



Leiv Solheim

Foreløpige tall i FoB2001

Utvalg, vektor, estimering
og usikkerhet

Notater

1. Innledning og sammendrag	3
1.1. Innledning.....	3
1.2. Utvalget til foreløpige tall.....	3
1.3. Prinsippene for å beregne vektene	4
1.4. Feilkilder og usikkerhet i tallene.....	5
1.5. Innholdet i notatet	8
2. Utvalget til foreløpige tall i Fob2001.....	9
2.1. Boligtellingen.....	9
2.2. Utvalget er Arbeidskraftundersøkelsen i 4. kvartal 2001(AKU-2001-4) med et tillegg.....	10
2.3. Egenskaper til utvalget av formelle boliger og husholdninger	11
2.4. Adressegrupper - konsekvenser for frafallet.....	12
3. Estimere antall boliger og husholdninger - foreløpige tall.....	15
3.1. Trekksannsynligheter for boligene og husholdningene	16
3.2. Beregne vekter på grunnlag av utvalgsdesignen.....	19
3.3. Bruk av statistiske modeller for å beregne vekter - data er alle familienumrene.....	21
3.4. Bruk av statistiske modeller for å beregne vekter - data er alle personene.....	25
3.5. Vekter for foreløpige tall - kalibrering av designvektene	28
3.6. Valg av betingelser for å kalibrere vektene	30
3.7. Hvordan beregne foreløpige tall for studentene.....	33
3.8. Opplegget for å beregne vekter.....	34
4. Usikkerhet	42
4.1. En enkel formel for å bestemme hva som skal publiseres	42
4.2. Beregne den faktiske usikkerheten ved bruk av de inverse trekksannsynlighetene.....	45
4.3. Usikkerheten til de kalibrerte tallene vi publiserer	46
4.4. Usikkerheten til antallet husholdninger i hvert fylke og for hele landet.....	48
5. Imputering.....	49
6. Referanser	50
Vedlegg A Beregne vekter på person- og husholdningsnivå - en sammenlikning.....	51
Vedlegg B Enkle person- og familievekter.....	58
Vedlegg C Detaljert fordeling av DSF familiene bosatt i boligbygg.....	60
Vedlegg D Beregnede befolkningstall basert på AKU+ utvalget sammenliknet med befolkningstallene per 3. november 2001	61
Vedlegg E Fordeling av husholdningene i utvalget etter antall DSF familier og fylker	62
De sist utgitte publikasjonene i serien Notater.....	63

1. Innledning og sammendrag

1.1. Innledning

Folke- og bolig tellingen 2001(FoB2001) er både en total skjemaundersøkelse av alle boliger og husholdninger og en registerundersøkelse av personene bosatt i Norge per 3. november 2001. De foreløpige tallene for FoB2001 gjelder kun husholdninger og boliger og tar utgangspunkt i det formelle bostedet - altså den adressen som alle personer er registrert bosatt på ifølge Det sentrale folkeregisteret per 3. november 2001.

Vi presenterer her det metodiske grunnlaget for de foreløpige tallene til Folke- og bolig tellingen 2001 som ble publisert 22. april 2002. Tabellene med foreløpige tall er basert på et utvalg av boligskjemaene. Dette utvalget bygger på en utvidelse av utvalget til Arbeidskraftundersøkelsene (AKU) for 4. kvartal 2001. Utvidelsen består av et tillegg blant de eldre - det betyr personer over 74 år. Utvalget til AKU er trukket som et selvveiende utvalg av familienumre, se definisjonen av DSF familier seinere, i hvert fylke og slik er også tillegget trukket. Boligskjemaene ble sendt til familier som enten var definert ved et felles familienummer eller der to familienumre var koblet sammen på grunn av at en kvinne og mann var samboere med felles barn, disse kalles også Besys familier.

Det er tre hovedmål med Boligtellingen: Det første er å etablere husholdningen som enten formelt bor i boligen eller faktisk bor i boligen. Det første betyr at vi bygger på det formelle bostedet ifølge Det sentrale folkeregisteret (DSF), mens det andre begrepet bygger på hvor personene normalt oppholder seg. Forskjellen blir særlig åpenbar for borteboende studenter som formelt sett ifølge DSF bor hjemme hos foreldrene. Det andre er å beskrive boligen husholdningene bor i og det tredje er å identifisere boligene med bolignummeret - det siste gjelder da alle boliger som før tellingen ikke hadde en unik adresse i betydningen at det kun finnes en bolig på denne adressen.

I statistikken basert på det faktiske bostedet vil vi behandle studenter forskjellig fra andre. For studenter vil den faktiske adressen gå foran den formelle, mens for andre gjelder den formelle adressen i Det sentrale folkeregisteret per 3. november som faktisk adresse. Det betyr at når en sammenlikner statistikk basert på faktisk adresse med statistikk basert på formelt bosted vil antall boliger, husholdninger og personer i kommuner med universiteter og høyskoler stort sett øke, mens for kommuner/fylker med få eller ingen studieplasser vil antallet personer avta. Virkningen vil være størst for de fire universitetsbyene Oslo, Bergen, Trondheim og Tromsø der både befolkningen, antall husholdninger og boliger øker. Totalt for Norge vil antallet husholdninger og boliger øke når vi går fra formelt bosted til faktisk bosted. Denne forskjellen er langt større enn utvalgsfeilen vi har i de foreløpige tallene på grunn av at de kun bygger på et utvalg av skjemaene.

1.2. Utvalget til foreløpige tall

Utvalget av utsendte boligskjemaer som vi bruker for å produsere foreløpige tall er en utvidelse av utvalget til Arbeidskraftundersøkelsen (AKU) for 4. kvartal 2001. Utvalget til AKU er et utvalg av DSF familier. En DSF familie består av personer med felles familienummer. Hver DSF familie har en referanseperson og det er hans eller hennes fødselsnummer som er DSF familiens familienummer. Dersom DSF familien er et ektepar er familienummeret lik mannens fødselsnummer, ellers er familienummeret lik fødselsnummeret til den som er lik enpersonfamilien eller aleneforelderen. I AKU deltar kun personer mellom 16 og 74 år. Utvalget til AKU er altså trukket blant DSF familier med en referanseperson mellom 16 og 74. For å gjøre utvalget til foreløpige tall representativt for alle DSF familier har vi trukket et tillegg etter samme opplegg som til AKU blant resten av DSF familiene, dvs. for familiene med en referanseperson utenfor aldersgruppa 16 til 74 år. I praksis betyr det blant DSF familier med referansepersoner over 74 år. Det totale utvalget av DSF familier(AKU + utvalget)

vi bruker til foreløpige tall er et selvveiende utvalg av DSF familier for hvert fylke. De fire minste fylkene etter folketallet, Finnmark, Aust-Agder, Sogn og Fjordane og Nord-Trøndelag, er overrepresentert i utvalget. De tre største fylkene, Oslo, Akershus og Hordaland, er underrepresentert i utvalget. Dette betyr at kvaliteten på fylkestallene varierer mindre enn forskjellene på befolkningstallene skulle tilsi.

I FoB2001 er Besys familier brukt som mottakere av boligskjema. En Besys familie består enten av en eller to DSF familier. En Besys familie består av to DSF familier dersom referansepersonene i disse to DSF familiene er samboere med felles barn. Mottaker av boligskjema var den eldste personen i Besys familien. Dessuten ble det også sendt boligskjema til alle studenter som hadde søkt om borteboerstipend for våren og/eller høsten 2001.

For å finne boligskjemaene som er sendt til DSF familiene i utvalget og eventuelt til andre DSF familier eller Besys familier som hører til de husholdningene som de uttrukne DSF familiene hører til har vi søkt blant alle DSF familier på de adressene som utvalget av DSF familier hører til på. Slik vil da en husholdning som vår uttrukne DSF familie hører til, bli registrert med innsendt skjema dersom minst en av Besys familiene i husholdningen har sendt inn skjema. Bruttoutvalget av husholdninger var på 16 957, mens nettoutvalget var på 15 782 husholdninger eller boliger. Det gir en total svarprosent i utvalget på 93,1 prosent.

1.3. Prinsippene for å beregne vektene

En grunnleggende statistisk enhet i beregningene av vektene til foreløpige tall er DSF familien, dessuten bruker vi også Besys familie i det siste trinnet i beregningene av vektene. Vi deler nemlig inn befolkningen i fem typer Besys familier: en person (DSF) familie, ektepar uten barn, ektepar med barn, alene forelder (med barn) og samboerpar med felles barn.

Prinsippene for å beregne vektene bygger på følgende fem trinn:

- Vi deler alle DSF familier og personer bosatt i Norge 3. november 2001 inn i 76 strata basert på de 19 fylkene og fire adressegrupper. Inndelingen i fire adressegrupper bygger på om adressen er unik(kun en bolig på adressen) eller ikke(flere boliger på adressen) og om det finnes en annen bolig i nærheten som med stor sannsynlighet er identisk, vi sier at det finnes en stedfortreder, eller om det ikke finnes en slik stedfortreder.
- For hvert stratum definert i forrige punkt beregner vi antall DSF familier i populasjonen og antallet husholdninger i utvalget. Dessuten må vi kjenne antall DSF familier til hver husholdning i utvalget.
- For hver husholdning i utvalget beregner vi en vekt, som kalles en designvekt, ved først å beregne forholdet mellom antallet DSF familier i populasjonen og antallet husholdninger i utvalget i det stratum husholdningen tilhører. Deretter dividerer vi dette forholdet med antallet DSF familier i husholdningen.
- For hvert stratum definert ved fylke og etter om adressen er unik eller ikke, i alt 38 grupper, beregnes antallet Besys familier fordelt på fem kategorier i populasjonen og i utvalget og dessuten antallet personer i populasjonen og i utvalget fordelt på åtte aldersgrupper for menn og kvinner hver for seg.
- Vi kalibrerer for hver av de 38 gruppene definert i foregående punkt disse vektene med hensyn på en inndeling i de fem typene Besys familier og de åtte aldersgruppene for menn og kvinner beskrevet i foregående punkt. Resultatet er da vekter som gjør at utvalget er konsistent med antallet Besys familier i de fem gruppene og antallet personer i de åtte aldersgruppene for menn og kvinner i hvert stratum i populasjonen som vi beregnet tall for i foregående ballpunkt.

Det siste ballpunktet betyr at dersom vi ved bruk av designvektene i tredje ballpunktet har for få samboerpar med felles barn vil slike husholdninger i utvalget i gjennomsnitt få økt vektene, mens resten av husholdningene i gjennomsnitt får redusert vektene. På liknende vis vil vi når design vektene gir oss for få menn over åtti år øke vektene til husholdninger med menn over åtti år. Når vi skal samtidig ta hensyn til fordelingen for de fem typene Besys familier og 16 aldersgrupper vil det bli forholdsvis komplisert å beskrive i detalj hva som faktisk skjer med vektene i den enkelte husholdning.

1.4. Feilkilder og usikkerhet i tallene

Tabellene med foreløpige tall for FoB2001 bygger på et bruttoutvalg på 16 957 husholdninger. Disse tallene vil være usikre på grunn av flere typer feil. For det første har vi utvalgsfeil siden det er et utvalg. Videre har vi feil på grunn av frafall. Produksjonen av tabellene bygger på et nettoutvalg på 15 782 husholdninger, altså for 1 175 husholdninger og boliger har vi ikke mottatt noe boligskjema. Til slutt er det målefeil i de enkelte variablene som vi produserer tallene fra. Generelt regner en med at utvalgsfeilen ofte utgjør kun halvdelen av den totale feilen og dette tillegget skyldes følgende fem feilkilder: frafallsfeil, målefeil, registerfeil, bearbeidingsfeil og modellfeil. Nedenfor er de forskjellige feiltypene beskrevet i større detalj.

Innsamlings- og bearbeidingsfeil

Vi må regne med målefeil når skjemaene fylles ut. Dette kan både skyldes at spørsmålene kan være vanskelige å forstå og at den som fyller ut rett og slett misforstår spørsmålet eller fyller ut feil. Eksempler på spørsmål som kan føre til målefeil er spørsmålene om hvor mange andre oppholdsrom har boligen (spørsmål 15) og hvor stor er boligen (spørsmål 18b). Siden ordet oppholdsrom ikke er forklart kan det fort regnes med rom som vi ikke har ment skulle regnes med eller omvendt en tar ikke med rom som skulle vært med. Samtidig kan det også tenkes at oppgavegiver overser det viktige ordet andre og regner med f.eks. soverom. Vi må regne med at det riktige antallet kvadratmeter på størrelsen av boligen ikke er like lett å gi for mange, men her bør vi regne med at målefeilen er klart mindre på spørsmål 18a der en kun skal krysse av i hvilket intervall arealet til boligen ligger i.

Vi må også regne med bearbeidingsfeil knyttet til optisk lesning. Dette skyldes at tegn tolkes feil, f.eks. at 7 tolkes som 1 og omvendt. Andre eksempler kan være at en ikke ser noe svar i det hele tatt - for svakt eller uklart tegn på skjemaet. Denne type feil forsøker en å redusere mest mulig gjennom kontroller, men det er selvsagt umulig å kunne forhindre denne type feil.

Vi har ikke et eget opplegg for å beregne størrelsen på denne type feil nå i forbindelse med publiseringen av foreløpige tall. Dette vil vi komme tilbake til når de endelige tallene publiseres.

Utvalgsfeil

Vi har utarbeidet et opplegg for å kunne beregne utvalgsfeilen i de foreløpige tallene vi publiserer. Et rimelig godt anslag på usikkerheten til det publiserte antallet boliger eller husholdninger med en bestemt egenskap finner en fra følgende formel for standardavviket gitt ved

$$\bullet \quad SD(\hat{T}) = \sqrt{\frac{\hat{T} \cdot (N - \hat{T})}{15782}}$$

der

- \hat{T} er det publiserte tallet
- N er det beregnede totale antallet husholdninger og boliger for hele landet per 3. november i foreløpige tall

Setter vi inn 2 millioner for antallet husholdninger og boliger finner vi i tabell 1.1. nedenfor noen utvalgte tall som kan brukes som veiledning i vurderingene av usikkerheten i de foreløpige tallene.

Vi ser da at dersom vi har beregnet at 50 000 husholdninger eller boliger har en egenskap kan vi med stor sikkerhet påstå at det samme antallet ligger mellom 45 000 og 55 000. Vi ser av tallene at den relative feilen blir mindre etter som det publiserte tallet vokser, dvs. forholdet mellom standardavviket og det publiserte tallet avtar når det publiserte tallet vokser. Dette forholdet kaller vi det relative standardavviket. Den absolutte feilen øker imidlertid.

Tabell 1.1. Usikkerheten til de foreløpige tallene over antallet husholdninger og boliger med bestemte egenskaper.

Publisert tall	Standardavviket	Relativt standardavvik	Usikkerhet = 2 standardavvik	95% Konfidensintervall
5 000	795	15,9%	1 590	[3 410,6 590]
10 000	1 123	11,2%	2 246	[7 754,12 246]
25 000	1 768	7,1%	3 537	[21 463,28 537]
50 000	2 486	5,0%	4 971	[45 029,54 971]
100 000	3 470	3,5%	6 939	[93 061,106 939]
250 000	5 265	2,1%	10 530	[239 470,260 530]
500 000	6 894	1,4%	13 787	[486 213,513 787]
1 000 000	8 960	0,9%	15 920	[984 080,1 015 920]

Vi har valgt å publisere foreløpige tall dersom minst 5 000 husholdninger eller boliger har en egenskap.

Videre har vi også utledet at antallet husholdninger regnes ut med en sikkerhet som svarer til at

- for små fylker må vi regne med et avvik på inntil 2 000 husholdninger i forhold til det endelige tallet
- for store fylker må vi regne med et avvik på inntil 4 - 5 000 husholdninger i forhold til det endelige tallet
- for hele landet må vi regne med et avvik på inntil 12 500 husholdninger i forhold til det endelige antallet

Vi understreker at denne usikkerheten er knyttet til de tenkte endelige tallene over formelle husholdninger eller boliger, altså hvor personene bor ifølge Det sentrale folkeregisteret. I de endelige tallene vi skal publisere i FoB2001 vil det være størst vekt på faktisk bosted og da vil som tidligere nevnt antallet husholdninger og boliger stige betydelig mer enn usikkerheten angir.

