



Anne Vedø og Dina Rafat

**Sammenligning av
utvalgsplaner i AKU**

Notater

Innhold

1. Innledning og oppsummering.....	2
2. Kvalitativ sammenligning.....	2
2.1. Familieutvalg.....	2
2.2. Personutvalg.....	3
2.3. Husholdningsutvalg.....	3
3. Simuleringer	4
3.1. Populasjonsfil.....	4
3.1.1. Noen sentrale populasjonstall.....	5
3.1.2. AKU registersyssetting kontra ”faktisk” sysselsetting.....	6
3.2. Familieutvalg.....	7
3.2.1. Dagens opplegg	7
3.2.2. Vår tilpasning	8
3.3. Husholdningsutvalg – trukket gjennom person.....	8
3.4. Husholdningsutvalg – trukket direkte	9
3.5. Trekking av FoB husholdning.....	9
3.6. Personer trukket direkte	10
3.7. Konklusjon total sysselsetting/ledighet.....	10
3.8. Sysselsetting fordelt på næring og yrkesstatus.....	10
Referanser	12
Vedlegg A	13
Programfiler (underområde prog)	13
Datafiler (underområde wk12).....	14
Populasjonsfila popfil.sas7bdat.....	14
De sist utgitte publikasjonene i serien Notater	16

1. Innledning og oppsummering

Det er to hovedformål med arbeidskraftundersøkelsen (AKU):

- Å kartlegge strukturen til arbeidsmarkedet.
- Å undersøke endringer i arbeidsmarkedet fra en periode til en annen.

Utvalgsdesignen kan ikke være optimal med hensyn til de to målene samtidig. Derfor må man finne et kompromiss for å tjene de to formål på best mulig måte samtidig når man konstruerer en utvalgsplan.

I dette notatet skal vi se på forskjellige utvalgsplaner og sammenligne de aktuelle utvalgsplaner i AKU. Vi skal ikke se på endringstall men konsentrere oss om strukturen på ett tidspunkt.

Det er tre mulige trekkenheter i AKU:

- Familier;
- Personer;
- Husholdninger.

I kapittel 2 studeres egenskaper ved de tre mulige utvalgene og det gjøres en kvalitativ sammenligning av dem. Kapittel 3 beskriver simuleringer som er gjort for alle de tre utvalgene. Først skisseres det konstruert populasjonsfil, så trekking og estimering av antall sysselsatte og antall ledige for hvert utvalg. På den basis beregnes det utvalgsskjevhet og utvalgsvarians for de forskjellige trekkemetodene. Til slutt sammenligner vi estimater for sysselsetting innen utvalgte næringer.

Rapporten viser at det ikke er store forskjeller mellom de ulike utvalgsplanene når det gjelder standardfeil, så med hensyn til utvalgsfeil er alle trekkenhetene anvendbare. Direkte trekking av personer gir beste resultat, mens husholdningsutvalg trukket gjennom person er minst like bra som familieutvalg. Av alle familie/husholdningsutvalgene er direkte trekking av husholdninger best.

2. Kvalitativ sammenligning

2.1. Familieutvalg

Dagens design bygger på et familieutvalg, det betyr at alle personer i familien mellom 16 og 74 år er intervjuet. Familieutvalg trekkes på følgende måte:

I Det Sentrale Folkeregisteret (DSF) er det definert en referanseperson for hver familie. Hans/hennes fødselsnummer brukes som familienummer for alle medlemmer av familien. (Her er samboere med felles barn ikke inkludert i familiebegrepet).

Et antall referansepersoner trekkes. Deretter inkluderes alle personer tilhørende referansepersonenes familienummer i utvalget.

Familieutvalget har følgende egenskaper:

- det er et selvveiende utvalg av familier i hvert fylke – alle DSF-familienumre får lik sannsynlighet for å bli trukket til utvalget per fylke;
- det trekkes systematisk slik at vi får spredt familiene geografisk proporsjonalt;
- det er ingen kontroll av andelen eller antallet familier av forskjellig type. Allikevel regnes det med at man i gjennomsnitt får en riktig fordeling etter familiestørrelse siden utvalget er stort (Solheim og Zhang, 2001).

Rafat (2002) viste at det er en stor sammenheng mellom ektefellers sysselsetting i en familie i visse aldersgrupper. Siden det er en positiv korrelasjon for sysselsetting mellom familiemedlemmene er et familieutvalg mindre effektivt enn et personutvalg, dvs. den effektive utvalgsstørrelsen er lavere i familieutvalg.

På den annen side er det billigere å gjennomføre en undersøkelse basert på familieutvalg enn på personutvalg, siden det er kostnadseffektivt å intervju alle mellom 16 og 74 år som bor under samme tak.

Siden familiebegrepet i AKU ikke inkluderer samboere og samboere med felles barn, dobles trekksannsynligheten for samboere, siden de består av to familier definert ved to forskjellige familienumre. Av samme grunn er enslige og enslige foreldre (med barn under 16 år) underrepresentert i AKU-utvalget (Solheim og Zhang, 2001). Man kan rette opp dette problemet noe ved å inkludere samboere med felles barn i AKUs familiebegrep.

2.2. Personutvalg

Det er to alternativer ved trekking av personutvalg:

- Rent personutvalg;
- Bruke personnummer som trekkenhet, så inkludere enten familien eller husholdningen til de uttrukne personene i det endelige utvalget.

