



Anne Vedø og Dina Rafat

**Redigering av husholdningsfilen
fra Kvalitetsundersøkelsen**

Notater

Innhold

1. Innledning	3
2. Kvalitetsundersøkelsen	3
2.1 Noen observasjoner fra Kvalitetsundersøkelsen	3
2.2 Hovedproblemet med Kvalitetsundersøkelsen	5
2.3 Opprydding av filen	6
3. Problemstillinger knyttet til husholdningsfilen	9
4. Referanser	9
Vedlegg A	10

1. Innledning

I forbindelse med omlegging av Arbeidskraftundersøkelsens (AKUs) utvalgsplan pga. krav fra Eurostat, ble det simulert trekking fra forskjellige utvalgsplaner (Vedø og Rafat, 2003). Til dette var det nødvendig med en trekkepopulasjon som for hver person inneholdt faktisk bohusholdning. Det ble valgt å bruke AKUs husholdningsundersøkelse fra 4. kvartal 2001, som var samordnet med kvalitetsundersøkelsen (KU) i folke- og bolig tellingen (FoB). I denne undersøkelsen ble husholdningene til personene i AKU-utvalget kartlagt for den delen av utvalget som ble intervjuet etter folke- og bolig tellingen.

Ved kontroll ble det oppdaget at husholdningsfilen inneholdt mange dubletter og måtte korrigeres. I dette notatet går vi gjennom detaljer om hvordan den filen ble bearbeidet.

Først gir vi en kort beskrivelse av Kvalitetsundersøkelsen og spørresekvensen som ble brukt for å identifisere husholdninger. Deretter presenteres det noen observasjoner fra KU. Hovedproblemet med KU diskuteres i kapittel 2.2 og algoritme for opprydding av filen presenteres i kapittel 2.3. Til slutt drøftes det problemstillinger som er knyttet til arbeidet med denne filen.

2. Kvalitetsundersøkelsen

Kvalitetsundersøkelsen ble foretatt for å kartlegge de faktiske husholdningene og ble bygget på AKU-utvalget. Arbeidskraftundersøkelsen (AKU) er et familieutvalg. Det betyr at alle personer i familien mellom 16 og 74 år blir intervjuet om sitt forhold til arbeidsmarkedet. De familiene i AKU som ble intervjuet etter folketellingen den 3. november 2001 fikk i tillegg spørsmål om husholdningen, husholdningsmedlemmenes tilknytning til intervjuobjektet (IO, vanligvis familiens "hovedperson"), felles kost, og om det var studenter i familien. Hvis det var studenter i familien, fikk de også husholdningsspørsmålene.

For å identifisere husholdninger i KU ble følgende spørresekvens brukt:

1. "Ifølge Folkeregisteret består familien din av følgende personer"
Personer listes opp.
2. "Hvem av disse bor ikke sammen med deg?"
3. "Er det andre som er fast bosatt i samme bolig som deg, men som ikke er nevnt?" (Fosen et al, 2001)

De i familien som ikke bodde på adressen ble tatt ut av listen, mens de andre som var fast bosatt i den boligen ble lagt til. I alt 7351 familier fikk disse ekstraspørsmålene.

Ut fra disse opplysningene ble det mulig å konstruere faktisk kost- og bohusholdning for alle IO i AKU som ble intervjuet etter 3.11.2001.

2.1 Noen observasjoner fra Kvalitetsundersøkelsen

Husholdningsfilen fra KU inneholder 8446 husholdninger med totalt 21 228 personer når man tar med alle aldre. Disse kan deles i 17 888 personer som har familienummer blant de utvalgte familienumrene, og 3 340 som har andre eller manglende familienummer. De sistnevnte er de som har blitt føyd til under intervjuet. Av disse er det 335 som mangler fødselsnummer. Andelen med manglende fødselsnummer blant personer som er kommet inn under intervjuet er derfor $335/3340 \approx 0.10$.

Blant de 21 228 personene er det kun 168 som ikke deler kost med IO personene (Tabell 1).

