



Liv Belsby

Vekter i Forbruksundersøkelsen

Notater

Forord

Arbeidet har blitt utført i nært samarbeide med seksjon 350 ved Eiliv Mørk. Datagrunnlaget er tilrettelagt av Per Olav Lande og Ove Larsen. Tilleggsinformasjon for vektingen er hentet fra Folke- og boligtellingsen 2001 ved hjelp av Espen Andersen. Inntektsdata har blitt tilrettelagt av Vidar Pedersen.

Johan Heldal har laget kalibreringsprogrammet som blir brukt til beregning av vektene. Han har i tillegg hjulpet meg med faglige innspill og kommentarer. Aslaug Hurlen Foss har hjulpet meg med opplysninger om Folke- og boligtellingsen 2001.

Innhold

1. Innledning	3
2. Estimeringsmetode	3
2.1. Responssannsynligheten.....	4
2.2. Justering for ulik representasjon over året.....	5
2.3. Kalibrering.....	6
3. Vurdering av vektene	8
3.1. Estimert fordeling i befolkningen for kosthusholdning.....	9
4. Mulige forbedringer av vektene	9
Referanser	10
Vedlegg	11
A. Oversikt over beregningsprogrammene	11
B. Programmer	14
De sist utgitte publikasjonene i serien Notater	27

1. Innledning

Siden 1974 har Statistisk sentralbyrå gjennomført årlige forbruksundersøkelser. Hovedformålet med undersøkelsene har vært å gi en detaljert oversikt over private husholdningers forbruk, som grunnlag for en eventuell ajourføring av konsumprisindeksen. Det er videre tatt sikte på å kartlegge forbruket i forskjellige grupper av husholdninger. Husholdningene klassifiseres etter husholdningstype (ny standard), inntekt, sosioøkonomisk status, landsdel og bostedsstrøk, størrelse av forbruksutgift mv.¹

Undersøkelsespopulasjonen til FU består av alle private kosthusholdninger i Norge (NOS C 677). Fra og med 1992 har personer i Norge i aldersgruppen 0-79 år vært trekkeenheter (trekkepopulasjon, utvalgsramme). Et utvalg av personer trekkes fra det sentrale personregisteret og kosthusholdningene som disse tilhører danner utvalget til undersøkelsen.

I dette notatet beskriver vi estimatoren som blir brukt for å beregne vektene. Vektene gir estimater for totalen for populasjonen og gjennomsnitt for husholdningene. Estimatoren tar hensyn til at det er skjevt frafall i forbruksundersøkelsen (FU). Dessuten justeres vektene slik at føringsperioden blir tilnærmet likt representert. Disse vektene kalibreres deretter med hensyn på husholdningsfordelingen for populasjonen basert på husholdningsdefinisjonen i Folke- og bolig tellingen 2001 (FoB2001). Justeringen for lik representasjon av føringsperiode, og kalibreringen, er en utvidelse i forhold til det eksisterende vektoplegget, Belsby (1995).

Uttesting av vektoplegget er gjort med data for 2001.

2. Estimeringsmetode

Som i tidligere forbruksundersøkelser, er husholdningenes vektor basert på en modifisert Horwitz-Thompson (HT) estimator, dvs. den inverse av

$$\Pr(\text{inkludert}) = \Pr(\text{trukket}) \cdot \Pr(\text{respons} | \text{trukket}),$$

hvor $\Pr(\text{trukket})$ er bestemt av utvalgsplanen. Som nevnt ovenfor er trekkeenheten i FU personer. De personene som tilhører samme kosthusholdning som personen som blir trukket, inkluderes i utvalget. Det medfører at trekkesannsynligheten er proporsjonal med antall personer i kosthusholdningen. La N være antallet i populasjonen, dvs. antallet personer i aldersgruppen 0-79 år for det aktuelle året, n antallet som er trukket (- avgang) og j antallet personer i kosthusholdningen. Da blir trekkesannsynligheten $(n/N) \cdot j$. For FU 2001 var størrelsen på trekkepopulasjonen $N = 4\,302\,862$ personer og utvalgsstørrelsen i undersøkelsespopulasjonen n (- avgang) = 2171 husholdninger. Merk at trekkeprosedyren gir overrepresentasjon av store husholdninger i forhold til fordelingen av husholdninger i populasjonen. (Mens trekkeprosedyren gir representativt utvalg i forhold til fordelingen av antall *personer* tilhørende husholdninger med varierende størrelse.)

Motivasjonen for å justere for responssannsynlighet er det store og til dels skjeve frafallet i FU. I 2001 var for eksempel frafallet 52%.

¹ Lodberg-Holm, A & Mørk, E. (2001). Forbruksundersøkelsen 1996-1968. Notater, 2001/22.

2.1. Responssannsynligheten

Responssannsynligheten estimeres fra bruttoutvalget. I tettbygde strøk svarte 46% , og i spredtbygde strøk svarte 56%. Denne markante, og fra år til år gjentagende forskjellen gjør at vi lar responssannsynligheten avhenge av tett- og spredtbygde strøk. Adressen til husholdninger bestemmer om husholdningen defineres å høre til i tett- eller spredtbygde strøk.

Tidligere har vi også tatt hensyn til at responssannsynlighet for husholdninger avhenger av sammensetningen i husholdningen, som for eksempel av antall personer i husholdningen. For nettoutvalget kjenner vi faktisk antall personer i husholdningen, mens vi bruker registrert antall for husholdninger hvor vi manglet faktisk antall. I FU 2001 svarer 31% av husholdningene med én person, mens svarprosenten er 52%-55% for husholdninger med flere personer. Vi vet ikke med sikkerhet hvor mange personer som bor i de husholdningene som ikke svarer. Dette øker usikkerheten for estimert svarsannsynlighet som funksjon av antall personer i husholdningen.

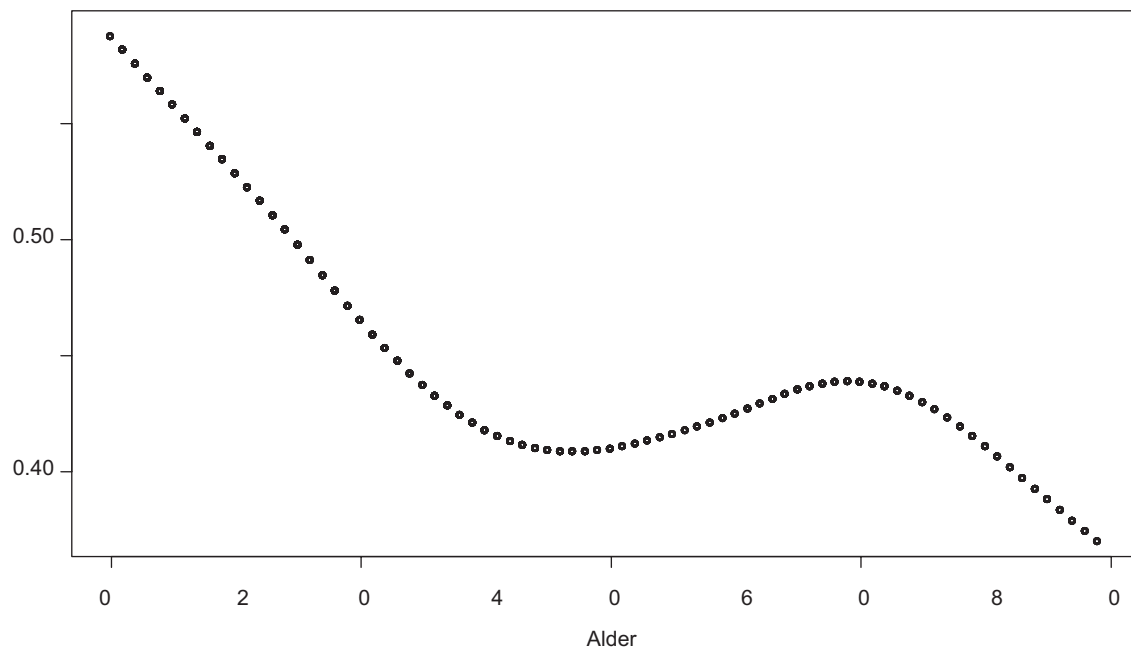
Selv om svarsannsynligheten avhenger av husholdningssammensetningen, lar vi ikke den estimerte responssannsynligheten avhenge av husholdningssammensetning, siden kalibreringen med hensyn på bohusholdning gir lignende effekt. Kalibreringen er beskrevet nedenfor.