I kapittel 4 er detaljene for utregningen av utvalgsfeilene beskrevet i større detalj.

Ikke-utvalgsfeil

For foreløpige feil kan vi tilnærmet se bort fra registerfeil siden vi i produksjonen kun bruker antallet DSF familier og antallet bosatte i Norge per 3. november. Noen endringer fram til publikasjonen av endelige tall kan vi nok regne med på grunn av forsinkete flyttemeldinger, men det vil ha en marginal betydning. Vi vil også hevde at modellfeil er irrelevant i denne sammenheng siden tabellene ikke er avhengig av modellberegninger.

Frafallsfeil må vi imidlertid regne med. Den totale svarprosenten er 93,1 prosent. Analyserer vi frafallet etter fylker og adressegrupper ser vi en varierende svarprosent. For de fire adressegruppene har vi følgende tall:

- For unike adresser med stedfortreder(eksempel er rekkehus) er frafallet på 6%
- For unike adresser uten stedfortreder(eksempel er eneboliger) er frafallet på 5%
- For ikke unike adresser med stedfortreder(eksempel er blokker) er frafallet på 10%
- For ikke unike adresser uten stedfortreder(eksempel er småhus) er frafallet på 9%

For fylkene varierer også frafallet en del med følgende to ytterpunkter:

- I Hedmark og Nord-Trøndelag er frafallet under 4%
- I Oslo og Finnmark er frafallet over 10%

Det avgjørende for kvaliteten på de tallene vi publiserer er om frafallet er rimelig tilfeldig fordelt innen hvert stratum. Vi har sett på estimatet av antallet personer i aldersgruppene og i fylkene for bruttoutvalget og nettoutvalget før vi kalibrerer nettoutvalget sammenliknet med den faktiske populasjonen vi bruker for situasjonen per 3. november 2001. Dessuten har vi også tatt med hva vi ville fått dersom vi kun brukte det opprinnelige AKU+ utvalget, dvs. kun de uttrukne DSF familier som var utgangspunktet for å hente fram innsendte boligskjemaer. I vedlegg D har vi sammenliknet AKU+ utvalget mot befolkningstallene fylkesvis. Vi finner da følgende fire tall for hele landet:

- Det sentrale folkeregisteret gir 4 520 947 personer
- AKU+ utvalget gir 4 524 823 personer
- Bruttoutvalget gir 4 572 906 personer
- Nettoutvalget gir 4 604 111 personer

AKU+ utvalget gir et befolkningstall som er utrolig nær det korrekte antallet personer i Norge siden det er knappe 4 000 for høyt. Vi ser av vedlegg D at avvikene er betydelig større i fylkene siden utvalget er langt mindre i hvert fylke. Generelt må en regne med en utvalgsfeil på ½ til 1 prosent for antallet personer i alt. Bruttoutvalget som er dannet ved å legge til andre DSF familier som hører til husholdningene ut fra hvordan boligskjemaene er fylt ut fører til at antallet personer øker med 48 000. Hvorfor denne økningen skjer kan skyldes flere faktorer. For det første skyldes nok dette at DSF familier er blitt korrigeret i forhold til status 3. november 2001 og dessuten kan en tenke seg at når to DSF familier kobles sammen i en husholdning vil faktisk antallet personer i den nye husholdningen øke med mer enn faktoren 2. Allikevel er jo økningen på 48 000 kun litt i overkant av en prosent. Økningen på drøyt 31 000 på grunn av frafallet virker også rimelig siden det er all grunn til å tro at små husholdninger har større sannsynlighet for frafall enn større. Økningen er imidlertid mindre enn utvalgsfeilen. Det er imidlertid ingen grunn til å være bekymret over de økningene som har skjedd og som vi vet allerede retter vi opp dette avviket når vi kalibrerer mot familiefordeling og aldersfordelingen for menn og kvinner.

Dersom vi ser nærmere på avviket i befolkningstallene mellom bruttoutvalget og nettoutvalget på fylker og etter alder ser vi at for fylkene er økningen på grunn av frafallet rimelig lik med unntak av ytterpunktene Nord-Trøndelag der det er ingen endring og for Troms der økningen er på nesten 3%. Ellers er økningen under 1 prosent. For aldersgruppene er det for de eldste vi ser forskjeller - spesielt for aldersgruppa 80 til 89 år der nettoutvalget gir over 15 000 flere enn vi finner fra bruttoutvalget. Dette er en indikasjon på at vi særlig for eldre har frafall blant aleneboende, men dette skal da kalibreringen rette på.

Det siste trinnet i beregningene av vektene retter som tidligere nevnt opp skjevheter i utvalget med hensyn på familiefordeling etter fem typer Besys familier og aldersfordelingen for menn og kvinner etter åtte aldersgrupper.

1.5. Innholdet i notatet

I kapittel 2 gir vi som tidligere nevnt en detaljert gjennomgang av utvalget som danner datagrunnlaget for de foreløpige tallene. Hovedsaken er at utvalget er trukket som et utvalg av DSF familier, mens skjemaene ble sendt til Besys familier. Det kan medføre at DSF familier i det utvalget vi trakk, AKU i 4. kvartal med et tillegg, ikke har fått tilsendt boligskjema, men den andre DSF familien i Besys familien (samboerpar med felles barn) har mottatt skjemaet.

Siden vi i kapittel 3.1 viser at husholdningene har en treksannsynlighet som er proporsjonal med antall DSF familier i husholdningen, er det viktig å ta hensyn til dette i beregningene av vektene eller oppblåsningsfaktorene. Utvalget vil altså inneholde for mange "store" husholdninger i betydningen at antall familienumre er minst to. Det vil f.eks. være relativt sett dobbelt så mange samboerpar som ektepar i utvalget og for få aleneboende i betydningen enperson husholdninger. Vi har både gjennom direkte å argumentere med hensyn på design av utvalget og på grunnlag av bruk av statistiske modeller beregnet vektene som bygger på en inndeling av utvalget etter antall familienumre i husholdningen. Dessuten argumenterer vi for at det er viktig å skille mellom unike adresser (trenger ikke bolignummer) og ikke unike adresser (trenger bolignummer). I tillegg må dette skillet kombineres med informasjon om det finnes en bolig som er lik den boligen vi ser på - en form for stedfortreder som kan utnyttes ved imputering for enhetsfrafall når vi ikke har mottatt noe utfylt boligskjema. Vi har beskrevet dette nærmere i kapittel 2.4 og kapittel 5. Vi bruker betegnelsen adressegruppe om denne inndelingen siden inndelingen er knyttet til familienumre og slik sett også til personene.

Vi har i kapittel 3 vist hvordan vi både indirekte (design argumentet) og direkte (modell argumentet) estimerer antallet familienumre og husholdninger som finnes av hver kombinasjon av adressegruppe og antall familienumre i samme husholdning. Til slutt i kapittel 3 viser vi hvordan vi kan kalibrere disse vektene enperson familie, aleneforelder (med barn), ektepar uten barn, ektepar med barn og samboere med felles barn. Dette kapitlet er til dels svært teknisk og detaljert i argumentasjonen. Særlig de tre første avsnittene er skrevet ut i stor detalj for å vise den formelle argumentasjonen med hensyn på utvalgsplanen og bruken av statistiske modeller for å beregne vektene.

I kapittel 4 ser vi på beregninger av usikkerheten til tallene som skal publiseres. Vi gir både enkle regler for hvilke tall som kan publiseres og mer komplekse formler for hvordan usikkerheten kan beregnes mer eksakt.

I kapittel 5 har vi som tidligere nevnt behandlet imputering av boligopplysninger og husholdnings-sammensetning for skjemaene i frafallet. Vår konklusjon er at vi ikke finner det mulig å imputere direkte for enhetsfrafall, men bruker etterstratifisering etter fylker og adressegrupper og kalibrering som en indirekte metode til å imputere for enhetsfrafall.

Til slutt er det gitt en sammenlikning mellom kalibrering av vektene på husholdningsnivå og personnivå i vedlegg A og noen spesielle utregninger av egenskapene til vektene brukt på familier og personer i vedlegg B. I vedlegg C er det gitt en detaljert tabell over fordelingen av DSF familiene på typer boligbygg, mens i vedlegg D sammenliknes AKU+ utvalget mot befolkningstallene per 3. november. Til slutt i vedlegg E har vi gitt en fylkesvis fordeling av husholdningene etter antall DSF familier i husholdningene.

Prinsippene for kalibrering er tidligere beskrevet av Heldal (1992) og Zhang (2000). Disse har også vært brukt i over ti år på de årlige inntekts- og formuesundersøkelsene og er dessuten tatt i bruk på Arbeidskraftundersøkelsene, se Heldal (2000).

2. Utvalget til foreløpige tall i Fob2001

I dette kapitlet går vi gjennom utvalget som ligger til grunn for foreløpige tall, publisert 22. april 2002, om husholdningene og boligene i Norge 3. november 2001. I avsnitt 2.1 gir vi først en presentasjon av hvordan bolig tellingen ble gjennomført, mens vi i avsnitt 2.2. forklarer egenskaper ved utvalget som vi bruker, nemlig AKU for 4. kvartal 2001 med et tillegg blant de eldre. I avsnitt 2.3. beskriver vi hvilke skjevheter det er i utvalget med hensyn på husholdninger og boliger som den statistiske enheten. Til slutt i avsnitt 2.4 kommer vi inn på hvordan vi kan skille mellom husholdningene og boligene etter om adressen inneholder flere boliger og om det finnes en svært lik bolig på samme adresse eller en nær adresse.

2.1. Boligtellingen

Boligtellingen er blitt gjennomført ved at det ble definert en populasjon av familier der en familie enten var en DSF familie, der DSF familie betyr personer med samme familienummer, eller en samboerfamilie med felles barn. Altså besto hver familie i denne populasjonen enten av de personene som hadde ett felles familienummer ifølge DSF (Det sentrale folkeregister) eller av to slike DSF familier som var koblet sammen ved at en kvinne med barn og en mann med eller uten barn i de to DSF familiene var samboere med minst et felles barn. Sett ut fra utsending av skjemaene kan vi derfor beskrive Boligtellingen som en familieundersøkelse. En referanseperson i hver (skjema)familie er valgt ut, den eldste personen i denne familien, og skjemaet ble sendt til denne. Vi vil bruke betegnelsen Besys familie videre i teksten på den familien som mottok ett boligskjema og DSF familie på familien slik den er definert i DSF ved familienummeret.

I tillegg ble det også sendt ut boligskjema til alle studenter som hadde søkt om borteboerstipend våren og/eller høsten 2001. Det betyr at til alle familier med studenter som har søkt om borteboerstipend for våren og/eller høsten 2001 ble det sendt boligskjemaer til disse studentene for å hente inn opplysningene om de boligene de faktisk bodde i på tellingstidspunktet 3. november 2001.

Boligtellingen har tre formål, nemlig undersøke boligene i Norge i 2001, undersøke husholdningene i Norge i 2001 og tildele alle personer i Norge en entydig adresse. Det siste er målet ved boligadresseprosjektet. Siden vi har sendt ut boligskjemaer til alle familier vil et betydelig antall husholdninger motta to eller flere skjemaer. Dessuten vil vi også skille mellom to typer husholdninger etter om vi bruker den formelle adressen eller den faktiske adressen. Dette gjelder særlig studentene som stort sett vil høre sammen med foreldrene dersom vi ser på den formelle adressen, men bo alene eller sammen med andre(studenter) når vi bruker deres faktiske adresse. For alle andre enn studenter og institusjonsbeboere vil også den formelle adressen, der en er registrert bosatt ifølge Det sentrale folkeregisteret, være den faktiske adressen.

Det er siste side i boligskjemaet som skal knytte familier og personer sammen til husholdninger. På denne siden er de andre familiemedlemmene til referansepersonen ført opp og en skulle da krysse ut de som ikke bodde i husholdningen og legge til de andre som i tillegg bodde i boligen eller om en vil, hørte til husholdningen. To typiske endringer ville derfor være å stryke de av barna som var studenter fordi de bodde i en egen bolig og føye til den andre DSF familien som utgjorde et samboerpar uten felles barn. I det siste tilfellet ville da begge DSF familier ha mottatt skjema.

Familienummeret er identisk med mannens fødselsnummer hos et ektepar. I andre DSF familier familienummeret lik fødselsnummeret til personen selv i en enperson familie eller til forelderens dersom en har barn. Det betyr at det finnes fem slike DSF familier:

- Enperson familie
- Ektepar uten barn
- Mor med barn
- Far med barn
- Ektepar med barn

Nå er det i tillegg selvsagt regler for når barn opphører å være barn i betydningen har samme familienummer som foreldrene (eventuelt mor eller far). Hovedprinsippet er at tre veier fører barnet ut av DSF familien til foreldrene, nemlig

- Ekteskap
- Barn
- Arbeid kombinert med flyttemelding

Det betyr at ugifte studenter uten barn som hovedprinsipp beholder familienummeret og er "barn", mens studenter som gifter seg eller får barn også får eget familienummer og føres ut av foreldrenes familie.

Det faktiske regelverket for tildeling av familienummer er noe mer detaljert enn prinsippene vi har beskrevet ovenfor, men dette er nok for å gi en beskrivelse av hvilke konsekvenser regelverket for familienummer får for arbeidet med foreløpige tall. Det kan f.eks. nevnes at fosterbarn har eget familienummer (lik fødselsnummeret).

Denne klassifiseringen av personer i familier betyr at et betydelig antall av faktiske husholdninger, de personene som har sitt ordinære opphold i samme bolig, består av flere formelle familier eller om en vil alle personene i husholdningen har ikke samme familienummer. Et eksempel er fosterbarn, men av langt større betydning er samboere. Dette gjelder både samboere med felles barn og uten felles barn. Alle slike faktiske familier eller husholdninger består da av to DSF familier. Altså sett ut fra familienummeret vil

- samboere uten barn består av en mann og en kvinne som hver er enperson familier
- samboere med felles barn kan være en alenemor og en mann som er en enperson familie(kun felles barn), en alenemor og en alenefar(felles barn og far har omsorgen for egne barn i tillegg) eller at både mor og far til felles barn har omsorgen for egne barn i tillegg
- samboere uten felles barn , men med omsorg for egne barn, dvs. består av enten en alenemor og en mann som er enperson familie, en kvinne som er enperson familie og en alenefar eller både en alenemor og en alenefar.

Ifølge Noack m. fl. (2001) var det drøyt 91 000 samboerpar med felles barn per 1. januar 2000. Det totale antallet husholdninger av samboere er i foreløpige tall beregnet til 212 000. Det andre ballpunktet - DSF familier med felles barn har vi da som beskrevet ovenfor forsøkt å fjerne dobbelttellingen av i utsendelsen av boligskjemaene, men det må nevnes at slike merkede samboerfamilier blir fratatt koblingen når alle barna har meldt utflytting. For samboere med felles barn har vi altså ikke en registrert overgang til samboere uten barn som for ektepar. Det betyr at samboere med felles barn ikke lenger er merket i Besys når barna er fløyet ut av redet. Selv om mange samboere med felles barn unngår å få to boligskjemaer vil en del slike samboerpar der barna er flyttet ut ha mottatt to skjemaer.

Mellom 10 og 20 prosent av boligene som var bebodd fikk minst to boligskjemaer. Utfyllingen, spørsmål 30 i skjemaet, skulle knytte disse sammen slik at det enten blir en faktisk familie og husholdning (enfamilie husholdning) eller flere faktiske familier i en husholdning(flerfamilie husholdning). Denne sammenknytningen skal i framtiden løses gjennom bolignummeret - alle boliger får et unikt bolignummer som identifiserer boligen. Se også Zhang(2001) og Standard for gruppering av familier og husholdningen, SSB(2001), for detaljer om familier og husholdninger.