Tabell 1. Opplysninger om kosthusholdning (alle personer)

	Antall	Prosent
Har felles kost	20 160	99,21
Har ikke felles kost	168	0,79

Samtidig er det 28 personer som har slektsforhold til IO lik ”losjerende, hybelboer”, 171 personer er ”annen person, ikke-slektning”, og 29 personer er ”ektefelle/registrert partner/samboer til person i kategori 9-17” (Tabell 2). Andre kategorier det kan være personer med separat kost i, er besteforeldre (12), foreldre/svigerforeldre (733) og søsken (386).

Tabell 2. Slektsforhold til IO (alle personer)

	Antall	Prosent
1. IO Selv	8 414	39,66
2. Ektefelle/ registrert partner	3 138	14,79
3. Samboer	1 761	8,30
4. Eget barn	6 279	29,59
5. Samboers barn	175	0,82
6. Stebarn	17	0,08
7. Ektefelle/ registrert partner til person i kategoriene 4-6	1	0,00
8. Samboer til person i kategoriene 4-6	8	0,04
9. Besteforeldre (ev. ektefelles/samboers besteforeldre)	12	0,06
10. Foreldre/svigerforeldre (ev. samboers foreldre)	733	3,45
11. Søsken/halvsøsken (ev. ektefelles/samboers søsken)	386	1,82
12. Fosterbarn	15	0,07
13. Barnebarn	16	0,08
14. Annen slektning (ev. ektefelles/samboers slektning)	30	0,14
15. Losjerende, hybelboer	28	0,13
16. Arbeidshjelp (f. eks praktikant, au-pair)	4	0,02
17. Annen person, ikke-slektning	171	0,81
18. Ektefelle/ registrert partner til person i kategoriene 9-17	17	0,08
19. Samboer til person i kategoriene 9-17	12	0,06
Totalt	21 217	100,00
Mangler opplysning om slekt	11	

Blant de som er utenfor familien er det 52,4% som er samboer med IO, og 23,5% er IOs barn. (Tabell 3).

Tabell 3. Slektsforhold blant de utenfor familien

	Med fnr	Uten fnr	Totalt	Prosent
2. Ektefelle/ registrert partner	53	4	57	1,71
3. Samboer	1 682	611	743	52,36
4. Eget barn	766	18	784	23,55
5. Samboers barn	166	5	171	5,14
6. Stebarn	11	3	14	0,42
7. Ektefelle/ registrert partner til person i kategoriene 4-6	1	0	1	0,03
8. Samboer til person i kategoriene 4-6	7	1	8	0,24
9. Besteforeldre (ev. ektefelles/samboers besteforeldre)	10	2	12	0,36
10. Foreldre/svigerforeldre (ev. samboers foreldre)	85	93	178	5,35
11. Søsken/halvsøsken (ev. ektefelles/samboers søsken)	55	23	78	2,34
12. Fosterbarn	15	0	15	0,45
13. Barnebarn	16	0	16	0,48
14. Annen slektning (ev. ektefelles/samboers slektning)	24	5	29	0,87
15. Losjerende, hybelboer	11	17	28	0,84
16. Arbeidshjelp (f. eks praktikant, au-pair)	2	2	4	0,12
17. Annen person, ikke-slektning	86	83	169	5,08
18. Ektefelle/ registrert partner til person i kategoriene 9-17	7	3	10	0,30
19. Samboer til person i kategoriene 9-17	7	5	12	0,36
Totalt	3 004	3253	329	100,00
Mangler opplysninger om slekt	1	10	11	

2.2 Hovedproblemet med Kvalitetsundersøkelsen

Hovedproblemet er at mange personer står oppført i mer enn en husholdning. Dette har sammenheng med at alle studenter i AKU-familiene skal ha fått husholdningssekvensen, uansett om foreldrene allerede har påstått at de bor hos dem. Det lages en rekord per intervju, og studenten, (og eventuelt foreldrene og resten av familien, hvis studenten oppgir å bo sammen med dem) blir dermed ført opp to steder. På husholdningsfilen er det 8446 husholdninger. I 1511 av dem er IO student. Av de 1511 studentene er 646 ført opp i bare en husholdning, mens 865 står oppført i to eller flere husholdninger. 580 av dem som står flere steder er 16 til 19 år.

