



Johan Heldal

Kalibrering av AKU

Dokumentasjon av metode og program

Notater

Innhold

1	Bakgrunn	1
2	Beskrivelse av metoden	2
2.1	Trinn 1	2
2.2	Trinn 2	2
2.3	Trinn 3	4
3	Programmet og bruk av programmet.	5
3.1	Brukerdelen	6
3.2	Beregningsdelen	8
3.3	Kalibreringsrutinen	15
4	Utskrift fra programmet	20
4.1	Personer med ugyldige fylkeskoder	20
4.2	Kontroll av endring i vekter	21
4.3	Kalibrerte vekter	26
5	Referanser	27

1 Bakgrunn

I 1996 ble utvalgsplanen for arbeidskraftundersøkelsene lagt om. Mens AKU tidligere var basert på byråets standardutvalgsplan som var selvveiende for hele landet, ble det fra 1/1-96 tatt i bruk en fylkesvist stratifisert utvalgsplan der de minste fylkene skulle overrepresenteres. Noe av bakgrunnen for den nye planen var et ønske om å kunne estimere sysselsettingstall på fylkesnivå i overenstemmelse med anbefalinger fra Eurostat (Council Regulation (EC) No 577/98). Det ble også ansett som ønskelig å kunne fordele sysselsettingstallene etter kjønn og 12 fem års aldersgrupper innen hvert fylke.

Den nye utvalgsplanen og ambisjonene for publisering stilte nye krav til estimeringsmetoden i AKU. Den gamle, dokumentert i Boije (1996), var basert på etterstratifisering etter kjønn, alder, og sysselsetting/næringskategori ifølge arbeidstaker/arbeidsgiverregisteret, i alt 93 etterstrata. Dette opplegget tilfredstilte behovet for å kunne publisere tall for antall personer etter sysselsetting/næring/status etc. etter kjønn og aldergrupperinger på landsbasis slik at estimatene ble kalibrert riktig i forhold til antall personer i hver av disse gruppene. Anvendt på landsnivå tok metoden imidlertid ikke hensyn til at at ulike fylker hadde forskjellige trekkeandeler etter den nye planen og heller ikke at frafallsandelen kan variere fra fylke til fylke. Metoden kunne ikke brukes innen hvert fylke siden utvalgene der var for små i forhold til antall etterstrata med mange tomme etterstrata som følge. Dette ble også påpekt i Boije (1996). Zhang (1998) foreslo å bruke kalibrering på fylkesnivå for å tilpasse estimater til fordelingen etter alder og kjønn i hvert fylke.

Kalibrering vil omgå problemet med tomme celler ved at den innen hvert fylke bare tar hensyn til de marginale og ikke den simultane fordelingen etter kjønn og alder. Metoden tar i trinn 1 (beskrevet nedenfor) hensyn til at den nye utvalgsplanen har ulike trekkeandeler og ulike (variable) frafallsandeler i ulike fylker. Metoden korrigerer ikke ekspisitt for frafall ut over dette, men vil implisitt gjøre det i den grad valget av etterstratifiserings- og kalibreringsvariable fanger det opp.

Dette notatet beskriver implementeringen av den fylkesvise kalibreringen. Metoden som er benyttet følger med noen få modifikasjoner det opplegget som er foreslått av Zhang. Metoden vil bli presentert i detalj i avsnitt 2. I avsnitt 3 presenteres og forklares programkoden og et tilpasset opplegg for hvordan brukeren skal kunne anvende programmet på en enklest mulig måte. I avsnitt 4 presenteres noen estimater generert ved anvendelse av metoden på AKU første kvartal 1998.

2 Beskrivelse av metoden

Metoden som beskrives er en tre trinns metode. På hvert trinn genereres et sett individvekter. Alle tre vektsettene tas vare på og kobles til utvalget. Tredje trinn kan betraktes som en generalisert regresjonsestimator (Särndal, Swensson og Wretman (1992)) med det forbehold at inputvektene ikke er inverse inklusjonssannsynligheter, men etterstratifiseringsvektene produsert i trinn 2.

2.1 Trinn 1

er en ren oppblåsing ved hjelp av fylkesvise utvalgsvekter. og gir opphav til vekter som vi vil kalle $W1$.

La N være populasjonsstørrelsen. Denne er fordelt på $F = 19$ fylker med N_{1+}, \dots, N_{F+} innbyggere. La n_{1+}, \dots, n_{F+} være den tilsvarende fordelingen av utvalget etter fylker. På dette grunnlag lages utvalgsvekter

$$W1_f = \frac{N_{f+}}{n_{f+}}.$$

for hvert av fylkene $f = 1, \dots, F = 19$. Dette gir vektsettet $W1$. $W1$ vektene korrigerer for den ikke-selveiende utvalgsplanen på landsbasis og er med som input til neste trinn.

2.2 Trinn 2

Dette er en etterstratifisering på landsbasis ved hjelp av en kryssklassifisering av variablene

1. Kjønn, 2 kategorier
2. Alder, 12 kategorier: 16-19 år og senere i 11 5-årskategorier 20-24,,70-74.
3. Sysselsatt/Ikke sysselsatt etter Arbeidstaker-Arbeidsgiverregisteret og næring etter Standard for Næringsgruppering 1994 (SN94) fordelt etter 3 kategorier: primærnæring, sekundærnæring og tertiærnæring, i alt 4 kategorier.

I alt gir dette 96 etterstrata. For å unngå tomme celler ble det for den eldste aldersgruppen (70-74 år) bare benyttet to kategorier av SN94, ikke

sysselsatt og sysselsatt. Dette reduserte antall etterstrata til 92. Etterstratifiseringen gir opphav til vekter som vil bli kalt $W2$.