2.2. Utvalget er Arbeidskraftundersøkelsen i 4. kvartal 2001(AKU-2001-4) med et tillegg

Arbeidskraftundersøkelsen er en kvartalsvis undersøkelse av arbeidsmarkedet der et utvalg på 24 000 personer mellom 16 og 74 år intervjues om sitt forhold til arbeidsmarkedet, om de er sysselsatt, ledige, skoleelever osv. Alle som trekkes ut til undersøkelsen deltar 8 kvartaler på rad slik at 3 000 nye

personer kommer til i hvert kvartal. Dette utvalget av personer er trukket som et utvalg av DSF familier - altså familier definert ved felles familienummer. Siden det kun er personer mellom 16 og 74 år som skal intervjues er det kun personene mellom 16 og 74 år i de uttrukne DSF familiene som er med i utvalget.

AKU- utvalget er innen hvert fylke trukket som et selvveiende utvalg av DSF familier blant alle DSF familier med en referanseperson mellom 16 og 74 år, altså den personen som definerer familienummeret. I et ektepar er referansepersonen mannen, ellers den voksne i DSF familien. Utvalgsprosenten varierer litt fra fylke til fylke. Noen små fylker er overrepresentert og noen store fylker er underrepresentert. I forhold til å bruke utvalget, mer presist boligskjemaene fra utvalget, som grunnlag for foreløpig bolig- og husholdningsstatistikk er ikke det noe alvorlig problem, men vi må ta hensyn til dette når vi skal beregne vektene eller oppblåsingsfaktorene.

Det er langt mer alvorlig at deler av boligene overhodet ikke er representert i AKU-2001-4. Alle DSF familier med en referanseperson utenfor aldersgruppa 16 til 74 år er ikke representert i AKU. Det betyr først og fremst at husholdninger og boliger som består av eldre personer - en eller alle over 74 år mangler i utvalget. Dette er rettet opp ved å trekke et tilleggsutvalg av DSF familier eller om en vil referansepersoner utenfor den populasjonen som AKU-2001-4 er trukket på grunnlag av.

Tillegget er trukket etter samme opplegg som for AKU i den populasjonen som er utenfor trekkgrunnlaget til AKU. Det samlede utvalget har da følgende egenskaper som vi må ta hensyn til når vi skal utarbeide estimeringsopplegget vårt eller om en vil beregne vektene til å blåse opp tallene med:

- Utvalgsprosenten varierer med fylke
- Utvalgsprosenten varierer med antall DSF familier i husholdningen som bor i boligene, noe vil komme tilbake til seinere

Vi vil bruke betegnelsen AKU+ for det utvalget som danner grunnlaget for de foreløpige tallene. Vi vil se på den situasjonen der vi skal beregne antallet formelle husholdninger og disse boligenes egenskaper.

I de to siste månedene i 4. kvartal er det også gjennomført en tilleggundersøkelse i AKU for å undersøke husholdningssammensetningen. Dette ble gjennomført for å undersøke kvaliteten på husholdningssammensetningen slik den innhentes på boligskjemaet. Det betyr at vi intervjuet AKU utvalget for november og desember for å finne alle personene som bor sammen med referansepersonen i den uttrukne DSF familien i boligen som han eller hun bor i. I tillegg skulle også personer i disse husholdningene som ifølge registeret bor sammen med referansepersonen fjernes dersom de ikke lenger bor der. Hensikten med intervjuet var å etablere bo- og kosthusholdningen til referansepersonen eller om en vil den uttrukne DSF familien. For flere detaljer om kost- og bohusholdninger og spørsmålssekvensen som er brukt i kontrollundersøkelsen vises det til Fosen et al(2001). Dersom studenter i den uttrukne DSF familien som en hentet inn opplysninger om, bodde utenfor boligen ble disse også intervjuet om den husholdningen de faktisk bodde i på studiestedet. Denne informasjonen har vi imidlertid ikke benyttet oss av i produksjonen av foreløpige tall. Dette er data sammen med andre kilder som primært skal brukes til estimere omfanget og størrelsen på målefeil i boligtellingsen.

2.3. Egenskaper til utvalget av formelle boliger og husholdninger

Vi vil nå kort se litt på egenskapene til utvalget vi skal bruke til å lage foreløpige tall av. Innen hvert fylke er utvalget et selvveiende utvalg av DSF familier. Det betyr at det utvalget av boliger og husholdninger som vi egentlig skal bruke som grunnlag for bolig- og husholdningsstatistikken vil ha trekk sannsynligheter som er proporsjonale med antallet DSF familier i boligen eller husholdningen. Det betyr f.eks. at samboerne vil ha dobbelt så stor trekk sannsynlighet siden begge DSF familier kan trekkes ut. Altså vil vi ha en overrepresentasjon av boliger og husholdninger som består av samboer-

par. På liknende vis vil også boliger der personene lever i et kollektiv være overrepresentert i utvalget. Se også Solheim og Zhang(2001) for flere detaljer om egenskapene til AKU- utvalget.

Det andre problemet vi må ta hensyn til når vi skal lage et estimeringsopplegg er frafallet. AKU+ utvalget ble fulgt opp svært godt i AKU og videre ble det stilt spørsmål som minnet de på FoB2001 slik at vi skulle forvente en høy svarprosent. I tilleggsutvalget hadde vi ikke samme mulighet til oppfølging som i AKU og det er også et spørsmål om hvor langt vi kunne gå i "plage" eldre for å få inn skjemaene utover den ordinære oppfølgingen i FOB2001. Det vil opplagt være en gevinst å ikke få for stort frafall.

For å summere opp er det grunn til å tro at

- Utvalget av skjemaer som vi skal lage foreløpige tall på grunnlag av, har en høy svarprosent for AKU-familiene og en noe lavere svarprosent for resten - først og fremst husholdninger over 75 år.
- Utvalget av boliger og husholdninger har trekkssannsynlighet proporsjonal med antall familier i boligen/husholdningen.

Svarprosenten i utvalget var som tidligere beskrevet drøyt 93 prosent, noe høyere enn den totale svarinngangen etter første påminning.

2.4. Adressegrupper - konsekvenser for frafallet

Prinsippet er at alle personer i befolkningen har en adresse og adressen er en bolig/bygningsadresse siden alle i prinsippet skal være (formelt sett) bosatt. En annen måte å se dette på er at alle skjemaer er sendt til en adresse og siden et skjema enten er til en DSF familie eller to (samboere med felles barn) vet vi dermed at det formelt sett finnes minst en eller to DSF familier på de adressene vi har sendt skjema til. Vi kan også dele inn adressene vi har sendt skjema til etter en helt annen dimensjon - nemlig etter om adressen er unik, det betyr at det finnes kun en bolig på denne adressen, eller at det finnes flere boliger og det er sendt ut bolignummer (minst to) til denne adressen. Med adresse mener vi de første 17 posisjonene i adressefeltet i DSF. Det betyr at vi bruker en adresse som består av

- Kommunenummer - fire posisjoner
- Gatekode/gårdsnummer - fem posisjoner
- Husnummer/bruksnummer - fire posisjoner
- Bokstav/festenummer - fire posisjoner

I GAB er alle bygninger klassifisert ifølge NS 3457 og gruppert i 8 hovedtyper:

- 1. Boligbygg, 2. Industri- og lagerbygg, 3. Kontor- og forretning, 4. Samferdsel/kommunikasjon, 5. Hotell og restaurant, 6. Kultur- og Forskning, 7. Helsebygning og 8. Fengsel og beredskapsbygning

I tabell 2.1 har vi presentert fordelingen av DSF familiene etter disse 8 hovedtypene når vi kobler DSF adressene mot GAB adressene. Vi har selvsagt at de fleste DSF familiene(1 995 379 av 2 249 314) er plassert i boligbygg der det kun finnes en bygning på adressen.. Av de 253 935 DSF familiene som ifølge koblingen med GAB ikke er bosatt i et boligbygg er det kun 32 112 som vi finner igjen i andre bygg, mens vi for de resterende 221 823 DSF familiene må konkludere med at enten er det flere bygninger på adressen, 87 182 DSF familier, eller vi får ikke kobling mellom DSF adresse og GAB adresse, dette gjelder de resterende 134 641 DSF familiene.

Tabell 2.1. Fordeling av DSF familiene etter boligtype og unik/ikke unik adresse

Boligtype	Totalt	Unik adresse	Ikke unik adresse
Totalt antall DSF familier	2 249 314	1 300 083	949 231
<i>Totalt - bygningstype 1-9</i>	<i>2 027 491</i>	<i>1 300 083</i>	<i>727 408</i>
1. Boligbygg	1 995 379	1 285 354	710 025
2. Industri og lagerbygning	1 289	738	551
3. Kontor- og forretning	12 323	3696	8 627
4. Samferdsel/kommunikasjon	206	95	111
5. Hotell og restaurant	1 226	526	700
6. Kultur og forskning	1 135	768	367
7. Helsebygning	5 314	1670	3 644
8. Fengsel, beredskapsbygning	39	12	27
9. Annet, gamle bygningskoder	10 580	7 224	3 356
Flere bygninger på adressen ¹	87 182	0	87 182
Ikke kobling mellom DSF og GAB ²	134 641	0	134 641

For boligbygningene har vi en ytterligere inndeling i 9 undergrupper:

- 11. Enebolig, 12. Tomannsbolig, 13. Rekkehus - kjedehus - andre småhus, 14. Boligblokk, 15. Bygning for bofellesskap, 16. Fritidsbolig, 17. Koie - seterhus og liknende, 18. Garasje og uthus til bolig og 19. Annen boligbygning

Det er de fem første typene som dekker de fleste av de aktuelle boligadressene i bolig tellingen, kun 23 237 av de 1 995 379 DSF familiene som vi har identifisert til å bo i boligbygg bor utenfor det vi kan karakterisere som normale helårsboliger. Flere detaljer er gitt i tabell 2.2. I tabellene 2.1 og 2.2 deler vi DSF familiene inn etter om adressen er unik, det finnes kun en bolig på adressen, og om adressen ikke er unik, det finnes flere boliger på adressen.

Tabell 2.2. Fordeling av DSF familiene etter boligtype og unik/ikke unik adresse for boligbygg

Boligtype	Totalt	Unik adresse	Ikke unik adresse
Totalt - boligbygg	1 995 379	1 285 354	710 025
11. Eneboliger	1 140 303	1 009 566	130 737
12. Tomannsboliger	183 360	85 416	97 944
13. Rekkehus - kjedehus - andre småhus	227 266	157 103	70 663
14. Boligblokk	407 252	20 406	386 846
15. Bygning for fellesskap	13 461	2 790	10 671
16. Fritidsbolig	3 228	3 178	50
17. Koie - seterhus - liknende	22	17	5
18. Garasje og uthus til bolig	341	262	79
19. Annen boligbygning	19 646	6 616	13 030

I vedlegg C er det laget en mer detaljert oversikt over fordelingen for boligbygg enn i tabell 2.2. Der ser vi da at boligtypene 111 Eneboliger (uten sokkelleilighet) og 113 Våningshus har omtrent alle unik

¹ Det finnes flere bygningstyper på denne adressen, f.eks. våningshus og kårbolig

² Vi har ikke oppnådd kobling mellom DSF familiens adresse i Det sentrale folkeregisteret og adressene i GAB

adresse, mens nesten alle 112 Eneboliger med hybel/sokkelleilighet er uten unik adresse. Dersom bygningstypen var feilfri skulle alle 111 Eneboliger (uten sokkelleilighet) hatt unik adresse og omvendt skulle ingen 112 Eneboliger med hybel/sokkelleilighet hatt unik adresse. Det er også en klar forskjell mellom vertikaldelte og horisontaldelte tomannsboliger. For de vertikaldelte har de fleste unik adresse, mens det er omvendt for de horisontaldelte. Andre klare forskjeller en ser er at de fleste rekkehusboligene har unik adresse, mens for boligene i terrassehusene og andre småhus er det en stor overvekt av ikke unike adresser. For blokker det relativt få unike adresser med unntak av blokker på kun to etasjer. Blokker på to etasjer kan antakelig i mange tilfeller like gjerne klassifiseres som rekkehus.

På grunnlag av disse to tabellene vil vi lage en inndeling i fire grupper der vi både skiller mellom unike adresser - det betyr at det finnes kun en bolig på denne adressen (17 første posisjoner i adressen i DSF) og om det finnes en bolig i nærheten som med stor sannsynlighet er lik den boligen vi ser på - dette kalles en stedfortreder.

1. Boliger i bygninger som har unike adresser og mangler stedfortredere - dette gjelder spesielt **111 Eneboliger (uten sokkelleilighet)**
2. Boliger i bygninger som har unike adresser og har stedfortredere - dette gjelder i særlig grad **rekkehus** og til en viss grad **tomannsboliger**
3. Boliger i bygninger som ikke har unik adresse og mangler stedfortredere - dette gjelder spesielt **eneboliger med sokkelleilighet**, en del **tomannsboliger** og **kjedehus - andre småhus**
4. Boliger i bygninger som ikke har unik adresse, men har stedfortreder - dette gjelder i særlig grad **boligblokker** og **bygninger for bofellesskap**

Vi kan i hvert fall trekke den konklusjonen at det er naturlig å dele

- befolkningens adresser og DSF familier i fire grupper etter unik adresse eller ikke og stedfortreder eller ikke og vi vil bruke betegnelsen adressegruppe om denne inndelingen. Vi vil også vite hvor mange DSF familier, Besys familier og personer fordelt etter alder og kjønn det finnes i hver adressegruppe.

Det gjenstår å definere stedfortreder og vi foreslår følgende definisjon:

- En bolig sies å ha en **stedfortreder** dersom boligens adresse er en bygning av en følgende tre bygningstyper: **12 Tomannsbolig**, **13 Rekkehus - kjedehus - andre småhus** eller **14. Boligblokk**.

Dersom vi klassifiserer alle boliger på en adresse med en annen bygningstype enn de tre ovenfor i ballpunktet til å **ikke ha en stedfortreder** har vi laget fullstendig inndeling av adressene til alle DSF familier i fire adressegrupper. Spesielt betyr dette at vi gir alle adresser i Det sentrale folkeregistret (DSF) karakteristikken ingen stedfortreder og ikke unik adresse dersom vi enten finner flere bygninger på adressen eller at adressen i DSF mangler kobling med GAB.

I tabell 2.3 har vi vist fordelingen til alle DSF familier etter denne inndelingen i fire adressegrupper. Vi ser da at gruppen stedfortreder med unik adresse kun omfatter drøyt 11 prosent av alle DSF familiene, mens gruppen unike adresser uten stedfortredere er klart størst med i overkant av 46 % av DSF familiene. Til tross for at disse fire gruppene er nokså ujevne i omfang skulle det ikke være noen fare med at de vil mangle representanter i noen av fylkene. Se også tabell 3.1 i avsnitt 3.7 for flere detaljer om fordelingen på disse fire adressegruppene fylkesvis.

Tabell 2.3. Fordelingen av DSF familiene etter disse adressegrupper

	Bygningstype	Totalt	Unik adresse	Ikke unik adresse
Totalt		2 249 314	1 300 083	949 231
Stedfortreder	12,13 og 14	817 378	262 925	554 453
Ikke stedfortreder	Resten	1 431 936	1 037 158	394 778

Denne inndelingen kan vi også nytte til å utforme ett opplegg for å imputere direkte boligopplysninger for de skjemaene som vi mangler i gruppe 2 og 4, mens vi for gruppe 1 og 3 ikke ser noen annen løsning enn å imputere indirekte gjennom opplegget for beregning av vekter. I avsnitt 3 vil vi gå gjennom opplegget for beregning av vekter, mens vi i avsnitt 5 vil se nærmere på muligheten og behovet for å imputere for frafallet.

For frafallet vil vi heller ikke kjenne til den formelle husholdningen eller den faktiske husholdningen for alle grupper, men her går forskjellen mellom unike og ikke unike adresser. For unike adresser vil vi kunne stille opp den riktige husholdningen med hensyn til formell adresse siden dette faktisk finnes i DSF, mens for **ikke unike** adresser må vi nøye oss med å bruke kun den preprintede listen dersom vi ser bort fra tilfeller der alle andre skjemaer fra adresser er levert unntatt for det skjemaet som hører til utvalget.

Vi kan fastslå at for skjemaer i frafallet er det slik at dersom vi kan imputere boligopplysninger med stor nøyaktighet er kontrollmuligheten av husholdningen liten, mens dersom vi har gode muligheter til å finne fram den korrekte husholdningen er det nærmet umulig å imputere boligopplysninger som med rimelig sikkerhet er riktige.