Vi vet at alder påvirker svarsannsynligheten. Siden enheten i FU er husholdning, må vi velge en funksjon av alderen til medlemmene i husholdningen. Vi har valgt yngste persons alder, som vi mener karakteriserer aldersstrukturen i husholdningen ganske godt, uten at vi har estimert dette eksplisitt. Alder til yngste person i husholdningen vil ofte identifisere barnefamilier. Dessuten vil trolig yngste barns alder ofte påvirke hvor mye familien er hjemme, som igjen påvirker svarsannsynligheten. For par uten barn vil ofte alderen til den eldste være korrelert med den yngstes alder. Figur 1 nedenfor illustrerer hvordan svarsannsynligheten avhenger av alder til yngste medlem i husholdningen. Svarsannsynligheten er estimert med en ikke-parametrisk modell, nærmere bestemt med en generaliserte additiv modell, ofte kalt GAM², som er implementert i analyse programmet S+. Vi ser at husholdninger med barn har størst svarsannsynlighet. Svarsannsynligheten varierer ikke så mye når alderen varierer mellom ca. 25 år og 60 år. Etter 60 år avtar svarsannsynligheten. Dette ligner på det mønsteret vi fant i Tidsbruksundersøkelsen 2000/01, Belsby(2003).

² Denne typen ikke-parametriske, moderne regresjonsmetoder gir bedre mulighet for grafisk fremstilling av data enn tradisjonelle metoder. I tradisjonelle logistiske regresjonsmetoder må man bestemme seg for funksjonsformen før estimeringen. Ofte er det en ulempe fordi man ikke har nok kunnskap om hva som er en god transformasjon. Ved å velge en vilkårlig transformasjon, kan en risikere å ikke finne en god tilpassning til data. Moderne regresjonsmetoder, se Venables og Ripley(1994), beregner en ikke-parametrisk funksjon ved hjelp av spline-metoder. Fordelen er at man slipper å transformere, inkludert lage grupper, basert på mer eller mindre godt funderte hypoteser om funksjonsformen.

Figur 1. Estimert svarsannsynlighet som funksjon av alder til yngste person i husholdningen

Svarsannsynlighet



Alderen til yngste medlem i husholdningen er heller ikke inkludert i svarsannsynlighet, som inngår i vektberegningene.

2.2. Justering for ulik representasjon over året

Det er 26 føringsperioder, hver på 14 dager. Disse grupperes i fem perioder av ulik lengde. Vi bestemte gruppeinndelingen ut i fra hva vi antok å være forholdsvis homogene handleperioder, og at denne justeringen ikke skulle gi ekstreme effekter på vektene. De justerte, modifiserte HT vektene justeres videre slik at summen av vektene for en handleperiode er proporsjonal med antallet opprinnelige føringsperioder, l , i handleperioden.

La w_{ij} betegne de modifiserte HT vektene for husholdning j i handleperiode i . De modifiserte HT vektene i handleperiode i med antall opprinnelige føringsperioder lik l , justeres ved å multiplisere vektene i perioden med konstanten

$$k_i(l) = \left(\sum_{kj} w_{kj} / 26 \right) \cdot \left(l / \sum_j w_{ij} \right).$$

Tabell 1 viser hvordan husholdningene fordeler seg i bruttoutvalget og nettoutvalget over de grupperte føringsperiodene. Tabellen viser også resultatet av justeringen.

Tabell 1. Fordeling på føringsperioder i det vektete nettoutvalget og uvektet fordeling i brutto- og nettoutvalget

Handleperiode	Antall opprinnelige føringsperioder	Vektet fordeling i %	Fordeling i bruttoutvalget	Fordeling i nettoutvalget
1/1 - 25/3	6	23,08	21,83	21,79
26/3 - 17/6	6	23,08	21,46	22,26
18/6 - 26/8	5	19,23	18,33	18,27
27/8 - 4/11	5	19,23	20,27	19,88
5/11-31/12	4	15,38	18,10	17,79
I alt	26	100,00	100,00	100,00

Den vektete fordelingen svarer til jevn fordeling av observasjoner for føringsperiodene. Husholdningene i bruttoutvalget skal i utgangspunktet fordeles likt over året, og derfor ha en fordeling omtrent som den vektete fordelingen. Likevel avviker fordelingen i bruttoutvalget fra den vektete fordelingen. Grunnen til dette er at husholdningene har anledning til å forskyve føringsperioden et par uker, av og til lenger. Det er ikke anledning til å forskyve føringsperioden til over nyttår. Det er trolig grunnen til at 18% fører i siste periode, mens det ville ha vært 15% om dagbokføringen hadde vært likt fordelt over året.

2.3. Kalibrering

Forbruket i en husholdning avhenger av antall voksne, antall barn, barnas alder o.l. Dessuten varierer svarsansynligheten for ulike typer husholdninger. Dette og det omfattende datamaterialet for husholdninger fra Fob2001 tilsier at de justerte, modifiserte HT vektene bør kalibreres med hensyn på husholdningsfordeling, dvs. justeres slik at vektene ``tvinges`` til å følge samme fordeling som i populasjonen.

Enheden i Forbruksundersøkelsen er *kosthusholdning*. Det er rimelig å anta at forbruksmønsteret avhenger sterkere av kosthusholdning enn av bohusholdning. Ideelt burde vektene derfor kalibreres med fordeling for kosthusholdninger i populasjonen. Dessverre gir Fob2001 ikke tall for kosthusholdning, men for *bohusholdning*³. Kalibreringsvariabel er bohusholdning basert på definisjonen i Fob. Tabellen nedenfor viser husholdningsfordelingen de justerte, modifiserte HT vektene kalibreres med.

³ Husholdning

Til samme *bohusholdning* regnes personer som er fast bosatt i samme bolig (boenhet) eller institusjon. FoB2001 gir ingen informasjon om *kosthusholdninger*, dvs. om personer har felles kost.

En *privathusholdning* består av personer som er bosatt i samme bolig, der denne boligen ikke er en institusjon. En *felleshusholdning* består av personer som er bosatt og får dekket sine behov for kost, pleie og omsorg på en institusjon. Ansatte som er bosatt i/ved en felleshusholdning (institusjon, militærforlegning mv.), er alltid regnet som bosatt i privathusholdning. FoB2001 gir ikke statistikk for felleshusholdninger, men viser tallet på personer som ikke bor i privathusholdninger. Gruppen kalles bosatte i *andre husholdninger* og består av personer bosatt i felleshusholdninger samt personer uten fast bopel.

I FoB2001 lages primært statistikk basert på formell bostedsadresse (dvs. adresse ifølge Folkeregisteret), men gjennom bolig tellingen er det også samlet inn opplysninger om faktisk adresse på studiestedet for borteboende studenter og institusjonsadresse for personer som faktisk bor på alders- og sykehjem (det er ikke samlet inn tilsvarende opplysninger for personer som bor på andre typer institusjoner). Sammensetningen til *husholdninger etter formell adresse* har skjedd ut fra opplysninger gitt på boligskjema og adresseopplysninger fra Folkeregisteret. Husholdninger kan bestå av én eller flere familier. Personer som tilhører samme familie, tilhører også samme husholdning. For at personer som ikke tilhører samme familie skal utgjøre én husholdning, må de være registrert på samme adresse i Folkeregisteret og ha oppgitt på spørreskjemaet at de bor sammen. Dette betyr at ugifte, borteboende studenter som er registrert på foreldrenes adresse, regnes med i foreldrenes husholdning. Bare personer som er registrert bosatt på en institusjon, regnes å tilhøre en felleshusholdning. Mange personer som faktisk bor på institusjoner, f. eks. alders- og sykehjem, er registrerte som bosatte i privatbolig (sammen med ektefelle).

Én svakhet er manglende fullstendig oversikt over de som bor på institusjon³, for eksempel de som "sitter inne".

Tabell 2. Befolkning i aldersgruppen 0-79 år 3.november 2001 fordelt på husholdningstype-bohusholdning

Husholdningstype - bohusholdning ²	Antall personer	Fordeling i %
Enslig	629 886	14,64
Par uten barn	769 221	17,88
Par med barn, yngste barn 0-6 år	1 015 675	23,60
Par med barn, yngste barn 7-19 år	955 933	22,22
Par med barn, yngste barn 20 år og over	338 761	7,87
Andre	593 386	13,79
I alt	4 302 862	100,00

Som tabellen antyder, kalibrers utvalget på personnivå. De justerte, modifiserte HT vektene er identiske for personer som tilhører samme husholdning (kosthusholdning). Men alle personene i husholdningen hører ikke nødvendigvis til samme bohusholdning. I så fall har kalibreringsvariabelen forskjellige verdier innen én husholdning. De kalibrerte vektene vil da avvike innen husholdningen. For å unngå dette benyttes en metode beskrevet på side 12 i Heldal (1992). La $Y_{ij}(h) = 1$ hvis person j i husholdning i tilhører en bohusholdning med størrelse h personer. Vi innfører hjelpevariabel $I_i(h) =$ gjennomsnittet av $Y_{ij}(h)$ over personene i husholdning i , dvs. andelen av personer i husholdning i som tilhører bohusholdning med størrelse h . Eksempelet nedenfor illustrerer metoden for en kosthusholdning med tre personer, hvorav én person tilhører en bohusholdningsgruppen 'enslig' og de to andre tilhører bohusholdningsgruppen 'par uten barn'. I eksempelet har vi antatt at vi, som i tabell 2, har seks bohusholdningsgrupper.