De $E = 92$ etterstrataene er i populasjonen representert med N_{+1}, \dots, N_{+E} personer. Det er N_{fe} personer i fylke nr f og etterstratum e . Tilsvarende inneholder utvalget n personer med n_{fe} personer i fylke nr. f og etterstratum nr. e . Antall personer i etterstratum e i fylke f i populasjonen, N_{fe} estimeres innen hver fylke ved å blåse opp n_{fe} med de fylkesvise vektene fra trinn 1:

$$\widehat{N1}_{fe} = W1_f n_{fe} = \frac{N_{f+}}{n_{f+}} n_{fe}.$$

Antall personer i hvert etterstratum på landsbasis, N_{+e} , estimeres så ved å summere over fylkene:

$$\widehat{N1}_{+e} = \sum_{f=1}^F \widehat{N1}_{fe} = \sum_{f=1}^F W1_f n_{fe} = \sum_{f=1}^F \frac{N_{f+}}{n_{f+}} n_{fe}.$$

Deretter genereres syntetiske fylkesvise etterstratifiseringsvekter som

$$W2_{fe} = W1_f \frac{N_{+e}}{\widehat{N1}_{+e}}.$$

som er vektsettet $W2$. Dette vektsettet sikrer at estimater over antall personer i hvert etterstratum på landsnivå blir riktige. La

$$x_i = \begin{cases} 1 & \text{hvis person nr. } i \text{ tilhører etterstratum nr. } e. \\ 0 & \text{ellers} \end{cases}$$

La s betegne utvalget. Estimat for antall personer i etterstratum e i fylke f blir

$$\widetilde{N2}_{fe} = \sum_{i \in s} W2_i x_i = W2_{fe} n_{fe} = W1_f \frac{N_{+e}}{\widehat{N1}_{+e}} n_{fe} = \frac{N_{+e}}{\widehat{N1}_{+e}} \widehat{N1}_{fe}$$

Estimatoren for totalt antall personer i etterstratum e i populasjonen med dette vektsettet blir estimert helt korrekt:

$$\widetilde{N2}_{+e} = \sum_{i \in s} \widetilde{N2}_{fe} = \frac{N_{+e}}{\widehat{N1}_{+e}} \sum_{f=1}^F W1_f n_{fe} = \frac{N_{+e}}{\widehat{N1}_{+e}} \widehat{N1}_{+e} = N_{+e}.$$

På den annen side blir ikke antall personer i hvert fylke estimert eksakt riktig slik de blir med $W1$ vektene.

2.3 Trinn 3

$W2$ -vektene danner basis forsom er den fylkesvise kalibreringen. La $f = 1, \dots, F = 19$ indeksere fylkene, $k = 1, 2$ indeksere de to kjønnene, $a = 1, \dots, 12$ representere de 12 aldersgruppene og $r = 1, 2$ representere sysselsatt og ikke-sysselsatt etter registeret. (I trinn 2 benyttet vi 4 kategorier fra registervariabelen SN94 i etterstratifisering. Her benytter bare to kategorier.) Vi lar N_{fkar} være antall personer i fylke f med kjønn k i alder a og med registerstatus r i populasjonen. La videre N_{fk++} være antall personer med kjønn k i fylke f , N_{f+a+} antall personer i aldersgruppe a i fylke f og N_{f++r} antall personer etter registerstatus r i fylke f . La

$$\mathbf{N}_f = (N_{f1++}, N_{f2++}, N_{f+1+}, \dots, N_{f+12+}, N_{f++1}, N_{f++2})'$$

være vektoren av marginalfordelinger i fylke f etter disse variablene. La videre

$$\widetilde{\mathbf{N}}_f = (\widetilde{N}_{f1++}, \widetilde{N}_{f2++}, \widetilde{N}_{f+1+}, \dots, \widetilde{N}_{f+12+}, \widetilde{N}_{f++1}, \widetilde{N}_{f++2})'$$

Vi ønsker nå å utvikle et sett med vekter $W3$ som estimerer disse størrelsene riktig innen hvert fylke og samtidig avviker så lite som mulig fra $W2$ vektene etter et bestemt avstandsmål. Valg av avstandsmål har betydning for estimatoren. Avstandsmålet som brukes her kan uttrykkes ved

$$D_f(W3, W2) = \sum_{i \in \mathbf{s}} \left(\frac{W3_i}{W2_i} - 1 \right)^2,$$

Hvis N_{fg} er en av komponentene i \mathbf{N}_f og

$$x_{fi} = \begin{cases} 1 & \text{hvis person nr. } i \text{ tilhører fylke } f \text{ og kategorien } g \\ 0 & \text{ellers,} \end{cases}$$

så skal

$$\widetilde{N}_{3fg} = \sum_{i \in \mathbf{s}} W3_i x_{fi} = N_{fg}.$$

Siden $N_{f1++} + N_{f2++} = N_{f+1+} + \dots + N_{f+12+} = N_{f++1} + N_{f++2} = N$ kan ikke alle kategoriene tas eksplisitt med i kalibreringen da det ville føre til kolinearitet. I praksis kuttes derfor den 12-te aldersgruppen og yrkesaktive etter registeret fra vektorene \mathbf{N}_f . La

$$\mathbf{x}_{fi} = (x_{fi1}, x_{fi2}, y_{fi1}, \dots, y_{fi11}, z_{fi1})'$$

der

$$x_{fi1} = \begin{cases} 1 & \text{Hvis person nr. } i \text{ i fylke } f \text{ er kvinne} \\ 0 & \text{ellers} \end{cases} = 1 - x_{fi2}$$

$$y_{fia} = \begin{cases} 1 & \text{Hvis person nr. } i \text{ i fylke } f \text{ er i aldersgruppe } a \\ 0 & \text{ellers} \end{cases}$$

og

$$z_{fi1} = \begin{cases} 1 & \text{Hvis person nr. } i \text{ i fylke } f \text{ er ikke-sysselsatt} \\ 0 & \text{ellers} \end{cases}$$

