



*Li-Chun Zhang*

**Estimeringsmetode for familie-/  
husholdningsfordeling**

# Notater

## **Forord**

Dette arbeidet følger opp et av delprosjektene i prosjektet "Definisjoner og standarder for gruppering av husholdninger og familier", der formålet har vært å finne en mer enhetlig måte å beregne familie- og husholdningsfordelingen på. Zhang (1999) skisserte en prinsipiell løsning basert på utvalgsundersøkelser. Implementeringen har blitt utsatt i over et år i påvente av datamaterialet fra AKU, 2. kvartal 1999.

Datagrunnlaget i dette studiet er forberedt av Seksjon 420. Her vil jeg takke Vidar Pedersen for mye hjelp. Familieregisteret brukt som tilleggsinformasjon i vekting er tilrett lagt av Seksjon 320, her er Ingvild Hauge kontaktpersonen. I tillegg har Arne Andersen fra Seksjon 350 vært behjelpelig med opplysninger om Levekårsundersøkelsen.

# 1 Innledning

I dette notatet beskriver og forklarer vi estimeringsmetoden for en statistikk over den private familie- og husholdningsfordelingen i den norske befolkning, etter den nye standarden for klassifisering av private kosthusholdninger<sup>1</sup>. Metoden består av 4 deler som følgende:

- bearbeiding av data- og registergrunnlag
- frafallsbehandling
- vekting
- variansberegning.

To problemer til må løses før metoden kan gi oss den omtalte statistikken:

- En husholdningsstatistikk har som målpopulasjon alle private husholdninger i Norge. Idag vet vi ikke hvor mange, eller hvilke, personer som lever i private husholdninger, bl.a. fordi at det ikke finnes en standard klassifisering av de ikke-private husholdningene. Det mangler her et standard arbeid. Det minste vi må vite om rammet er hvor mange personer hører til målpopulasjonen. (I dette studiet har vi imidlertid brukt hele den norske befolkningen som målpopulasjonen.)
- Spørsmålskvensen i husholdningsundersøkelser må være istand til å identifisere husholdningene etter den nye standarden for husholdningsklassifisering. Det er i denne forbindelsen satt i gang et prosjekt knyttet til den årlige innsamlingen av husholdningsdata i AKU. (I dette studiet har vi imidlertid brukt husholdningsklassifiseringen i Inntekts- og Formuesundersøkelsen 1999.)

Vi gjør oppmerksomhet på at selve estimeringsmetoden ikke er påvirket av løsningen på de to overnevnte problemene. Derimot må den justeres mht. følgende to punkter:

- Begrepet registerfamilie brukt i dette studiet tilsier at samboere med felles barn behandles som en familie. I AKU behandler man idag et silkt tilfelle som to registerfamilier. Inntil videre må man derfor justere begrepet i estimeringen slik at det stemmer overens med praksisen i AKU.
- Familieregisteret brukes som tilleggsinformasjon i estimeringen. Dette skal erstattes av et boligregister når det siste er etablert i forbindelsen med Boligtelling 2001.

Problemstillingen, nemlig å estimere familie- og husholdnings totaler i populasjonen basert på et personutvalg, har vært aktuell helt siden FOB90. Men frafallet var mye lavere i Folke- og Boligtelling enn i de utvalgsdataene vi nå har. Resten av dette notatet er disponert som følgende. I avsnitt 2 presenterer vi den estimerte husholdningsfordelingen i 1999, samt det estimerte standardavviket på de tallene. I avsnitt 3 til 6 går vi gjennom de 4 delene av den estimeringsmetoden. I avsnitt 7 diskuterer vi mer detaljert en del valg vi har tatt underveis. De mest tekniske detaljene er lagt til appendikset.

---

<sup>1</sup>Rapp 2000/5. *Standard for begreper og kjennemerker knyttet til familie- og husholdningsstatistikken*. Ingvild Hauge, Coen Hendriks, Øystein Hokstad, Anne Gro Hustoft

## 2 Statistikk over familie-/husholdningsfordelingen

Begrepet familie-/husholdningsfordeling refererer til antall familier eller husholdninger fordelt etter type og størrelse, slik som definert hos Zhang (1999). Tabell 1 viser et eksempel av den estimerte husholdningsfordelingen i 1999. Resultatet er basert på 18940 husholdninger i AKU (2. kvartal 1999) og Levekårpanelet 1999. Blant disse var det 15712 som svarte. Frafallet er imputert under antagelsen om at sannsynligheten for frafall er uavhengig av familiestørrelsen ifølge registeret gitt den faktiske husholdningsstørrelsen og alderen til referansepersonen. Oppblåsningsfaktorene, dvs. vektene, er basert på etterstratifisering m.h.t. familiestørrelsen ifølge registeret, alderen til referansepersonen og det yngste hjemmeboende barn ifølge registeret. Tabell 2 viser tilsvarende estimerte standardavvik. I variansberegningen tar vi hensyn til både utvalgsvariasjonen og usikkerheten i imputering.

Tabell 1: Estimert husholdningsfordeling i den norske befolkningen i 1.1.1999 (i tusen).

Énfamiliehusholdning	Antall personer i husholdning					Total
	1	2	3	4	≥ 5	
Aleneboende under 45 år	396	-	-	-	-	396
Aleneboende 45 - 64 år	190	-	-	-	-	190
Aleneboende 65 år og over	298	-	-	-	-	298
Par u. hj. barn, eldste person under 45 år	-	107	-	-	-	107
Par u. hj. barn, eldste person 45 - 64 år	-	190	-	-	-	190
Par u. hj. barn, eldste person 65 år og over	-	202	-	-	-	202
Par m. hj. barn, yngste barn 0 - 6 år	-	-	77	103	64	244
Par m. hj. barn, yngste barn 7 - 17 år	-	-	49	95	46	190
Par m. hj. barn, yngste barn 18 år og over	-	-	69	20	2	91
Mor m. yngste hj. barn 0 - 17 år	-	43	26	7	1	77
Far m. yngste hj. barn 0 - 17 år	-	7	4	1	0	11
Mor/far m. yngste hj. barn 18 år og over	-	32	4	0	0	36
Flerfamiliehusholdning						
To eller flere enpersonsfamilier	-	39	4	0	0	43
Andre u. yngste hj. barn 0 - 17 år	-	-	9	1	0	11
Andre m. yngste hj. barn 0 - 17 år	-	-	8	7	7	22
Total	884	620	249	235	121	2108

## 3 Data- og registergrunnlag

I IF99 finnes det 5 undersøkelser med husholdningsdata:

- IF-panelet (IFP) er et panelutvalg for IF. F.o.m. år 2001 legges IFP ned.
- Levekårpanelet (LKP) er et tilnærmet selveiende personutvalg som består av ca. 5000 personer fra 16 år og over. Opplysninger om tilknyttet husholdning samles inn i 2. kvartal hvert år. Andelen frafall er ca. 25%. LKP skal gjennomføres minst fram til år 2007.
- Tidsnyttingsundersøkelsen (TU), eller Levekårtverrsnitt, består av ca. 3500 husholdninger. Undersøkelsen foretas om høsten. Frafallsandelen ligger i underkant av 50%.
- IF-tillegg av selvstendig næringsdrivende (IFN) er et spesielt tilleggsutvalg i IF.