3. Estimere antall boliger og husholdninger - foreløpige tall

I dette kapitlet skal vi beskrive hvordan antallet husholdninger med bestemte egenskaper fra utvalget som er beskrevet i foregående kapittel er blitt estimert i foreløpige tall. Dette inkluderer også det totale antallet boliger og husholdninger i fylkene og for hele landet i alt. Hovedlinjene i opplegget er først å ta hensyn til designen i utvalgsplanen og innarbeide en stratifisering eller oppdeling av hele populasjonen som tar hensyn til dette. Deretter skal vi se på hvordan annen informasjon fra populasjonen kan brukes til å rette opp avvik i utvalget i forhold til hele populasjonen. Avvikene som rettes opp er bestemt av denne informasjonen vi bruker. Nyten av disse korreksjonene er altså avhengig av sammenhengen mellom statistikkvariablene og denne informasjonen vi bruker.

Vi skiller mellom private og andre husholdninger. Foreløpige tall bygger på en gjennomgang av alle husholdninger i utvalget for å klassifisere disse som private og andre husholdninger. Til støtte for arbeidet med å beskrive de andre husholdningene har SSB et eget opplegg der vi har hentet inn data direkte fra institusjonene. Vi mangler foreløpig et fullstendig bilde av alle husholdningene i Norge siden side fire i boligskjemaet ikke er ferdig bearbeidet for alle skjemaer. Vi har derfor i denne omgangen valgt å bruke kun tilgjengelig informasjon for alle husholdninger vi har etablert omkring de uttrukne DSF familiene i utvalget som vi har beskrevet i kapittel 2. Dette betyr at vi baserer oss på et bilde av den totale populasjonen gjennom DSF familiene og personene som vi regner som bosatte i Norge per 3. november 2001. I tillegg vil vi også bruke informasjon om Besys familiene, der vi i tillegg til familenummer også bruker informasjon om samboermerket (samboere med felles barn). Vi vil altså estimere antallet private og andre husholdninger, antallet DSF familier, Besys familier og personer som bor i henholdsvis private og andre husholdninger.

Vi går ut fra at målefeil i foreløpige tall er som for endelige tall i FoB2001, men vil i de foreløpige tallene kun regne på størrelsen av utvalgsfeilene og vurdere om tall skal publiseres eller ikke ut fra beregningene av utvalgsfeilene. Dette kapittelet danner grunnlaget for hvordan utvalget skal blåses opp til totaltall for hele populasjonen - altså hvordan beregnes vektene til husholdningene og personene i utvalget. Vi skal i de endelige tallene i FoB2001 beregne omfanget av målefeil, se Solheim, Hurlen Foss, Fosen og Zhang(2002).

I avsnitt 3.1 viser vi hvordan trekksannsynligheten til husholdningene øker proporsjonalt med antallet DSF familier i husholdningene, mens vi i avsnitt 3.2 viser hvordan disse trekksannsynlighetene brukes til å lage designbaserte vektorer. I avsnittene 3.3 og 3.4 viser vi hvordan vi kan komme fram til de samme vektene gjennom å arbeide direkte med statistiske modeller for skjemavariablene gjennom DSF familiene og personene i husholdningene. Disse to avsnittene er forholdsvis tekniske og kanskje mer appetittvekkende for metodestatistikere enn statistikere. I avsnitt 3.5 beskriver vi hvordan vi kalibrerer designvektene slik at utvalget er konsistent med antallet Besys familier fordelt på fem grupper og personene fordelt etter åtte aldersgrupper for menn og kvinner hver for seg. I avsnitt 3.6 beskriver vi hvilke kalibreringsvariabler vi skal bruke, mens vi i avsnitt 3.7 går gjennom hvordan tabeller over studentenes faktiske bosted kan produseres. Til slutt i avsnitt 3.8 beskriver vi prosessen fram til de endelige vektene, beskrevet i avsnitt 3.5, som ble brukt for å produsere de foreløpige tallene i FoB2001, publisert 22. april 2002.

3.1. Trekksannsynligheter for boligene og husholdningene

Vi må nå innføre litt notasjon for å beskrive opplegget mer nøyaktig. Først tre hovedstørrelser vi skal ta hensyn til i beregningene. De familiene som boligskjemaene ble sendt ut til kan vi dele inn i grupper etter to kriterier, nemlig hvilket fylke en bor i og hva slags adresse en har.

- r fylke, dvs. 19 verdier
- s Inndeling i 4 adressegrupper etter om en DSF familie har unik adresse eller ikke og om det finnes en bolig som kan brukes som stedfortreder eller ikke på denne adressen, se inndelingen i 4 adressegrupper vi gjorde i avsnitt 2.4

Denne inndelingen gjelder for hele populasjonen.

I tillegg kan vi for utvalget legge til en inndeling av husholdningene eller om en vil boligene:

- f inndeling etter antall DSF familier i husholdningen, dvs. $f = 1, 2, 3, \dots$

I private husholdninger er f normalt lik 1 eller 2, mens i andre husholdninger kan f være langt større.

Vi skal seinere se på forholdet til andre registervariabler definert ved egenskaper til DSF familiene, Besys familiene og personene. Dette tar vi opp i større detalj i avsnitt 3.6.

Vi må innføre notasjon for noen størrelser vi trenger i det videre arbeidet:

- M_r - totalt antall DSF familier i fylket r
- m_r^{AKU+} - antall DSF familier som ble trukket ut til utvalget i AKU+ i fylket r
- π_r^{AKU+} - sannsynligheten for å trekke ut et familienummer i fylke r
- $M_{r,s}$ - totalt antall DSF familier i fylket r for adressegruppe s
- $m_{r,s}^{AKU+}$ - antall DSF familier i utvalget i fylket r for adressegruppe s - antallet familier vi trakk
- $\pi_{r,s}^{AKU+}$ - sannsynligheten for å trekke ut et familienummer for adressegruppe s i fylke r

Vi kan da først stille opp en sammenheng mellom de fire typene DSF familier etter adressegrupper og alle DSF familiene i fylket r:

$$(3.1) \quad M_r = M_{r,1} + M_{r,2} + M_{r,3} + M_{r,4} = \sum_{s=1}^4 M_{r,s}$$

Vi har videre per definisjon følgende sammenheng mellom de tre første størrelsene ovenfor fordi sannsynligheten for å trekke ut en familie er lik forholdet mellom antallet i utvalget(gunstige) og antallet i populasjonen(mulige) innen hvert fylke

$$(3.2) \quad \pi_{r,s}^{AKU+} = \pi_r^{AKU+} = \frac{m_r^{AKU+}}{M_r}$$

Dersom en husholdning består av flere DSF familier øker sannsynligheten for at en slik husholdning er kommet med i AKU+ utvalget og den øker proporsjonalt med antallet DSF familier. Siden AKU+ utvalget består av knappe 17 000 DSF familier og det finnes om lag to og en kvart million DSF familier i alt er sannsynligheten for å trekke ut ett familienummer lik i underkant av 1/132 , altså i underkant av 0,76%. Det er lite sannsynlig å trekke ut begge familienumrene i en samboerhusholdning eller en annen type husholdning som består av to DSF familier, faktisk er den under 0,005%, som betyr at vi i AKU+ utvalget ikke burde finne mer enn høyst ett slikt tilfelle. Altså kan vi regne som om antallet DSF familier trukket ut til AKU+ vil være lik antallet husholdninger vi sitter tilbake med i bruttoutvalget.

Nå er også ett av målene ved bolig tellingen å plassere alle personer i rett bolig eller om en heller vil sammen med de rette personene til den korrekte husholdningen. Det vil imidlertid være to typer slike husholdninger, nemlig formelle og faktiske. Den formelle husholdningen tar utgangspunkt i hvor personer er bosatt ifølge Det sentrale folkeregister, mens den faktiske skal ta utgangspunkt i hvor en faktisk bor. Et typisk eksempel på forskjell er at studenter ofte er formelt bosatt hos foreldrene på hjemstedet, mens de faktisk bor på studiestedet. Typisk kan det være at to studenter er formelt bosatt i en enfamilie husholdning på fem, mens de faktisk bor i henholdsvis en trefamilie husholdning som består av tre personer og en femfamilie husholdning som består av fem personer: Alle de faktiske familiene er altså lik enpersonfamilier.

Dersom vi nå ser på antall DSF familier som formelt bor i en bolig kan vi innføre følgende notasjon:

- $M_{r,f}^{hush}$ - totalt antall husholdninger eller boliger i fylket r som består av f DSF familier
- $m_{r,f}^{hush}$ - antall husholdninger eller boliger i utvalget i fylket r som består av f DSF familier
- $\pi_{r,f}^{hush}$ - sannsynligheten for å trekke en husholdning som består av f DSF familier

Da har vi på liknende vis som for (3.1) en sammenheng mellom alle DSF familier i fylket r og husholdningene i fylket gitt ved

$$(3.3) \quad M_r = \sum_f f \cdot M_{r,f}^{hush}$$

Antallet husholdninger som består av f antall DSF familier er ukjent i populasjonen og det første spørsmålet vi skal svare på i neste avsnitt er hvordan dette antallet skal estimeres fra utvalget.

Det er også slik at sannsynligheten for å trekke husholdningene er som tidligere nevnt proporsjonal med antallet DSF familier i husholdningen, altså har vi

$$(3.4) \quad \pi_{r,f}^{hush} = f \cdot \pi_r^{AKU+}$$

La oss nå til slutt koble sammen DSF familiene og adressegruppene i en husholdning. Da må vi først innføre en notasjon som kobler disse tre enhetene sammen, nemlig

- $M_{r,s,f}^{hush}$ - totalt antall husholdninger i fylket r av adressegruppe s som består av f antall DSF familier
- $m_{r,s,f}^{hush}$ - antall husholdninger i utvalget i fylket r av adressegruppe s som består av f antall DSF familier
- $\pi_{r,s,f}^{hush}$ - sannsynligheten for å trekke en husholdning i fylket r av adressegruppe s som består av f antall DSF familier

Vi har da følgende sammenheng mellom antallet DSF familier i alt i et fylke r av adressegruppe s og antallet husholdninger etter antall familienumre i disse:

$$(3.5) \quad M_{r,s} = \sum_f f \cdot M_{r,s,f}^{hush}$$

Vi kjenner venstre side fra registeret, mens størrelsene på høyre siden i summen er ukjente og vil bli estimert fra beregningene. For utvalget har vi videre, dersom vi ser bort fra at det ble trukket ut flere DSF familier i samme husholdning, følgende sammenheng:

$$(3.6) \quad m_{r,s}^{AKU+} = \sum_f m_{r,s,f}^{hush}$$

La oss til slutt skrive opp sannsynlighetene for å trekke de forskjellige typer husholdninger i utvalget

$$(3.7) \quad \pi_{r,s,f}^{hush} = f \cdot \pi_{r,s}^{AKU+} = f \cdot \pi_r^{AKU+}$$

siden antallet DSF familier i husholdningen som er proporsjonal med sannsynligheten for å trekke husholdningen. Dette er imidlertid en apriori sannsynlighet uten at vi tar hensyn til informasjon om at vi faktisk vet antallet DSF familier av de fire typene vi operer med. Dersom vi utnytter denne informasjonen vil vi i stedet bruke følgende etterstratifiserte trekksannsynligheter (mer rett er det egentlig å kalle disse for betingede trekksannsynligheter gitt adressegruppe) for å ha trukket husholdningene

$$(3.8) \quad \hat{\pi}_{r,s,f}^{hush} = f \cdot \frac{m_{r,s}^{AKU+}}{M_{r,s}}$$

siden vi som forklart ovenfor kjenner alle størrelsene på høyre side i (3.8).

Vi vil også til slutt understreke at de etterstratifiserte trekksannsynlighetene vi har beregnet i (3.8) gjelder formelle husholdninger eller boliger. Nå viser det seg imidlertid at en faktisk husholdning kan sees på som en egenskap ved de formelle husholdningene slik at vi kan bruke (3.8) som grunnlaget for vektene vi beregner etter hvert utover i kapitlet også som vektene til de faktiske husholdningene. I avsnitt 3.7 har vi spesielt brukt dette som argument for å kunne produsere tall om studentenes faktiske

boliger og husholdninger dersom det var ønskelig. Vi valgte imidlertid å sløyfe denne muligheten i publiseringen av de foreløpige tallene.

3.2. Beregne vekter på grunnlag av utvalgsdesignen

Vi skal i dette avsnittet se på en enkel metode for å beregne vekter eller oppblåsingsfaktorer for utvalget basert på trekkssannsynlighetene eller om en vil designen på utvalgsplanen. I seinere avsnitt skal vi gå videre å utnytte egenskaper i populasjonen som vi ønsker at utvalget også skal ha.

På grunnlag av (3.8) kan vi derfor definere vekter for alle husholdninger i utvalget i et fylke r av type s som består av f DSF familier ved bruk av den inverse etterstratifiserte trekkssannsynligheten

$$(3.9) \quad w_{h,r,s,f}^{hush} = \frac{1}{\hat{\pi}_{r,s,f}^{hush}} = \frac{M_{r,s}}{f \cdot m_{r,s}^{hush}}$$

Vi bruker altså for hver husholdning i utvalget informasjon om antallet DSF familier i husholdningen og antallet DSF familier det finnes i populasjonen av hver adressegruppe.

La oss først undersøke egenskapene til disse vektene. Vi finner da ved å summere for en fast f og s over husholdningene i en slik undergruppe et estimat for antallet husholdninger i fylket r av adressegruppe s med f antall familienumre på formen

$$(3.10) \quad \hat{M}_{r,s,f}^{hush} = \sum_{h \in utv_{r,s,f}^{hush}} w_{h,r,s,f}^{hush} \cdot 1 = m_{r,s,f}^{husf} \frac{M_{r,s}}{f \cdot m_{r,s}^{hush}} = \frac{1}{f} M_{r,s} \frac{m_{r,s,f}^{hush}}{m_{r,s}^{hush}}$$

når vi betegner utvalget med $utv_{r,s,f}^{hush}$.

Formel (3.10) kunne faktisk vært tatt som utgangspunkt for å beregne antallet husholdninger som består av f DSF familier. Siden utvalget av DSF familier i AKU+ utvalget er selvveiende er forholdet

$$\bullet \quad \frac{m_{r,s,f}^{hush}}{m_{r,s}^{hush}}$$

et anslag på andelen DSF familier som hører til husholdninger som består av f familier. Dersom denne andelen multipliseres med antallet DSF familier i fylket r for adressegruppe s , $M_{r,s}$, finner vi estimatet for antallet DSF familier som hører til husholdninger som består av f slike er lik

$$\bullet \quad M_{r,s} \frac{m_{r,s,f}^{hush}}{m_{r,s}^{hush}}$$

Vi finner da antallet husholdninger som består av f antall DSF familier ved å dividere dette tallet med f slik siste uttrykket i (3.10) viser. På dette viset kunne en fra det siste uttrykket i (3.10) argumentert bakover til vekten i (3.9) ved å vise til at det finnes $m_{r,s,f}^{hush}$ husholdninger med f DSF familier vi skal summere over i utvalget. Vi kunne derfor ha forenklet deler av det formelle apparatet vi har utviklet og argumentert innenfor i det foregående avsnitt, dette avsnittet og neste avsnitt, men har valgt å skrive ut argumentene svært nøyaktig for å skille klart mellom bruken av utvalgsplanen og bruken av statistiske modeller.

Dersom vi multipliserer resultatet med f , dvs. teller antall DSF familier i hver husholdning finner vi videre siden

$$(3.11) \quad m_{r,s}^{hush} = \sum_f m_{r,s,f}^{hush}$$

at

$$(3.12) \quad \sum_f \sum_{h \in utv_{r,s,f}^{hush}} w_{h,r,s,f}^{hush} \cdot f = \sum_f m_{r,s,f}^{hush} \frac{M_{rs}}{f \cdot m_{r,s}^{hush}} \cdot f = \frac{M_{rs}}{m_{r,s}^{hush}} \sum_f m_{r,s,f}^{hush} = \frac{M_{r,s}}{m_{r,s}^{hush}} m_{r,s}^{hush} = M_{r,s}$$

Vi har vist at disse vektene er konsistent med antallet DSF familier av type s i hvert fylke r . Vi har da brukt at det kun er trukket ut en DSF familie per husholdning vi har etablert på grunnlag av utvalget, men det kan være flere boligskjemaer som er sendt ut til husholdningene i utvalget. Vi har ovenfor argumentert for at det er må være svært få husholdninger i utvalget der flere DSF familier er trukket ut til AKU+, slik at denne antakelsen er godt begrunnet.