være vektoren av kalibreringsvariable for person nr. i i fylke nr. f . La $\mathbf{X}_f = (\mathbf{x}_{f1}, \dots, \mathbf{x}_{fn_f})'$ være matrisen med individvektorene som rader. La $\mathbf{n}_f = (n_{f1++}, n_{f2++}, n_{f+1+}, \dots, n_{f+11+}, n_{f++1})'$ være motstykket til \mathbf{N}_f i utvalget og la til slutt $\mathbf{W2}$ være diagonalmatrisen med vektene $W2_i$ på diagonalen. Løsningen for vektene $W3_i$ kan uttrykkes på matriseform ved

$$W3_i = W2_i(1 + \mathbf{x}'_{fi}(\mathbf{X}'_f \mathbf{W2} \mathbf{X}_f)^{-1}(\mathbf{N}_f - \widetilde{\mathbf{N2}}_f)).$$

Det er imidlertid ikke mulig å uttrykke $W3_i$ på en enkel og oversiktlig form. Matrisen $\mathbf{X}'_f \mathbf{W2} \mathbf{X}_f$ vil imidlertid bestå av blokker av alle mulige krysstabuleringer av de tre kalibreringsvariablene i fylke f slik de blir estimert med $W2$ -vektene. Diagonalen vil bestå av alle totaltall for kategoriene til de tre kalibreringsvariablene slik de blir estimert med $W2$ -vektene.

3 Programmet og bruk av programmet.

Kalibreringsprogrammet er delt inn i to deler, en bruker-del og en beregnings del.

1. En brukerdel hvor brukeren må sette nødvendige styringsparametre og passe på at datasettene blir lest riktig. To datasett skal leses, utvalg sfilen med den nødvendige informasjonen fra utvalget og en fil med tabell over de totaltall for antall personer etter FYLKE x KJØNN x ALDERSGRUPPER x SN94 som danner utgangspunkt for fasit for etterstratifisering og kalibrering.
2. Beregningsdelen. Dette ligger på en egen fil som kalles opp av innlesingsdelen og skal ikke røres av bruker. Her beregnes alle sett med vektorer.

Inputparametre til kalibrering som skal holdes faste og ikke skal endres fra gang til gang settes her. I dette programmet kalles også selve den generelle kalibreringsrutinen.

3.1 Brukerdelen

I brukerdelen spesifiseres filene som leses og skrives. Malen for denne kan kopieres fra filen /ssb/lupus/a8/aku/prog/metode/kalib.sas. Filrefene må ikke endres. Sluttproduktet VEKTER inneholder identifikatorer for kalibreringsenhet (f.eks. fødselsnummer), vektene W_1 , W_2 og W_3 samt opplysninger om år, kvartal, og måned. Hvis kalibreringen skjer for et kvartal settes måned lik 00. Programmalen inneholder tekst som forklarer programmet.

```
LIBNAME AKU '$AKU/data/tilkalibs';
FILENAME AKUFIL '$AKU/data/tilkalib/utug98k01.txt';
                                                    /* Utvalgsfil */
FILENAME MARGINAL '$AKU/data/tilkalib/marginalg98k01.txt';
                                                    /* Marginaltabellfil */
FILENAME VEKTER '$AKU/data/tilkalib/opblg98k01.txt';

options obs=100000;

/* ID er en makrovariabel som identifiserer nivaa et utvalget
skal kalibreres paa. Hvis recordene representerer personer
og kalibreringen skal finne sted paa personnivaa, maa ID
vaere en variabel som identifiserer personene entydig i
utvalget, f.eks foedselsnummer eller loepenummer for person.
Hvis ID er et nummer som identifiserer familie eller hus-
holdning, vil kalibreringen foregaa som om alle individdata
som inngaar i kalibreringen (kjoenn, alder og sn94 fra
register) var erstattet med deres ''gjennomsnitt'' for familien
/husholdningen. Dette er ekvivalent en kalibrering paa
familie/husholdningsnivaa.
*/
/* UT = navn paa datasettet som produseres ved kalibreringen.
Datasettet bestaar av datasettet AKUUTV generert nedenfor med
tillegg av tre sett vekter, W1, W2 og W3 omtalt i dokumentasjon-
```



```

    en av metoden. W3 er de kalibrerte vektene.
*/
%LET ID=f_nr;
%LET UT=AKU.test; /* Sasdatasett som dannes under kalibreringen */
%LET AAR='1998';
%LET KVARTAL='1';
%LET MND='00';

/* A. Lager SAS datasett av AKU utvalget. Datasettet MAA inneholde
foelgende variable som brukes i kalibreringen:

    kjonn(2 kat.), aldgr(12 kat), sn94(4 kat) og fylke.

I tillegg maa variabelen spesifisert til ''ID =' ' ovenfor
vaere med. ''fylke'' og ID-variabelen behandles som karakter-
variable. Variablene ''funreg'' og ''svar'' brukes til aa ekstra-
here relevante respondenter fra bruttoutvalget. Andre variable
er frivillige og blir med paa UT-datasettet.
*/

data AKUUTV;
infile AKUFIL missover lrecl=2846;
length erio $ 1 gangmed 3 trekkaar 3 trekkkv 3 trekkuke 3 uke 3
    aar $ 4 kvartal $ 2 mnd $2 svar 3 funreg 3 kjonn 3 aldgr 3
    sn94 3 naceireg 3 fylke $ 2;
input f_nr $ 1-11 erio $ 12-12 gangmed 19-19 trekkaar 20-21
    trekkkv 22-22 trekkuke 23-24 uke 25-26 aar $ 30-33
    kvartal $ 34-34 svar 35-35 mnd $ 36-37 funreg 40-40
    kjonn 41-41 aldgr 43-44 sn94 45-45 naceireg 46-46
    fylke $ 47-48;
if (funreg eq 2 or svar eq 2) then delete; /* Frafall slettes */
if ((aar=&AAR) and (kvartal=&KVARTAL)) then output;
/* Periode selekteres */

/* B. Lager SAS datasett av tabell for fylke*etterstrata for hele
populasjonen. Dette maa ogsaa inneholde variablene ''kjonn,
alder, sn94 og fylke'' I tillegg maa den inneholde variabelen
''total'' som angir antall individer i hver celle for