Rent personutvalg har sine fordeler og ulemper. Det trenger færre intervjuobjekter (IO), gir bedre kontroll siden antallet indirekte intervju minskes.

På den andre siden er ulemper ved et rent personutvalg følgende:

- Det vil være dyrere å intervju et personutvalg enn dagens familieutvalg;
- Det er betydelig fare for et større frafall, særlig i små husholdninger;
- Det viktigste argumentet kontra et rent personutvalg er at husholdningsdata må kartlegges pga. EØS krav.

Personutvalg bygger på at alle personer mellom 16 og 74 år behandles likt siden vi trekker et selvveiende utvalg - alle har lik sannsynlighet for å trekkes til utvalget. Man kan så lage det endelige utvalget ved å inkludere enten familien eller husholdningen til de uttrukne personene. Hvis man velger å inkludere familien/husholdningen, vil trekksannsynligheten til en familie/husholdning bli proporsjonal med antall personer mellom 16 og 74 år i familien/husholdningen. Trekksannsynligheten til en person blir da lik trekksannsynligheten til familien/husholdningen personen bor i.

Trekking av familie/husholdning gjennom person fører til overrepresentasjon av store familier/husholdninger. Antakeligvis har store husholdninger mer stabilitet på arbeidsmarkedet, de er ikke så mobile som små husholdninger. Utvalget blir dermed mindre egnet til å estimere endringer.

2.3. Husholdningsutvalg

Det er en mulighet å trekke husholdninger gjennom bolignumre i Boligregisteret som bør være realistisk fra 1. kvartal 2004. I denne studien har vi også anledning til å trekke ekte husholdninger direkte, noe som er umulig i virkeligheten, men som kan sammenlignes med trekking av bolignumre. Analyser av en slik sammenligning kan svare på spørsmålet om boligregisteret er et godt grunnlag for trekking av husholdningene.

Allerede nå, uten å analysere egenskaper ved et husholdningsutvalg kan vi nevne noen fordeler med den typen utvalg:

- Likestiller forskjellige typer samlivsformer, dvs. ektepar og samboere;
- Gir en riktig fordeling etter husholdningsstørrelse;
- En undersøkelse basert på husholdningsutvalg kan brukes til mange formål. For eksempel, kan det være et utvidet datagrunnlag for Inntekts- og formueundersøkelsene av personer og husholdninger. Den kan også tjene som løpende kvalitetsundersøkelse av boligregisteret (hvis utvalget er trukket gjennom bolignumre).

3. Simuleringer

For å gjøre en kvantitativ sammenligning av de aktuelle trekkeenhetene, har vi gjennomført simuleringer. Først konstruerte vi en kunstig populasjonsfil, der alle relevante variable er kjent for alle personer. Vi trakk deretter 10 000 utvalg fra hver utvalgsplan. For hvert utvalg estimerte vi antall sysselsatte og antall ledige etter den samme estimeringsmetoden som brukes i AKU i dag. Dette bruker vi til å beregne utvalgsskjevhet og utvalgsvarians for de forskjellige trekkemetodene.

Vi har også sammenlignet estimerer for sysselsetting innen utvalgte næringer basert på 1000 simuleringer.

Populasjonen vi trekker fra består av 14 392 personer, og er mye mindre enn den faktiske populasjonen det trekkes fra i AKU. Våre utvalg er også mindre enn de ekte utvalgene; ca. 2 200 personer mot 24 000. Standardavvikene i våre simuleringer er derfor ikke de samme som i virkeligheten, men vi kan anta at rangeringen dem imellom er riktig.

3.1. Populasjonsfil

Vår populasjonsfil er laget med utgangspunkt i fila fra kvalitetsundersøkelsen (KU) i FoB. Denne undersøkelsen var samordnet med AKUs husholdningsundersøkelse. KU inneholder husholdningsopplysninger for alle IO i AKU som ble intervjuet etter 3.11.2001. Det spesielle med KU i forhold en ordinær AKU husholdningsundersøkelse er at alle studenter i utvalget får husholdningsspørsmålene. Vanligvis er det bare en person i hver husholdning som får disse spørsmålene. Ekstra oppfølging av studenter ble gjort fordi det i denne gruppen er stort avvik mellom registrert og faktisk husholdning. Dette førte imidlertid til at foreldre og studenter ofte ga motstridende opplysninger om hvor studenten faktisk bodde. Vi har derfor måttet revidere KU en del for å få hver person plassert i bare en husholdning. Dette er dokumentert i et eget notat.

Populasjonsfila består av alle personer i alle husholdninger på den reviderte versjonen av KU, som i tillegg har fødselsnummer og familienummer og er mellom 16 og 74 år.

Under følger en beskrivelse av de viktigste begrepene på populasjonsfila. En mer utfyllende variabeliste finnes i vedlegg om programmer og datafiler.