De 8446 husholdningene inneholder 15 890 personer med fødselsnummer mellom 16 og 74 år. 14 420 er forskjellige, mens 1387 er dubletter og 83 er triplerter.

Figur 1 illustrerer det mest vanlige eksemplet på dubletter og triplerter. Det er tatt ut rett fra husholdningsfilen. Her er det sannsynligvis far som oppgir at kona og to barn bor sammen med ham. Kona er student og får dermed sin egen rekord og gir de samme opplysningene som mannen. Et av barna er student og sier at han bor alene, mens foreldrene opplyste at han bor hos dem.

Hushold. IO nummer	Navn på person 1	Navn på person 2	Navn på person 3	Navn på person 4	Er IO student/ skoleelev	Pers_i hush_ aku	barn_ i_hus _aku
11033	Berger, Lars 1961	Berger, Ann Kristin 1960	Berger, Lars Tore 1981	Berger, Andreas 1988		4	2
11034	Berger, Ann Kristin 1960	Berger, Lars 1961	Berger, Andreas 1988	Berger, Lars Tore 1981	S	4	2
11035	Berger, Lars Tore 1981				S	1	0

Figur 1. Eksempel på dubletter og tripletter.

2.3 Opprydding av filen

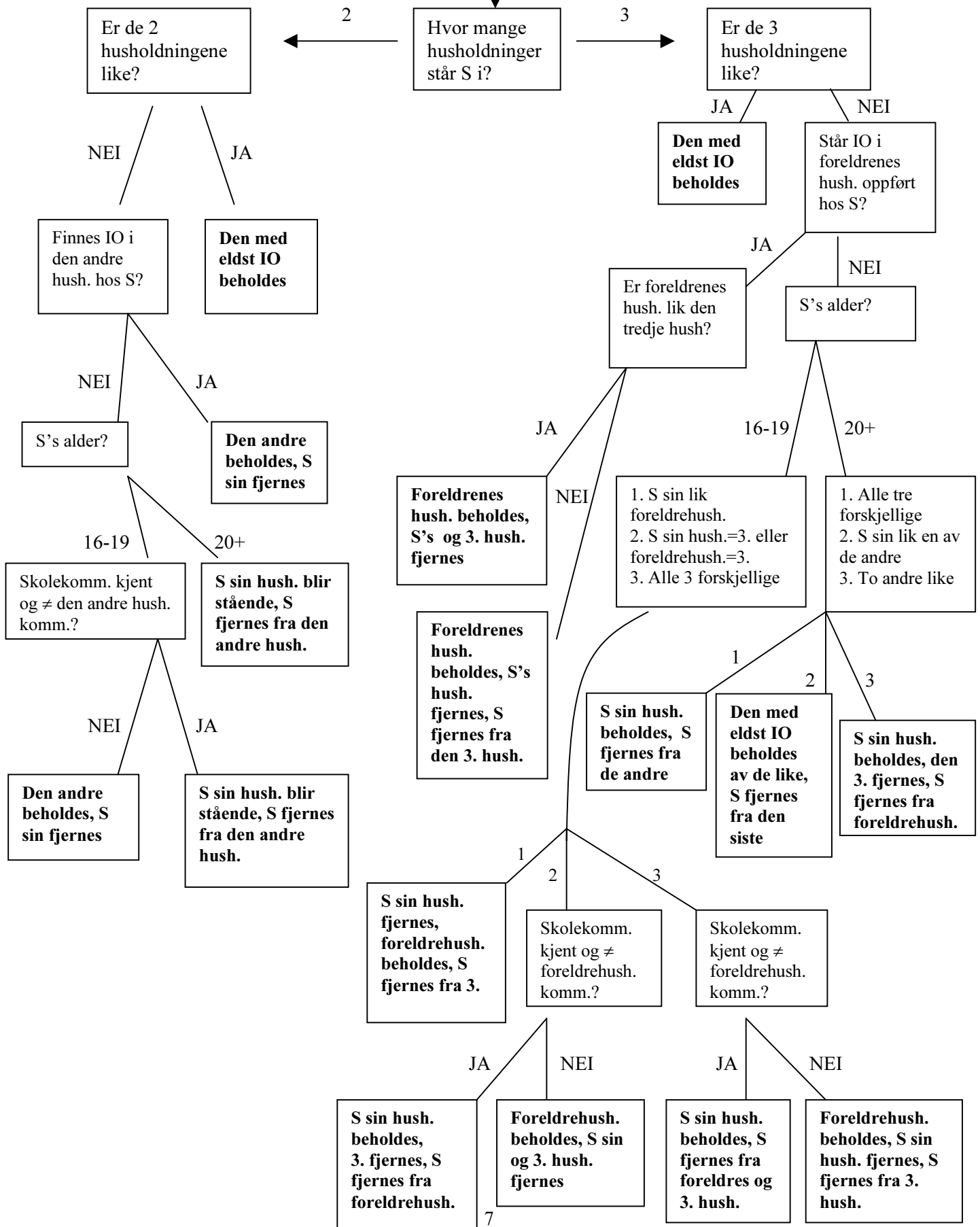
Det minst problematiske tilfellet oppstår når far/mor og studenten begge oppgir samme husholdning. Dette kan enkelt rettes opp ved å fjerne den ene husholdningen. 215 husholdninger kan fjernes på denne måten.