```

```

                                                                    kjonnn x alder x sn94 x fylke.
*/
data STRATAB;
infile MARGINAL missover;
length kjonnn 3 aldgr 3 sn94 3 fylke $ 2 aar $ 4 kvartal $ 1 mnd $ 2;
input kjonnn 1-1 aldgr 3-4 sn94 6-6 fylke $ 8-9 total 30-34
       aar $ 40-43 kvartal $ 44-44 mnd $ 45-46;
if fylke='99' then delete;
if ((aar=&AAR) and (kvartal=&KVARTAL)) then output;
run;

%INCLUDE '$AKU/prog/metode/kalibaku.sas';

```

3.2 Beregningsdelen

Også denne delen av programmet tar sikte på å være mest mulig selvforklarende. Noen kommentarer er likevel på sin plass.

Det er laget et eget program for å beregne $W2$ -vektene mens ”standard” kalibreringsprogrammet benyttes for å lage $W3$ -vektene. For å lage $W3$ -vektene er det laget en løkke som kalibrerer ett fylke om gangen. Standard kalibreringsprogrammet ligger på en egen fil og kalles opp hver gang løkken gjennomgås. Standard kalibreringsprogrammet kunne i prinsippet også ha vært benyttet til å beregne $W2$ -vektene, men ville ikke kunne håndtere 0-celler i etterstratifiseringen.

Det er satt en sperre på hvor mye vektene kan tillates å forandre seg fra $W2$ til $W3$ på en slik måte at $0.33 \leq W3/W2 \leq 10$ for alle individer i alle fylker. Gir kalibreringen, slik den er beskrevet ovenfor, endringer som er utenfor disse grensene, vil kalibreringsprogrammet starte en iterativ prosess som tvinger alle endringene til å holde seg innenfor grensene samtidig som $W3$ -vektene fortsatt kalibrerer som ønsket. Spesifikasjonene til dette kravet gis i programkoden nedenfor, mens den iterative prosedyren foregår innenfor standardkalibreringsprogrammet.

```

proc freq data=AKUUTV;
  table sn94;

```

```

data AKUUTV;

```

```

    set AKUUTV;
    if (aldgr=12 and sn94<=3) then sn94=1;

proc freq data=AKUUTV;
    table sn94;

data STRATAB;
    set STRATAB;
    if (aldgr=12 and sn94<=3) then sn94=1;

proc sort data=AKUUTV;
    by fylke kjonn aldgr sn94;

/* 2. Lager tabell for fordeling etter fylke*etterstrata, n_{ef}
                                     i utvalget. */

proc freq data=AKUUTV;
    table fylke*kjonn*aldgr*sn94 / noprint nopercnt norow nocol
                                     sparse out=T0;

/* 2A. Lager tabell for fordeling etter fylke, n_{+f} */

proc freq data=T0;
    table fylke / noprint nopercnt norow nocol out=FYLKUTV; /* n_{+f} */
    weight count;
run;

/* 3A. Sorterer tabellen etter fylke*etterstrata. */

proc sort data=STRATAB;
    by fylke kjonn aldgr sn94;

/* 3B. Lager tabell for fordeling etter fylke N_{+f}
                                     og etter etterstrata, N_{e+}, i populasjonen. */

proc freq data=STRATAB;
    table fylke / noprint nopercnt norow nocol out=FYLKER; /* N_{+f} */

```

```

table kjonn*aldgr*sn94 / noprint nopercnt norow nocol
                                sparse out=TRE;
weight total;                                /* N_{e+} */

data FYLKER;
  set FYLKER;
  FYLKPOP = count;          /* N_{+f} */
  keep fylke FYLKPOP;

data TRE;
  set TRE;
  CPOP = count;            /* N_{e+} */
  drop count percent;

proc sort data=TRE;
  by kjonn aldgr sn94;
run;

/* 4. Lager fylkesvise oppblaasningsvektorer i henhold til utvalgsdesignen. */

data FYLKER;
  merge FYLKER FYLKUTV;
  by fylke;
  W1 = FYLKPOP/count;      /* W1_{f} = N_{+f}/n_{+f} */
  keep fylke W1;
run;

/* 5. Kobler de fylkesvise W1-vektene til personene i utvalget. */

data AKUUTV;
  merge AKUUTV FYLKER;
  by fylke;
run;

/* 6. Estimerer etterstratastoeerrelse 'hat{N}_{fe}'
      paa grunnlag av utvalget og oppblaasingsvektorer. */

data T0;

```

```

merge TO FYLKER;
by fylke;
WC =W1*count;          /* WC_{fe} = \hat{N}_{fe} = W1*n_{fe} */
drop percent;
run;

proc freq data=TO;
  /* COUNT_{e} = \hat{N}_{+e} = \sum_{f=1}^F \hat{N}_{fe} */
  table kjonn*aldgr*sn94 / noprint nopercent norow nocol sparse
                        out=STRATUTV;
  weight WC;

proc sort data=STRATUTV;
  by kjonn aldgr sn94;
run;

/* 7. Kobler sammen estimerte og faktiske ettersstratatoerrelser. */

data STRATUTV;
  merge STRATUTV TRE;
  by kjonn aldgr sn94;
  if COUNT=0 then QW = 0;
  else QW = CPOP/COUNT;          /* QW_{e} = N_{+e}/\hat{N}_{+e} */
  drop percent COUNT CPOP;
run;

/* 8. Beregner etterstratifiserte vekter for hele utvalget. */

proc sort data=AKUUTV;
  by kjonn aldgr sn94;

data AKUUTV;
  merge AKUUTV STRATUTV;
  by kjonn aldgr sn94;
  W2=W1*QW;          /* W2_{fe} = W1_{f}*QW_{e} */
run;