- Familie – Definert ved hjelp av familienummer i folkeregisteret.
- Husholdning – Definert ved hjelp av husholdningsnummer i KU/AKU's husholdningsundersøkelse. Dette definerer personer som *faktisk* bor sammen.
- FoB husholdning – Definert ved hjelp av FoB husholdningsnummer, som er basert på formell adresse.
- AKU registersysselsetting - Dette er den registersysselsettingsvariabelen som brukes i kalibreringen i AKU. Den er basert på sysselsettingsopplysninger fra register, men ikke registeropplysninger om ledighet.
- FoB registersysselsetting – Dette er FoBs forbedrede registersysselsettingsvariabel. Den er laget på grunnlag av flere registeropplysninger enn AKU registersysselsetting. Inneholder en kategori for ledig.
- AKU faktisk sysselsetting/ledighet – Dette er basert på opplysninger fra AKU intervjuet, og er personens faktiske sysselsettingsstatus.
- ("Faktisk") sysselsetting/ledighet – Dette er vår tilnærming til faktisk sysselsetting/ledighet på populasjonsfila. Den er lik AKUs faktiske sysselsetting/ledighet der disse finnes, ellers er verdien tatt fra FoBs forbedrede registersysselsettingsvariabel. Det er disse variablene vi bruker i simuleringene.

3.1.1. Noen sentrale populasjonstall

Tabell 1 viser antall ledige, sysselsatte og personer utenfor arbeidsstyrken.

Tabell 1. Sysselsetting/ledighet

	Antall	Prosent av arbeidsstyrken
Ledige	324	2,89
Sysselsatte	10 871	97,11
Utenfor arbeidsstyrken	3 197	-
	14 392	

Det er litt lavere ledighet på vår populasjonsfil enn publisert ledighet for fjerde kvartal 2001, som var på 3,3 prosent. Dette er antakeligvis fordi vi bare bruker svarutvalget fra husholdningsundersøkelsen.

Tabell 2, 3 og 4 viser familiestørrelse og husholdningstørrelse for AKU og FoB henholdsvis.

Tabell 2. Familiestørrelse

Familie- størrelse	Antall Personer	Prosent
1	4 514	31,36
2	3 404	23,65
3	2 696	18,73
4	2 350	16,33
5+	1 428	9,92
	14 392	

Tabell 3. Husholdningsstørrelse AKU

	Antall personer	Prosent
1	2 314	16,08
2	4 720	32,80
3	3 169	22,02
4	2 711	18,84
5+	1 478	10,27
	14 392	

Tabell 4. Husholdningsstørrelse FoB

	Antall personer	Prosent
1	2 367	16,45
2	4 066	28,25
3	3 190	22,17
4	3 010	20,91
5+	1 759	12,22
	14 392	

Vi hadde forventet flere enpersonhusholdninger i AKU enn i FoB, fordi studenter står registrert på sin formelle adresse i FoB, dvs. de står ofte oppført hos foreldrene, selv om de bor alene. Dette veies imidlertid opp av at en del av de som formelt bor alene, faktisk bor i en flerpersonhusholdning.

Tabell 5. Studenter fordelt på AKU- og FoB-husholdning

AKU/ FoB	1	2	3	4	5+	
1	165	35	4	1	0	205
2	50	92	5	0	1	148
3	148	33	66	4	2	253
4	163	19	12	45	1	240
5+	122	18	4	5	23	172
	648	197	91	55	27	1 018

Det er 483 studenter som bor i flerpersonholdninger i FoB, men alene i AKU. Bare 40 studenter bor i flerpersonhusholdning i AKU, men alene i FoB.

Tabell 6. Hele populasjonsfila fordelt på AKU- og FoB-husholdning

AKU/ FoB	1	2	3	4	5+	
1	1 783	425	93	41	25	2 367
2	86	3 856	96	21	7	4 066
3	158	331	2 630	56	15	3 190
4	165	84	296	2 420	45	3 010
5+	122	24	54	173	1 386	1 759
	2 314	4 720	3 169	2 711	1 478	14 392

Totalt sett er det 531 personer som bor i flerpersonhusholdninger i FoB, men alene i AKU, og 584 personer som bor i flerpersonhusholdning i AKU, men alene i FoB.

3.1.2. AKU registersysseletting kontra ”faktisk” sysselsetting

Vi mangler opplysninger om faktisk sysselsetting/ledighet for noen personer på populasjonsfila. Noen er ikke intervjuobjekter i AKU, fordi de er i husholdningen til et IO i AKU uten å være i familien. Andre kan være frafall i AKU. For disse personene er ”faktisk” sysselsetting/ledighet tatt fra FoBs registersysselettingsvariabel. Denne bygger delvis på den samme variabelen som AKU bruker i kalibreringsprosedyren. Vi risikerer dermed å få en kunstig høy sammenheng mellom AKU registersysseletting og ”faktisk” sysselsetting på vår populasjonsfil, hvilket vil føre til at kalibreringsestimatorene vil virke bedre enn den gjør i virkeligheten. Under følger noen tabeller over sammenhengen mellom AKUs registersysselettingsvariabel som brukes i kalibreringen og ”faktisk” sysselsetting/ledighet i svarutvalget, utenfor svarutvalget og totalt.