Situasjonen er mer komplisert når studentene oppgir en annen husholdning enn foreldrene. Vi har da valgt å bruke alder og skolekommune. Litt forenklet har vi flyttet 16-19-åringene hjem med mindre de har skolekommune forskjellig fra foreldrenes kommune, mens studenter som er over 19 år ble plassert i den husholdningen de selv oppgir. 259 16-19 –åringer ble flyttet hjem på den måten.

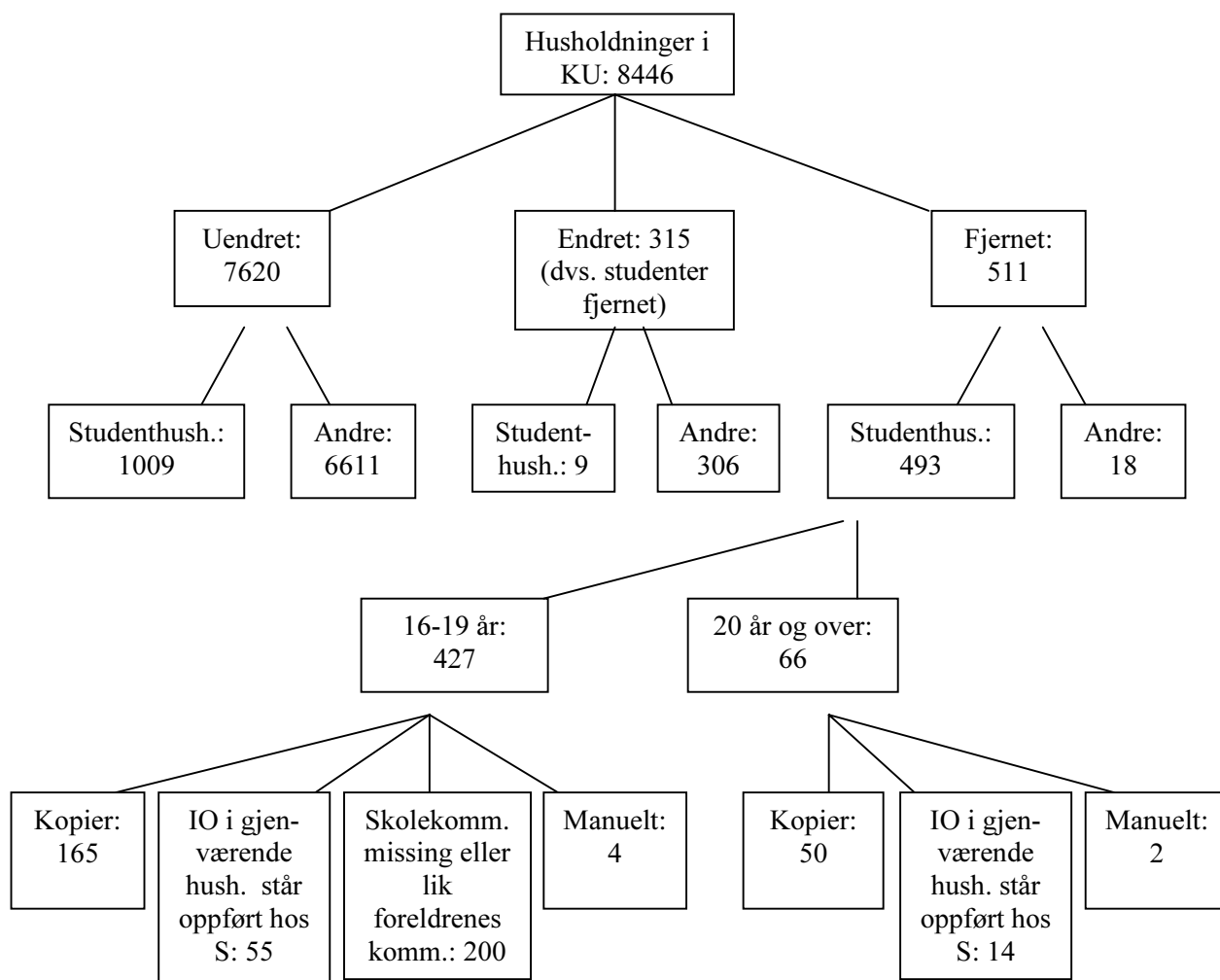
Figur 2 viser “step-by-step” diagram over opprydding av filen og mer detaljert beskrivelse (med koder) finnes i Vedlegg 1.