```

```

%MACRO FYLKER;

%LET INDATA=IN;
%LET UTDATA=IN;

%DO FYLK=1 %TO 20;
    %IF &FYLK=13 %THEN %GO TO NESTE;

%PUT FYLK = &FYLK;

data STRA2;
    set STRATAB;
    if fylke ne &FYLK then delete;

/* Lager fasiter til kalibrering for fylke nr. &FYLK. */

proc freq data = STRA2;
    table kjonn / noprint nopercnt norow nocol sparse out=KJONN;
    table aldgr / noprint nopercnt norow nocol sparse out=ALDGR;
    table sn94 / noprint nopercnt norow nocol sparse out=SN94;
    weight total;

data FASIT;
    set KJONN ALDGR SN94;
    fasitall = count;
    keep fasitall;
/* Initialiserer kalibreringen, fylke nr. &FYLK. */

%LET VEKT=W2;
%LET NYVEKT=W3;
%LET LAV=0.333;          /* Nedre tillatte grense for W3/W2. */
%LET HOY=10.0;          /* Oevre tillatte grense for W3/W2. */
%LET KRIT=0.001;        /* Konvergenskriterium. */
%LET MAXIT=50;          /* Max antall iterasjoner. */
%LET Q=1/nid;
%LET FASITVAR=
    MANN KVINNE A16_19 A20_24 A25_29 A30_34 A35_39 A40_44 A45_49 A50_54
    A55_59 A60_64 A65_69 A70_74 NA1 NA2 NA3 NA4;

```

```

%LET VELG=1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 ;

DATA &INDATA;
  SET AKUUTV;
  IF fylke ne &FYLK then delete;

DATA A;
  SET &INDATA;
  ARRAY ALD(12) A16_19 A20_24 A25_29 A30_34 A35_39 A40_44 A45_49
                A50_54 A55_59 A60_64 A65_69 A70_74;
  ARRAY SN94(4) NA1 NA2 NA3 NA4;

  if kjonn=1 then mann=1; else mann=0;
  kvinne=1-mann;

  do i=1 to 12;
    if aldgr=i then do;
      ALD(i)=1;
    end;
    else do;
      ALD(i)=0;
    end;
  end;

  do k=1 to 4;
    if sn94=k then do;
      SN94(k)=1;
    end;
    else do;
      SN94(k)=0;
    end;
  end;

keep &ID W1 W2 &FASITVAR;

RUN;

/* Kalibreringsrutinen kalles. */

```

```

%INCLUDE '$AKU/prog/metode/nyvektAKU.sas';

%IF &FYLK=1 %THEN %DO;
    data &UT;
        set IN;
    %END;
%ELSE %DO;
    data &UT;
        set &UT IN;
    %END;
%NESTE: %END;

%MEND FYLKER;

%FYLKER
run;

proc sort data=&UT;
by &ID;
run;

data test;
    set &UT;
FILE VEKTER; /* Tekstfil som ''put'' skriver ID nummer,
               vekter og periode for undersoekelsen til. */
mnd = &MND;
q23 = W3/W2;
q13 = W3/w1;
q12 = w2/w1;
put &ID 1-11 @12 W1 F14.9 @26 W2 F14.9 @40 W3 F14.9 @55 aar $4.
      @60 kvartal $1. @62 mnd $2.;
run;

PROC UNIVARIATE DATA=test;
    VAR w2 w3 q23;
run;

```



```

*/ Skriver ut records hvor W3/W2=0.333 eller >=3 /*

data test2;
    set test;
if ((q23 <= 0.333) or (q23 >= 3.0)) then output;
keep &ID aar mnd kvartal fylke kjonn aldgr nace1 W1 W2 W3 q23;
run;

proc print data=test2;
run;

```

3.3 Kalibreringsrutinen

Til sist i dette avsnittet presenteres programmet som utformer kalibreringen slik den er utformet for AKU i filen \$AKU/prog/metode/nyvektAKU.sas.

```

%MACRO VELGER(A,B,C);
%*
    Denne macroen tar for seg en macrovariabel (A) som består av en
    liste med karakterstrenger adskilt med blanke og lager en ny
    liste med et uvalg av disse karakterstrengene. For } bestemme
    hvilke strenger som skal velges ut, brukes makrovariabelen B
    som består av 0-ere og 1-ere adskilt med blanke, svarene til
    karakterstrengene i A. 1-tall indikerer at variabelen skal
    velges ut, 0 at den ikke skal velges ut. Resultatet legges i C.
    ;
%GLOBAL &C;
%LET A=&A DUMMY;
%LET F=;
%DO I=1 %TO 1000;
    %LET D=%SCAN(&A, &I);
    %IF &D=DUMMY %THEN %GOTO C1;
    %LET E=%SCAN(&B, &I);
    %IF &E=1 %THEN %DO;
        %LET F=&F &D;
    %END;
%END;

```

```

%C1:%LET &C=&F;
%MEND VELGER;

%VELGER(&FASITVAR,&VELG,XTXT)
*%PUT XTXT = &XTXT;

%MACRO ANTALL(A,B);
%*
    Denne macroen tar inn en liste med variabelnavn (&B) og teller
    opp hvor mange variabelnavn det er i listen.
;
    %GLOBAL &B;
    %LET A=&A DUMMY;
    %DO I=1 %TO 1000;
        %LET F=%SCAN(&A,&I);
        %IF &F=DUMMY %THEN %GOTO C1;
        %END;
%C1: %LET &B=%EVAL(&I-1);
%MEND ANTALL;

%ANTALL(&XTXT,NVAR)
*%PUT NVAR = &NVAR
*
    ***** STEG 2. *****
    Nytt datastep som genererer klyngedatasett p}
    grunnlag av individdatasettet A;

PROC SORT data=A;
    by &ID;

DATA B;
    set A;
    by &ID;
*
    Definerer array x(.) av variable som vi |nsker } aggregere over
    husholdninger samt hjelpearray y(.). (Disse aggregerte st|rrelsene
    kan eller skal brukes til } gruppere husholdningene.)