Vi har til nå kun brukt DSF familiene. En videre behandling av vektene må skje gjennom å kontrollere mot viktige registervariabler knyttet til egenskapene til personene i utvalget og boligene i utvalget der det finnes totaler i registrene. Mens beregningene som fører fram til vektene definert i (3.9) bygger på en stratifisering av utvalget og populasjonen skal gå videre og se på en kalibrering av vektene betinget de egenskapene vi har vist i (3.12), se avsnitt 3.5.

Vi avslutter dette avsnittet med å formulere beregningene av vektene i (3.9) gjennom følgende tre punkter:

- Beregn for hvert fylke antallet DSF familier i hver adressegruppe
- Beregn i utvalget for hvert fylke antallet husholdninger i hver adressegruppe.
- Hver husholdning/bolig i utvalget gis en vekt som er forholdet mellom antallet DSF familier beregnet i første ballpunkt dividert med produktet av antallet DSF familier i husholdningen og antallet husholdninger i det andre punktet

Målet med vektene vi beregnet i (3.9) er å beregne foreløpige tall for egenskaper ved boliger og husholdninger ved å bruke informasjonen i data. La oss betegne data i utvalget hentet inn på skjemaene for y , dvs. vi har

$$(3.13) \quad y_{h,r,s,f}$$

er enten en egenskap (har barn eller har kjøkken på over 6 kvm) eller en verdi (antall barn eller antall soverom) ved husholdningen eller boligen. Vi kan da enten estimere antallet husholdninger eller boliger med den bestemte egenskapen eller den totale summen av verdiene i fylket r med adressegruppe s ved å summere produktet av vekten og y -verdien over alle husholdninger i utvalget

$$(3.14) \quad \hat{t}_{r,s} = \sum_h w_{h,r,s,f}^{hush} y_{h,r,s,f} = \sum_f \sum_{h \in utv_f^{hush}} w_{h,r,s,f}^{hush} y_{h,r,s,f} = \sum_f \hat{t}_{r,s,f}$$

og denne formelen definerer da en metode for å beregne alle foreløpige tall. Metoden utnytter lite av den informasjon vi har om personene som utgjør alle husholdningene i utvalget, kun antallet familier disse husholdningene består av. Legg derfor også merke til i (3.14) at vi egentlig beregner totaler for hver kombinasjon av fylke, adressegruppe og antall familier i husholdningene. Det er god grunn til å

tro at vi kan forbedre estimeringsopplegget ved å utnytte mer informasjon om DSF familiene, personene og boligene i utvalget og totalene for egenskapene til disse i populasjonen. Dette skal vi se nærmere på i de neste avsnittene.

3.3. Bruk av statistiske modeller for å beregne vekter - data er alle familienumrene³

I forrige avsnitt studerte vi hvordan vektene til boligene/husholdningene beregnes ved å bruke kun designen på utvalget AKU+ og sammensetningen av husholdningene i en bolig, dvs. antallet DSF familier. Vi skal nå forsøke å utlede vekten definert i (3.10) ved bruk av statistiske modeller for en egenskap eller verdi til husholdningen som er målt ved bruk av boligskjemaet.

Dersom vi hadde gjennomført en fullstendig bearbeiding av side fire på alle boligskjemaer ville vi hatt en fullstendig oversikt for hele befolkningen over

- alle husholdninger med hensyn til DSF familier og personer i disse

Denne kunne vi brukt til lage et estimeringsopplegg som tok hensyn til forskjellige krav til konsistens mellom utvalget og populasjonen som var naturlig eller påkrevd å forlange. Et slikt opplegg ville selvsagt også gi at vi ikke estimerer noen egenskaper ved husholdningene som kan utledes fra husholdningssammensetningen, men bruker disse som betingelser i forhold til vektene vi beregner for utvalget. Dette skal vi komme tilbake til i avsnitt 3.5, mens vi foreløpig bare skal utnytte at vi antar at vi kjenner alle husholdninger og da særlig adressegruppen og antallet familier i hver husholdning.

Vi kan også formulere dette som vi har en liste over alle DSF familier i hele populasjonen og hvilken husholdning hver av disse tilhører. Vi antar fortsatt en stratuminndeling av familiene med hensyn på fylket r , adressegruppen s og antall familier f i husholdningen. Vi tenker oss videre at en egenskap om boligen eller husholdningen som måles på hvert skjema kan beskrives ved en tilfeldig variabel definert ved

$$(3.15) \quad Y_{i,h,r,s,f} = \mu_{r,s,f} + \varepsilon_{h,r,s,f}$$

der i betegner familienummer og h husholdning. Indeksen i for DSF familie, betegner kun at vi gjentar husholdningsinformasjonen for hver DSF familie i husholdningen.

Modellen i (3.15) har altså en forventet verdi som er avhengig av fylket, adressegruppen og antallet DSF familier i husholdningen. Dessuten antar vi at avvikene, dersom vi kun ser på husholdningene, fra denne forventede verdien er uavhengige og identisk fordelte omkring null, dvs. at

$$(3.16) \quad \begin{aligned} E(\varepsilon_{h,r,s,f}) &= 0 \\ \text{Var}(\varepsilon_{h,r,s,f}) &= \sigma_{r,s,f}^2 \end{aligned}$$

Det kan være en egenskap til husholdningen eller boligen vi måler, f.eks. om det er barn i husholdningen eller om boligen har kjøkken på over 6 kvm, eller det er en verdi til husholdningen eller boligen vi måler, f.eks. antall personer i husholdningen eller antall kvm bruksareal i boligen. Vi kan også knytte egenskapene eller verdiene til hver DSF familie i utvalget og resten av populasjonen, men det betyr at dersom vi summerer Y -verdiene for alle DSF familiene i en husholdning må vi dividere denne summen med antall DSF familier i husholdningen. Altså skal vi summere alle Y -verdiene over DSF familiene må vi dividere hver Y -verdi med antall DSF familier i husholdningen. Dette kan vi

³ Dette og neste avsnitt er forholdsvis tekniske.

skrive opp som en formell formel for hvert stratum definert ved fylke r, adressegruppe s og antall DSF familier gitt ved f

$$(3.17) \quad T_{r,s,f} = \sum_{i,h \in utv_{r,s,f}^{fam}} \frac{1}{f} Y_{i,h,r,s,f} + \sum_{i,h \notin utv_{r,s,f}^{fam}} \frac{1}{f} Y_{i,h,r,s,f} = \sum_{i,h \in utv_{r,s,f}^{fam}} \left[\frac{Y_{i,h,r,s,f}}{f} \right] + \sum_{i,h \notin utv_{r,s,f}^{fam}} \left[\frac{Y_{i,h,r,s,f}}{f} \right]$$

der $i, h \in utv_{r,s,f}^{fam}$ og $i, h \notin utv_{r,s,f}^{fam}$ betyr henholdsvis summen over alle DSF familier og husholdninger i utvalget og utenfor utvalget for et stratum definert av fylket r og adressegruppen s. I den siste versjonen av (3.17) kan vi tolke beregningene som en sum av husholdningens andel av DSF familien og utføre alle beregninger på familienivået. Fordelen med (3.17) er selvsagt at vi da kan lage statistikk med familie som enhet og ikke husholdning. Tilbake til husholdningen som enhet i (3.17) finner vi på dette viset antallet husholdninger som har enn bestemt egenskap eller den totale verdien av en egenskap til husholdningen eller boligene som er målt på skjemaet. Setter vi inn antallet DSF familier for Y eller setter Y lik 1 ser vi at vi finner henholdsvis antallet DSF familier og husholdninger fra (3.17). Det viser at (3.17) er konsistent med antallet DSF familier og husholdninger i populasjonen. Den første summen på høyre side i (3.17) kjenner vi fra utvalget, mens den andre summen i (3.17) må vi anslå ved å bruke modellen i (3.15 -16) sammen med data i den første summen i (3.17). Vi estimerer da den forventede verdien lik

$$(3.18) \quad \hat{\mu}_{r,s,f} = \frac{1}{m_{r,s,f}^{hush}} \sum_{i,h \in utv_{r,s,f}^{fam}} \frac{1}{f} Y_{i,h,r,s,f} = \frac{1}{m_{r,s,f}^{hush}} \sum_{h \in utv_{r,s,f}^{hush}} Y_{h,r,s,f}$$

siden

- $m_{r,s,f}^{hush} = \sum_{i,h \in utv_{r,s,f}^{fam}} \frac{1}{f} = \sum_{h \in utv_{r,s,f}^{hush}} 1$

er antallet husholdninger i utvalget. Helt tilsvarende kan vi også som nevnt ovenfor skrive antallet husholdninger i populasjonen ved

- $M_{r,s,f}^{hush} = \sum_{i,h \in utv_{r,s,f}^{fam}} \frac{1}{f} + \sum_{i,h \notin utv_{r,s,f}^{fam}} \frac{1}{f}$

Dersom vi predikerer Y for hvert familienummer i en husholdning utenfor utvalget ved hjelp av (3.18),

$$(3.19) \quad \hat{Y}_{i,h,r,s,f} = \hat{\mu}_{r,s,f}$$

kan vi predikere den totale verdien til Y-ene i fylke r for adressegruppe s når det er f antall DSF familier i husholdningene ved formelen

$$(3.20) \quad \begin{aligned} \hat{T}_{r,s,f} &= \sum_{i,h \in utv_{r,s,f}^{fam}} \frac{1}{f} Y_{i,h,r,s,f} + \sum_{i,h \notin utv_{r,s,f}^{fam}} \frac{1}{f} \hat{Y}_{i,h,r,s,f} = \sum_{i,h \in utv_{r,s,f}^{fam}} \frac{1}{f} Y_{i,h,r,s,f} + \sum_{i,h \notin utv_{r,s,f}^{fam}} \frac{1}{f} \hat{\mu}_{r,s,f} \\ &= m_{r,s,f}^{hush} \hat{\mu}_{r,s,f} + (M_{r,s,f}^{hush} - m_{r,s,f}^{hush}) \hat{\mu}_{r,s,f} = M_{r,s,f}^{hush} \hat{\mu}_{r,s,f} = M_{r,s,f}^{hush} \frac{1}{m_{r,s,f}^{hush}} \sum_{i,h \in utv_{r,s,f}^{fam}} \frac{1}{f} Y_{i,h,r,s,f} \\ &= \sum_{i,h \in utv_{r,s,f}^{fam}} \left[\frac{M_{r,s,f}^{hush}}{f \cdot m_{r,s,f}^{hush}} \right] Y_{i,h,r,s,f} = \sum_{i,h \in utv_{r,s,f}^{fam}} w_{i,h,r,s,f}^{fam} Y_{i,h,r,s,f} \end{aligned}$$

der vekten

$$(3.21) \quad w_{i,h,r,s,f}^{fam} = \frac{M_{r,s,f}^{hush}}{f \cdot m_{r,s,f}^{hush}}$$

er av samme form som vekten vi beregnet i (3.9) basert på trekk sannsynlighetene, men der vi bruker mer detaljert informasjon om antallet husholdninger i populasjonen og utvalget. Det er derfor et problem ved vektene i (3.21) - vi kjenner ikke antallet husholdninger i populasjonen siden vi ikke vil ha tilgang til et ferdig bearbeidet datagrunnlag av side 4 for alle boligskjemaer. Dersom vi setter Y lik antallet DSF familier i husholdningene betyr (3.20) at vi predikerer antallet slike familier i alt

$$(3.22) \quad \hat{M}_{r,s,f} = \sum_{i,h \in \text{utv}_{r,s,f}^{fam}} \frac{M_{r,s,f}^{hush}}{f \cdot m_{r,s,f}^{hush}} f = \sum_{i,h \in \text{utv}_{r,s,f}^{fam}} \frac{M_{r,s,f}^{hush}}{m_{r,s,f}^{hush}} = \frac{M_{r,s,f}^{hush}}{m_{r,s,f}^{hush}} \sum_{i,h \in \text{utv}_{r,s,f}^{fam}} 1 = \frac{M_{r,s,f}^{hush}}{m_{r,s,f}^{hush}} m_{r,s,f}$$

siden vi summerer over alle DSF familier i utvalget.

Løser vi (3.22) med hensyn på antallet husholdninger i populasjonen finner vi at

$$(3.23) \quad M_{r,s,f}^{hush} = \frac{\hat{M}_{r,s,f}}{m_{r,s,f}} m_{r,s,f}^{hush}$$

La oss bruke (3.23) og heller predikere antall husholdninger fra antallet DSF familier i populasjonen og utvalget sammen med antall husholdninger i utvalget. Da får vi at antallet husholdninger er lik

$$(3.24) \quad \hat{M}_{r,s,f}^{hush} = \frac{M_{r,s,f}}{m_{r,s,f}} m_{r,s,f}^{hush} = \frac{M_{r,s,f}}{f}$$

siden $m_{r,s,f} = f \cdot m_{r,s,f}^{hush}$.

Setter vi den første likheten i (3.24) inn i formelen for vekten i (3.21) finner vi estimert vekt lik

$$(3.25) \quad \hat{w}_{i,h,r,s,f}^{fam} = \frac{\frac{M_{r,s,f}}{m_{r,s,f}} m_{r,s,f}^{hush}}{f \cdot m_{r,s,f}^{hush}} = \frac{M_{r,s,f}}{f \cdot m_{r,s,f}}$$

Problemet med denne vekten er fortsatt at vi heller ikke kjenner antall DSF familier i husholdninger med et bestemt antall av disse i hver husholdning. Siden vi tar med alle DSF familier i den husholdningen vi har etablert er det opplagt at forholdet mellom antallet familienumre i utvalget og i populasjonen vokser med antallet familier i husholdningene - altså burde vi forvente helt presist at

$$(3.26) \quad \frac{m_{r,s,f}}{M_{r,s,f}} = f \cdot \frac{m_{r,s,1}}{M_{r,s,1}}$$

og løser vi (3.26) med hensyn på antallet familienumre totalt i populasjonen som hører til i husholdninger med f familienumre vil vi finne en sammenheng mellom det totale antallet familienumre og antallet familienumre som utgjør en husholdning aleine gitt ved

$$(3.27) \quad M_{r,s} = \sum_f M_{r,s,f} = \sum_f \frac{m_{r,s,f}}{f \cdot m_{r,s,1}} M_{r,s,1} = \left\{ \sum_f \frac{m_{r,s,f}}{f \cdot m_{r,s,1}} \right\} M_{r,s,1}$$

Kombinerer vi nå (3.26) og (3.27) får vi en estimator for antallet familienumre i husholdninger som består av f antall familienumre

$$(3.28) \quad \hat{M}_{r,s,f} = \frac{m_{r,s,f}}{f \cdot m_{r,s,1}} \frac{M_{r,s}}{\sum_g \frac{m_{r,s,g}}{g \cdot m_{r,s,1}}} = \frac{m_{r,s,f}}{f \cdot m_{r,s,1}} \frac{M_{r,s}}{\frac{1}{m_{r,s,1}} \sum_g \frac{m_{r,s,g}}{g}} = m_{r,s,f} \frac{M_{r,s}}{f \cdot m_{r,s}^{hush}}$$

siden

$$\bullet \quad \sum_f \frac{m_{r,s,f}}{f} = m_{r,s}^{hush}$$

er lik antallet husholdninger i utvalget og det siste uttrykket i (3.28) svarer helt til (3.10).