Person j	$Y_{ij}(1)$	$I_i(1)$	$Y_{ij}(2)$	$I_i(2)$	$Y_{ij}(3)$	$I_i(3)$	$Y_{ij}(4)$...
1	0	1/3	1	2/3	0	0	0	...
2	0	1/3	1	2/3	0	0	0	...
3	1	1/3	0	2/3	0	0	0	...
Sum	1	1	2	2	0	0	0	...

Hjelpevariabel $I_i(h)$ brukes i kalibreringen.

3. Vurdering av vektene

Vi har estimert aldersfordeling og inntektsfordeling og sjekket de estimerte fordelingene mot fordelingen for befolkning i aldersgruppen 0-79 år pr. 3. november 2001. Vi har ikke beregnet varianser og konfidensintervall for estimatene. Derfor baseres kommentarene bare på sammenligninger mellom selve estimatene og ikke på om forskjellene er signifikante.

Tabell 3. Aldersfordeling for befolkning i aldersgruppen 0-79 år 31.12.2001 og aldersfordeling i det vektete utvalget for FU2001

Aldersgruppe	I befolkningen	Kalibrerte, føringsperiode justerte, modifiserte HT vekter	Føringsperiode justerte, modifiserte HT vekter	Antall observasjoner i nettoutvalget
0-9	13,84	14,87	14,29	676
10-19	13,22	13,84	13,38	606
20-29	13,63	12,06	12,19	303
30-39	16,02	16,51	16,97	576
40-49	14,61	15,76	15,75	548
50-59	13,28	12,72	12,30	296
60-69	8,23	8,50	8,82	156
70-79	7,16	5,40	5,98	85
80- ¹		0,35	0,32	8
I alt	100,00	100,00	100,00	3 254

¹ Er med fordi de tilhører husholdninger som har blitt trukket ut via personer i aldersgruppen 0-79 år.

Hovedinntrykket fra tabellen er at ingen av vektene gir særlig skjev aldersfordeling siden de estimerte fordelingene samsvarer ganske bra med fordelingen populasjon. Estimaterne for aldersgruppen 30-39 år og 40-49 tyder på at disse aldersgruppene er litt overrepresentert. Mens personer som er 70 år eller mer er underrepresentert. Dette kan forklares av den estimerte svarsannsynligheten vist i figur 1 ovenfor. FU inkludere ikke personer på institusjon. Institusjonsbeboere er bare delvis fjernet i beregningene av aldersfordelingen i populasjonen. Dette betyr at estimert andel i aldersgruppen 70-79 år i utvalget bør være litt lavere enn i populasjonen. Videre ser vi at kalibreringen har liten effekt på den estimerte aldersfordelingen. Det er størst avvik mellom estimert fordeling og fordelingen i populasjonen eldre.

Tabell 4. Samlet inntektsfordeling for befolkningen 0-79 år for 1999 og det vektete utvalget for FU2001

Inntektsdesil i %	Intervall i kr	Estimert andel i % Kalibrerte, føringsperiode justerte, modifiserte HT vekter	Estimert andel i % Føringsperiode justerte, modifiserte HT vekter	Antall observasjoner
Missing	.	2,80	2,93	114
0-30	0 - 55 484	30,06	31,13	1286
30 -40	55 485 - 107 809	7,26	7,08	183
40-50	107 810 - 153 012	8,96	8,71	225
50-60	153 013 - 199 511	8,73	8,54	230
60-70	199 512 - 241 025	10,23	10,09	289
70-80	241 026 - 284 128	10,54	10,46	322
80-90	284 129 - 355 368	10,49	10,31	290
90-100	355 369 -	10,92	10,76	315
Alle		100,00	100,00	3254

Ved beregning av desilene hadde vi ikke mulighet for å utelate personer som bor på institusjon. Ideelt sett burde disse vært utelatt siden de ikke er med i FU. De tre første desilene er slått sammen siden det er så mange personer med 0 i samlet inntekt. I denne gruppen er det mange barn og unge. Vektene gir så å si ikke avvik fra populasjonen i denne gruppen. Dette virker rimelig siden barnefamilier har større svarsansynlighet enn andre grupper. De tre neste desilene er underrepresentert. Delvis kan dette forklares med at en del minstepensjonister bor på sykehjem og derfor ikke deltar i populasjonen. De fire øverste desilene er litt overrepresentert. Dette mønsteret er konsistent med frafallsanalysen for Tidsbruksundersøkelsen 2000/01, Belsby(2003).

Til tross for stort frafall er det vektete utvalget ganske representativt, i hvert fall med hensyn på alders- og inntektsfordeling.

3.1. Estimert fordeling i befolkningen for kosthusholdning

Siden vi ikke kjenner fordelingen for kosthusholdning i befolkningen har vi ikke noen fordeling å sammenligne den estimerte fordelingen mot. Tabellen nedenfor viser derfor bare de estimerte fordelingene.

Tabell 5. Befolkning i aldersgruppen 0-79 år 3.november 2001 etter husholdningstype-bohusholdning

Husholdningstype -bohusholdning ²	Estimert med <i>kalibrerte</i> , føringsperiode justerte, modifiserte HT vekter		Estimert med <i>ukalibrerte</i> , føringsperiode justerte, modifiserte HT vekter	
	Antall	Fordeling i %	Antall	Fordeling i %
Enslig	649 239	15,09	564 711	13,11
Par uten barn	975 390	22,67	984 766	22,86
Par med barn, yngste barn 0-6 år	1 101 442	25,60	1 173 865	27,24
Par med barn, yngste barn 7-19 år	959 654	22,30	1 004 597	23,32
Par med barn, yngste barn 20 år og over	170 511	3,96	151 332	3,51
Andre	446 626	10,38	429 428	9,97
I alt	4 302 862	100,00	4 308 699	100,00

Vi ser at kalibreringen øker estimert antall enslige. Det er rimelig siden enslige har høyt frafall og kalibreringen skal justere for dette. Par med barn kalibreres ned. Det er også rimelig siden de har lavere frafall.

4. Mulige forbedringer av vektene

Vi så fra tabell 3 og 4 at det er noe avvik mellom fordelingen i det vektete utvalget og befolkning for alders- og inntektsfordelingen. Muligens kan representasjonen forbedres ved også å kalibrere med hensyn på alder og inntekt. Hovedproblemet for alder kan muligens være det lave antallet i aldersgruppen 70-79år. En kalibrering vil kunne påvirke vektene sterkt. Har de eldre som svarer et forbruksmønster som er forskjellig fra de som ikke svarer, vil dette kunne føre til skjevheter. Det vil også kunne bli stor variasjon fra år til år pga. tilfeldigheter.

For både alders- og inntektsfordelingen i befolkningen er det et problem at registrene ikke gir mulighet til å skille ut institusjonsbeboere. M.a.o. dekker ikke registeret populasjonen korrekt.

Referanser

Belsby, L. (1995): Forbruksundersøkelsen. Vektmetoder, Frafallskorrigerering og intervjuer-effekt. Notater 95/18, Statistisk sentralbyrå.

Heldal, J. (1992): A Method for Calibration of Weights in Sample Surveys. Arbeidsnotat fra Avdeling for personstatistikk, metoder for innsamling og analyse 3/1992, Statistisk sentralbyrå.

Statistisk sentralbyrå (2002): Standard for gruppering av familier og husholdninger. Norges Offisielle Statistikk (NOS C 677).

S-Plus 5, Guide to Statistical and Mathematical Analysis. Math Soft(1998).

S-Plus 5, User's Guide. Math Soft(1998).

Venables, W.N and Ripley, B.D. (1997): Modern Applied Statistics with S-PLUS. Springer

Oversikt over beregningsprogrammene

Eksemplet bruker fil- og variabelnavn fra FU2002.

Merger på Fob bohusholdning på Fu-fil med personer i de intervjuede husholdninger.