Tabell 7. Antall personer i svarutvalget og utenfor svarutvalget

	Antall
Svarutvalg	11 874
Utenfor svarutvalg	2 518
	14 392

Tabell 8. Sysselsetting

	Faktisk sysselsetting					
	Svarutvalg		Utenfor svarutvalg		Alle	
AKU regsyss.	0-ikke syss.	1-syss.	0-ikke syss.	1-syss.	0-ikke syss.	1-syss.
0 – ikke syss.	70,94	29,06	63,65	36,35	69,74	30,26
1 - sysselsatt	5,79	94,21	1,88	98,12	5,09	94,91

Tabell 9. Ledighet

	Faktisk ledighet					
	Svarutvalg		Utenfor svarutvalg		Alle	
AKU regsyss.	0-ikke ledig	1-ledig	0-ikke ledig	1-ledig	0-ikke ledig	1-ledig
0 – ikke syss.	94,09	5,91	92,22	7,78	93,79	6,21
1 - sysselsatt	99,36	0,64	99,83	0,17	99,44	0,56

Det er ikke helt den samme sammenhengen i og utenfor svarutvalget, men totalt sett er populasjonsfila relativt lik svarutvalget.

3.2. Familieutvalg

3.2.1. Dagens opplegg

Trekking

Det trekkes DSF familienumre stratifisert etter fylke. Det er litt forskjellige utvalgsandeler i fylkene. Det er følgende forhold mellom utvalgsandelene:

Tabell 10. Forhold mellom utvalgsandelene

Oslo, Akershus, Hordaland	0,871
Aust-Agder	1,484
Sogn og Fjordane	1,398
Nord-Trøndelag	1,151
Finnmark	1,851
Andre	1

Alle personer med de uttrukne familienumrene blir med i personutvalget. Hvis personutvalget blir for stort/lite trekkes det på nytt.

Estimering

Dagens estimeringsopplegg bruker en blanding av etterstratifisering og kalibrering. Metoden er beskrevet i Heldal,2000. Den er delt inn i tre trinn. I trinn 1 beregnes det som kalles utvalgsvektene. Disse regnes ut som antall personer i fylket i populasjonen delt på antall personer i fylket i utvalget (Strengt tatt er ikke dette helt det samme som de faktiske utvalgsvektene, som for hver person er den inverse av familienes trekk sannsynlighet). I trinn 2 justeres utvalgsvektene med faktoren N_e / \hat{N}_e , der N_e og \hat{N}_e er henholdsvis korrekt antall personer og estimert antall personer i etterstratum e i populasjonen. \hat{N}_e er etterstratifiseringsestimatorene, der etterstrataene dannes ved å kryssklassifisere etter kjønn, alder (12 grupper) og AKUs registersysselsetting/næring (4 grupper). Etterstratifiseringen foregår på landsbasis. I trinn 3 kalibreres det på kjønn, alder (12 kategorier) og registersysselsetting (2 kategorier) innen hvert fylke. Vektene fra trinn 2 brukes som initialvekter.

3.2.2. Vår tilpasning

Trekking

Trekksannsynligheten i hvert stratum er 0,15 multiplisert med forholdstallene i tabellen over.

Estimering

Dagens opplegg er beregnet for en utvalgsstørrelse på 24 000. I våre simuleringer er utvalgene på rundt 2 200 personer. Når vi kjører dagens opplegg på våre utvalg, får vi derfor mange tomme etterstrata. De to øverste aldersgruppene blir også fort tomme i noen fylker. Vi har forsøkt både å redusere antall etterstrata og å droppe hele etterstratifiseringen, og fant at det var veldig liten forskjell. Vi har derfor utelatt etterstratifiseringen i estimeringsopplegget. I tillegg har vi slått sammen de tre øverste aldersgruppene (60-64, 65-69, 70-74) i kalibreringen. Estimeringsmetoden er dermed ren kalibrering med 10 aldersgrupper for hvert kjønn og utvalgsvekt som initialvekt.

Resultater

Under har vi regnet ut gjennomsnitt og standardfeil for de 10 000 estimatene basert på familieutvalgene, samt for utvalgsstørrelsene.

Tabell 11. Gjennomsnitt og standardfeil basert på familieutvalgene

	Gjennomsnitt	Standardfeil
Antall sysselsatte	10 855,50	100,64
Antall ledige	327,91	46,66
Prosent ledige	2,93	0,41
Utvalgsstørrelse	2 183,58	28,86

De sanne verdiene for antall sysselsatte og ledige er henholdsvis 10 871 og 324. Standardfeilen til

$\frac{1}{10000} \sum_{i=1}^{10000} \hat{\theta}$ er $\frac{1}{100} Sd(\hat{\theta})$, så et konfidensintervall for den faktiske forventningen til

kalibreringsestimatorene for antall sysselsatte er $10855,50 \pm 2 \times 1,0064$, og for antall ledige er det $327,91 \pm 2 \times 0,4666$. Disse inneholder ikke den sanne verdien, så vi kan påstå at familieutvalgene faktisk gir en skjevhet i estimatene.

3.3. Husholdningsutvalg – trukket gjennom person

Trekking

Personer trekkes stratifisert etter fylke. Trekksannsynligheten i hvert fylke er 0,076 multiplisert med forholdstallene i Tabell 10. Tallet 0,076 er valgt slik at gjennomsnittlig utvalgsstørrelse ble omtrent lik som i simuleringene av familieutvalg. Personutvalget består av alle personer som er i husholdningen til de uttrukne personene.