Figur 2. Diagram over opprydding av filen

Studenter som står oppført flere steder



Figur 3 viser hvor mange husholdninger som er henholdsvis uendret, endret og fjernet, fordelt på studenter og andre. For fjernede studenthusholdninger vises også hvorfor de er fjernet.



Figur 3. Oversikt over hva som er gjort med husholdningene på KU. Grunnlag for fjerning for fjernede studenthusholdninger.

3. Problemstillinger knyttet til husholdningsfilen

Gjennom bearbeiding av husholdningsfilen ble det avdekket noen problemstillinger som trenger videre oppmerksomhet.

Hvorfor svarte foreldre og studenter forskjellig? Vi tror dette delvis har sammenheng med måten spørresekvensen er utformet på. Først listes de registrerte familiemedlemmene opp, og deretter spørres det ”Hvem av disse bor ikke sammen med deg?”. Det virker som en del foreldre har problemer med å eksplisitt fjerne barna fra lista. Kanskje de tolker ”bor” som ”hører til” eller ”er registrert bosatt”. Det er mulig man bør presisere at det spørres etter hvem som *faktisk* bor i husholdningen.

Videre trenger SSB en standard for behandling av studenter. Skal det brukes faktisk eller formell adresse? Skal formell adresse baseres på flere registre enn folkeregisteret? Svaret på dette spørsmålet avgjør standard for husholdningsspørsmålet.

Ut fra vårt behov valgte vi en litt forenklet måte å plassere studentene på. De som skal jobbe videre med filen kan bruke borteboendestipend, studielån og arbeidsinntekt i tillegg for å plassere studentene mer nøyaktig.

4. Referanser

Fosen Johan, Anne Gro Hustoft og Bengt Oscar Lagerstrøm (2001). Ny Spørresekvens for å identifisere husholdninger i utvalgsundersøkelser. Notater 2001/25

Vedø Anne og Rafat Dina (2003). Sammenligning av utvalgsplaner i AKU. Notater 2003/56

Opprydding av husholdningsfil, teknisk beskrivelse

Programmet finnes på /ssb/ovibos/a1/metoder/akuomlegging/prog/revider_ku.sas. Vi tar utgangspunkt i husholdningsfila fra KU, som finnes på /ssb/ovibos/a1/metoder/akuomlegging/wk12/ku.sas7bdat. For hver student som er registrert flere steder (865 stykker), finner vi alle husholdninger studenten står i. I første omgang setter vi koder på husholdningene som indikerer om de skal beholdes, fjernes eller editeres. Kodene settes på følgende måte:

Når studenten står i nøyaktig to husholdninger:

Hvis de to husholdningene er like: Den med eldst IO beholdes (koder: b2l=beholdes blant to like, f2l=fjernes av to like)

Hvis de to husholdningene ikke er like:

Hvis IO i den andre husholdningen finnes i egen husholdning, beholdes den andre husholdningen og egen husholdning fjernes (koder: b2f=beholdes blant to forskjellige, f2f=fjernes av to forskjellige)

Ellers deles det i to etter alder:

16-19 år: Disse antas å bo i den andre husholdningen med mindre de har skolekommune forskjellig fra missing og forskjellig fra den andre husholdningens kommune. Hvis det bestemmes at de bor for seg selv, fjernes de fra den andre husholdningen (koder: b2f1619, f2f1619, e2f1619=editert av 2 forskjellige, tilfelle 16-19 år)

20 år og over: Disse blir stående i den husholdningen de oppgir selv, og blir fjernet fra den andre husholdningen. (koder: b2f20, f2f20, e2f20)

Når studenten står i nøyaktig tre husholdninger:

Det lages en variabel "egen" som er 1 for studentens egen husholdning, 2 i foreldrehusholdningen (dvs. hvis studenten står som eget barn, samboers barn, stebarn, fosterbarn, barnebarn), 3 ellers.

Hvis de tre husholdningene er like: Den med eldst IO beholdes (koder: b3l=beholdes av tre like, f3l=fjernes av tre like)

Hvis ikke alle tre er like:

Hvis studenten har ført opp IO i foreldrenes husholdning i sin husholdning, beholdes foreldrenes husholdning, og egen husholdning fjernes. Studenten fjernes fra den tredje husholdningen. Hvis den tredje husholdningen er lik foreldrenes husholdning, fjernes hele den tredje husholdningen. (koder: b3f, f3f, e3f)

Ellers:

Hvis studenten er 16-19 år:

Det velges mellom egen og foreldres husholdning på grunnlag av skolekommune som før. Hvis det bestemmes at studenten bor hos foreldrene fjernes egen husholdning, og studenten tas ut av den tredje husholdningen. Hvis det bestemmes at han/hun bor i egen husholdning, fjernes han/hun fra de to andre husholdningene.

Hvis to av de tre husholdningene er like, fjernes en av de like. Foreldrenes husholdning beholdes fremfor egen og egen fremfor den tredje ved likhet.

Koder: b3f1619, f3f1619, e3f1619

Hvis studenten er 20 år eller over:

Hvis alle tre husholdninger er forskjellige, beholdes studentens egen husholdning og studenten fjernes fra de to andre.

Hvis egen husholdning er lik en av de to andre, beholdes den med eldst IO (av de to like). Studenten fjernes fra den tredje husholdningen.

Hvis de to husholdningene som ikke er egen er like, fjernes den ene helt. Studentens egen husholdning beholdes, og studenten fjernes fra den gjenværende av de to andre.

Koder: b3f20, f3f20, e3f20

Resultatet av denne kodingen ligger på

/ssb/ovibos/a1/metoder/akuomlegging/wk12/kumedrevkode1.sas7bdat.

Hvis det er flere studenter i samme husholdning, kan man få forskjellige koder på samme husholdning. Vi har prøvd å unngå dette ved å velge den husholdningen med eldst IO ev. foreldres husholdning i størst mulig grad. Husholdninger som får kode både behold og fjern, blir beholdt. Husholdninger som får kode både editer og fjern blir fjernet.

Etter alt dette var det 35 personer som sto oppført i flere husholdninger. Disse har vi redigert manuelt etter skjønn. Man kan se nøyaktig hva som er gjort ved å lese programmet revider_ku.sas. Vi har lagret en fil som inneholder både de automatisk genererte kodene og de manuelt valgte kodene på /ssb/ovibos/a1/metoder/akuomlegging/wk12/kumedrevkode2.sas7bdat.

Den ferdig redigerte fila, dvs. der husholdningene er fjernet/redigert ifølge kodene, ligger på

/ssb/ovibos/a1/metoder/akuomlegging/wk12/kurev2.sas7bdat.

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.