Fu-fil, inneholder personene i nettoutvalget.

```
$$S303LES/wk24/lbe/g200  
2/pers_i_hush.sas7bdat
```

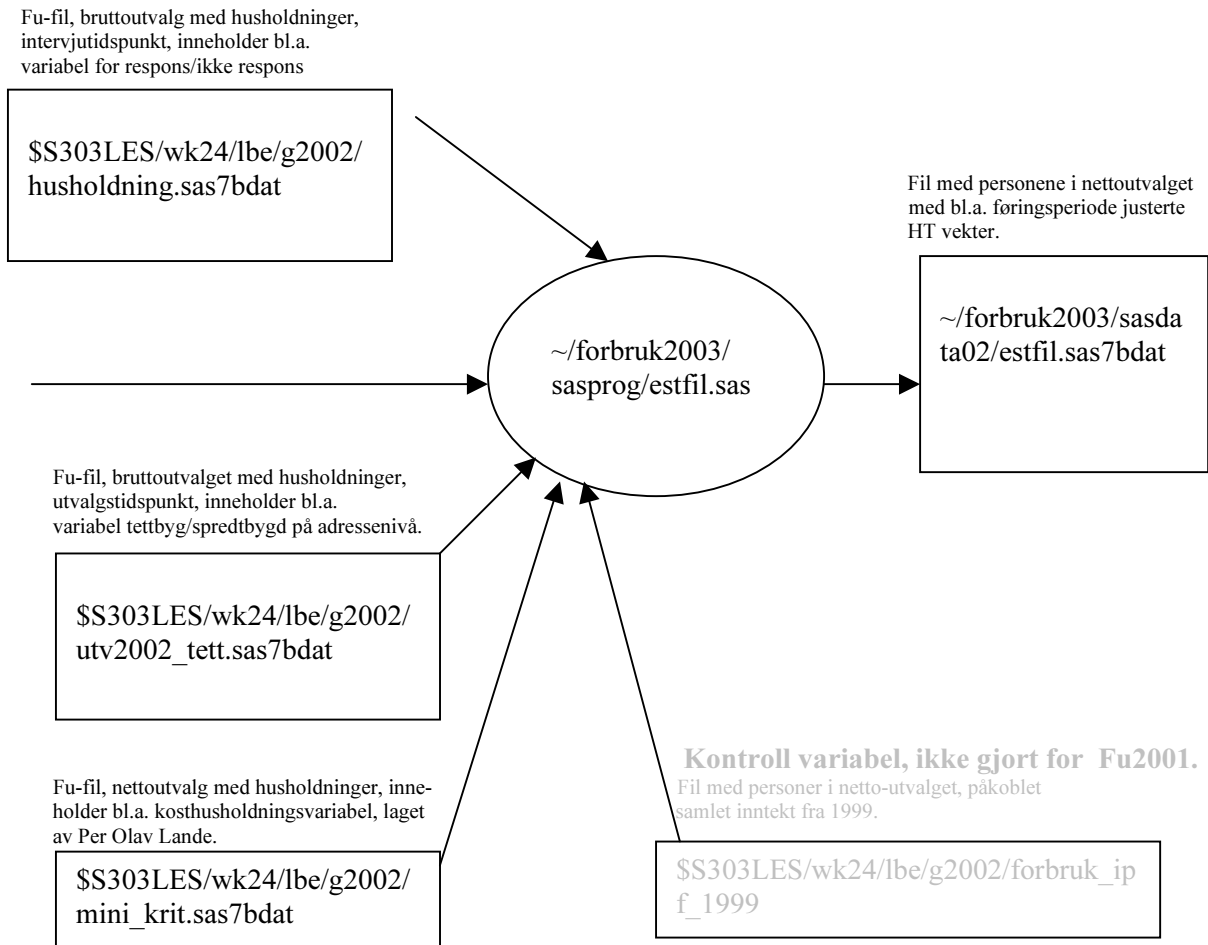
Registerfil fra Fob2001 bohusholdningsvariabel.

```
$DATA370/tilgang/wk2/fob  
2001_forbruk.sas7bdat
```

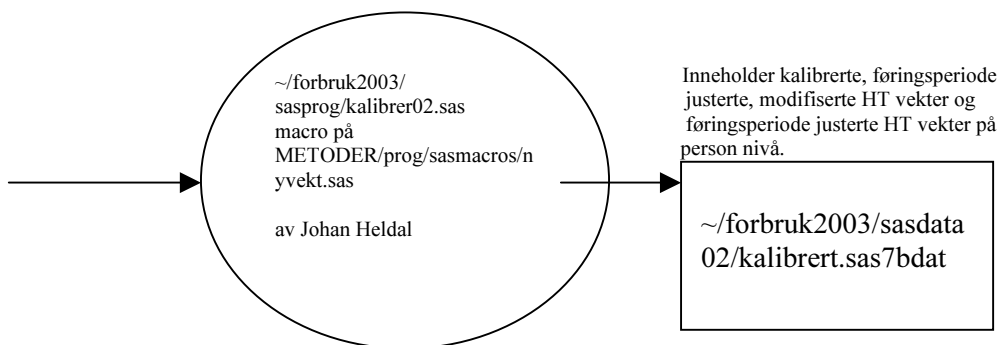
```
~/forbruk2003/  
sasprog/fob200  
1_persihus.sas
```

```
~/forbruk2003/sasdata02  
/fob_persihh.sas7bdat
```

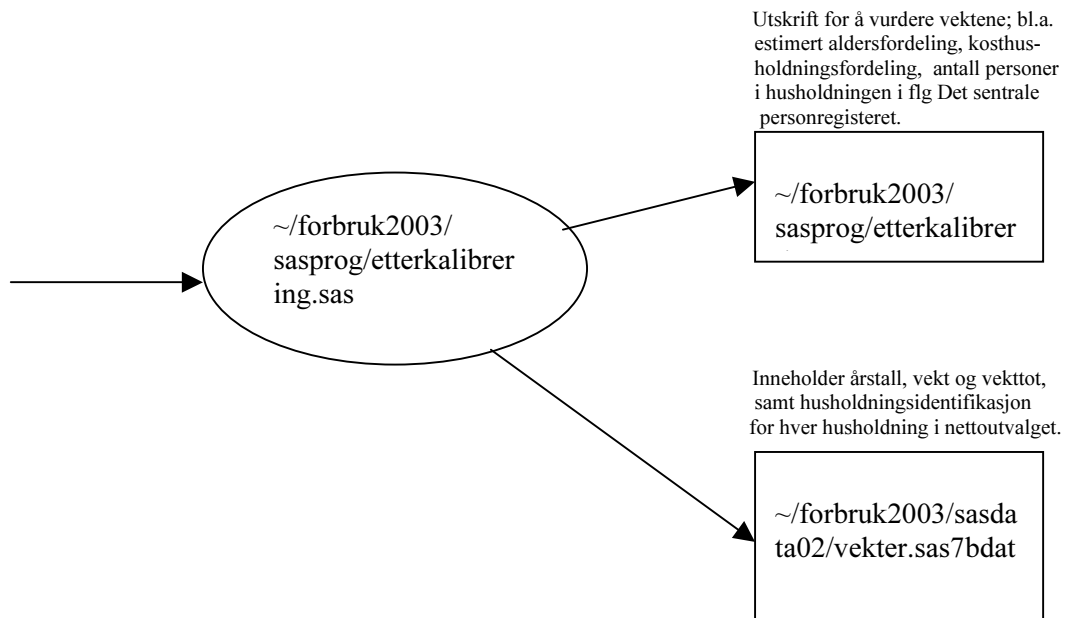
Kobler sammen variable som er nødvendige for å beregne
 føringsperiode justerte, modifiserte HT vekter, samt kalibrerings
 variabelen for Fob bohusholdning.



Kalibrerer de føringsperiodejusterte, modifiserte HT vektene



Lager datasett for husholdningene ved å plukke ut en observasjon for hver husholdning og skalerer de vektene, s.a. variabelen *vekt* summerer seg til antall husholdninger i bruttoutvalget. Summen av *vekttot* er et estimat for antall husholdninger i befolkningen.