Estimering

Trekksannsynlighetene er regnet ut som $1 - (1 - p)^{N_{hush}}$, der p er utvalgsandelen i fylket og N_{hush} er antall personer i husholdningen. Dette fordi sannsynligheten for at ingen av personene i husholdningen trekkes ut er tilnærmet lik $(1 - p)^{N_{hush}}$. Den nøyaktige sannsynligheten er

$$\frac{N_{fylke} - n_{fylke}}{N_{fylke}} \cdot \frac{N_{fylke} - n_{fylke} - 1}{N_{fylke} - 1} \cdots \frac{N_{fylke} - n_{fylke} - (N_{hush} - 1)}{N_{fylke} - (N_{hush} - 1)} \leq (1 - p)^{N_{hush}}$$

Likhet inntreffer når $N_{hush} = 1$, ellers er uttrykket til venstre mindre enn det til høyre. Det vil si at vi beregner trekksannsynlighetene litt for små. Summen av beregnede trekksannsynligheter i

populasjonen er 2237,27. Dette skal være lik forventet utvalgsstørrelse, som ut fra simuleringene er ca. 2256, så denne feilen er liten, og har ingenting å si for de kalibrerte estimatene.

Tabell 12. Resultater

	Gjennomsnitt	Standardfeil
Antall sysselsatte	10 886,96	101,63
Antall ledige	324,67	47,33
Prosent ledige	2,90	0,42
Utvalgsstørrelse	2 258,08	25,86

Disse utvalgene overestimerer antall sysselsatte omtrent like mye som familieutvalgene underestimerer det. Vi får imidlertid ingen signifikant skjevhet på antall ledige her. Standardfeilene er omtrent de samme som for familieutvalg.

3.4. Husholdningsutvalg – trukket direkte

Trekking og estimering

Husholdninger trekkes direkte innen hvert fylke. Treksannsynligheten er 0,15 multiplisert med forholdstallene. Utvalgsvektene er antall personer i fylket i populasjonen delt på antall personer i fylket i utvalget.

Tabell 13. Resultater

	Gjennomsnitt	Standardfeil
Antall sysselsatte	10 868,69	96,24
Antall ledige	323,45	45,11
Prosent ledige	2,89	0,40
Utvalgsstørrelse	2 236,79	24,41

Disse utvalgene gir ingen skjevhet å snakke om, og standardfeilene er litt mindre enn for de to foregående trekkeenhetene.

3.5. Trekking av FoB husholdning

Trekking og estimering

FoB husholdninger (fremtidig bolignummer) trekkes direkte innen hvert fylke. Treksannsynligheten er 0,15 multiplisert med forholdstallene. Utvalgsvektene er antall personer i fylket i populasjonen delt på antall personer i fylket i utvalget.

Tabell 14. Resultater

	Gjennomsnitt	Standardfeil
Antall sysselsatte	10 868,50	95,75
Antall ledige	323,72	44,85
Prosent ledige	2,89	0,40
Utvalgsstørrelse	2 240,08	26,52

Disse resultatene er veldig like direkte trekking av faktisk husholdning over.

3.6. Personer trukket direkte

Trekking og estimering

Personer trekkes direkte innen hvert fylke. Trekksansynligheten er 0,15 multiplisert med forholdstallene. Utvalgsvektene er antall personer i fylket i populasjonen delt på antall personer i fylket i utvalget.

Tabell 15. Resultater

	Gjennomsnitt	Standardfeil
Antall sysselsatte	10 870,21	91,40
Antall ledige	323,35	43,09
Prosent ledige	2,89	0,38
Utvalgsstørrelse	2 229,51	4,90

Disse utvalgene gir ingen skjevhet, og har den minste standardfeilen av alle trekkenhetene vi har undersøkt.

3.7. Konklusjon total sysselsetting/ledighet

Tabell 16 og 17 viser estimater for antall sysselsatte og ledige basert på 5 forskjellige utvalgstyper.

Tabell 16. Estimater for antall sysselsatte

	Gjennomsnitt	Standardfeil	Skjevhet
Familie	10 855,50	100,64	-15,50
Husholdning gjennom person	10 886,96	101,63	15,96
Husholdning direkte	10 868,69	96,24	-2,31
FoB husholdning	10 868,50	95,75	-2,50
Personer direkte	10 870,21	91,40	-0,79

Tabell 17. Estimater for antall ledige

	Gjennomsnitt	Standardfeil	Skjevhet
Familie	327,91	46,66	3,91
Husholdning gjennom person	324,67	47,33	0,67
Husholdning direkte	323,45	45,11	-0,55
FoB husholdning	323,72	44,85	-0,28
Personer direkte	323,35	43,09	-0,65

Trekking av personer direkte gir, ikke uventet, den laveste standardfeilen og ingen påviselig skjevhet. Dagens familieutvalg og trekking av husholdning gjennom person gir både størst skjevhet og standardfeil. Familieutvalgene overestimerer ledigheten med 1,2 prosent av sant nivå. Imidlertid gir ingen av trekkeenhetene ubrukelig stor skjevhet, og forskjellen i standardfeil er ikke veldig stor, så med hensyn til utvalgsfeil er alle trekkeenhetene anvendbare. Trekking av FoB-husholdning og AKU-husholdning gir veldig like resultater.

3.8. Sysselsetting fordelt på næring og yrkesstatus

Det kan tenkes at forskjellene mellom utvalgsplanene blir større hvis man brytter tallene ned på næring eller andre mindre grupper. Dette er fordi det kan være korrelasjon innen familien. For eksempel, hvis man trekker en lege til et familie- eller husholdningsutvalg får man ofte to, fordi

ektefellen også er lege. Det kan derfor hende at direkte personutvalg vil komme ut som en klarere vinner når tallene fordeles på mindre grupper. For å teste dette har vi kjørt 1000 simuleringer, og estimert sysselsettingen i noen utvalgte grupper.