Tabell 2: Estimert standardavvik til husholdningsfordelingen i 1.1.1999 (i tusen).

Énfamiliehusholdning	Antall personer i husholdning					Total
	1	2	3	4	≥ 5	
Aleneboende under 45 år	8	-	-	-	-	8
Aleneboende 45 - 64 år	4	-	-	-	-	4
Aleneboende 65 år og over	5	-	-	-	-	5
Par u. hj. barn, eldste person under 45 år	-	3	-	-	-	3
Par u. hj. barn, eldste person 45 - 64 år	-	3	-	-	-	3
Par u. hj. barn, eldste person 65 år og over	-	3	-	-	-	3
Par m. hj. barn, yngste barn 0 - 6 år	-	-	2	2	2	3
Par m. hj. barn, yngste barn 7 - 17 år	-	-	2	2	2	3
Par m. hj. barn, yngste barn 18 år og over	-	-	2	1	0	2
Mor m. yngste hj. barn 0 - 17 år	-	2	1	1	0	2
Far m. yngste hj. barn 0 - 17 år	-	1	1	0	0	1
Mor/far m. yngste hj. barn 18 år og over	-	2	1	0	-	2
Flerfamiliehusholdning						
To eller flere enpersonsfamilier	-	2	1	0	-	2
Andre u. yngste hj. barn 0 - 17 år	-	-	1	0	0	1
Andre m. yngste hj. barn 0 - 17 år	-	-	1	1	1	1
Total	10	6	3	3	2	7

- Arbeidskraftundersøkelsen (AKU) er et familieutvalg som består av ca. 14000 familier. Opplysninger om husholdninger hentes inn i 2. kvartal hvert år, med unntak av år 2001 p.g.a. Folke- og Boligtelling. Frafallsandelen i husholdningsdelen av AKU er ca. 15%.

Til etablering av familie-/husholdningsfordelingen er det mest hensiktsmessig å bruke AKU og LKP, som er de to største undersøkelsene og har det minste frafallet. Siden data samles inn i løpet av det første halvåret, kan man allerede om høsten produsere statistikken for det inneværende året.

Datamaterialet fra AKU og LKP samles i en arbeidsfil. Den ordnes etter en og kun en referanseperson fra hver trukket husholdning. Til disse legger vi til følgende opplysninger:

- frafallsindikator (ja eller nei),
- alder (til referansepersonen),
- alder til det yngste hjemmeboende barn ifølge registeret,
- familiestørrelse ifølge registeret,
- familietype ifølge registeret, og
- husholdningsstørrelse ifølge intervjuet hvis ikke frafall.

I tillegg henter vi fra familierregisteret antall personer i populasjonen, innen hvert etterstratum som er definert i Tabell 5.

## 4 Imputering av frafall

Frafall i husholdningsundersøkelsene er bl.a. skjevt fordelt m.h.t. størrelsen av husholdningen. Andelen små husholdninger er høyere i frafallet enn de som svarer. Estimering basert på nettoutvalget alene vil derfor gi oss for få små husholdninger. Spesielt gjelder dette for antall en-persons husholdninger.

Det finnes idag 3 typer behandling av enhetsfracfall i SSBs person-/husholdningsundersøkelser:

- Etterstratifisering m.h.t. tilleggsvARIABLER som er korrelert med frafallet. Metoden brukes bl.a. i AKU og LKP. Den fjerner all skjevhet forårsaket av frafallet dersom frafall er uavhengig av undersøkelses variabler innen hvert etterstratum. I realiteten er det sjelden tilfellet at denne antagelsen er fullt gjeldende. Det er derfor viktig å søke etter bedre løsninger.
- Å imputere for frafallet med lignende verdier som finnes i registeret. Metoden brukes i IF, der register opplysninger om familien imputeres for husholdningen i tilfellet frafall.
- Imputering under eksplisite modeller om effekten av undersøkelses variabler på frafallet. Belsby and Bjørnstad (1997) studerte flere modeller for estimering av husholdningsfordelingen basert på Forbruksundersøkelsen (FU). Opplegget i FU i dag er en variasjon av denne metoden.

Under den siste fremgangsmåten definerer vi 9 *fracfallsklasseer* i Tabell 3 etter husholdningsstørrelsen og alderen til referansepersonen. (Her deles alderen til referansepersonen inn i tre grupper: (i) under 45, (ii) mellom 45 og 64, og (iii) over 64. Grensen mellom gruppene svarer til klassifiseringen av husholdninger brukt i dette studiet. Når den nye standarden tas i bruk, skal terskelen 64 endres til 67.) Vi antar at sannsynligheten for frafall er uavhengig av familiestørrelsen i registeret gitt frafallsklassen.

Tabell 3: Definisjon av frafallsklasser.

Nr.	Klasse	Nr.	Klasse
I	1-persons husholdning, referanseperson under 45 år	VII	3-persons husholdning
II	1-persons husholdning, referanseperson 45 - 64 år	VIII	4-persons husholdning
III	1-persons husholdning, referanseperson over 64 år	IX	Andre
IV	2-persons husholdning, referanseperson under 45 år		
V	2-persons husholdning, referanseperson 45 - 64 år		
VI	2-persons husholdning, referanseperson over 64 år		

Tabell 4 viser resultater av imputering for AKU og LKP 1999 med forskjellige metoder. Uten imputering blir andelen en-persons husholdninger for lav, også innen delutvalg av like store familier. Imputering (I) av registerverdier ga nokså like marginale andeler av husholdninger (fordelt etter husholdningsstørrelsen) som imputering (II), dvs. under antagelsen om at frafallet er uavhengig av familiestørrelsen gitt frafallsklassen. Forskjellen mellom de to metodene kommer mer tydelig fram når vi sammenligner fordelingen etter både familie- og husholdningsstørrelsen:

- Ta først personer som ifølge registeret hører til en-persons familier. Blant dem som svarte finnes det en betydelig andel (ca. 28%) som faktisk lever i to-persons husholdninger. Den samme andelen er ca. 2.5% i imputering (I), og ca. 20% i imputering (II). Siden andelen en-persons husholdninger er høyere blant frafallet, må andelen to-persons husholdninger der være lavere enn blant dem som svarte. Likevel synes 2.5% å være altfor lav.