Programmer

```

/* Program fob2001_persihus.sas */
/* merger på fobvariable fra filen $data370/tilgang/wk2/fob2001_forbruk */
/* på personer som er i pers_i_hush */

/* Forbruksfil personer, intervjutispunkt */
data pers_i_hush;
set sasd02.pers_i_hush;
fnr=f_nummer;

proc sort data=pers_i_hush;

by fnr;

/* Fob data */
data nyfil2;
set fobvar.fob2001_forbruk;
ald311201=alder;
drop alder;
label
ald311201="alder pr 31.12.01";

proc sort data=nyfil2;
by fnr;

/* Lager datasett som pers_i_hush, i tillegg fobhgr hushtyp2_forbr */
data sasd02.fob_persihh;
merge pers_i_hush(in=x) nyfil2(in=y);
by fnr;
if x;
keep fnr hushnr ald311201 hushtyp2_forbr;

data utvpers;
set sasd02.utvpers;
fnr=fodselsnr;

proc sort data=utvpers;
by fnr;

proc sort data=nyfil2;
by fnr;

data sasd02.fob_utvpers;
merge utvpers(in=x) nyfil2(in=y);
by fnr;
if x;
keep fnr hushnr ald311201 hushtyp2_forbr;
run;

/* Program estfil.sas */
/* starter med husholdningsfil som inneholder både respons og ikke respons */
/* skal beregne trekke og svarsh */

quit;
options ls=80;
data husholdning;
set sasd02.husholdning;
/* husholdningsnivaa både respons og ikke-respons */
/* respons = 1 betyr ikke-fracfall */

```

```

if bfracfall='1' then respons=1;
else if bfracfall='3' or bfracfall='4' then respons=0;

/* avgang fjernes */
if bfracfall='5' then delete;

/* hjelpe variabel */
if respons=1 then r=0;
else if respons=0 then r=1;

/* Grupperer de opprinnelige føringsperiodene i 5 grupper */
if fper='01' or fper='02' or fper='03' or fper='04' or fper='05' or fper='06' then fper='01';
else if fper='07' or fper='08' or fper='09' or fper='10' or fper='11' or fper='12' then
fper='02';
else if fper='13' or fper='14' or fper='15' or fper='16' or fper='17' then fper='03';
else if fper='18' or fper='19' or fper='20' or fper='21' or fper='22' then fper='04';
else if fper='23' or fper='24' or fper='25' or fper='26' then fper='05';

proc freq;
table respons;
table respons*anthhmed;
table fper;
run;

/* Fu-fil med bruttoutvalget, inneholder tett/spredtbyggsvariablene ts_kode på adresse-nivå,
ts_stor og dessuten ant_pers */

data data2;
set sasd02.utv2002_tett;
husholdningsnr=io_nummer;
* if ant_pers gt 4 then ant_pers=2;
keep husholdningsnr ts_kode ts_stor ant_pers;

proc sort data=husholdning;
by husholdningsnr;

proc sort data=data2;
by husholdningsnr;

data husholdning;
merge husholdning(in=x) data2(in=y);
by husholdningsnr;
if x;
if ts_kode=' ' then ts_kode='u';

proc freq;
table respons*ts_kode/chisq;
table respons*ts_stor/chisq;
table respons*ant_pers/chisq;
table respons*anthhmed/chisq;

data husholdning;
set husholdning;
antp=anthhmed;
if antp='.' or antp=. then antp=ant_pers;
if antp ge 2 then antp=2;

proc freq;
table antp ts_kode r;

/* estimerer svarsannsynlighet for tettbygd/spredtbygd, kan eventuelt prøve med andre variable
*/

proc probit data=husholdning;
* class r respons bostr ts_stor ts_kode fpergr;
class r ts_kode antp ;
model r=ts_kode/ noint;
output out=ut
p=prob;

proc sort data=ut;
by husholdningsnr;

```



```

proc sort data=husholdning;
by husholdningsnr;

data husholdning;
merge husholdning ut;
by husholdningsnr;

/* trekkevekt = (ant hush trukket-avgang) / (pop str)= 2165 / 4334897 */
/* antall husholdninger som svarer er 1097 */
/* lager delvis husholdninggruppe variabelen hhgr i dette steget */

trekksh=(2165/ 4334897)*anthhmed;

/* vekt 0 baseres på gjennomsnt svarsh */

svarsh=1097/2165;
vekt0=1/(trekksh*svarsh);
* if r=. then vekt0=.;

/* vekt 1 baseres på estimert svarsh ved proc probit */

* if prob=. then prob=0.4841;

svarsh=prob;
vekt1=1/(trekksh*svarsh);

/* dropper de frafallet */
if respons=1;
keep husholdningsnr htype anthhmed respons trekksh svarsh vekt0 vekt1 prob fper ant_pers;
label
prob="estimert svarsh av antp i hh faktisk eller registr og spredtb/ttettb"
vekt0="vekt justerer for trekke-sh"
vekt1="vekt just. for tr.sh. og svarsh fu av antp i hh faktisk eller registr og
spredtb/ttettb";

/* kobler på kost husholdningsgrupper fra filen $$S303LES/wk24/lbe/g2002/mini_krit.sas7bdat
*/
/* Per Olav Lande har laget denne */

data data1;
set sasd02.mini_krit;
husholdningsnr=hushnr;
htypegr=htype_2;
if htype_2='01' or htype_2='02' or htype_2='03' then htypegr='01';
else if htype_2='04' or htype_2='05' or htype_2='06' then htypegr='02';
else if htype_2='10' or htype_2='11' then htypegr='10';
else if htype_2='13' or htype_2='14' or htype_2='15' then htypegr='13';

keep htype_2 husholdningsnr barnald htypegr;
label htype_2="kosthusholdning"
htypegr="kosthusholdning";

proc format;

value $htype
'01' = 'Enslig under 45 år'
'02' = 'Enslig 45-64 år'
'03' = 'Enslig 65-79 år'
'04' = 'Par uten barn, eldste under 45 år'
'05' = 'Par uten barn, eldste 45-64 år'
'06' = 'Par uten barn, eldste 65-79 år'

'07' = 'Par med barn, yngste barn 0-6 år'
'08' = 'Par med barn, yngste barn 7-19 år'
'09' = 'Par med barn, yngste barn 20 år og over'

'10' = 'Mor med barn, yngste barn 0-19 år'
'11' = 'Far med barn, yngste barn 0-19 år'
'12' = 'Mor/Far med barn, yngste barn 20 år og over'
'13' = 'To eller flere enslige'
'14' = 'Andre husholdninger uten barn 0-19 år'
'15' = 'Andre husholdninger med barn 0-19 år' ;

```

```

value $htypegr
  '01' = 'Enslig'
  '02' = 'Par uten barn'
  '07' = 'Par med barn yngste barn 0-6 år'
  '08' = 'Par med barn yngste barn 7-19 år'
  '09' = 'Par med barn, yngste barn 20 år og over'
  '10' = 'Mor/far med barn, yngste barn 0-19 år'
  '12' = 'Mor/Far med barn, yngste barn 20 år og over'
  '13' = 'Andre';

value $fobhhgr
  '01' = 'Enslig'
  '02' = 'Par uten barn'
  '03' = 'Par med barn yngste barn 0-6 år'
  '04' = 'Par med barn yngste barn 7-19 år'
  '05' = 'Par med barn, yngste barn 20 år og over'
  '06' = 'Mor/far med barn, yngste barn 0-19 år'
  '07' = 'Mor/Far med barn, yngste barn 20 år og over'
  '08' = 'Andre';

value $fper
  '01'='01/01-25/03'
  '02'='26/03-17/06'
  '03'='18/06-26/08'
  '04'='27/08-04/11'
  '05'='05/02-31/12';

proc sort data=data1;
by husholdningsnr;

proc sort data=husholdning;
by husholdningsnr;

data husholdning;
merge data1 husholdning;
by husholdningsnr;

/* summerer vektene */
proc means data=husholdning n sum;
var vekt0 vekt1 prob;

/* estimert fordeling for husholdningsgruppe, antall personer basert på registret og observert
antall personer i hh */
/* husholdningsstype fra filen /ssb/ovibos/a1/s303les/wk24/lbe/hhgrp_2 */

title 'Estimater for husholdninger - ikke vektet for lik tidsperiode ';
proc tabulate data=husholdning missing f=8.;
class htypegr ant_pers anthhmed;
format htypegr $htypegr.;
var vekt0 vekt1;
table htypegr all, vekt0*(sum pctsum*f=6.2) vekt1*(sum pctsum*f=6.2 n*f=6.);
table ant_pers all, vekt0*(sum pctsum*f=6.2) vekt1*(sum pctsum*f=6.2 n*f=6.);
table anthhmed all, vekt0*(sum pctsum*f=6.2) vekt1*(sum pctsum*f=6.2 n*f=6.);

/* skalere s.a. alle perioder veier likt */
/* nyvekt i periode p = k* gammel vekt = ( gjnst. vekt for periodene / sum vekt i periode p )
* gammel vekt */
/* k(periode med l opprinnelige føringsperioder)=(sum vektene / 26)*1*(1/sum vektene i
perioden) */

proc means n sum;
var vekt0 vekt1;

proc sort;
* format fper $fper.;
by fper;

proc means noprint;
* format fper $fper.;
by fper;