Vi har samtidig estimert antall husholdninger for hver f også ved å dele med antallet familienumre i husholdningene, altså har vi at

$$(3.29) \quad \hat{M}_{r,s,f}^{hush} = \frac{\hat{M}_{r,s,f}}{f} = m_{r,s,f}^{hush} \frac{M_{r,s}}{f^2 \cdot m_{r,s}^{hush}}$$

La oss imidlertid sette inn (3.28) i (3.25) som gir et nytt uttrykk for vekten med hensyn på DSF familiene:

$$(3.30) \quad \hat{w}_{i,h,r,s,f}^{fam} = \frac{\hat{M}_{r,s,f}}{f \cdot m_{r,s,f}} = \frac{\frac{m_{r,s,f}}{f \cdot m_{r,s,1}} \frac{M_{r,s}}{\sum_g \frac{m_{r,s,g}}{g \cdot m_{r,s,1}}}}{f \cdot m_{r,s,f}} = \frac{M_{r,s}}{f^2 m_{r,s,1}} \frac{1}{\sum_g \frac{m_{r,s,g}}{g \cdot m_{r,s,1}}} = \frac{M_{r,s}}{f^2 m_{r,s}^{hush}}$$

Vekten vi har beregnet i (3.30) bygger på at vi fra modellen i (3.15-16) har kommet fram til en etterstratifisering av utvalget etter

- fylke - varierende trekksannsynligheter i AKU+
- adressegruppe - skiller mellom unike/ikke unike adresser kombinert med stedfortrederboliger eller ikke
- antall familier - deler husholdningene etter antall familier

Vi må imidlertid huske at et avgjørende punkt er beregningen av antallet DSF familier i husholdningen som hører til husholdninger som består av et bestemt antall slike. Effekten av dette grepet avhenger selvsagt av hvor representativt utvalget er i betydningen riktig fordeling mellom typer DSF familier med hensyn på antallet i husholdningene i utvalget sammenliknet med populasjonen. I kapittel 4 har vi vist at den relative usikkerheten til antallet DSF familier av forskjellig type er på 2-3 prosent i de minste fylkene og 1-2 prosent i de største fylkene - se kapittel 4 for flere detaljer. Vi vil imidlertid ikke kunne holde fast ved denne etterstratifiseringen når vi skal gå videre og kalibrere mot andre størrelser.

Vi kan derfor betrakte det estimerte antallet husholdninger i (3.29) som et estimat der vi kun tar hensyn til trekkssannsynligheten og det kan derfor være interessant å følge utviklingen av disse tallene etter hvert som vi legger til kalibreringskrav i forhold til personene, familienumrene og boligopplysningene fra registrene.

La oss også kombinere vekten i (3.30) med formel (3.20) til å stille opp den tilhørende husholdningsvekten ved å multiplisere familievekten i (3.30) med antallet familier i husholdningen som gir

$$(3.31) \quad \hat{w}_{h,r,s,f}^{hush} = f \cdot \hat{w}_{i,h,r,s,f}^{fam} = \frac{M_{r,s}}{f \cdot m_{r,s}^{hush}} = w_{h,r,s,f}^{hush}$$

som viser at vi har utledet eksakt samme husholdningsvekt som i (3.9), men nå ut fra modellen definert i (3.15-16) og opplegget for å beregne totalene i (3.17).

3.4. Bruk av statistiske modeller for å beregne vekter - data er alle personene

I foregående avsnitt brukte vi informasjon om DSF familiene for å lage vekter. Dessuten var disse vektene konstruert slik at vi brukte alle DSF familiene til husholdningene som datagrunnlag for å produsere foreløpige tall. I dette avsnittet skal vi arbeide videre med å utnytte annen informasjon om personene i husholdningene for å justere videre på vektene vi beregnet i forrige avsnitt for familiene og husholdningene.

Vi skal nå se hvordan vi kan utnytte informasjon om de enkelte personer i husholdningene til å beregne vekter som tar hensyn til kjente totaler for relevante egenskaper til personene i utvalget. Vi skal bygge videre på vekten vi utledet i (3.25) i foregående avsnitt. Vi antar nå at vi har en datafil over alle personene i utvalget der vi også kjenner til DSF familiene.

Vi skal bare se på hvordan vi kan justere vektene i (3.25) slik at vi beholder egenskapen at de summerer til det riktige antall familier. Vi definerer da noen størrelser vi trenger:

- $N_{r,s}$ - antall personer i populasjonen i fylke r og for adressegruppe s
- $n_{r,s,f}$ - antall personer i utvalget i fylke r , for adressegruppe s og antall familienummer i husholdningen lik f
- $n_{h,r,s}$ - antall personer i den husholdningen person i tilhører og vi inkluderer også fylket og adressegruppe i notasjonen

Vi skal først gå tilbake til modellen for egenskaper ved husholdningen definert i foregående avsnitt, formlene (3.15-16), og totalene vi skal beregne, definert i (3.17). Vi skal endre litt på opplegget ut fra at enheten nå er personer og ikke familier og gjennomføre en argumentasjon av samme type som i foregående avsnitt slik at vi ender opp med vekter som svarer til (3.30).

Vi definerer derfor først en enkel statistisk modell for egenskapene i husholdningene etter samme lest som modellen (3.15-16) i forrige avsnitt. Det vil si at vi antar for hver person i som hører til husholdningen h at

$$(3.32) \quad Y_{i,h,r,s,f} = \mu_{r,s,f} + \varepsilon_{i,h,r,s,f}$$

egenskapen i husholdningen er avhengig av fylket, adressegruppe og antall DSF familier. Vi antar at feilleddene tilfredstiller

$$(3.33) \quad \begin{aligned} E(\varepsilon_{h,r,s,f}) &= 0 \\ \text{Var}(\varepsilon_{h,r,s,f}) &= \sigma_{r,s,f}^2 \end{aligned}$$

Vi skal predikere den totale verdien for husholdningene i et stratum definert av fylket og adressegruppe. Siden vi skal summere over alle personer i henholdsvis utvalget og i resten av utvalget må vi dividere egenskapen med antall personer i personens husholdning:

$$(3.34) \quad T_{r,s,f} = \sum_{i \in \text{utv}_{r,s,f}^{\text{person}}} \frac{1}{n_{h,r,s}} Y_{i,h,r,s,f} + \sum_{i \notin \text{utv}_{r,s,f}^{\text{person}}} \frac{1}{n_{h,r,s}} Y_{i,h,r,s,f}$$

Vi finner da ved å sette inn estimatet for den forventede verdien

$$(3.35) \quad \hat{\mu}_{r,s,f} = \frac{1}{m_{r,s,f}^{\text{hush}}} \sum_{i \in \text{utv}_{r,s,f}^{\text{person}}} \frac{1}{n_{h,r,s}} Y_{i,h,r,s,f} = \frac{1}{m_{r,s,f}^{\text{hush}}} \sum_{h \in \text{utv}_{r,s,f}^{\text{hush}}} Y_{h,r,s,f} = \bar{Y}_{r,s,f}^{\text{hush}}$$

for Y- verdiene utenfor utvalget at

$$(3.36) \quad \begin{aligned} \hat{T}_{r,s,f} &= \sum_{i \in \text{utv}_{r,s,f}^{\text{person}}} \frac{1}{n_{h,r,s}} Y_{i,h,r,s,f} + \sum_{i \notin \text{utv}_{r,s,f}^{\text{person}}} \frac{1}{n_{h,r,s}} \hat{Y}_{i,h,r,s,f} \\ &= m_{r,s,f}^{\text{hush}} \hat{\mu}_{r,s,f} + (M_{r,s,f}^{\text{hush}} - m_{r,s,f}^{\text{hush}}) \hat{\mu}_{r,s,f} = M_{r,s,f}^{\text{hush}} \hat{\mu}_{r,s,f} = \sum_{i \in \text{utv}_{r,s,f}^{\text{person}}} \frac{M_{r,s,f}^{\text{hush}}}{n_{h,r,s} \cdot m_{r,s,f}^{\text{hush}}} Y_{i,r,s,f} \end{aligned}$$

Fra (3.36) har vi da en vekt for hver person som er lik for alle personer i samme husholdning:

$$(3.37) \quad w_{i,h,r,s,f}^{\text{person}} = \frac{M_{r,s,f}^{\text{hush}}}{n_{h,r,s} \cdot m_{r,s,f}^{\text{hush}}}$$

Vi ser at formen på personvekten i (3.37) er lik formen på familievekten i (3.21). Dersom vi nå forsøker å estimere antall DSF familier i fylket r av type s i husholdninger som består av f antall DSF familier finner vi, når bruker $Y_{i,h,r,s,f} = f$, at

$$(3.38) \quad \begin{aligned} \hat{M}_{r,s,f}^{\text{person}} &= \sum_{i \in \text{utv}_{r,s,f}^{\text{person}}} w_{i,h,r,s,f}^{\text{person}} \cdot f = \sum_{i \in \text{utv}_{r,s,f}^{\text{person}}} \frac{M_{r,s,f}^{\text{hush}}}{n_h \cdot m_{r,s,f}^{\text{hush}}} f = \sum_{h \in \text{utv}_{r,s,f}^{\text{hush}}} \frac{M_{r,s,f}^{\text{hush}}}{n_h \cdot m_{r,s,f}^{\text{hush}}} n_h \cdot f \\ &= \sum_{h \in \text{utv}_{r,s,f}^{\text{hush}}} \frac{M_{r,s,f}^{\text{hush}}}{m_{r,s,f}^{\text{hush}}} f = \frac{M_{r,s,f}^{\text{hush}}}{m_{r,s,f}^{\text{hush}}} \sum_{h \in \text{utv}_{r,s,f}^{\text{hush}}} f = \frac{M_{r,s,f}^{\text{hush}}}{m_{r,s,f}^{\text{hush}}} m_{r,s,f} = \hat{M}_{r,s,f} \end{aligned}$$

som er det estimerte antallet familier vi gikk ut fra i forrige avsnitt når vi regnet oss fram til samme vekt i (3.31) som vi hadde argumentert for i (3.9) ut fra trekk sannsynligheten. Dersom vi fra (3.38) gjennomfører de samme argumentene som vi brukte fra (3.22) fram til familievekten (3.30) i forrige avsnitt vil vi her også komme fram til en familievekt

$$(3.39) \quad \hat{w}_{i,h,r,s,f}^{person} = \frac{M_{r,s}}{n_{h,r,s} \cdot f \cdot m_{r,s}^{hush}}$$

og vi ser forbindelsen med familievekten i foregående avsnitt gjennom likningen

$$(3.40) \quad \begin{aligned} \sum_{i,h \in \text{utv}_{r,s,f}^{person}} \hat{w}_{i,h,r,s,f}^{person} Y_{i,h,r,s,f} &= \sum_{i,h \in \text{utv}_{r,s,f}^{person}} \frac{M_{r,s}}{n_{h,r,s} \cdot f \cdot m_{r,s}^{hush}} Y_{i,h,r,s,f} \\ &= \sum_{i,h \in \text{utv}_{r,s,f}^{fam}} \frac{M_{r,s}}{f \cdot f \cdot m_{r,s}^{hush}} Y_{i,h,r,s,f} = \sum_{i,h \in \text{utv}_{r,s,f}^{fam}} \hat{w}_{i,h,r,s,f}^{fam} Y_{i,h,r,s,f} \end{aligned}$$

Det hele bygger egentlig på en enda enklere observasjon dersom vi sammenlikner å summere husholdningsegenskapens fordeling på henholdsvis personene og familiene i husholdningene:

$$(3.41) \quad \sum_{i,h \in \text{utv}_{r,s,f}^{person}} \frac{Y_{i,h,r,s,f}}{n_h} = \sum_{h \in \text{utv}_{r,s,f}^{hush}} Y_{h,r,s,f} = \sum_{i,h \in \text{utv}_{r,s,f}^{fam}} \frac{Y_{i,h,r,s,f}}{f}$$

Personvekten vi har utledet i (3.39) er på samme vis som familievekten i (3.30) koblet til den opprinnelige husholdningsvekten definert i (3.09).

Vi vil avslutte dette avsnittet ved å understreke at vekten i (3.39) ikke gir en garanti for at vi får korrekt antall personer verken for hver adressegruppe eller fylke siden vi har at

$$(3.42) \quad \hat{N}_{r,s} = \sum_f \sum_{i,h \in \text{utv}_{r,s,f}^{person}} \hat{w}_{i,h,r,s,f}^{person} n_h = \sum_f \sum_{i,h \in \text{utv}_{r,s,f}^{person}} \frac{M_{r,s}}{f \cdot m_{r,s}^{hush}} = \sum_f \frac{M_{r,s}}{f \cdot m_{r,s}^{hush}} n_{r,s,f} = \frac{M_{r,s}}{m_{r,s}^{hush}} \sum_f \frac{n_{r,s,f}}{f}$$

Dette viser at vi anslår antallet personer i populasjonen som produktet av

- forholdet mellom antallet DSF familier i populasjonen og det opprinnelig utrukne utvalget

og

- det forventede antallet personer i de mulige familiene som kunne trekkes ut gitt de utrukne husholdningene.

Det betyr at i gjennomsnitt gir metoden det korrekte antallet personer, men det varierer fra utvalg til utvalg.

I neste avsnitt skal vi se på hvordan vi kan forbedre vektene ved å kalibrere mot antallet personer i forskjellige grupper etter alder og kjønn.

3.5. Vekter for foreløpige tall - kalibrering av designvektene

Vi har i de foregående avsnittene vist hvordan vi ved å definere en passende modell for husholdnings-egenskapene som tar hensyn til antall DSF familier når data er representert ved henholdsvis familie eller personer får vi tilbake vekten definert i (3.9) kun basert på trekkssannsynligheten for husholdningene. Vi kan sammenfatte resultatene i avsnittene 3.2 -3.4 ved likningen

$$(3.43) \quad \hat{T}_{Y,r,s,f} = \sum_{h \in utv_{r,s,f}^{hush}} w_{h,r,s,f}^{hush} Y_{h,r,s,f} = \sum_{i, h \in utv_{r,s,f}^{fam}} w_{h,r,s,f}^{hush} \frac{Y_{h,r,s,f}}{f} = \sum_{i, h \in utv_{r,s,f}^{person}} w_{h,r,s,f}^{hush} \frac{Y_{h,r,s,f}}{n_{h,r,s}}$$

der

$$(3.44) \quad w_{r,s,f}^{hush} = \frac{M_{r,s}}{f \cdot m_{r,s}^{hush}}$$

Dette viser at vi kan bruke samme vekt for husholdninger, DSF familier og personer, men der vi må justere husholdnings egenskapen enten ved å dividere med antall familienumre eller antall personer.

I dette avsnittet skal vi se på hvordan vektene i (3.44) kan justeres slik at vi tar hensyn til designen og et antall betingelser for personene og DSF familiene som vi vil ha oppfylt. Dette er egenskaper som karakteriserer husholdningen både med hensyn til DSF familiene og personene i familien. Vi vet allerede at vektene i (3.44) gir korrekt antall DSF familier av type s i fylke r dersom vi setter inn husholdningsegenskapen antall familienumre i (3.43) siden

$$(3.45) \quad \sum_f \sum_{h \in utv_{r,s,f}^{hush}} w_{r,s,f}^{hush} f = \sum_f \sum_{h \in utv_{r,s,f}^{hush}} \frac{M_{r,s}}{f \cdot m_{r,s}^{hush}} f = \sum_f \sum_{h \in utv_{r,s,f}^{hush}} \frac{M_{r,s}}{m_{r,s}^{hush}} = \sum_f \frac{M_{r,s}}{m_{r,s}^{hush}} m_{r,s}^{hush} = M_{r,s}$$

I Heldal(1992) er det vist hvordan vi kan beregne vekter for personene som tar hensyn til et antall betingelser vi legger på personene slik at vi har samme vekt for alle personer i samme husholdning, men der vektene kan variere fra husholdning til husholdning. I vedlegg A har vi vist prinsippet for dette ved at dersom en ønsker å kalibrere vekter som i utgangspunktet gir riktig antall personer kan vi kalibrere disse vektene slik at vi også får riktig antall menn ved å bruke andelen menn i husholdningen som en kalibreringsvariabel. Siden andelen menn i husholdningen har samme verdi for alle personer i husholdningen gis alle personer i husholdningen samme endring i forhold til den opprinnelige vekten. Vi vil da sitte tilbake med kalibrerte vekter som er like for alle personer i husholdningen og denne vekten er også vekten for husholdningen. Vi kan derfor både bruke vekten når vi skal beregne tall for personene og husholdningene.