Tabell 18. Estimerer for antall sysselsatte i jordbruk

	Gjennomsnitt	Standardfeil	Skjevhet
Familie	344,80	50,25	-7,20
Husholdning gjennom person	354,37	50,92	2,37
Husholdning direkte	349,66	52,27	-2,34
FoB husholdning	353,49	50,78	1,49
Personer direkte	351,66	44,35	-0,34

Tabell 19. Estimerer for antall sysselsatte i industri

	Gjennomsnitt	Standardfeil	Skjevhet
Familie	1 419,45	90,27	4,45
Husholdning gjennom person	1 426,79	92,53	11,79
Husholdning direkte	1 420,85	91,55	5,85
FoB husholdning	1 421,11	92,88	6,11
Personer direkte	1 412,97	85,69	-2,03

Tabell 20. Estimerer for antall sysselsatte i undervisning

	Gjennomsnitt	Standardfeil	Skjevhet
Familie	882,17	76,23	-7,83
Husholdning gjennom person	885,73	76,24	-4,27
Husholdning direkte	887,30	77,00	-2,70
FoB husholdning	886,86	74,23	-3,14
Personer direkte	893,22	72,09	3,22

Tabell 21. Estimerer for antall sysselsatte i helse- og sosialtjenester

	Gjennomsnitt	Standardfeil	Skjevhet
Familie	2 043,94	104,12	9,94
Husholdning gjennom person	2 037,34	109,17	3,34
Husholdning direkte	2 033,40	100,09	-0,60
FoB husholdning	2 033,67	101,18	-0,33
Personer direkte	2 028,01	96,25	-5,99

Tabell 22. Estimerer for antall selvstendige

	Gjennomsnitt	Standardfeil	Skjevhet
Familie	661,62	63,07	-11,38
Husholdning gjennom person	671,75	69,12	-1,25
Husholdning direkte	669,64	66,55	-3,36
FoB husholdning	670,57	64,05	-2,43
Personer direkte	670,92	61,34	-2,08

Tabell 23. Estimater for antall lønnstakere

	Gjennomsnitt	Standardfeil	Skjevhet
Familie	10 139,11	106,11	-4,89
Husholdning gjennom person	10 162,79	113,58	18,79
Husholdning direkte	10 139,01	111,82	-4,99
FoB husholdning	10 147,57	106,85	3,57
Personer direkte	10 148,08	97,92	4,08

Trekking av personer direkte er klart best, men forskjellen fra de andre utvalgsplanene er ikke noe større enn den var for hovedtallene. Forskjellene mellom de fire andre er ikke så klar som den var for hovedtallene, men også her kommer familie og husholdning gjennom person stort sett dårlig ut, enten med hensyn til skjevhet eller standardfeil eller begge deler.

Referanser

Heldal, Johan (2000). Kalibrering av AKU. Dokumentasjon av metode og program. Notater 2000/7.

Rafat, Dina (2002). Analyse av sammenheng mellom ektefellers sysselsetting i en familie. Notater 2002/35.

Solheim, Leiv og Zhang, Li Chun (2001). AKU: Utvalgsplanen, utvalget og husholdninger. Upublisert notat.

Filer på Unix

Filene ligger på området \$METODER/akuomlegging.

Programfiler (underområde prog)

libnames.sas - definerer alle libnames

Populasjon:

- revider_ku.sas - velger husholdning for studenter som er ført opp flere steder
- lagpopfil.sas - lager populasjonsfil
- divtab - underområde med programmer som lager nyttige frekvenstabeller på grunnlag av populasjonsfila

Simulering:

Følgende programmer trekker gjentatte utvalg fra populasjonsfila etter forskjellige utvalgsplaner. For hvert utvalg estimeres sentrale størrelser.

- simulerfam_dagens.sas - trekker familieutvalg slik AKU-utvalg trekkes i dag. Estimerer etter dagens estimeringsopplegg.
- kalibfam_dagens.sas - Program som lager AKU's vektorer. Kalles fra simulerfam_dagens.sas. Dette er Heldals kalib.sas og kalibaku.sas slått sammen til en fil.
- simulerfam_grov2.sas, kalibfam_grov2.sas - Som simulerfam_dagens.sas og kalibfam_dagens.sas, men med 10 aldersgrupper i kalibreringen og 6 aldersgrupper i etterstratifiseringen. Selv i kalibreringen ble 12 aldersgrupper for mye, fordi utvalget lett blir tomt for personer i aldersgruppe 12 i noen fylker.
- simulerfam_10aldgr_uten2.sas, kalibfam_10aldgr_uten2.sas - Som simulerfam_dagens.sas og kalibfam_dagens.sas, men utelater trinn 2, dvs. etterstratifiseringen, i estimeringsopplegget og bruker 10 aldersgrupper i stedet for 12 (slår sammen de tre øverste aldersgruppene). Vi får da ren kalibrering med utvalgsvekt som initialvekt.
- simulerfam.sas, kalibfam.sas – Bruker valgt metode, dvs. kalibrering med 10 aldersgrupper og utvalgsvekt som initialvekt. Beregner også et vanlig HT-estimat.
- simulerfam_utenstrat – Som simulerfam.sas, bare uten stratifisering på familiestørrelse.
- simulerhushpers.sas, kalibhushpers.sas - Lager husholdningsutvalg ved å trekke personer og inkludere husholdningen. Kalibrering med 10 aldersgrupper og utvalgsvekt som initialvekt. Beregner også et vanlig HT-estimat.
- simulerhushdir.sas, kalibhushdir.sas - Lager husholdningsutvalg ved å trekke husholdninger direkte. Kalibrering med 10 aldersgrupper og utvalgsvekt som initialvekt. Beregner også et vanlig HT-estimat.
- simulerpers.sas, kalibpers.sas - Lager personutvalg ved å trekke personer direkte. Kalibrering med 10 aldersgrupper og utvalgsvekt som initialvekt. Beregner også et vanlig HT-estimat.