```

```

var vekt0 vekt1;
output out=ut1
sum=sumperv0 sumperv1
n=antpv0 antpv1;

proc print;
format fper $fper.;
var fper sumperv0 sumperv1;

proc means data=ut1;
var sumperv0 sumperv1;
output out=ut2
sum=sumv0 sumv1
n=antperv0 antperv1;

data ut2;
merge ut1 ut2;
retain hjelp0 0 hjelp1 0 hjelp2 0 hjelp3 0;

if sumv0 ne . then hjelp0=sumv0;
else sumv0=hjelp0;

if sumv1 ne . then hjelp1=sumv1;
else sumv1=hjelp1;

if antperv0 ne . then hjelp2=antperv0;
else antperv0=hjelp2;

if antperv1 ne . then hjelp3=antperv1;
else antperv1=hjelp3;

if fper='01' or fper='02' then k0=(sumv0/26)*6/sumperv0;
else if fper='03' or fper='04' then k0=(sumv0/26)*5/sumperv0;
else if fper='05' then k0=(sumv0/26)*4/sumperv0;

if fper='01' or fper='02' then k1=(sumv1/26)*6/sumperv1;
else if fper='03' or fper='04' then k1=(sumv1/26)*5/sumperv1;
else if fper='05' then k1=(sumv1/26)*4/sumperv1;

proc print;
* format fper $fper.;
var fper k0 k1 antpv0 antpv1 sumperv0 sumperv1 sumv0 sumv1 antperv0 antperv1;
run;

proc sort data=ut2;
* format fper $fper.;
by fper;

proc sort data=husholdning;
* format fper $fper.;
by fper;

data husholdning;
merge ut2 husholdning;
* format fper $fper.;
by fper;
v0likper=k0*vekt0;
v1likper=k1*vekt1;
fperhjelp=fper*1;

label
v0likper="som v0 men justert s.a. føringper. er likt vektet"
v1likper="som v1 men justert s.a. føringper. er likt vektet";

proc sort data=husholdning;
by husholdningsnr;

title 'Estimater for husholdninger test ';
proc tabulate missing f=8.;
* format fper $fper.;
class fper;

```

```

var v0likper v1likper;
table fper all, v0likper*(sum pctsum*f=6.2) v1likper*(sum pctsum*f=6.2 n*f=6.);
run;

/* filen sasd01.forbruk_ipf_1999 er på person-niva, bare respons, pers_i_hus fra 2001 dessuten
merget på samlet inntekt fra person- og
familieregisteret 1999 */
/*
data person;
set sasd01.forbruk_ipf_1999; */

/* for data for 2002 får jeg ikke merget på inntekt, kan gjøres senere, bruker derfor
sasd02.pers_i_hush */

data person;
set sasd02.pers_i_hush;

proc sort data=person;
by husholdningsnr;

proc sort data=husholdning;
by husholdningsnr;

data person;
merge husholdning(in=x) person(in=y);
by husholdningsnr;
fnr=f_nummer;
if y;

/* merger på fob hhgr fra filen sasd01.fob_persihh, denne brukes til kalibreringen */
proc sort data=sasd02.fob_persihh;
by fnr;

proc sort data=person;
by fnr;

data person;
merge person sasd02.fob_persihh;
by fnr;
fobhhgr=hushtyp2_forbr;

* proc sort;
* by husholdningsnr;

* proc print;
* var fobhhgr htypegr husholdningsnr;

data person;
set person;
if fobhhgr=' ' or fobhhgr='0' or fobhhgr=0 or fobhhgr='.' then do;
if htypegr='01' then fobhhgr='1';
else if htypegr='02' then fobhhgr=2;
else if htypegr='07' then fobhhgr='3';
else if htypegr='08' then fobhhgr='4';
else if htypegr='09' then fobhhgr='5';
else if htypegr='10' then fobhhgr='6';
else if htypegr='12' then fobhhgr='7';
else if htypegr='13' then fobhhgr='8';
end;

* disse hh manglet både htypegr og fobhhgr, har sjekket manuelt antallet, event. alder i hh;
if husholdningsnr='1665' then fobhhgr='1';
if husholdningsnr='0935' then fobhhgr='1';
if husholdningsnr='1152' then fobhhgr='1';

* proc sort;
* by fobhhgr;
* proc print;
* var fobhhgr htypegr husholdningsnr;
* proc contents;

* run;
label

```

```

ant_pers="antall personer i husholdninger fra det sentrale personregisteret"
fobhhgr="fob-bo husholdning";

proc format;
value hhgr
1='enslige'
2='par uten barn'
3='par med barn 0-19'
4='enslige med barn 0-19'
5='andre';

/*
proc format;
value hhgr
1='enslige'
2='par uten barn'
3='par med barn 0-19'
4='enslige med barn 0-19'
5='enfamiliehh med barn 20-'
6='husholdninger med flere familiekjerner';
*/

/* trenger innt10, osv for eventuell kalibrering */
data person;
set person;
innt30=0; innt40=0; innt50=0; innt60=0; innt70=0; innt80=0; innt90=0; innt10=0;
if wsaminnt le 55484 then innt30=1;
else if wsaminnt gt 55484 and wsaminnt le 107809 then innt40=1;
else if wsaminnt gt 107809 and wsaminnt le 153012 then innt50=1;
else if wsaminnt gt 153012 and wsaminnt le 199511 then innt60=1;
else if wsaminnt gt 199511 and wsaminnt le 241025 then innt70=1;
else if wsaminnt gt 241025 and wsaminnt le 284128 then innt80=1;
else if wsaminnt gt 284128 and wsaminnt le 355368 then innt90=1;
else if wsaminnt gt 355368 then innt10=1;

/* blir ellers tull tar vekk obs uten vekt */
if vekt1 ne '.';
keep husholdningsnr vekt0 vekt1 v0likper v1likper prob htype
anthhmed ant_pers bostrok fper wsaminnt fperhjelp
innt30 innt40 innt50 innt60 innt70 innt80 innt90 innt10 ald htype_2 htypegr fobhhgr;

/* summerer vektene */
proc means data=person n sum;
var vekt0 vekt1 prob;

/* summere for antall personer basert på registret */

/* summerer for husholdningstype fra filen /ssb/ovibos/al/s303les/wk24/lbe/hhgrp_2 */

data person;
set person;
alder=ald*1;

proc format;
value aldersgr
0-9='0-9'
10-19='10-19'
20-29='20-29'
30-39='30-39'
40-49='40-49'
50-59='50-59'
60-69='60-69'
70-79='70-79'
80-99='80-99';

value inntgr
low-55484='0-55484'
55485-107809='55485-107809'

```

```

107810-153012='107810-153012'
153013-199511='153013-199511'
199512-241025='199512-241025'
241026-284128='241026-284128'
284129-355368='284129-355368'
355369-high='355369-';

title 'Estimater for perdoner - ikke vektet for lik tidsperiode ';
proc tabulate missing data=person format=8.;
class alder htypegr ant_pers wsaminnt fper fobhhgr;
format alder aldersgr. htypegr $htypegr. wsaminnt inntgr. fobhhgr $fobhhgr.;
var vekt0 vekt1;
table alder all,vekt0*(sum pctsum*f=6.2) vekt1*(sum pctsum*f=6.2 n*f=6.);
table htypegr all,vekt0*(sum pctsum*f=6.2) vekt1*(sum pctsum*f=6.2 n*f=6.);
table fobhhgr all,vekt0*(sum pctsum*f=6.2) vekt1*(sum pctsum*f=6.2 n*f=6.);
table ant_pers all, vekt0*(sum pctsum*f=6.2) vekt1*(sum pctsum*f=6.2 n*f=6.);
table wsaminnt all, vekt0*(sum pctsum*f=6.2) vekt1*(sum pctsum*f=6.2 n*f=6.);
table fper all, vekt0*(sum pctsum*f=6.2) vekt1*(sum pctsum*f=6.2 n*f=6.);

* proc print;
* var fper antpv0 k0 sumv0 antperv0 sumperv0 antpv1 k1 sumv1 antperv1 sumperv1;

title 'Estimater for personer - vektet for lik tidsperiode ';
proc tabulate missing data=person f=8.;
class alder htypegr ant_pers wsaminnt fper fobhhgr;
format alder aldersgr. htypegr $htypegr. wsaminnt inntgr. fobhhgr $fobhhgr.;
* fper $fper.;
var v0likper v1likper;
table alder all,v0likper*(sum pctsum*f=6.2) v1likper*(sum pctsum*f=6.2 n*f=6.);
table htypegr all,v0likper*(sum pctsum*f=6.2) v1likper*(sum pctsum*f=6.2 n*f=6.);
table fobhhgr all,v0likper*(sum pctsum*f=6.2) v1likper*(sum pctsum*f=6.2 n*f=6.);
table ant_pers all, v0likper*(sum pctsum*f=6.2) v1likper*(sum pctsum*f=6.2 n*f=6.);
table wsaminnt all, v0likper*(sum pctsum*f=6.2) v1likper*(sum pctsum*f=6.2 n*f=6.);
table fper all, v0likper*(sum pctsum*f=6.2) v1likper*(sum pctsum*f=6.2 n*f=6.);

data person;
set person;
if fobhhgr=06 or fobhhgr=07 then fobhhgr=08;
if htypegr=10 or htypegr=12 then htypegr=13;

fobhhgr1=0; fobhhgr2=0; fobhhgr3=0; fobhhgr4=0; fobhhgr5=0; fobhhgr6=0; fobhhgr7=0;
fobhhgr8=0;
if fobhhgr=1 then fobhhgr1=1;
if fobhhgr=2 then fobhhgr2=1;
if fobhhgr=3 then fobhhgr3=1;
if fobhhgr=4 then fobhhgr4=1;
if fobhhgr=5 then fobhhgr5=1;
if fobhhgr=6 then fobhhgr6=1;
if fobhhgr=7 then fobhhgr7=1;
if fobhhgr=8 then fobhhgr8=1;

proc freq data=person;
table fobhhgr htypegr;

/* lage kalibreringsvariabel som har samme verdi for hele husholdningen */

proc sort;
by husholdningsnr;

proc means noprint;
by husholdningsnr;
var fobhhgr1-fobhhgr8;
output out=ut
sum=fob1-fob8
n=ant1-ant8
mean=fobgr1-fobgr8;

proc sort data=ut;
by husholdningsnr;

proc sort data=person;

