærnæring	62,3	62,4	62,1	61,7	50,0	49,2	48,9	48,3	79,8	81,1	80,7	81,6
Varehandel	17,7	17,6	18,3	17,2	13,9	13,9	14,2	13,7	23,1	22,7	23,9	22,5
Transport .	8,7	7,8	8,5	7,9	11,0	9,7	10,7	9,7	5,3	5,1	5,5	5,2
Bank	5,2	5,3	5,2	5,4	4,7	4,9	4,8	4,9	5,8	6,0	5,8	6,1
Tjenesteyting	30,8	31,7	30,2	31,2	20,3	20,8	19,2	20,0	45,6	47,3	45,5	47,9

kategori for ikke yrkesaktive samt en kategori for uoppgitte og personer det ikke har vært mulig å klassifisere i en bestemt næringsgruppering, jfr. Børke (1984).

Dette medfører at tallene i rubrikken "annet" har følgende betydning. Første rad er 10 ganger antallet ikke yrkesaktive og uoppgitte fra FoB delen som i registerdelen er klassifisert som yrkesaktive. Rad 2 er 10 ganger antallet som i registerdelen er klassifisert som ikke yrkesaktive i næringskoden, men som likevel har oppgitt årlig arbeidstid på over 100 timer i registerdelen, samt en restkategori som blant annet omfatter selvstendige utenfor primærnæringene. For disse er det for få opplysninger til å foreta en nøyaktig næringsklassifisering. Rad 3 er antallet som havner i henholdsvis kategori ikke yrkesaktiv og uoppgitt ved metode I. Rad 4 er tilsvarende for metode II.

Hvis man sammenligner resultatene fra rad 3, d.v.s. estimatene basert på metode I, med resultatene i tabell 9 i Børke (1984) er bildet noe motstridende. For en god del av næringene er det ingen gevinst ved veiing. Derimot ser det ut til at der det forekommer meget store avvik mellom resultatene basert på FoB delen og registerdelen, vil veiing redusere avviket. Dette kan ses på estimatene for antall selsatte i primærnæringene, i bygg/anlegg (bortsett fra for kvinner) og i transport. Vi minner igjen om at samplet som benyttes ved veiingen trolig er skjeivt, og at det kan forklare en del av aviket mellom rad 1 og 3.

Tabell 5 er basert på tallene fra tabell 4. Prosenttallene for de ulike næringene er beregnet på grunnlag av tallene for alle med kjent næring i tabell 3. Resultatene bekrefter stort sett konklusjonene fra forrige avsnitt.

5. APPENDIX

Det vil her bli gitt en mer formell beskrivelse av estimeringsmetoden som er benyttet. Dessuten vil forventning og varians bli utledet. Resultatene er en omskriving av tilsvarende resultater fra Tenenbein (1972).

Anta vi har en populasjon på N individer som kan klassifiseres etter en rekke kjennetegn. Disse kjennetegn har indeksemengde I som kan uttrykkes som en disjunkt union $I = \bigcup_{i=1}^k I_i$. For et sample av størrelse n observeres det nøyaktige kjennetegn (j_i, i) , $j \in I_i$, $i \in \bar{k} = \{1, \dots, k\}$. For de resterende $N-n$ observeres bare kjennemerket $i \in \bar{k}$.

Anta observasjonene kommer fra en multinomisk forsøksrekke der sannsynligheten for kjennetegn (j_i, i) er $p_{j_i i}$ der $\sum_i \sum_{j \in I_i} p_{j_i i} = 1$.

La $\{n_{j_i i} : j_i \in I_i, i \in \bar{k}\}$ og $\{N_i : i \in \bar{k}\}$ være de observerte frekvensene i samplet og i hele populasjonen henholdsvis. Hvis vi lar $n_{.i} = \sum n_{j_i i}$, vil den betingede fordelingen til $\{n_{j_i i} : j_i \in I_i\}$ gitt $n_{.i}$ være multinomisk $i=1, \dots, k$. Fordelingen til $\{n_{.i} : i \in \bar{k}\}$ gitt $\{N_i : i \in \bar{k}\}$ er superhypergeometrisk, og fordelingen til $\{N_i : i \in \bar{k}\}$ er multinomisk.