- `simulertfobhushnr.sas`, `kalibfobhushnr.sas` - Lager "husholdnings"utvalg ved å trekke FoB husholdningsnumre direkte.
- `Skript10x1000` – Unix script som kjører 1000 simuleringer 10 ganger.
- `Andrevar`, `andrevar2` – Underområder som inneholder programmer for simulering innen næring og yrkesstatus. Det er de samme programmene som på hovedområdet, bare kjørt på en versjon av populasjonsfila der variablene `faktsyss` og `ledig` er redefinert til `sysselsatt` innen en viss næring/yrkesstatus.

Analyse av simuleringene

- `analysersim.sas` - regner ut standardavvik og gjennomsnitt for 1000 simuleringer fra de forskjellige utvalgsplanene.
- `analysersim10x1000.sas` – setter sammen og analyserer resultater fra 10 ganger 1000 simuleringer.

Datafiler (underområde wk12)

Populasjon

- `ku.sas7bdat` - FoB/AKU 4. kvartal 2001, Kvalitetsundersøkelsen, fil fra Aslaug Foss
- `kurev2.sas7bdat` – Resultat av `revider_ku.sas`. Dette er vår endelige reviderte husholdningsfil.
- `kumedrevkode1.sas7bdat` – Som `ku`, men med koder for hva som skal gjøres med problematiske records. Se eget notat for beskrivelse av koder. Alle valg er tatt maskinelt. Fremdeles er det noen få dubletter igjen.
- `kumedrevkode2.sas7bdat` – Som `kumedrevkode1`, men med noen flere koder lagt på manuelt for å fjerne dubletter fra `kumedrevkode1`.
- `personer.sas7bdat` - Et (tidkrevende) mellomresultat på veien fra `ku.sas7bdat` til `popfil.sas7bdat`.
- `popfil.sas7bdat` - Vår populasjonsfil, basert på `ku.sas7bdat`

Resultater fra simuleringene på underområdet `simest`:

`simestfam.sas7bdat`
`simestfam_10aldgr_uten2.sas7bdat`
`simestfam_grov2.sas7bdat`
`simestfam_utenstrat.sas7bdat`
`simesthushpers.sas7bdat`
`simesthushdir.sas7bdat`
`simestpers.sas7bdat`
`simestfobhush.sas7bdat`

Disse nummereres fra 1 til 10, ev. også med et suffiks `ht` for Horvitz-Thompson.

Populasjonsfila `popfil.sas7bdat`

Personer

Populasjonsfila er basert på vår reviderte versjon av fila fra kvalitetsundersøkelsen i FoB. Alle personer mellom 16 og 74 år på denne fila er tatt med, bortsett fra de med manglende fødselsnummer eller familienummer. Fila består av 14 392 personer.

Variable

`fnr` - fødselsnummer
`alder`
`aldgr` - 12 kategorier: 16-19 år og senere i 11 5-årskategorier 20-24, ... ,70-74
`kjonn` - kjønn
`fylke`

dsf_famnr - familienummer i folkeregisteret
personkode - nummererer dsf familiemedlemmer. Kode 1 er familiens referanseperson
antfam - antall personer i familien 16-74 år
famstr - som antfam, bare at 4 eller flere er slått sammen (alle får kode 4)
antfamalle - antall personer i familien totalt
famstralle - som antfamalle, bare at 5 eller flere er slått sammen (alle får kode 5)

hushnr – husholdningsnummer for den husholdningen personen er medlem av i KU/AKU's husholdningsundersøkelse.
anthush - antall personer i husholdningen 16-74 år
hushstr - som anthush, bare at 4 eller flere er slått sammen (alle får kode 4)
anthushalle - antall personer i husholdningen totalt
hushstralle - som anthushalle, bare at 5 eller flere er slått sammen (alle får kode 5);

fobhushnr - fob husholdningsnummer
fobanthush - antall personer i fob-husholdningen 16-74 år
fobhushstr - som fobanthush, bare at 4 eller flere er slått sammen (alle får kode 4)
fobanthushalle - antall personer i fobhusholdningen totalt
fobhushstralle - som fobanthushalle, bare at 5 eller flere er slått sammen (alle får kode 5)

regnace1 – Dette er den registersyssesttingsvariabelen som brukes i kalibreringen i AKU, basert på sysselsettingsopplysninger fra register, men ikke ledighet. Vi har laget denne på samme måte som de lager den i AKU, nemlig ved hjelp av programmet hentopp.sas. Den skal i teorien være lik utnace1 der utnace1 finnes (dvs. for IO i AKU), men det er noen små avvik. Det er 4 kategorier: 1, 2, 3 er sysselsatt i henholdsvis primær, sekundær og tertiærnæring og 4 er ikke sysselsatt.