```

```

by husholdningsnr;

data sasd02.estfil;
merge person ut;
by husholdningsnr;

* proc print data=sasd02.estfil;
* var husholdningsnr fobhhgr1-fobhhgr5 fobgr1-fobgr5;

* proc contents data=sasd02.estfil;
run;

```

```

/* Program kalibrer02.sas */

```

```

/* Programmert av Johan Heldal */

```

```

* libname sasd02 '~/forbruk2003/sasdata02';
* LIBNAME IF90 '~/intdata/if90';
OPTIONS DQUOTE PAGESIZE=500;
*PTIONS NOREPLACE; *MPRINT;
*

```

```

PROGRAM FOR ENDRING AV OPPBL}SNINGSVEKTER FOR HUSHOLDNINGER

```

Dette programmet kan endre faste oppbl}sningsvekter i en utvalgsunders}kelse ved hjelp av kalibrering mot st}rrelser som er kjente for hele populasjonen. Metoden som programmet virker etter, er beskrevet i "Arbeidsnotat fra avdeling for personstatistikk. Metoder for innsamling og analyse." nr. 3/1992 med tittelen "A Method for Calibration of Weights in Sample Surveys" av Johan Heldal. Denne vil bli referert til som JH92.

Denne versjonen av programmet virker slik b}de dersom enhetene er trukket direkte (f.eks personer i rene personutvalg) eller som klynger (f.eks husholdninger) og alle enhetene i klyngen (d.v.s. personene i husholdningen) er med i utvalget. Metoden som programmet virker etter, krever at alle enheter i samme klynge initielt har de samme oppbl}sningsvekter. De nye vektene som programmet genererer vil da ogs} bli like innen samme klynge. Dette gj}r det mulig } generere et utvalg med klyngene (husholdningene) som enheter ved } aggregere over enhetene i klyngene. Klyngene vil da f} samme oppbl}sningsvekt som enhetene i den.

Metoden som anvendes virker slik at spesifiserte kjente totaltall i befolkningen blir ans}tt eksakt riktig ved det nye vektsystemet. De variable som brukes til dette kalles FASITVARIABLE og m} spesifiseres nedenfor.

Programmet krever at brukeren spesifiserer en del informasjon. Denne spesifiseres som innholdet av SAS makro variable i %LET satsene nedenfor.

Programmet krever at det er laget et SAS datasett av utvalgsfilen p} forh}nd. Oppgi f}rste ledd i navnet p} dette datasettet. Dette m} v}re lik filreferansen til utvalgsfilen i LIBNAME spesifisjonen for datasettet.

```

;
%LET INDATA=sasd02.estfil;
*

```

Skriv filreferansen til det datasett og member hvor du }nsker de nye vektene plassert. Hvis f}rste ledd i navnet er forskjellig fra det du oppgav i %LET satsen ovenfor, m} du spesifisere et eget LIBNAME.

```

;
%LET UTDATA=sasd02.kalibrert;
*

```

Oppgi navnet p} den variabelen i datasettet som inneholder de opprinnelige vektene for oppbl}sing av utvalget og navn p} ny vekt som genereres.

```

;
%LET VEKT=vllikper;
%LET NYVEKT=vektk;
%LET Q=1/nid;
*

```

Oppgi navnet p} variabel som identifiserer klyngene.

```

;
%LET ID=husholdningsnr;
*

```

```

    Gi navn til fasitvariable. Listen kan inneholde s) mange
fasitvariable som vi har tilgjengelige. Variablene i FASITVAR m) ogs}
finnes i utvalgsfilen eller kunne genereres i datasteget nedenfor.
    Variabelnavnene i listen m) adskilles med minst en blank.
;
/* %LET FASITVAR= hhgr1 hhgr2 hhgr3 hhgr4 hhgr5 innt30 innt40 innt50 innt60 innt70 innt80
innt90; */
* %LET FASITVAR= fobhhgr1 fobhhgr2 fobhhgr3 fobhhgr4 fobhhgr5 fobhhgr6 fobhhgr7 fobhhgr8
innt30 innt40
innt50 innt60 innt70 innt80 innt90;

%LET FASITVAR= fobgr1 fobgr2 fobgr3 fobgr4 fobgr5 fobgr6 fobgr7 fobgr8 innt30 innt40
innt50 innt60 innt70 innt80 innt90;

*
    Gi verdier til fasitvariablene. Disse verdiene m) v{re totaltall
(sumtall) for hele befolkningen for de fasitvariablene som er gitt i
%LET FASITVAR= satsen ovenfor. Verdiene m) v{re hentet fra andre
kilder enn utvalget selv, f.eks. fra registre. Det m) v{re eksakt
overensstemmelse mellom listen FASITVAR og tallene i FASITALL.
;
DATA FASIT;
    input fasitall;
    cards;
634629
775080
1023036
963214
341156
1
1
597782
1300469
4334897
4334897
4334897
4334897
4334897
4334897
4334897
;
*
*
Vektoren VELG nedenfor inneholder en posisjon (0 eller 1) for hver
av fasitvariablene i FASITVAR. 1 angir at variabelen skal benyttes
ved en gitt kalibrering og 0 angir at variabelen ikke skal benyttes.
Totaltall med gamle og nye vekter blir beregnet for alle variablene.
Estimatene for totaltallene for de fasitvariablene som ikke benyttes
kan brukes til evaluering av det nye vektsystemet fungerer.
;
%LET VELG= 1 1 1 1 1 0 0 1 0 0 0 0 0 0 0;

%LET LAV=0.33;
%LET HOY=3;
%LET MAXIT=50;
%LET KRIT=00.01;

*
    Programmet kan arbeide med tre varianter av kalibreringsestimatorer og
velges ved } angi KALIBMET = 1, 2 eller 3 nedenfor. Metode 1 kalles i
JH92 for Regresjonsmetoden. Metode 2 kalles for Prediksjonsmetoden.
Metode 3 er ikke omtalt i JH92 og kan kalles Differensestimatormetoden.
;
*
JH92, avsnitt 5, beskriver to metoder for } lage initielle vekter for
klyngene. Velg variant ved } sette makrovariabelen KLYNGMET. Ved
KLYNGMET = 1 velges metoden til Lemaitre og Dufour (se JH92). Ved
KLYNGMET = 2 velges den andre metoden.
;
*
*** STEG 1 ***

    I dette f}rste steget omkodes variable fra personfilen til en form
som er hensiktsmessig n)r variablene skal aggregeres til husholdnings-
niv}. Brukeren kan fritt foreta omkodinger av eksisterende variable
i datasteget nedenfor.