```

```

;
array x(&NVAR) &XTXT;
array y(&NVAR) y1-y&NVAR;
*
Aggregerer ulike st|rrelser over husholdningen.
;
if first.&ID then do;
    nid=0;
    end;
nid+1;
do i=1 to &NVAR;
    if first.&ID then y(i)=0;
    y(i)=x(i)+y(i);
    if last.&ID then x(i)=y(i);
    end;
if last.&ID then do;
    q=&Q;
    dvkt=&VEKT;
    output;
    end;
retain y1-y&NVAR;
keep &XTXT &ID q dvkt;
run;

```

***** STEG 3. *****

Genererer felles faktorer som a1 - a&NX som de benyttede fasit-
variablene x1-x&NX m} multipliseres med for } danne grunnlag for bereg-
ning av nye vekter.

```

;
* Proc Print DATA=B (obs=50);

DATA FASIT2;
set FASIT;
velg = scan("&VELG",_n_);
if velg = 1 then output;
keep fasitall;

```

```

PROC IML;
  USE B var _all_;
  READ all var {&XTXT} into X;
  READ all var {dvkt} into D;
  READ all var {q} into Q;
  USE FASIT2 var _all_;
  READ all var {fasitall} into FA;

  DX = hdir(D,X);
  QX = hdir(Q,X);
  DQX = hdir(D,QX);
  n = nrow(X);
  XDQX =DQX'*X;
  La = solve(XDQX,FA-DX[+,]');

  F = 1 + QX*La;
  W = D#F;
  if any((W<D*&LAV)|(W>D*&HOY)) then do;
    iter = 0;
    lav = &LAV;
    hoy = &HOY;
    print "FYLKE =" &FYLK;
    print lav hoy;
    F = (&LAV<>F)><&HOY; /* Trunkerer F */
    XDQX = (DQX#((F>=&LAV)#(F<=&HOY)))'*X;
    print iter;
    do until (iter>&MAXIT);
      iter = iter + 1;
      Psi = hdir(D#F,X)[+,]' - FA;
      De = solve(XDQX,-Psi);
      La = La + De;
      F = 1 + QX*La;
      F = (&LAV<>F)><&HOY;
      W2 = D#F;
      krit =sqrt((W2-W)##2/(D#Q))[+,];
      print iter krit;
      W = W2;
      if (krit < &KRIT) then goto cont;
  end;

```

```

                XDQX = (DQX#((F>=&LAV)#(F<=&HOY))) '*X;
            end;
        end;
cont: NYVEKT = symget('NYVEKT');
        mattrib W colname=NYVEKT;
        create W from W[colname=NYVEKT];
        append from W;

DATA C;
        merge B W;
        keep &ID &NYVEKT;

* PROC PRINT data=C(obs=20);

                ***** STEG 4 *****

        Kobler filen generert i steg 3 sammen med filen generert i steg 1
        ved hjelp av &ID. Den nye filen tas vare p} og kan kalles separat.
;
PROC SORT data=C;
        by &ID;

PROC SORT data=&INDATA;
        by &ID;

DATA R;
        merge &INDATA C;
        by &ID;

PROC SORT data=R;
        by &ID;

DATA &UTDATA;
        set R;
        by &ID;

PROC DATASETS MT=DATA;
        DELETE R B C;

```

```
run;
```

4 Utskrift fra programmet

Det følgende viser en utskrift fra en månedskjøring, nærmere bestemt fra utvalget i april måned 1999, spesifisert i filen kalib.sas ved spesifikasjonen

```
LIBNAME AKU '$AKU/data/tilkalibs';
FILENAME AKUFIL '$AKU/data/tilkalib/utug99m04.txt';
                                                    /* Utvalgsfil */
FILENAME MARGINAL '$AKU/data/tilkalib/marginalg99m04.txt';
                                                    /* Marginaltabellfil */
FILENAME VEKTER '$AKU/data/tilkalib/opblg99m04.txt';

%LET ID=f_nr;
%LET UT=AKU.test; /* Sasdatasett som dannes under kalibreringen */
%LET AAR='1999';
%LET KVARTAL='2';
%LET MND='04';
```

4.1 Personer med ugyldige fylkeskoder

Det forekommer at personer ikke har gyldig fylkeskode i register, nærmere bestemt i Marginaltabellfilen. Disse blir skrevet ut og deretter ignorert. Utskriften viser at det i april 1999 var 7 slike, fordelt på 6 menn og en kvinne, alle i aldersgruppen 9-12 og alle unntatt 1 med NACE1=4 (ikke registerselsatt).

The SAS System

OBS	KJONN	ALDGR	NACE1	FYLKE	AAR	KVARTAL	MND	TOTAL
1	1	9	2	99	1999	2	04	1
2	1	9	4	99	1999	2	04	1

3	1	10	4	99	1999	2	04	1
4	1	11	4	99	1999	2	04	1
5	1	12	4	99	1999	2	04	2
6	2	10	4	99	1999	2	04	1

4.2 Kontroll av endring i vekter

Det er lagt inn at forholdet $W3/W2$ tillates større enn 10 og ikke mindre enn $1/3$. Dette kommer til uttrykk ved verdiene LAV og HOY. Brytes dette kravet starter en iterativ prosedyre som tvinger vektene til å tilpasse seg grenseverdiene. I april 1999 ble dette nødvendig i fylke 5, Oppland. Det var denne gangen nødvendig med 10 iterasjoner for å få konvergens.