Tabell 4: Referanspersoner fordelt etter familie- og husholdningsstørrelsen i AKU og LKP 1999: imputering (I) av registerverdier, og imputering (II) under antagelsen om at frafallet er uavhengig av familiestørrelsen gitt frafallsklassen.

Nettoutvalg	Antall personer i husholdning				
	1	2	3	4	≥ 5
Antall personer i familie					
1	3092	1333	161	76	29
2	200	3040	268	57	19
3	150	478	2039	146	42
4	105	171	296	2372	73
5	39	57	47	140	1282
<b>Imputering (I)</b>					
1	1474	38	13	5	0
2	21	632	15	2	1
3	15	28	392	8	7
4	6	8	20	316	6
5	5	2	3	12	199
<b>Imputering (II)</b>					
1	1196	300	22	7	4
2	77	548	37	6	3
3	56	91	283	14	6
4	39	37	41	229	10
5	15	13	7	14	173
Husholdninger fordelt etter antall personer (i prosent)					
Nettoutvalg	22.8	32.3	17.9	17.8	9.2
Nettoutvalg og Imputering (I)	27.0	30.6	17.2	16.5	8.8
Nettoutvalg og Imputering (II)	26.2	32.0	16.9	16.2	8.7

- Ta så personer som hører til to-persons familier. Blant dem som svarte finnes det ca. 6% som faktisk lever i en-persons husholdninger. Denne andelen synker til 3% i imputering (I), men øker til 11% i imputering (II). Her synes imputering (I) å ha gått i feil retning.

På bakgrunn av slike betraktninger ovenfor, konkluderer vi med at resultater av imputering virker å være best under antagelsen om at frafallet er uavhengig av familiestørrelsen ifølge registeret gitt frafallsklassen. Vi skal komme tilbake til flere vurderinger om frafallsbehandlingen i avsnittet "Diskusjon og oppsummering". Vi refererer til appendiks A for en nøyaktig beskrivelse av metoden.

## 5 Vekting

### 5.1 Initiale vekter som justering for frafall

Istednfor å imputere for interessevariabler som f.eks. husholdningstype i tilfellet frafall, skal vi omsette resultatene i Tabell 4 til vekter av referansepersoner i nettoutvalget. En referanseperson som har familiestørrelse  $i$  og frafallsklasse  $j$  sies å høre til den *initiale klassen*  $(i,j)$ . Vi vet antallet personer i alle de initiale klassene i nettoutvalget. Imputering gir oss antallet slike personer blant frafallet. Den

initiale vekten til en referanseperson som hører til den initiale klassen  $(i, j)$  er da gitt som

$$a_{ij} = \frac{\text{Antall referansepersoner i den initiale klassen } (i, j) \text{ i nettutvalget eller i frafallet}}{\text{Antall referansepersoner i den initiale klassen } (i, j) \text{ i nettutvalget}}. \quad (1)$$

Dersom vi legger sammen den initiale vekt til alle referansepersoner i nettutvalget, får vi antallet referansepersoner i hele bruttoutvalget, slik at referansepersonene i frafallet nå er 'representert' av referansepersonene i nettutvalget.

## 5.2 Etterstratifisering som justering for utvalgsplan

Det er spesielt to sider ved utvalgsplanen til AKU og LKP som trenger justering:

- I AKU trekkes familier med en lik sannsynlighet innen hvert fylke uansett størrelsen til familien. Dette gjør at det forekommer flere referansepersoner fra små familier i AKU enn i LKP. Vi må derfor ta hensyn til register familiestørrelsen når utvalget skal vektet opp mot populasjonen.
- Alderen til referansepersonene i AKU ligger mellom 16 og 74 år. Selv om det ikke finnes en øverste aldersgrense i LKP, er også der de eldste personer i populasjonen under-representert. Vi må derfor ta hensyn til alderen i oppblåsing. Siden eldste personer i de fleste tilfellene lever i en- eller to-persons familier, deler vi kun alderen i tre grupper der. (Vi skal komme tilbake til mer om dette i avsnittet "Diskusjon og oppsummering".)

I tillegg ønsker vi å kontrollere de vektene mot register opplysninger om hjemmeboende barn som er under 18 år. Tabell 5 viser definisjonen av etterstrata. Her deles alderen til referansepersonen inn i tre grupper, der grensen mellom gruppene svarer til klassifiseringen av husholdninger. Når den nye standarden tas i bruk, skal terskelen 64 endres til 67. For etterstratum  $h$ , der  $h$  kan være fra 1 til 17, beregner vi en *oppblåsningsfaktor* som følgende:

$$b_h = \frac{\text{Antall personer i etterstratum } h \text{ i populasjonen}}{\text{Summen av initiale vekter til referansepersoner i etterstratum } h \text{ i nettutvalget}}. \quad (2)$$

## 5.3 Vekting

Vekten til en referanseperson i nettutvalget, som hører til initial klasse  $(i, j)$  og etterstratum  $h$ , er gitt som produkt av den initiale vekten  $a_{ij}$  og oppblåsningsfaktoren  $b_h$ , dvs.

$$w_{ij,h} = a_{ij} \times b_h. \quad (3)$$

Denne vekten gir oss korrekte antallet familier etter størrelsen ifølge registeret, og antallet familier med hjemmeboende barn under 18 år. Vi får nesten korrekte antallet hjemmeboende barn under 18 år, dvs. korrekt med unntak av de som bor i familier med 5 eller flere personer. På den måten blir familierregisteret i mye større grad tatt i bruk enn i dagens utvalgsbasert husholdningsstatistikk.