```



```

;
DATA A;
  SET &INDATA;
/*
  if indikat=0 then delete;
  blaas = blaas*5072/3740;
  pgivl = (k019+k020+k021+k022+k023+k039+k040+k041+k042+k043
          +k058+k059+k060+k061+k062)/1000000;
  pgivj = (k102+k103+k104)/100000;
  pgivh = (k105+k106)/100000;
  en = 1;
  if klasse=. or klasse=0 then k10=1;else k10=0;
  if klasse=1 then k11=1;else k11=0;
  if klasse=2 then k12=1;else k12=0;
*
Koding av grupper etter alder.
;
  if dalder<=6 then a0_6=1; else a0_6=0;
  if 7<=dalder<=12 then a7_12=1; else a7_12=0;
  if 13<=dalder<=15 then a13_15=1; else a13_15=0;
  if 16<=dalder<=19 then a16_19=1; else a16_19=0;
  if 20<=dalder<=29 then a20_29=1; else a20_29=0;
  if 30<=dalder<=39 then a30_39=1; else a30_39=0;
  if 40<=dalder<=49 then a40_49=1; else a40_49=0;
  if 50<=dalder<=59 then a50_59=1; else a50_59=0;
  if 60<=dalder<=66 then a60_66=1; else a60_66=0;
  if dalder>=67 then a67_=1; else a67_=0; */

*proc sort;
*by &ID;

%INCLUDE '$METODER/prog/sasmacros/nyvekt.sas';

DATA G;
set &utdata;
* SET &NYUTVFIL..&NYMEMBER;
  diff2 = (&VEKT - &NYVEKT)**2/&VEKT;
run;

proc univariate data=G vardef=wgt;
  var &VEKT &NYVEKT diff2;
  weight &VEKT;

data sasd02.kalibrert;
set G;

* proc contents data=sasd02.kalibrert;
run;

/* Program etterkalibrering.sas */
/* Etterkalibrering.sas plukker ut en observasjon for hvre husholdning, dvs tilpasser
strukturen til Fu-data */
/* Dessuten skaleres vektene s.a. vekt, nederst i programmet, summererer seg til antall
husholdninger i brutto-utvalget */
/* vektot summerers seg til estimert antall husholdninger i populasjonen */

/* Skriver ut vekten */

/* proc sort data=sasd02.kalibrerhh;
by husholdningsnr;

proc print data=sasd02.kalibrerhh;
var husholdningsnr vsumbr;
run; */

data kalibrerpers;
set sasd02.kalibrert;

label
vektk='kalibrert vekt';

```

```

proc format;
value aldersgr
0-9='0-9'
10-19='10-19'
20-29='20-29'
30-39='30-39'
40-49='40-49'
50-59='50-59'
60-69='60-69'
70-79='70-79'
80-99='80-99';

value inntgr
low-55484='0-55484'
55485-107809='55485-107809'
107810-153012='107810-153012'
153013-199511='153013-199511'
199512-241025='199512-241025'
241026-284128='241026-284128'
284129-355368='284129-355368'
355369-high='355369-';

value $htypegr
'01' = 'Enslig'
'02' = 'Par uten barn'
'07' = 'Par med barn yngste barn 0-6 år'
'08' = 'Par med barn yngste barn 7-19 år'
'09' = 'Par med barn, yngste barn 20 år og over'
'10' = 'Mor/far med barn, yngste barn 0-19 år'
'12' = 'Mor/Far med barn, yngste barn 20 år og over'
'13' = 'Andre';

value $fobhhgr
'1' = 'Enslig'
'2' = 'Par uten barn'
'3' = 'Par med barn yngste barn 0-6 år'
'4' = 'Par med barn yngste barn 7-19 år'
'5' = 'Par med barn, yngste barn 20 år og over'
'6' = 'Mor/far med barn, yngste barn 0-19 år'
'7' = 'Mor/Far med barn, yngste barn 20 år og over'
'8' = 'Andre';

title 'Estimater for personer - vektet for lik tidsperiode for hh ';
proc tabulate missing f=8.2;
class alder htypegr ant_pers wsaminnt fper fobhhgr;
format alder aldersgr. htypegr $htypegr. wsaminnt inntgr. fobhhgr $fobhhgr.;
var v0likper v1likper vektk;
table alder all,vektk*(sum*f=9. pctsum*f=6.2) v1likper*(sum*f=9. pctsum n*f=6.)/row=float;
table htypegr all,vektk*(sum*f=9. pctsum) v1likper*(sum pctsum n*f=5.);
table fobhhgr all,vektk*(sum*f=9. pctsum) v1likper*(sum pctsum n*f=5.);
table ant_pers all, vektk*(sum*f=9. pctsum) v1likper*(sum pctsum n*f=5.);
table wsaminnt all, vektk*(sum*f=9. pctsum) v1likper*(sum pctsum n*f=5.);
* table fper all, vektk*(sum pctsum) v1likper*(sum pctsum n*f=5.);

/* Plukker ut en record for hver husholdning */
proc sort data=kalibrerpers;
by husholdningsnr;

* proc print data=kalibrerpers;
* var husholdningsnr vektk fobhhgr;

data kalibrerhh;
set kalibrerpers;
by husholdningsnr;
if first.husholdningsnr then output;

title 'Estimater for husholdninger - vektet for lik tidsperiode for hh ';
proc tabulate missing data=kalibrerhh format=8.;
class htypegr ant_pers fper fobhhgr;
format htypegr $htypegr. fobhhgr $fobhhgr.;

```

```

var v0likper v1likper vektk;
table htypegr all,vektk*(sum pctsum*F=6.2 ) v1likper*(sum pctsum*F=6.2  n*F=5.);
table fobhhgr all,vektk*(sum pctsum*F=6.2 ) v1likper*(sum pctsum*F=6.2  n*F=5.);
table ant_pers all, vektk*(sum pctsum*F=6.2 ) v1likper*(sum pctsum*F=6.2  n*F=5.);
table fper all, vektk*(sum pctsum*F=6.2 ) v1likper*(sum pctsum*F=6.2  n*F=5.);

proc means data=kalibrerhh;
var vektk v1likper vekt0 vekt1;
output out=ut1
sum=sumvk sumv1 sumv0 sumv1a
n=antvk antv1 antv0 antv1a;

data sasd02.kalibrerhh;
merge ut1 kalibrerhh;
retain hjelpk 0 hjelp1 0 hjelpv0 0 hjelpa;

if sumvk ne . then hjelpk=sumvk;
else sumvk=hjelpk;

if sumv1 ne . then hjelp1=sumv1;
else sumv1=hjelp1;

if sumv0 ne . then hjelpv0=sumv0;
else sumv0=hjelpv0;

if sumv1a ne . then hjelpa=sumv1a;
else sumv1a=hjelpa;

* skalerre s.a. summen av vektene blir summen av antall husholdninger i bruttoutvalget;

vsumbr=(vektk/sumvk)*2165;
vsumbrv1=(v1likper/sumv1)*2165;
vsbrv0=(vekt0/sumv0)*2165;
vsbrv1a=(vekt1/sumv1a)*2165;
endring=vsumbr/vsumbrv1;

label
vektk="kalibrert, mod.ht,kalenderjustert,gir totaltall for fob pop=4 334897"
vsumbr="kalibrert,mod.ht,kalenderjustert,skalert s.a. sum er størrelsen på bruttoutvalget"
vsumbrv1="mod.ht,kalenderjustert,skalert s.a. sum er størrelsen på bruttoutvalget"
vsbrv0="ht.,skalert s.a. sum er størrelsen på bruttoutvalget"
vsbrv1a="mod.ht,skalert s.a. sum er størrelsen på bruttoutvalget"
endring="Forholdet mellom mod ht kal.just og den kalibrerte versjonen";

/*
proc print data=sasd02.kalibrerhh;
var husholdningsnr vektk vsumbr vsumbrv1 fper; */

proc univariate data=sasd02.kalibrerhh;
var vsumbr vsumbrv1 vsbrv0 vsbrv1a endring;

title 'Estimater for husholdninger - vektet for lik tidsperiode for hh ';
proc tabulate missing data=sasd02.kalibrerhh format=8.;
class htypegr ant_pers fper fobhhgr;
format htypegr $htypegr. fobhhgr $fobhhgr.;
var vsumbr vektk;
table htypegr all,vsumbr*(sum pctsum*f=5.) vektk*(sum pctsum*F=6.2);
table fobhhgr all,vsumbr*(sum pctsum*F=6.2 n*f=5.) vektk*(sum pctsum*F=6.2);
table ant_pers all, vsumbr*(sum pctsum*F=6.2 n*f=5.) vektk*(sum pctsum*F=6.2);
table fper all, vsumbr*(sum pctsum*F=6.2 n*f=5.) vektk*(sum pctsum*F=6.2);

/* lage sasdatasett med vekter */
data sasd02.vekter;
set sasd02.kalibrerhh;
aar=2002;
vekt=vsumbr;
vekttot=vektk;
keep aar husholdningsnr vekt vekttot;
label
vekttot="kalibrert, mod.ht,kalenderjustert,gir totaltall for fob pop=4 334897"
vekt="kalibrert,mod.ht,kalenderjustert,skalert s.a. sum er størrelsen på bruttoutvalget";

keep aar husholdningsnr vekt vekttot;
run;

```

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 levkår. 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: Levkår 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 oljevirksomheten 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 bolig tellingen. 32s.
- 2003/48 E. Siig Meen og O. Rognstad: Jordbrukstelling 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.