Noen måneder kommer prosedyren ikke til anvendelse og det skrives ingen rapport. Andre ganger vil det være behov for å bruke den i flere fylker. KRIT er konvergenstestvariabelen Prosedyren har konvergert når KRIT blir liten nok.

The SAS System

FYLKE = 5

LAV	HOY
0.333	10

ITER
0

ITER	KRIT
1	47.948671

ITER	KRIT
2	9.6845394

```
ITER      KRIT
  3 2.8049435
```

```
ITER      KRIT
  4 0.8120435
```

```
ITER      KRIT
  5 0.2350844
```

```
ITER      KRIT
  6 0.0680562
```

```
ITER      KRIT
  7 0.0197021
```

```
ITER      KRIT
  8 0.0057037
```

```
ITER      KRIT
  9 0.0016512
```

```
ITER      KRIT
 10 0.000478
```

Som en del av kontrollen med vektene skrives det ut statistikker over hvordan vektene W_2 , W_3 og $Q_{23} = W_3/W_2$ varierer. Slike utskrifter vises nedenfor.

The SAS System

Univariate Procedure

Variable=W2

Moments

N	6564	Sum Wgts	6564
Mean	484.3234	Sum	3179099
Std Dev	98.74764	Variance	9751.096
Skewness	0.252626	Kurtosis	2.10187
USS	1.6037E9	CSS	63996440
CV	20.38878	Std Mean	1.218828
T:Mean=0	397.3682	Pr> T	0.0001
Num ^= 0	6564	Num > 0	6564
M(Sign)	3282	Pr>= M	0.0001
Sgn Rank	10773165	Pr>= S	0.0001

Quantiles(Def=5)

100% Max	1229.542	99%	757.2729
75% Q3	540.6565	95%	645.1249
50% Med	481.398	90%	596.2641
25% Q1	431.1879	10%	368.2637
0% Min	123.2568	5%	318.5043
		1%	237.7469
Range	1106.286		
Q3-Q1	109.4686		
Mode	562.0657		

Extremes

Lowest	Obs	Highest	Obs
123.2568(5081)	947.49(1521)
123.2568(2538)	953.0819(4925)
123.2568(1217)	954.9676(5533)

140.0038(5882)	1093(2835)
140.0038(5428)	1229.542(5214)

The SAS System

Univariate Procedure

Variable=W3

Moments

N	6564	Sum Wgts	6564
Mean	484.7334	Sum	3181790
Std Dev	141.6342	Variance	20060.26
Skewness	0.531596	Kurtosis	1.137538
USS	1.674E9	CSS	1.3166E8
CV	29.219	Std Mean	1.748171
T:Mean=0	277.2803	Pr> T	0.0001
Num ^= 0	6564	Num > 0	6564
M(Sign)	3282	Pr>= M	0.0001
Sgn Rank	10773165	Pr>= S	0.0001

Quantiles(Def=5)

100% Max	1210.418	99%	911.2096
75% Q3	570.6114	95%	720.2332
50% Med	476.6824	90%	655.5807
25% Q1	396.7932	10%	309.2004
0% Min	99.30352	5%	269.0815
		1%	177.4726
Range	1111.115		
Q3-Q1	173.8183		
Mode	480.9923		

Extremes

Lowest	Obs	Highest	Obs
99.30352(5081)	1078.39(3047)
99.30352(2538)	1119.234(5214)
99.30352(1217)	1132.504(3582)
139.2202(702)	1132.504(3869)
141.0981(5797)	1210.418(1521)

The SAS System

Univariate Procedure

Variable=Q23

Moments

N	6564	Sum Wgts	6564
Mean	1.001049	Sum	6570.888
Std Dev	0.215051	Variance	0.046247
Skewness	0.876535	Kurtosis	4.476118
USS	6881.303	CSS	303.5187
CV	21.48256	Std Mean	0.002654
T:Mean=0	377.1363	Pr> T	0.0001
Num ^= 0	6564	Num > 0	6564
M(Sign)	3282	Pr>= M	0.0001
Sgn Rank	10773165	Pr>= S	0.0001

Quantiles(Def=5)

100% Max	2.719861	99%	1.625077
75% Q3	1.11348	95%	1.368138
50% Med	0.983896	90%	1.254469
25% Q1	0.869694	10%	0.758681
0% Min	0.333	5%	0.680822
		1%	0.534341
Range	2.386861		

Q3-Q1 0.243786
 Mode 0.773806

Extremes

Lowest	Obs	Highest	Obs
0.333(5797)	2.635941(2198)
0.333(4020)	2.709176(1592)
0.333(2507)	2.709176(3199)
0.333(96)	2.719861(1010)
0.339911(3183)	2.719861(5798)

Individer som har $Q_{23} = 1/3$ eller $Q_{23} \geq 3$ skrives ut med verdier på kalibreringsvariablene og vekter. Fødselsnumrene er her sensurerte. Alle hadde $Q_{23} = 0.333$ og var fra Oppland. Uten den iterative prosedyren ville de ha $Q_{23} < 0.333$.

The SAS System

```

          K
          V
          A      K  A  N  F
          R      J  L  A  Y      F
O  A  T  M  O  D  C  L      -      Q
B  A  A  N  N  G  E  K      N      W      W      W      2
S  R  L  D  N  R  1  E      R      1      2      3      3

1  1999  2  04  1  1  2  05  *****  494.701  601.702  200.367  0.333
2  1999  2  04  1  1  2  05  *****  494.701  601.702  200.367  0.333
3  1999  2  04  2  1  3  05  *****  494.701  423.718  141.098  0.333
4  1999  2  04  2  1  3  05  *****  494.701  423.718  141.098  0.333

```

4.3 Kalibrerte vekter

skrives for hele utvalget ut på filen opblg99m04 som inneholde fødselsnum-
 mer, W_1 , W_2 og W_3 , år kvartal og måned og ser ut som følger (personnum-

mer sensurert). Denne filen kobles senere til resten av AKU.