For å estimere antall personer i populasjonen som er av en bestemt type, trenger vi bare å legge sammen vektene til alle referansepersonene i nettutvalget av samme typen. F.eks. antallet personer som lever i 3-persons husholdninger fås ved å legge sammen vekten til alle referansepersonene i nettutvalget som lever i 3-persons husholdninger. Deler vi etterpå det tallet med 3, får vi et estimat for antallet 3-persons husholdninger i populasjonen. Spesielt bruker vi den gjennomsnittlige familiestørrelsen til familiene med 5 eller flere personer, som den gjennomsnittlige husholdningsstørrelsen til husholdningene med 5 eller flere personer, slik Belsby and Bjørnstad (1997) gjorde.



Tabell 5: Definisjon av etterstrata.

01. 1-persons familie, person under 45 år
02. 1-persons familie, person 45 - 64 år
03. 1-persons familie, person 65 år og over
04. 2-persons familie, par uten hjemmeboende barn, referanseperson under 45 år
05. 2-persons familie, par uten hjemmeboende barn, referanseperson 45 - 64 år
06. 2-persons familie, par uten hjemmeboende barn, referanseperson 65 år og over
07. 2-persons familie, alene mor/far med hjemmeboende barn, yngste barn under 18 år
08. 2-persons familie, alene mor/far med hjemmeboende barn, yngste barn 18 år og over
09. 3-persons familie, par med hjemmeboende barn, yngste barn under 18 år
10. 3-persons familie, alene mor/far med hjemmeboende barn, yngste barn under 18 år
11. 3-persons familie, par med hjemmeboende barn, yngste barn 18 år og over
12. 3-persons familie, alene mor/far med hjemmeboende barn, yngste barn 18 år og over
13. 4-persons familie, par med hjemmeboende barn, yngste barn under 18 år
14. 4-persons familie, alene mor/far med hjemmeboende barn, yngste barn under 18 år
15. 4-persons familie med hjemmeboende barn, yngste barn 18 år og over
16. 5- eller flere enn 5-persons familie med hjemmeboende barn, yngste barn under 18 år
17. 5- eller flere enn 5-persons familie med hjemmeboende barn, yngste barn 18 år og over

Tabell 6: Andel av variansen til husholdningsfordelingen som skyldes imputering (i prosent).

Énfamiliehusholdning	Antall personer i husholdning				
	1	2	3	4	≥ 5
Aleneboende under 45 år	17	-	-	-	-
Aleneboende 45 - 64 år	16	-	-	-	-
Aleneboende 65 år og over	11	-	-	-	-
Par u. hj. barn, eldste person under 45 år	-	9	-	-	-
Par u. hj. barn, eldste person 45 - 64 år	-	8	-	-	-
Par u. hj. barn, eldste person 65 år og over	-	6	-	-	-
Par m. hj. barn, yngste barn 0 - 6 år	-	-	1	1	2
Par m. hj. barn, yngste barn 7 - 17 år	-	-	2	2	1
Par m. hj. barn, yngste barn 18 år og over	-	-	5	2	1
Mor m. yngste hj. barn 0 - 17 år	-	3	1	1	0
Far m. yngste hj. barn 0 - 17 år	-	0	0	0	0
Mor/far m. yngste hj. barn 18 år og over	-	3	1	0	-
Flerfamiliehusholdning					
To eller flere enpersonsfamilier	-	5	1	0	-
Andre u. yngste hj. barn 0 - 17 år	-	-	2	2	1
Andre m. yngste hj. barn 0 - 17 år	-	-	0	0	1

## 6 Variansberegning

I variansberegningen (Tabell 2) har vi tatt hensyn til både utvalgsvariasjon og imputering. Tabell 6 viser andelen til den totale variansen som skyldes imputering. Ikke overraskende er den høyest for en-persons husholdninger, der justeringen pga. imputeringen er størst. Vi viser til Appendiks B og C for beskrivelser av metoder for variansberegning.

## 7 Diskusjon og oppsummering

I den foregående beskrivelsen av estimeringsmetoden har vi forsøket å unngå unødvendige avsporinger i fremstillingen. Alle valg tatt underveis er ikke like tilstrekkelig begrunnet. I dette avsnittet kommer vi tilbake til en del ting som synes å trenge nærmere forklaring.

### 7.1 Hvorfor er frafallsklassen slik definert?

Tabell 7 viser hvordan svarandelen varierer med alderen og familiestørrelsen til referansepersonen i AKU og LKP. I begge undersøkelsene øker svarandelen etter familiestørrelsen. Alderen har en tilleggseffekt blant familier med 1 eller 2 personer. Det er to ting som gjør alderseffekten uklar blant familier med 3 eller flere personer: (a) det er få personer, og (b) mønsteret der er ikke monotont. Pga. den sterke sammenhengen mellom familie og husholdning, antar vi at variasjonen i svarandelen som vist i Tabell 7 finnes også når vi fordeler referansepersonene etter husholdningsstørrelsen og alderen. Spørsmålet nå er hvilken antagelse er mest fornuftig: (i) frafallet er uavhengig av husholdningsstørrelsen gitt familiestørrelsen, eller (ii) frafallet er uavhengig av familiestørrelsen gitt frafallsklassen? Vi valgte den siste siden det er vanskelig å forestille seg at frafallet heller skal avgjøres av registeret enn det faktiske forholdet. Modellen er ikke ment å gi tilstrekkelig forklaringer på all frafall. Den er først og fremst et redskap til å redusere skjevheten forårsaket av frafallet.

Tabell 7: Svarandelen etter familiestørrelsen og alderen til referansepersonen (i prosent, antall i parenteser viser hvor mange personer andelen er basert på.)

AKU	Antall personer i familie				
	1	2	3	4	≥ 5
Under 45 år	75.8 (2866)	85.6 (848)	88.7 (1453)	92.5 (1676)	90.0 (880)
Mellom 45 - 64 år	78.5 (1561)	89.3 (1502)	91.5 (883)	92.0 (721)	92.1 (316)
Over 64 år	82.3 (547)	90.0 (652)	83.5 (79)	58.8 (17)	100 (4)
<b>LKP</b>					
Under 45 år	71.4 (625)	76.2 (265)	77.4 (517)	83.8 (722)	81.4 (474)
Mellom 45 - 64 år	66.6 (311)	74.7 (581)	78.1 (329)	79.3 (237)	81.9 (116)
Over 64 år	62.0 (316)	72.4 (410)	80.4 (51)	100 (4)	0 (0)

### 7.2 Kan vi imputere for AKU og LKP sammen?

Generelt har AKU høyere svarandel enn LKP. I tillegg viser Tabell 7 forskjellige alderseffekter. I AKU øker svarandelen blant en-persons familie ettersom alderen til referansepersonen stiger, mens den synker i LKP. I imputeringen (II) i Tabell 4 har vi slått sammen AKU og LKP før imputeringen. Vi kan også imputere for AKU og LKP hver for seg, og deretter slå sammen den imputerte AKU og

LKP. Dette gir oss et annerledes bruttoutvalg enn før. Tabell 8 viser at det kun er små forskjeller mellom de to måtene å imputere på.