Fra FoB (eller utledet bare fra FoB):

yrkstat - forbedret registersyssestting. Det er 5 kategorier: 1 og 2 er sysselsatt (lønnstager og selvstendig henholdsvis), 3 er arbeidsledig og 0 og 4 er utenfor arbeidsstyrken.
fobsys - sysselsatt/ikke sysselsatt i FoB, utledet fra yrkstat
fobledig - ledig/ikke ledig, utledet fra yrkstat

Fra AKU-utvalg (eller utledet bare herfra):

utnace1 - registersyssestting som de bruker i AKU, skal være lik regnace1.
sstat1 - 1. siffer i statuskode for sysselsetting
utfaktsys - faktisk sysselsetting (0/1) utledet fra sstat1
utledige - arbeidsledig (0/1), utledet fra sstat1

Egendefinerte sysselsettingsvariable:

nace1 - utnace1 der den finnes, deretter regnace1, til slutt 4 (de som mangler begge er alle 16-19 år og har sannsynligvis aldri vært sysselsatt)
faktsys - utfaktsys der den finnes, ellers fobsys
ledige - utledig der den finnes, ellers fobledig

De sist utgitte publikasjonene i serien Notater

- 2003/26 A. Akselsen, S. Lien og T. Sandnes: FD - Trygd. Dokumentasjonsrapport. Pensjoner. Grunn og hjelpestønader. 1992-2001. 113s.
- 2003/27 E. Eng Eikebak og R. Johannessen: Forventningsindikator - konsumprisene. November-mai 2003. 17s.
- 2003/28 A. K. Mevik: Usikkerhet i konjunkturbarometeret. 50s.
- 2003/29 A. Finstad og K. Rypdal: Bruk av helse- og miljøfaglige produkter i hisholdningene - et forprosjekt. 33s.
- 2003/30 T. Jørgensen: Dokumentasjon av prosjektet "Overgang utdanning-arbeid". Årgangene 1999-2000. 54s.
- 2003/31 Å. Cappelen og L.S. Stambøl: Virkninger av å fjerne regionale forskjeller i arbeidsgiveravgiften og noen mulige mottiltak. 35s.
- 2003/32 A. Rognan: Forprosjekt om studenters levekår. 31s.
- 2003/33 S. Vatne Pettersen: Bosettningsmønster og segregasjon i storbyregionene. Ikke-vestlige innvandrere og grupper med høy og lav utdanning. Utredninger til Storbymeldingen, del 1. 71s.
- 2003/34 A. Barstad og M.I. Kirkeberg: Levekår og ulikhet i storby. Utredninger til Storbymeldingen, del 2. 95s.
- 2003/35 E.H. Nymoen, L. Østby og A. Barstad: Flyttinger og pendling i storbyregionene. Utredninger til Storbymeldingen del 3. 75s.
- 2003/36 A. Andersen, T. Løwe og E. Rønning: boforhold i storby. Utredninger til Storbymeldingen, del 4. 82s.
- 2003/37 D. Sve: Seksualitet og helse. Dokumentasjon av datafangsten. 19s.
- 2003/39 C. Nordseth og T. Sandnes: FD - Trygd. Dokumentasjonsrapport. 1992-2001. 110s.
- 2003/40 A. Langørgen og R. Åserud: Faktorer bak kommunale variasjoner i utgifter til sosialhjelp i 2000. 20s.
- 2003/41 T.M. Normann: Omnibusundersøkelsen februar/mars 2003. Dokumentasjonsrapport. 35s.
- 2003/42 D.E. Somervoll: TROLL kan temmes. Kort innføring i Trollprogrammering. 13s.
- 2003/43 Å. Cappelen, T. Eika, P.R. Johansen og J.A. Jørgensen: Makroøkonomiske konsekvenser av lavere aktivitet i oljevirkksomheten framover. 30s.
- 2003/44 L. Østby: Innvandring fra nye EU-land; fortid, nåtid og mulig framtid. 44s.
- 2003/45 T. Dale, H. Høie og A-K.Johnsen: Evaluering av "Naturressurser og miljø" 30s.
- 2003/46 L. Solheim: Foreløpige landstall i KOSTRA. Prinsipper, metoder, produksjon og eksemper. 76s
- 2003/47 A. Hurlen Foss: kvaliteten i boligdelen av Folke- og boligtellingen. 32s.
- 2003/48 E. Siig Meen og O. Rognstad: Jordbruksstilling 1999- dokumentasjon. 105s.
- 2003/49 L.Rogstad: Statistiske temakart og PX-Map. 32s.
- 2003/50 E. Holmøy: Velferdsregnskap - et mulig teoretisk rammeverk.35s.
- 2003/51 C. Wiecek: Undersøkelse om fremtidsplaner, familie og samliv. Dokumentasjonsrapport. 59s.
- 2003/52 KOSTRA: Arbeidsgrupperapporter 2003. 153s.
- 2003/53 A. Haglund: Rapport fra arbeidsgruppa om forslag til arbeidsdeling mellom Brønnøysundregistrene (BR) og Statistisk sentralbyrå (SSB). 40s.
- 2003/54 E. Eng Eibak: Forventningsindikator - konsumprisene. Mai - november 2003. 19s.