Utgangspunktet for det videre arbeidet med å kalibrere vektene er at betingelsen vi har gjengitt i formel (3.45) ligger fast - altså skal vektene beholde den egenskapen at de gir riktig antall DSF familier i hvert fylke for hver adressegruppe. Vi vil imidlertid gjennomføre kalibreringen ved å justere vektene for hver kombinasjon av fylke og om adressen er unik eller ikke - i alt 38 grupper slik at vektene innenfor hver gruppe justeres slik at et sett av betingelser er oppfylt. Da vil de kalibrerte vektene kun stemme med de opprinnelige for kombinasjon av fylke og om adressen er unik eller ikke. La oss innføre en betegnelse for disse betingelsene:

$$(3.46) \quad X_{r,s,k} \quad \text{der } k=1,2,\dots,K$$

Dette kan være antall personer i alt, antall familienumre i alt, antall menn, antall menn over 67 år osv., se avsnitt 3.6 for en oversikt over betingelsene som ble brukt.

La videre

$$(3.47) \quad x_{h,r,s,f,k} \quad \text{der } k=1,2,\dots,K$$

er verdien i egenskapen k i husholdningen, f.eks. antall personer eller antall familier i husholdningen. Sett videre

$$(3.48) \quad z_{i,h,r,s,f,k} = \frac{x_{h,r,s,f,k}}{n_{h,r,s}} \quad \text{der } k=1,2,\dots,K$$

er andelen av egenskapen som hver person i husholdningen har.

Da antar vi at dersom vi på liknende vis definerer Y 's egenskap på hver person i husholdningen ved

$$(3.49) \quad U_{i,h,r,s,f} = \frac{Y_{h,r,s,f}}{n_{h,r,s}}$$

så vil vi formulere følgende modell for U med hensyn på z -ene gitt ved

$$(3.50) \quad U_{i,h,r,s,f} = \alpha_{r,s} + \sum_{k=1}^K \beta_{r,s,k} z_{i,h,r,s,f,k} + \varepsilon_{i,h,r,s,f}$$

Vi skal nå estimere parametrene ved minste kvadraters metode fra størrelsen

$$(3.51) \quad Q(\alpha_{r,s}, \beta_{r,s,1}, \beta_{r,s,2}, \dots, \beta_{r,s,K}) = \sum_f \sum_{i,h \in \text{utv}_{r,s,f}^{\text{Person}}} w_{h,r,s,f}^{\text{hush}} (U_{i,h,r,s,f} - \alpha_{r,s} - \beta_{r,s,1} z_{i,h,r,s,f,1} - \beta_{r,s,2} z_{i,h,r,s,f,2} + \dots + \beta_{r,s,K} z_{i,h,r,s,f,K})^2$$

dvs., minimere uttrykket i (3.51) med hensyn på parametrene. Dersom vi bruker antall familienumre som en av betingelsene vil vi få et sett av nye vekter som oppfyller betingelsene i (3.46) dersom vi predikerer alle U -er i populasjonen ved

$$(3.52) \quad \hat{U}_{i,h,r,s,f} = \hat{\alpha}_{r,s} + \sum_{k=1}^K \hat{\beta}_{r,s,k} z_{i,h,r,s,f,k}$$

og vi setter inn og gjennomfører regnestykket

$$(3.53) \quad \sum_f \sum_{i,h} \hat{U}_{i,h,r,s,f} = \sum_f \sum_{i,h \in \text{utv}_{r,s,f}^{\text{Person}}} v_{i,h,r,s,f}^{\text{Kalibrert}} U_{i,h,r,s,f}$$

der

de kalibrerte vektene oppfyller

$$(3.54) \quad v_{i,h,r,s,f}^{\text{Kalibrert}} = v_{h,r,s,f}^{\text{Kalibrert}}$$

dvs., samme vekt for alle personer i husholdningen og dessuten er formen på de kalibrerte vektene gitt ved formen

$$(3.55) \quad v_{i,h,r,s,f}^{Kalibrert} = w_{i,h,r,s,f}^{hush} \cdot \left[\sum_{k=1}^K \sum_{l=1}^K z_{i,h,r,s,f,k} A_{k,l} X_{r,s,l} \right]$$

der

- $A_{k,l}$

er avhengig av de opprinnelige vektene i (3.44) og z -ene definert i (3.48). I Heldal(1992) er detaljene for beregningene gitt og det er også vist at betingelsene er oppfylt, dvs. at

$$(3.56) \quad \sum_f \sum_{i,h \in utv_{r,s,f}^{Person}} v_{i,h,r,s,f}^{Kalibrert} z_{i,h,r,s,f,k} = X_{r,s,k}$$

og

$$(3.57) \quad \sum_f \sum_{h \in utv_{r,s,f}^{hush}} v_{h,r,s,f}^{Kalibrert} = \sum_f \sum_{h \in utv_{r,s,f}^{hush}} v_{h,r,s,f}^{hush} = \hat{M}_{r,s}^{hush}$$

Vi endrer med andre ord ikke estimatet av antall husholdninger totalt i hvert fylke og for gruppeinndelingen etter om adressen er unik eller ikke, men vi fordeler husholdningene på en annen måte i forhold til de betingelsene vi brukte for å kalibrere vektene.

Formelen i (3.55) garanterer ikke at vekten er større enn eller lik 1. For å sikre at en får meningsfulle vekter kan en legge inn slike betingelser i beregningene. Da regnes ikke vektene ut direkte, men det kan bli behov for en iterasjon fram mot løsningen. Det kan også legges inn krav om at de kalibrerte vektene ikke må variere for mye i forhold til de opprinnelige vektene. I den faktiske utregningen av vektene for foreløpige tall satte en som betingelse at forholdet mellom de opprinnelige og kalibrerte vektene skulle ligge mellom 1/3 og 3, se avsnitt 3.8 for flere detaljer om resultatet av kalibreringen.

3.6. Valg av betingelser for å kalibrere vektene

Vi skal nå stille opp de betingelsene vi vil at vektene skal oppfylle. Disse kravene er særlig knyttet til aldersfordeling for menn og kvinner og egenskaper ved Besys familiene. For de to gruppene av variabler som brukes som kalibreringsvariabler nedenfor må det utarbeides tabeller for alle personene i populasjonen slik at en finner hvor mange personer totalt i hvert stratum (bestemt av fylke og om adressen er unik eller ikke) som har de forskjellige egenskapene i utvalget. Vi har ingen kalibreringsvariabler som skiller mellom boligtyper siden dette er tatt hånd om allerede i inndelingen i adressegrupper som tilnærmet skiller mellom de fire hovedtypene eneboliger, småhus, rekkehus og blokker.

Inndelingen i fem familietyper for Besys familiene og de åtte aldersgruppene for menn og kvinner hver for seg tar for seg viktige forskjeller mellom husholdningene etter livsfaser, alder og kjønn utover den betydning antallet DSF familier har for treksannsynlighetene til utvalget. Hensikten med de variablene vi har tatt ut til kalibrering er både å korrigere åpenbare skjevheter i utvalget i forhold til populasjonen, men også bruke disse til å redusere usikkerheten i de publiserte tallene under forutsetning av at de forklarer godt variasjonene på forskjellige boligeenskaper i utvalget og populasjonen.

Vi foreslår da følgende variabler:

Besys familier

Vi klassifiserer alle personer i populasjonen og utvalget til en av fem familietyper, dvs. at hver person hører til en og kun en av fem mulige typer Besys familier (det normale er da at alle personer i husholdningen i utvalget får samme klassifisering, men ikke nødvendigvis):

- enperson familie
- Ektepar uten barn
- aleneforelder (med barn)
- Ektepar med barn
- Samboere med felles barn

Personer

Vi klassifiserer alle personer i utvalget og populasjonen i en og kun en av 8 aldersgrupper for enten menn eller kvinner:

- Mann, -19
- Mann, 20-29
- Mann, 30-39
- Mann, 40-49
- Mann, 50-59
- Mann, 60-69
- Mann, 70-79
- Mann, 80-
- Kvinne, -19
- Kvinne, 20-29
- Kvinne, 30-39
- Kvinne, 40-49
- Kvinne, 50-59
- Kvinne, 60-69
- Kvinne, 70-79
- Kvinne, 80-

Dette svarer også til at vi for hver av de 21 variablene, 5 familieegenskaper og 8 aldersegenskaper for menn og kvinner, teller vi opp antallet i hver husholdning i utvalget.

Vi må da for hver av de 38 gruppene definert ved fylke og unik/ikke unik adresse beregne antallet personer som tilhører de fem Besys familie gruppene og 16 persongruppene i populasjonen. Disse 21 størrelsene er da de 21 X-verdiene som vi definerte i (3.46) for hver kombinasjon av fylke og adressegruppe.

Første kontroll er derfor for hvert stratum å undersøke hvor mange personer vi finner i hver av de 5 familiegruppene og 16 gruppene etter alder og kjønn for både utvalget og populasjonen. Det vi forlanger for å kunne gjennomføre kalibreringen er at det finnes minst en person i hver gruppe det finnes personer i populasjonen innen. I tabell 3.1 er det gitt en oversikt over antallet personer i populasjonen og utvalget for hver av de 76 strata som finnes.

Vi ser fra tabell 3.1 betydelige forskjeller mellom fylkene med hensyn til fordelingen av DSF familier på de fire adressegruppene. Totalt er hver fjerde DSF familie på adressegruppen for stedfortredere med ikke unik adresse og av disse var det over 407 000 i boligblokker. Vi ser at for Oslo finnes over 70 % av DSF familiene på en slik adresse. Oslo skiller seg også ut ved langt mindre DSF familier enn i resten av landet. Mens antallet personer i DSF familiene i gjennomsnitt er på om lag 2,4 er antallet personer per DSF familie i Oslo omtrent 1,7.

Tabell 3.1. Oversikt over DSF familier i populasjonen etter fylker og adressegrupper⁴

Fylke	Totalt	s=1	s=2	s=3	s=4
I Norge totalt	2 249 314	262 925	1 037 158	555 453	393 778
01 Østfold	126 614	14 571	69 005	28 713	14 325
02 Akershus	224 833	47 350	100 176	45 344	31 963
03 Oslo	308 543	18 715	35 420	217 553	36 855
04 Hedmark	95 332	7 110	61 164	9 656	17 402
05 Oppland	91 331	6 566	58 018	7 562	19 185
06 Buskerud	119 661	14 186	61 754	22 565	21 156
07 Vestfold	107 132	13 986	60 789	18 941	13 416
08 Telemark	82 070	7 980	51 047	11 096	11 947
09 Aust-Agder	47 876	3 214	32 427	2 779	9 456
10 Vest-Agder	72 281	10 080	37 322	11 168	13 711
11 Rogaland	169 951	21 486	86 331	30 548	31 586
12 Hordaland	211 581	31 359	84 606	58 402	37 214
14 Sogn og Fjordane	48 909	4 414	25 431	2 977	16 087
15 Møre og Romsdal	113 792	9 093	58 767	18 671	27 261
16 Sør-Trøndelag	135 340	19 829	50 220	37 867	27 424
17 Nord-Trøndelag	61 145	5 836	35 397	4 728	15 184
18 Nordland	119 122	12 298	68 885	12 841	25 098
19 Troms	76 299	9 613	38 257	10 423	18 006
20 Finmark	37 502	5 239	22 142	3 619	6 502

Tabell 3.2. Oversikt over antallet husholdninger i bruttoutvalget etter fylker og adressegrupper

Fylke	Totalt	s=1	s=2	s=3	s=4
I Norge totalt	16 959	2 033	8 059	4 000	2 865
01 Østfold	960	112	527	220	101
02 Akershus	1 494	309	674	326	185
03 Oslo	1 977	125	187	1 447	218
04 Hedmark	720	52	462	80	126
05 Oppland	707	60	439	59	149
06 Buskerud	901	113	464	175	149
07 Vestfold	807	117	463	130	97
08 Telemark	633	51	411	95	76
09 Aust-Agder	529	40	334	41	114
10 Vest-Agder	551	89	285	82	95
11 Rogaland	1 291	162	706	217	206
12 Hordaland	1 417	208	570	396	243
14 Sogn og Fjordane	511	56	255	32	168
15 Møre og Romsdal	856	69	447	139	201
16 Sør-Trøndelag	1 048	151	381	297	219
17 Nord-Trøndelag	531	48	309	41	133
18 Nordland	900	105	519	91	185
19 Troms	607	95	308	84	120
20 Finmark	517	71	318	48	80

⁴ s=1 betyr stedfortreder og unik adresse, s=2 betyr ikke stedfortreder med unik adresse, s=3 betyr stedfortreder med ikke unik adresse og s=4 betyr ikke stedfortreder med ikke unik adresse

Tabell 3.3. Oversikt over antallet husholdninger i nettoutvalget etter fylker og adressegrupper

Fylke	Totalt	s=1	s=2	s=3	s=4
I Norge totalt	15 783⁵	1 909	7 667	3 606	2 601
01 Østfold	907	103	506	207	91
02 Akershus	1 408	295	645	296	172
03 Oslo	1 774	117	180	1290	187
04 Hedmark	686	52	437	74	123
05 Oppland	662	58	414	55	135
06 Buskerud	843	110	437	159	137
07 Vestfold	746	113	435	118	80
08 Telemark	595	48	395	84	68
09 Aust-Agder	494	38	316	39	101
10 Vest-Agder	515	85	266	74	90
11 Rogaland	1 209	144	682	196	187
12 Hordaland	1 314	197	550	345	222
14 Sogn og Fjordane	476	48	246	31	151
15 Møre og Romsdal	803	67	431	125	180
16 Sør-Trøndelag	974	141	358	274	201
17 Nord-Trøndelag	511	46	297	41	127
18 Nordland	850	98	498	84	170
19 Troms	557	87	287	74	109
20 Finmark	459	62	287	40	70

3.7. Hvordan beregne foreløpige tall for studentene

I dette avsnittet skal vi kort skissere hvordan vi kan bygge videre på de foregående avsnittene for å produsere tall for studenter og særlig hvor de bor. Utgangspunktet er at vi vet hvor mange studenter i hvert fylke og for hver adressegruppe som har fått tilsendt eget skjema og vi kan innføre en egen notasjon for disse:

- $N_{r,s}^{Stud}$ - antallet studenter i fylke r og adressegruppe s som har mottatt boligskjema

Disse studentene må da ha en trekk-sannsynlighet for å komme med i utvalget som svarer til trekk-sannsynligheten til den husholdningen de hører til i formelt - altså den husholdningen de hører til ut fra sammenhengen mellom de som var preprintet på boligskjemaet og den husholdningssammensetningen vi har kommet fram til nå vi bruker den formelle adressen eller husholdningen om en vil.

Vi kan derfor legge til en variabelgruppe til på opplegget for kalibrering i foregående avsnitt, nemlig om det er sendt eget studentskjema i tillegg:

Studentskjema:

Vi klassifiserer alle personer i husholdningen til en av to mulige kategorier:

- minst en student i husholdningen har mottatt eget boligskjema
- ingen student i husholdningen har mottatt eget boligskjema

⁵ Antallet husholdninger i nettoutvalget reduseres til 15 782 etter kalibreringen siden vi måtte fjerne en husholdning ved koblingene som måtte gjøres ved kalibrering.

Ved å bruke vektorer som også er kalibrert etter dette kriteriet vil de kalibrerte vektene gi oss et riktig antall utsendt studentskjemaer fra utvalget.

Vi kan lage bolig- og husholdningsstatistikk over studentenes faktiske bosted og boligstandard som er konsistent med studentenes formelle bosted og boligstandard. Vi bruker vektene for husholdningene studentene hører til formelt på Y-variabler som beskriver deres faktiske boliger og husholdninger, herunder hvor de bor. Det eneste problemet som kan oppstå er at vi teller dobbelt dersom flere studenter i utvalget som kommer fra forskjellige formelle husholdninger faktisk bor i samme husholdning. Dette siste er antakelig svært lite sannsynlig og lett å kontrollere for.

Det er et problem at det var et betydelig frafall blant studentskjemaene og derfor vanskelig å avgjøre om frafall betyr at studenten også faktisk bor i foreldrenes bolig. For å produsere tall for studentene basert på utvalget måtte vi ha laget et opplegg for å imputere for enhetsfrafallet av studentskjemaer. Vi valgte derfor å ikke lage statistikk over studentenes faktiske husholdninger og boliger i de foreløpige tallene som ble publisert 22. april.