Tabell 8: Differens mellom bruttoutvalg i prosent, der AKU og LKP imputeres sammen eller separat under antagelsen om at frafallet er uavhengig av familiestørrelsen gitt frafallsklassen.

Antall personer i familie	Antall personer i husholdning				
	1	2	3	4	≥ 5
1	0.6	-1.4	0	-1.1	-2.9
2	0.7	0.1	-0.6	-1.5	-4.4
3	1.9	0.9	-0.1	-0.5	-2.0
4	2.1	0	1.2	-0.2	-1.1
≥ 5	3.7	0	1.9	0.6	-0.2

### 7.3 Kan vi vekte de eldste personer sammen?

Både AKU og LKP er nesten selvveiende mht. alderen til referansepersonen for de som er under 75 år. Personer som er 75 år og over er derimot sterkt under-representert. Vi kan etterstratifisere de eldste personene, dvs. over 74 år, som et eget stratum. Mens skjevheten i estimatoren minskes for denne aldersgruppen, vil variansen øke betydelig. Alternativt kan vi bruke etterstratifiseringen i Tabell 5 under antagelsen om at husholdningsfordelingen betinget på familiestørrelsen varierer lite blant de eldste i populasjonen. Tabell 9 viser situasjonen i netto-LKP. Forskjellen i sannsynligheten for en-persons husholdning gitt en-persons familie er ikke signifikant mellom de to aldersgruppene. Det samme gjelder sannsynligheten for to-persons husholdning gitt to-persons familie. Legg merke til at vi kun etterstratifiserer etter alderen til referansepersonen fra en- eller to-persons familier, siden ca. 90% av alle referansepersoner mellom 64 og 74 i populasjonen lever i en- eller to-persons familier ifølge registeret. Den samme andelen øker til ca. 95% for de som er over 74 år.

Tabell 9: Referansepersoner fordelt etter alderen, familie- og husholdningsstørrelsen i netto-LKP.

Mellom 64 og 74 år	Antall personer i husholdning					Total
	1	2	3	4	≥ 5	
Antall personer i familie						
1	104	12	0	0	0	116
2	3	185	3	0	0	191
<b>Mellom 75 og 84 år</b>						
1	75	4	1	0	0	80
2	0	105	0	1	0	106

### 7.4 Hva er forskjellen mellom AKU og LKP?

Både utvalgsplan og datainnsamling er forskjellige mellom AKU og LKP. Spesielt har AKU følgende problem: trekkenheten i AKU er familie. I en familie der medlemmene lever i adskilte husholdninger, er det svært vanskelig å få til korrekt registrering av alle husholdningene. I slike tilfeller har man stort sett valgt å kun ta med husholdningen til referansepersonen. Dette gjør at en del av populasjonen faktisk har ingen sannsynlighet til å være med i utvalget. Siden den delen trolig også består av flere små

husholdninger enn i populasjonen ellers, får vi lett en under-representasjon av små husholdninger i AKU. Tabell 10 viser husholdningsfordelingen (etter husholdningsstørrelsen) basert på AKU eller LKP under forskjellige metoder. For en bedre sammenligning har vi kun tatt med de referansepersonene som er under 75 år i utvalget, og vektet dem opp mot den tilsvarende populasjonen. Som ventet ble andelen en-persons husholdninger estimert 1 - 2 prosent lavere på AKU enn LKP.

Tabell 10: Husholdninger med referanseperson under 75 år fordelt etter størrelsen basert på AKU eller LKP (i prosent): metode (I) etterstratifisering på nettoutvalget, metode (II) etterstratifisering på nettoutvalget og imputeringen av registerverdier, og metode (III) etterstratifisering på nettoutvalget og imputeringen under antagelsen om at frafallet er uavhengig av familiestørrelsen gitt frafallsklassen.

AKU	Antall personer i husholdning				
	1	2	3	4	≥ 5
Metode (I)	35.4	30.4	13.9	13.5	6.9
Metode (II)	37.7	28.6	13.7	13.3	6.8
Metode (III)	37.6	29.3	13.4	13.0	6.8
<b>LKP</b>					
Metode (I)	37.2	30.0	13.7	12.9	6.3
Metode (II)	39.2	28.1	13.6	12.8	6.4
Metode (III)	38.6	29.7	13.3	12.3	6.1

## 7.5 Oppsummering

Tabell 11 viser effekter av datagrunnlag, frafallsbehandling og vekting på husholdningsfordelingen. Vi har også tatt med resultater fra IF98, som var basert på undersøkelser i 1998 og vektet mot populasjonen i 1.1.1999. IF bruker et annet vektingsopplegg der vektene er kalibrert m.h.t. aldersfordelingen og en del inntekts- og formuestotaler i populasjonen. Til sammenligning har vi også inkludert den husholdningsfordelingen basert på famileregisteret i Tabell 12.

Tabell 11: Husholdninger i populasjonen i 1.1.1999 fordelt etter størrelsen: Inntekts- og formuesundersøkelsen 1998, metode (I) etterstratifisering på nettoutvalget, metode (II) etterstratifisering på nettoutvalget og imputeringen av registerverdier, og metode (III) etterstratifisering på nettoutvalget og imputeringen under antagelsen om at frafallet er uavhengig av familiestørrelsen gitt frafallsklassen.

Antall i tusen	Antall personer i husholdning					Total
	1	2	3	4	≥ 5	
IF98	829	583	263	242	131	2049
Metode (I)	832	626	254	240	121	2073
Metode (II)	888	598	253	240	122	2101
Metode (III)	884	620	249	235	121	2108
<b>Andel i prosent</b>						
IF98	40.5	28.5	12.8	11.8	6.4	100
Metode (I)	40.1	30.2	12.3	11.6	5.9	100
Metode (II)	42.3	28.5	12.1	11.4	5.8	100
Metode (III)	41.9	29.4	11.8	11.1	5.7	100

Tabell 12: Énfamiliehusholdnings fordeling ifølge familieregisteret i 1.1.1999 (i tusen).