Vi er interessert i størrelsen $p = \sum_{i=1}^k p_{j_i i}$. Som estimator er det naturlig å benytte

$$\hat{p} = \frac{1}{N} \sum_{i=1}^k \frac{n_{j_i i}}{n_{.i}} N_i$$

der det er underforstått at hvis en av de marginale hyppighetene i samplet er null skal også $n_{j_i i}/n_{.i}$ settes lik null.

Estimatoren \hat{p} har følgende egenskaper:

Forventning. $E[\hat{p}] = p$. Dette vises ved gjentatt bruk av regelen for dobbeltforventning.

Varians. For å beregne variansen trengs følgende mellomresultater

$$\begin{aligned} \text{var} \left[\frac{n_{j_i i}}{n_{.i}} \right] &= E \left[\text{var} \left[\frac{n_{j_i i}}{n_{.i}} \right] \right] + \text{var} \left[E \left[\frac{n_{j_i i}}{n_{.i}} \right] \right] \\ &= E \frac{1}{n_{.i}} \frac{p_{j_i i}}{p_{.i}} \left(1 - \frac{p_{j_i i}}{p_{.i}} \right) \\ &= \frac{N}{n} E \frac{1}{N_i} \frac{p_{j_i i}}{p_{.i}} \left(1 - \frac{p_{j_i i}}{p_{.i}} \right) + o\left(\frac{1}{n}\right) \\ &= \frac{1}{n} \frac{1}{p_{.i}} \frac{p_{j_i i}}{p_{.i}} \left(1 - \frac{p_{j_i i}}{p_{.i}} \right) + o\left(\frac{1}{n}\right) + o\left(\frac{1}{N}\right) \end{aligned}$$

$$\text{var} \left[\frac{N_i}{N} \right] = \frac{1}{N} p_{.i} (1 - p_{.i})$$

$$\text{cov} \left(\frac{N_i}{N}, \frac{N_l}{N} \right) = -p_{.i} p_{.l} / N$$

Ved Taylorutvikling av \hat{p} rundt $p = \sum_i p_{.i} \frac{p_{j_i i}}{p_{.i}}$ fås følgende tilnærming til $\text{var}(\hat{p})$

$$\frac{1}{N} \sum_i p_{.i}^2 \frac{p_{j_i i}}{p_{.i}} \left(1 - \frac{p_{j_i i}}{p_{.i}} \right) / n p_{.i}$$

$$+ \frac{1}{N} \sum_i \left(\frac{p_{j_i i}}{p_{.i}} \right)^2 p_{.i} (1 - p_{.i})$$

$$- \frac{1}{N} \sum_i \sum_{l \neq i} \frac{p_{j_i i}}{p_{.i}} \frac{p_{j_l l}}{p_{.l}} p_{.i} p_{.l}$$

$$= p (1-p) \left[(1-K_p)/n + K_p/N \right]$$

$$\text{der } K_p = \frac{p}{1-p} \sum_i \left\{ \left[\frac{p_{j_i i}}{p} \right]^2 / p_{.i} - 1 \right\}$$

Ved å spesifere indeksemengden fås nå følgende spesialtilfeller.

1. Utvalg og register

$$I = I_1 \times J_1$$

$$\hat{p}_{.i} = \sum_j \frac{n_{ij}}{n_{.j}} N_j$$

2. Utvalg og registre delt opp etter kjennetegn som bestemmes uten feil.

$$I = I_1 \cdot J_1 \cdot L_1$$

$$\hat{p}_{i.l} = \sum_j \frac{n_{ijl}}{n_{.ie}} N_{jl}$$

3. Utvalg og registre delt opp etter kjennetegn der klassifikasjonsvariabelen kan inneholde feil.

$$I = I_1 \cdot J_1 \cdot K_1 \cdot L_1$$

$$P_{i.k.} = \sum_{j1} \frac{n_{ijk1}}{n_{.j.1}} N_{j1}.$$

REFERANSER

Børke, S. (1984). Tilleggsundersøkelsen til Folke- og Boligtellingen 1980. Rapporter 84/22, Statistisk Sentralbyrå.

Statistisk Sentralbyrå (1982). Folke og Boligtelling 1980. Hefte II: Sysselsettingsstatistikk. NOS B299.

Tenenbein, A. (1972). A double sampling scheme for estimating from misclassified multinomial data with applications to sampling inspection. *Technometrics* 14 187-202.