Fnr.	W1	W2	W3
År k mm			
***** 547.104918033	586.775324815	458.296977585	1999 2 04
***** 475.447887324	462.549901697	499.533821999	1999 2 04
***** 471.149171271	400.079290496	363.715994353	1999 2 04
***** 471.149171271	542.379124061	493.082164237	1999 2 04
***** 448.468013468	403.224932425	395.011006519	1999 2 04
***** 448.468013468	437.036785983	477.716044312	1999 2 04
.....

5 Referanser

- Boije, L (1996). Estimering og etterstratifisering i AKU (boi,25.03.96). Upublisert notat.
- Deville, J-C. og Särndal, C-E. (1992): Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, vol 87, no 418 pp 376-382.
- COUNCIL REGULATION (EC) no 577/98 of 9 March 1998 on the organisation of a labour force sample survey in the community.
- Heldal, J. (1992) A method for calibration of weights in sample surveys. *Arbeidsnotat fra avdeling for personstatistikk 3/1992*.
- Särndal, C-E., Swensson, B. og Wretman, J. (1992): *Model Assisted Survey Sampling*. Springer Verlag
- Zhang, L-C (1998): Dokumentasjonsrapport: Den nye estimeringsmetoden for Arbeidskraftundersøkelsen (AKU). Fylkesvist kalibrering med landsetterstratifiserte vektorer med som startverdier. *SSB: Notater 98/1*.

De sist utgitte publikasjonene i serien Notater

- 1999/69 R. Eriksen: Inntekts- og formuesundersøkelsen for selskaper skattlagt med hjemmel i petroleumsskatteloven for årene 1994, 1995 og 1996: Dokumentasjon. 19s.
- 1999/70 B.O. Lagerstrøm: Bostøtteordningen i Husbanken: Dokumentasjonsrapport. 20s.
- 1999/71 L. Wiker og E. Knutsen: Inntekts- og formuesundersøkelsen for aksjeselskaper 1997: Dokumentasjon. 30s.
- 1999/72 K. Ibenholt: Framskrivning av avfall og tilhørende utslipp ved bruk av MSG6: Teknisk dokumentasjon. 45s.
- 1999/73 A. Langørgen og R. Aaberge: Like kommuner. 19s.
- 1999/74 S. Kristoffersen: Aksjestatistikk 1995, 1996 og 1997: Dokumentasjon. 15s.
- 1999/75 B. Halvorsen og M.I. Hansen: Dokumentasjon av utdrag fra skattestatistikken 1974-1994 for kobling mot forbruksundersøkelsen. 38s.
- 1999/76 H.P. Dahlslett og E. Engelen: Sentrumsstatistikk for Oslo og Akershus: Et pilotprosjekt. 50s.
- 1999/77 R. Eriksen og B. Haveråen: Inntekts- og formuesundersøkelsen for personlig næringsdrivende 1996: Dokumentasjon. 52s.
- 1999/78 P.M. Holt, S. Kristoffersen og V. Pedersen: Beregning av vekter til inntekts- og formuesundersøkelsen 1997. 26s.
- 1999/79 P.M. Holt og T. Vevle: Skattestatistikk for rederier 1996 og 1997: Dokumentasjon. 26s.
- 1999/80 T. Bye, Ø. Døhl og J. Larsson: Klimagasskvoter i kraftintensive næringer. Konsekvenser for utslipp av klimagasser, produksjon og sysselsetting. Regionale konsekvenser. 11s.
- 1999/81 B. Mathisen: Flyktninger og arbeidsmarkedet 4. kvartal 1998. 39s.
- 1999/82 Ø. Kleven, E. Dalheim og D. Roll-Hansen: Innvandreres utdanning: - en pilotundersøkelse. 61s.
- 1999/83 E. Fidjestøl og I. Håland: Yrkeskatalog: Pr. desember 1999. 136s.
- 1999/84 T. Solberg: Virkning av revisjon på Avlingsstatistikk for jordbruksvekster i 1998. 24s.
- 1999/85 R. Choudhury, T. Eika og L. Haakonsen: KVARTS i praksis II: Systemer og rutiner i den daglige driften. 66s.
- 1999/86 G. Frøiland: Økonometrisk modellering av husholdningenes konsum i Norge: Demografi og formuesefferter. 55s.
- 1999/87 Y. Li: Beregning av elementæraggater i konsumprisindeksen ved hjelp av generalisert gjennomsnitt. 41s.
- 1999/88 L. Rogstad og S.T. Vikan: Kobling av adresseregistrene i DSF og GAB 1999: Dokumentasjon av samsvar og avvik. 31s.
- 1999/89 E. Dalheim, J-A. S. Lie og D. Roll-Hansen: En skjemabasert komplettering av registeret over befolkningens høyeste utdanning - forprosjekt med fokus på innvandrere. 60s.
- 1999/90 K-A. Hovland og Å. Nossum: Flyreiser i konsumprisindeksen. 39s.
- 2000/1 E. Rønning: Utenlandske statsborgere og kommunestyrevalget 1999: Dokumentasjonsrapport. 34s.
- 2000/2 M. Bråthen: Personer registrert som yrkeshemmet i SOFA-søkerregisteret. 25s.
- 2000/3 A.K. Johnsen og Ø. Hokstad: FoB2001: Kvalitativ testing av boligskjema - prøveundersøkelse 1999: Dokumentasjonsnotat. 32s.
- 2000/4 C. Hendriks, Ø. Hokstad og R. Sønsterudbråten: FoB2001: Boligtelling - prøveundersøkelse 1999: Dokumentasjonsnotat. 60s.