	Antall personer i husholdning					Total
	1	2	3	4	≥ 5	
Aleneboende under 45 år	436	-	-	-	-	436
Aleneboende 45 - 64 år	216	-	-	-	-	216
Aleneboende 65 år og over	320	-	-	-	-	320
Par u. hj. barn, eldste person under 45 år	-	30	-	-	-	30
Par u. hj. barn, eldste person 45 - 64 år	-	131	-	-	-	131
Par u. hj. barn, eldste person 65 år og over	-	177	-	-	-	177
Par m. hj. barn, yngste barn 0 - 6 år	-	-	73	108	74	256
Par m. hj. barn, yngste barn 7 - 17 år	-	-	37	101	54	193
Par m. hj. barn, yngste barn 18 år og over	-	-	99	39	5	143
Mor m. yngste hj. barn 0 - 17 år	-	62	38	10	2	113
Far m. yngste hj. barn 0 - 17 år	-	9	4	1	0	15
Mor/far m. yngste hj. barn 18 år og over	-	55	10	1	0	67
Total	972	464	262	261	136	2095

- Forskjellen mellom metode (I), (II) og (III) ligger i frafallsbehandlingen. Uten imputering underestimerer vi andelen en-persons husholdninger. Belsby and Bjørnstad (1997) påviste det samme. Faktisk var forskjellen der enda større pga. enda høyere frafallsandel. Marginalt sett er ikke forskjellen mellom metode (II) og (III) veldig stor. Men som vi tidligere har vært inne på, imputeringen av registerverdier vil nødvendigvis innføre en del skjevhet når det gjelder forholdet mellom familier i registeret og den faktiske husholdningen.
- Forskjellen mellom IF98 og Metode (II) ligger i datagrunnlaget og vektingen. Det er imidlertid vanskelig helt klart å skille mellom de to effektene, til tross for at vi kjenner til en del forskjell mellom AKU og LKP. Vi vet nemlig ikke om de andre undersøkelsene i IF98, dvs. uten å ta med LKP, hadde gitt lavere andel små husholdninger enn AKU eller ikke. At andelen en-persons husholdninger ifølge IF98 er omtrent den samme som metode (I), kan tyde på at den lave andelen i IF98 skyldes i hvert fall delvis vektingen. Dette fordi
  - i IF98 har man også brukt imputering av registerverdier som i metode (II), og
  - vektene i IF98 ga færre en-persons familier sammenlignet med familierregisteret.

## A Imputering

I dette avsnittet beskriver vi imputeringsmetoden under antagelsen om at frafallet er uavhengig av familiestørrelsen gitt frafallsklassen. Betegn med  $X$  familiestørrelsen, med  $Y$  frafallsklassen, og med  $R$  frafallet. La  $R = 1$  betegne "svar", og  $R = 0$  "ikke-svar". Anta at  $R$  er uavhengig av  $X$  gitt  $Y = y$ . En slik modell om den simultane fordelingen til  $(X, Y, R)$  er kjent om en grafisk modell (Lauritzen, 1996), som igjen er en underklasse av de såkalte log-linære modellene (Bishop, Fienberg, and Holland, 1975). Vi har

$$f(x, y, r) = f(x, y)f(r|x, y) = f(x, y)f(r|y). \quad (4)$$

La  $n_{xy}$  være antallet referansepersoner i nettoutvalget som har  $(X, Y) = (x, y)$ . La  $m_{xy}^*$  være det tilsvarende antallet blant frafallet. Vi observerer ikke  $m_{xy}^*$  bortsett fra marginalen  $m_x = \sum_y m_{xy}^*$ . Tabellen under viser dataene definert på denne måten.

	Observert ( $R = 1$ )				<i>(Frafall)</i>	Frafall ( $R = 0$ )			
	$Y = 1$	$Y = 2$	$\dots$	$Y = H$		$Y = 1$	$Y = 2$	$\dots$	$Y = H$
$X = 1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1H}$	$(m_1)$	$m_{11}^*$	$m_{12}^*$	$\dots$	$m_{1H}^*$
$X = 2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2H}$	$(m_2)$	$m_{21}^*$	$m_{22}^*$	$\dots$	$m_{2H}^*$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$X \geq K$	$n_{K1}$	$n_{K2}$	$\dots$	$n_{KH}$	$(m_H)$	$m_{K1}^*$	$m_{K2}^*$	$\dots$	$m_{KH}^*$

For imputering under modellen (4) benytter vi følgende iterativ algoritme. For nåværende  $\{m_{xy}^*\}$ ,

- estimer  $\hat{f}(x, y) = (n_{xy} + m_{xy}^*) / (n + m)$ , og  $\hat{f}(r = 0|y) = m_{\cdot y}^* / (n_{\cdot y} + m_{\cdot y}^*)$ , der  $n = \sum_{x,y} n_{xy}$ ,  $m = \sum_x m_x$ ,  $n_{\cdot y} = \sum_x n_{xy}$  og  $m_{\cdot y}^* = \sum_x m_{xy}^*$ ; og  $\hat{f}(x, y, r = 0) = \hat{f}(x, y)\hat{f}(r = 0|y)$ ;
- oppdater  $m_{xy}^* = m_x \hat{f}(y|x, r = 0)$ , der  $\hat{f}(y|x, r = 0) = \hat{f}(x, y, r = 0) / \sum_y \hat{f}(x, y, r = 0)$ .

Algoritmen ovenfor skal gi oss det maksimum likelihood estimatet under modellen (4). For å kontrollere dette kan vi beregne log-likelihooden etter hver iterasjon, og se om den er monoton stigende,

$$l = \sum_{xy} n_{xy} \log \hat{f}(x, y, r = 1) + \sum_x m_x \log \left\{ \sum_y \hat{f}(x, y, r = 0) \right\}.$$

## B Variansberegning (I)

Anta at  $T$  er antallet en-persons husholdninger i populasjonen. La  $\hat{T}$  være et linært estimat basert på vektene  $\{w_i; i \in s\}$  der  $s$  er nettoutvalget. Generelt kan  $\hat{T}$  uttrykkes som

$$\hat{T} = \sum_{i \in s} w_i z_i,$$

der  $Z_i = 1$  hvis person  $i$  lever i en-persons husholdning or  $Z_i = 0$  ellers. (I tilfellet  $T$  står for en annen populasjonestotal, endres kun definisjonen på  $Z_i$ . Ellers er beregningsmetoden den samme som beskrevet under.) La  $s^*$  betegne frafallet. Betrakt nå  $\hat{T}(s, s^*)$  som en funksjon av både  $s$  og  $s^*$ .

La  $E_s$  og  $Var_s$  betegne forventning og varians overfor  $s$ . La  $E_{s^*}$  og  $Var_{s^*}$  betegne forventning og varians overfor  $s^*$ . Variansen til  $\hat{T}$  er gitt som

$$Var(\hat{T}) = E_{s^*}[Var_s(\hat{T}|s^*)] + Var_{s^*}(E_s[\hat{T}|s^*]).$$

La  $\hat{s}^* = \hat{E}[S^*|s]$  være det estimerte betingete forventete frafallet. Vi utvikler  $g(s^*) = Var_s(\hat{T}|s^*)$ , som er en funksjon av  $s^*$ , rundt  $\hat{s}^*$ , og tar med det første ledd, slik at

$$E_{s^*}[Var_s(\hat{T}|s^*)] \doteq E_{s^*}[g(\hat{s}^*)] = g(\hat{s}^*) = Var_s\{\hat{T}(S, \hat{s}^*)\}.$$

La  $s_h$  og  $s_h^*$  betegne henholdsvis etterstratum  $h$  i nettoutvalget og frafallet, slik som definert i Tabell 5. Så lenge  $s^*$  holdes fast ved  $\hat{s}^*$ , er den initiale vekten (1) uforandret, slik at vi kan betinge på størrelsen til  $s_h$  og  $s_h^*$ . Dette gjør vekten  $w_i$  til konstant, og variasjonen kommer fra  $Z_i$  alene. La  $s_{h,1}$  betegne alle personer i etterstratum  $h$  i nettoutvalget som lever i en-persons husholdninger. La  $a_i$  være den initiale vekten (1) til referanseperson  $i$ . Andelen

$$\hat{p}_1 = \left( \sum_{i \in s_{h,1}} a_i \right) / \left( \sum_{i \in s_h} a_i \right)$$

er et estimat for sannsynligheten at en vilkårlig trukket person fra populasjonstratum  $h$  lever i en-persons husholdninger, siden alle der er trukket med like stor sannsynlighet. Vi har,

$$\begin{aligned} E_{s^*}[Var_s(\hat{T}|s^*)] &\doteq Var_s\{\hat{T}(S, \hat{s}^*)\} = Var\left(\sum_h \sum_{i \in s_h} w_i Z_i\right) \\ &= \sum_h \left\{ \sum_{i \in s_h} w_i^2 Var(Z_i) \right\} = \sum_h \left( \sum_{i \in s_h} w_i^2 \right) \hat{p}_1 (1 - \hat{p}_1). \end{aligned}$$

Det gjenstår å beregne  $Var_{s^*}(E_s[\hat{T}|s^*])$ . Anta at  $\hat{T}(S, s^*)$  er forventningsrett for gitt  $s^*$ . Da kan vi ganske enkelt estimere  $E_s[\hat{T}|s^*]$  med  $\hat{T}(s, s^*)$  for gitt  $s$  og  $s^*$ . Vi har, dermed,

$$Var_{s^*}(E_s[\hat{T}|s^*]) \doteq Var_{s^*}\{\hat{T}(s, s^*)\} \doteq (B-1)^{-1} \sum_{k=1}^B \{t_{(k)} - \bar{t}_{(\cdot)}\}^2,$$

der  $\bar{t}_{(\cdot)} = B^{-1} \sum_{k=1}^B t_{(k)}$ , og  $t_{(1)}, \dots, t_{(B)}$  som kommer fra den følgende Bootstrap algoritmen. La  $B$  være antallet Bootstrap, der  $B$  anbefales å være minst 250. For  $k = 1, \dots, B$ ,

- trekk  $m_{xy}^*$  fra den estimerte multinomiske fordelingen betinget på  $\sum_y m_{xy}^* = m_x$ ;
- beregn  $t_{(k)} = \hat{T}(s, s_{(k)}^*)$  for  $s$  og  $s_{(k)}^*$  trukket ovenfor.

## C Variansberegning (II)

Metoden beskrevet i Appendix B deler den totale variansen i to komponenter, som henholdsvis skyldes variasjonen i utvalget og usikkerheten i imputeringen. Men den kan ikke brukes til å estimere variansen til det totale antallet husholdninger i populasjonen. Betrakt en linær estimator i formen  $\sum_{i \in s} w_i z_i$ , der  $Z_i$  indikerer om enhet  $i$  har en bestemt karakter eller ikke. Betinging på størrelsene til etterstrata i utvalget gjør  $w_i$  til konstant, slik at all variasjon kommer fra  $Z_i$  alene. Men  $Z_i \equiv 1$  hvis det er selve størrelsen til populasjonen vi estimerer, i hvilket tilfelle metoden bryter ned.

Efron (1994) beskrevet en ikke-parametrisk Bootstrap metode for et enkelt tilfeldig utvalg med frafall. Vi kan anvende Efrons metode betinget på familiestørrelsen ifølge registeret. La  $n_{xy}$  og  $m_x$  være som definert i Appendiks A. La  $\hat{f}_x$  være den empiriske tettheten betinget på  $X = x$ , dvs.

$$\hat{f}_x = (n_{x1}, \dots, n_{xH}, m_x) / \left( \sum_y n_{xy} + m_x \right).$$

Siden alle personer trekkes med like stor sannsynlighet betinget på  $X = x$ , er  $\hat{f}_x$  et estimat for parametrene til den multinomiske fordelingen med frafall som en egen kategori, dvs. i tillegg til  $Y = 1, \dots, H$ . Vi kan bruke den estimerte multinomiske fordelingen til å simulere den prosessen som genererer data  $(n_{x1}, \dots, n_{xH}, m_x)$  gjennom utvalg og frafall.

La  $B$  være antallet Bootstrap, der  $B$  anbefales å være minst 250. For  $k = 1, \dots, B$ ,

- trekk  $(n_{x1}^*, \dots, n_{xH}^*, m_x^*)$  fra den betingete multinomiske fordelingen  $(\sum_y n_{xy} + m_x, \hat{f}_x)$ ;
- imputer  $(m_{x1}^*, \dots, m_{xH}^*)$  der  $\sum_y m_{xy}^* = m_x^*$  som beskrevet i Appendiks A;
- lag den kryss-tabellen i bruttoutvalget etter familie- og husholdningsstørrelsen basert på både  $n_{xy}^*$  og  $m_{xy}^*$ , betegnet med  $u_{xj}^*$  der  $j = 1, \dots, 4$  eller  $\geq 5$ ;
- estimer antallet personer i populasjonen fra en  $x$ -persons familie, men lever i en  $j$ -persons husholdning, med  $(N_x u_{xj}^*) / (\sum_j u_{xj}^*)$ , og dermed antallet  $j$ -persons husholdninger i populasjonen, betegnet med  $t_{(k,j)}$ , og det totale antallet husholdninger, betegnet med  $t_{(k)} = \sum_j t_{(k,j)}$ .

Variansen til antallet  $j$ -persons husholdninger i populasjonen beregnes som vanlig basert på Bootstrap  $(t_{(1,j)}, \dots, t_{(B,j)})$ , og lignende for variansen til det totale antallet husholdninger. Legg merke til at metoden tar hensyn til variasjonen både i utvalget og frafallet, med uten å dele variansen i de to.

Strengt tatt burde vi ha trukket  $n_{xy}^*$  husholdninger blant de  $n_{xy}$  husholdningene i nettoutvalget ved hver Bootstrap iterasjon. Deretter skulle vi etterstratifisere den simulerte nettoutvalget ifølge Tabell 5, for å estimere antallet husholdninger. Bootstrap simuleringen vår er derfor en forenkling, som bruker vesentlig mindre tid. For AKU og LKP i 1999, fikk vi følgende standardavvik (i tusen) for antallet husholdninger etter størrelsen: 11, 7, 4, 3, og 2, som stemmer bra med metoden i Appendiks B (Tabell 2). Vi fikk 7 tusen som standardavviket til det totale antallet husholdninger.

## Referanser

- Belsby, L. and Bjørnstad, J.F. (1997). Modeling and estimation methods for household size in the presence of nonresponse. Technical report, Statistics Norway (Discussion Papers 206).
- Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W. (1975). *Discrete Multivariable Analysis: Theory and Practice*. Cambridge, MA: MIT Press.
- Efron, B. (1994). Missing data, imputation, and the Bootstrap (with discussion). *J. Amer. Statist. Assoc.*, **89**, 463–479.
- Lauritzen, S.L. (1996). *Graphical Models*. Clarendon Press, Oxford.
- Zhang, L.-C. (1999). Opplegg til en statistikk over familie- og husholdningsfordelingen i den norske befolkningen. Technical report, Statistisk sentralbyrå (Notater 99/62).



## De sist utgitte publikasjonene i serien Notater

- 2001/20 B. Indahl, D.E. Sommervoll og J. Aasness: Virkninger på forbruksmønster, levestandard og klimagassutslipp av endringer i konsumentpriser. 27s.
- 2001/21 A. Barstad: På vei mot det gode samfunn? Utredning til Finansdepartementet i forbindelse med arbeidet med nytt Langtidsprogram, 2002-2005. 363s.
- 2001/23 L. Østby: Beskrivelse av nyankomne flykningers vei inn i det norske samfunnet. Notat til Lovutvalget som skal utrede og lage forslag til lovgivning om stønad for nyankomne innvandrere. 32s.
- 2001/24 T. Nøtnæs: Innføring i bruk av fokusgrupper. 22s.
- 2001/25 J. Fosen, A.G. Hustoft og B.O. Lagerstrøm: Ny spørresekvens for å identifisere husholdninger i utvalgsundersøkelser. 29s.
- 2001/26 H.C. Hougen: Undersøkelse om folat-kunnskap blant kvinner i fertil alder: Dokumentasjonsrapport. 17s.
- 2001/27 Ø. Kleven og O.F. Vaage: Medieundersøkelsen 1999: Dokumentasjonsrapport. 49s.
- 2001/28 J. Heldal og J. Fosen: Statistisk konfidensialitet i SSB: Et diskusjonsnotat. 41s.
- 2001/29 B.O. Lagerstrøm: Bruk av folkehøgskoler, 2000/2001. 77s.
- 2001/30 R. Nygaard Johnsen: Undersøking om foreldrebetaling i barnehagar, januar 2001. 40s.
- 2001/31 R. Choudhury: Brukerveiledning for AMEN. 100s.
- 2001/32 R. Choudhury: Datagrunnlaget for AMEN: Teknisk dokumentasjon. 20s.
- 2001/33 G. Dahl og J. Johansen: FD - Trygd: Dokumentasjonsrapport. Sysselsetting. 1992-1997. 98s.
- 2001/34 L. Vågane: Samordnet levekårsundersøkelse 2000 - tverrsnittundersøkelsen: Dokumentasjonsrapport. 82s.
- 2001/35 J. Holmøy: Årsrapport 2000: Kontaktutvalget for helse- og sosialstatistikk. 36s.
- 2001/36 B. Astad og A.L. Brathaug: Kommunenes utgifter til primærlegetjenesten 1999: Evaluering av fastlegeordningen - førdata om kommunenes utgifter. 54s.
- 2001/37 B. Rosnes: Kommunale gebyrer knyttet til bolig . Januar 2001. 29s.
- 2001/38 K.I. Bøe og S. Lien: FD - Trygd: Dokumentasjonsrapport. Attføringspenger. 1992-1999. 80s.
- 2001/39 P.E. Lilleås, og R. Nygaard Johnsen: Foreldrebetaling i kommunale og private barnehager i perioden 1992-2001. 23s.
- 2001/40 B.R. Joneid: KOSTRA GenRev 2000. Malverk for generelt revisjonssystem - KOSTRA-data. Del 1: Håndbok for bruk av malverket ved generering av applikasjoner. Del 2: Systemdokumentasjon av GenRev-malverket. 46s.
- 2001/41 T.M. Normann: Bostedets betydning. Dokumentasjonsrapport. 36.
- 2001/42 B.A. Holth og J.T. Prangerød: Lederskapsundersøkelsen 2000. Dokumentasjonsrapport. 245s.
- 2001/43 T. Fæhn, J.-A. Jørgensen, T. Åvitsland, W. Drzwi: Næringsfordelte skatteutgifter. Dokumentasjon av beregningsgrunnlaget i ERA-beregningene 1998. 47s.
- 2001/44 KOSTRA - VAR-rapport 2001. 34s.
- 2001/45 KOSTRA - Kulturminne, natur og nærmiljø. 38s.
- 2001/46 KOSTRA: Arbeidsgrupperapporter 2001 - hefte 1. 58s.
- 2001/47 KOSTRA: Arbeidsgrupperapporter 2001 - hefte 2. 46s.
- 2001/48 Rapport fra arbeidsgruppa for KOSTRA - Samferdsel. 27s.