3.8. Opplegget for å beregne vektorer

I dette avsnittet beskriver vi i detalj trinnene for å beregne vektorer fram til de kalibrerte vektene vi brukte for å produsere tabellene som Statistisk sentralbyrå skal publiserte som foreløpige tall i FoB2001 22. april 2002. Vi valgte en prosess i tre trinn som ga en mulighet til å se hvordan tallene endret seg etter hvert som vi brukte mer informasjon og forhåpentligvis forbedret resultatene i tillegg.

Hele landet er et stratum

Vi utnytter kun at vi kjenner antall DSF familier i hele landet, antallet husholdninger i utvalget og antallet DSF familier hver husholdning i utvalget. Vi beregner da vekten til hver husholdning i utvalget ved brøken

$$\bullet \quad vekt = \frac{M}{f \cdot m^{hush}}$$

der

- M - er antall DSF familier i alt i Norge
- f - er antall DSF familier i husholdningen
- m^{hush} - er antall husholdninger i utvalget

Disse vektene gir det riktige antall DSF familier for hele landet.

Vi kan i tillegg til å kjøre ut tabellene til foreløpige tall også beregne

- Antallet DSF familier hvert fylke ved å bruke antall DSF familier i husholdningene som variabel
- Antallet personer i hvert fylke ved bruke antall personer i husholdningene som variabel

Ved sammenlikne med de faktiske populasjonstallene kan vi se hvor mye vi bommer i hvert fylke med denne enkle vekten.

For hvert fylke er hver adressegruppe et stratum

Vi utnytter at vi faktisk kjenner antall DSF familier i hvert fylke for alle fire adressegrupper, se tabell 3.1. for hele populasjonen, alle husholdninger i utvalget for de samme strata og antall DSF familier i hver husholdning i utvalget. Vektene beregnes nå ved brøken

$$\bullet \quad vekt = \frac{M_{r,s}}{f \cdot m_{r,s}^{hush}}$$

der

- $M_{r,s}$ - er antall DSF familier i alt i Norge
- f - er antall DSF familier i husholdningen
- $m_{r,s}^{hush}$ - er antall husholdninger i utvalget

Disse vektene gir det riktige antall DSF familier i hver adressegruppe for hvert fylke.

Disse vektene gir i tillegg til å produsere tabellene og sammenlikne med resultatene under det første trinnet, grunnlaget for å beregne følgende fordelinger i hvert fylke, eventuelt også for adressegruppene innen hvert fylke:

- Antallet Besys familier fordelt på fem grupper, se avsnitt 3.6, når vi bruker antallet Besys familier av de fem typene som husholdningsegenskap, se også tabell 3.4 for detaljer.
- Antallet menn og kvinner i åtte aldersgrupper, se avsnitt 3.6, når vi bruker antallet menn eller kvinner i de åtte aldersgruppene som husholdningsegenskap, se også tabellene 3.5 og 3.6 for detaljer.

Disse tabellene kan da sammenliknes med fasiten ifølge populasjonen av bosatte per 3. november i 2001 for å se hvor mye vi med disse vektene, designvektene, bommer før vi kalibrerer vekk disse avvikene i de endelige vektene som skal brukes i produksjonen av endelige tall.

Kalibrere vektene innen hvert stratum bestemt av fylke og adressegruppe

Vi skal nå bruke vektene i foregående avsnitt og kalibrere disse mot fordelingen av Besys familier i fem grupper og aldersfordelingen for menn og kvinner i åtte aldersgrupper, se avsnitt 3.6, innen fylke og om adressen er unik eller ikke. Vi har altså slått sammen to og to adressegrupper. Dette har vi gjort både fordi frafallet er omtrent det samme for stedfortreder og ikke stedfortreder for henholdsvis unike og ikke unike adresser. Dessuten manglet det også data i utvalget for noen grupper i populasjonen når vi delte inn etter stedfortreder/ikke stedfortreder i tillegg. Altså må vi for hver av de 38 kombinasjonene av fylke og adressegruppe beregne i populasjonen:

- Antallet Besys familier som er enperson familier, ektepar uten barn, ektepar med barn, aleneforelder (med barn) og samboere med barn
- Antallet menn og kvinner i hver av de åtte aldersgruppene

Videre må vi for hver husholdning i utvalget kjenne antallet av hver av de fem Besys familietyper og de åtte aldersgruppene for menn og kvinner hver for seg. For hvert stratum kjøres da et kalibreringsprogram som justerer vektene fra trinn 2 slik at

- de nye vektene gir nå at antallet Besys familier er korrekt for alle fem typer
- de nye vektene gir nå at antallet menn og kvinner i de åtte aldersgruppene er korrekt

Vi kontrollerer for ordens skyld at de to ballpunktene holder og kjører ut de endelige tabellene slik at vi kan se hvilke endringer som har skjedd fra trinn 2.

I tabell 3.4 til 3.6 har vi vist avviket mellom populasjonstallene for Besys familier, aldersgruppene for kvinner og menn og de tilsvarende estimerte tallene fra utvalget basert på de etterstratifiserte vektene.

Tabell 3.4 Prosentvis relativt avvik mellom populasjonstall og estimerte utvalgstall for Besys familiegrupper på grunnlag av etterstratifiserte vekt

Områder	Adresstyper	Totalt	1-person familie	Ektepar uten barn	Ektepar med barn	Enslige foreldre	Samboerpar med barn
Hele landet		0,1	4,2	-6,1	-6,7	4,6	3,9
Østfold	Unike	0,0	6,6	-2,1	-3,1	-14,9	9,2
	Ikke unike	0,0	-1,8	4,1	6,3	-8,1	23,4
Akershus	Unike	0,2	4,6	-6,9	-1,4	3,0	3,0
	Ikke unike	1,1	3,6	-2,2	-15,8	21,6	-11,3
Oslo	Unike	-0,2	-1,9	-4,0	-1,8	10,9	20,3
	Ikke unike	0,4	5,9	-17,3	-8,3	-9,2	-12,3
Hedmark	Unike	-0,8	10,1	-20,3	-3,5	-10,5	21,2
	Ikke unike	0,9	13,4	-5,9	-20,6	-32,3	-11,8
Oppland	Unike	1,0	14,1	-6,6	-13,9	7,9	-13,0
	Ikke unike	-0,1	-2,2	13,5	-17,2	22,5	8,4
Buskerud	Unike	-0,5	1,6	4,2	-7,1	-8,2	18,0
	Ikke unike	1,0	2,3	8,4	-9,3	6,6	-23,1
Vestfold	Unike	0,1	4,9	-6,8	-10,5	22,9	5,2
	Ikke unike	-0,5	7,8	-25,1	-15,5	-19,6	37,1
Telemark	Unike	0,4	10,7	-12,3	-5,5	6,4	-6,0
	Ikke unike	0,5	-1,2	27,3	-21,3	9,0	10,5
Aust-Agder	Unike	-0,1	13,3	4,6	-19,7	-3,6	10,1
	Ikke unike	0,2	-4,4	0,1	11,7	9,2	-5,2
Vest-Agder	Unike	-0,6	0,7	-13,0	-3,8	25,9	19,7
	Ikke unike	1,3	-1,2	38,4	-6,9	-8,4	-71,7
Rogaland	Unike	1,2	2,8	1,0	-0,1	12,3	-27,0
	Ikke unike	-0,4	2,1	8,7	-24,8	12,3	25,6
Hordaland	Unike	-0,2	-1,4	-9,2	0,3	15,6	9,5
	Ikke unike	-0,9	1,2	-13,4	-15,5	10,5	27,1
Sogn og Fjordane	Unike	-0,5	15,4	-17,2	-20,4	22,5	6,5
	Ikke unike	-0,9	0,7	-10,4	-1,4	-7,7	20,6
Møre og Romsdal	Unike	0,8	7,1	1,2	-4,6	2,5	-10,5
	Ikke unike	0,0	-1,1	-25,8	14,3	9,9	8,1
Sør-Trøndelag	Unike	0,8	2,4	7,9	-5,9	6,9	-7,4
	Ikke unike	1,0	-0,6	0,6	-6,4	30,2	-16,9
Nord-Trøndelag	Unike	-1,6	1,4	3,9	-23,3	11,0	22,2
	Ikke unike	-0,8	12,6	-57,1	9,4	-29,3	35,3
Nordland	Unike	-0,2	2,4	-6,7	-10,0	17,2	6,5
	Ikke unike	-2,3	5,2	-30,7	-33,1	2,1	53,2
Troms	Unike	0,8	12,7	-8,1	-9,0	-1,6	-10,7
	Ikke unike	0,2	11,4	-42,0	16,0	-40,6	-2,1
Finnmark	Unike	-1,8	-3,3	-21,2	-10,1	18,7	32,5
	Ikke unike	3,2	12,3	55,5	-38,4	-5,9	-55,5
Minimum avvik	Fylkene	-2,3	-4,4	-57,1	-38,4	-40,6	-71,7
Maksimum avvik	Fylkene	3,2	15,4	55,5	16,0	30,2	53,2

I tabell 3.4 ser vi at det er betydelige forskjeller mellom det estimerte antallet Besys familier i de 38 gruppene sammenliknet med de faktiske tallene for enperson familier, ektepar uten barn, ektepar med barn, aleneforeldre og samboere med felles barn.

Vi ser at det totale antallet Besys familier basert på utvalget er litt for lavt, 0,1 prosent. Det estimerte antallet ektepar uten barn og ektepar med barn er for høyt sammenliknet med populasjonen, hhv 6,1 og 6,7 prosent, mens for enperson familier, aleneforeldre og samboere med felles barn er det motsatt, utvalget gir for lave tall.

For alle fem typer er det betydelige variasjoner mellom fylker og adressetyper. For enperson familiene er avvikene minst. Unike adresser i Sogn og Fjordane gir det laveste antallet enperson familier sammenliknet med populasjonen, 15,4 prosent for lavt antall. Ikke unike adresser i Aust-Agder ligger i den andre enden av skalaen - 4,4 prosent for høyt antall enperson familier.

Ektepar uten barn og samboerpar med felles barn har de største avvikene mellom utvalget og populasjonen. Altfør høye tall i utvalget har vi for **ikke unike adresser** i Nord-Trøndelag(57,1 prosent for mange ektepar uten barn) og for **ikke unike adresser** i Vest-Agder(71,7 prosent for mange samboerpar med felles barn). Det motsatte avviket finner vi henholdsvis for **ikke unike adresser** i Finnmark(55,5% for få ektepar uten barn) og for **ikke unike adresser** i Nordland(53,2 prosent for få samboerpar med felles barn).

Tabell 3.5 Prosentvis relativt avvik mellom populasjonstall og estimerte utvalgstall for aldersgrupper for kvinner på grunnlag av etterstratifiserte vekter

Områder	Adresser	I alt	-20	20-29	30-39	40-49	50-59	60-69	70-79	80+
Hele landet		-2,3	0,3	1,7	-0,9	-3,0	-9,2	0,9	-9,3	-5,3
Østfold	Unike	-3,3	0,0	-10,5	2,6	-12,6	-1,9	9,2	-20,6	5,5
	Ikke unike	1,9	7,8	13,5	20,0	-11,1	-5,9	-19,4	-3,2	-0,8
Akershus	Unike	-2,8	-0,3	3,5	-5,5	-4,2	-5,7	14,6	-22,3	-17,9
	Ikke unike	-0,5	9,2	13,5	-10,1	-0,2	-2,5	-8,0	-14,9	-7,4
Oslo	Unike	0,7	9,4	6,2	-4,6	6,6	-31,0	30,6	-4,2	-19,3
	Ikke unike	-5,5	-4,5	8,0	-9,4	-3,5	-2,8	-24,8	-11,2	-15,2
Hedmark	Unike	-6,6	-8,8	-14,3	5,8	4,2	-5,8	-27,1	-16,8	11,0
	Ikke unike	-12,3	-13,3	-17,2	2,3	-38,6	-27,9	6,2	-15,0	11,7
Oppland	Unike	-6,4	-21,9	5,5	-26,4	6,2	5,2	0,1	-12,6	25,1
	Ikke unike	3,9	2,5	19,5	14,3	6,5	-0,6	-25,2	2,6	1,7
Buskerud	Unike	0,4	2,7	11,9	7,3	-9,9	-7,3	13,0	-3,8	-22,1
	Ikke unike	-0,3	19,3	1,8	-12,4	14,3	-19,7	-1,5	10,2	-28,5
Vestfold	Unike	0,4	5,4	-7,8	20,0	-0,1	-18,3	-4,6	1,9	-8,3
	Ikke unike	-3,3	-0,7	-10,2	1,9	0,7	-5,5	-29,8	-11,4	18,6
Telemark	Unike	-5,1	-6,3	1,5	21,7	-18,0	6,1	-39,0	-22,6	7,1
	Ikke unike	0,7	3,8	10,6	25,9	-26,0	21,6	61,9	-28,0	-50,4
Aust-Agder	Unike	-12,9	-24,4	-12,9	-6,8	-18,2	-25,7	9,9	9,0	14,9
	Ikke unike	1,1	24,4	-9,7	23,2	-4,2	18,0	30,6	-63,0	-36,2
Vest-Agder	Unike	2,3	7,3	-4,2	-10,6	33,0	-19,2	-1,5	8,0	-14,1
	Ikke unike	-9,7	-19,6	10,4	24,0	-47,8	-16,9	40,5	-47,2	-19,6
Rogaland	Unike	1,2	-1,5	-1,7	-1,4	0,8	-1,0	26,3	5,9	-5,4
	Ikke unike	-8,0	-8,9	0,6	-22,6	-7,5	-20,0	20,2	9,0	-20,5
Hordaland	Unike	0,5	4,9	15,9	5,1	3,9	-16,2	-8,5	-10,0	-14,2
	Ikke unike	1,3	8,5	4,1	0,7	-8,5	-15,0	18,8	-16,6	11,8
Sogn og Fjordane	Unike	-5,9	-11,7	2,9	-15,4	-11,5	1,5	4,9	-6,1	8,2
	Ikke unike	-1,0	17,3	18,3	7,9	-37,5	-22,2	23,7	-56,6	5,5
Møre og Romsdal	Unike	2,1	3,3	-3,6	1,1	-4,2	-0,1	8,5	14,7	5,6
	Ikke unike	3,4	17,2	11,4	-2,3	10,7	-8,5	-5,1	0,9	-18,1
Sør-Trøndelag	Unike	-3,4	-0,1	-10,3	-2,7	-11,4	-2,3	35,3	-34,5	-16,2
	Ikke unike	0,1	1,5	-4,2	9,0	-6,6	-17,8	-6,1	19,8	5,5
Nord-Trøndelag	Unike	-7,3	-1,2	-16,2	-4,6	-19,9	-19,1	2,7	20,5	-23,5
	Ikke unike	-0,9	2,0	17,5	-3,0	13,2	-11,0	-61,5	3,9	11,1
Nordland	Unike	-2,2	0,3	2,0	-9,3	14,3	-12,4	8,6	-29,1	-0,2
	Ikke unike	-5,8	7,2	-19,0	12,8	-8,2	-41,9	-23,9	-8,1	14,5
Troms	Unike	-2,7	3,5	-1,6	5,3	1,6	-22,8	3,8	-31,2	12,4
	Ikke unike	-10,8	-22,1	-3,0	6,6	-11,0	-20,6	-23,4	17,5	-31,2
Finnmark	Unike	2,2	28,2	-8,5	9,8	-5,7	-20,5	-26,6	-5,6	-4,6
	Ikke unike	-14,3	-49,2	-34,7	-18,3	-10,8	36,9	15,0	-9,0	48,8
Minimum avvik	Fylkene	-14,3	-41,2	-34,7	-26,4	-47,8	-41,9	-61,5	-63,0	-36,2
Maksimum avvik	Fylkene	3,4	28,2	19,5	25,9	14,3	36,9	61,4	20,5	48,8

Utslagene er mindre for både ektepar med barn og aleneforeldre. Ektepar med barn ligger mellom 16 prosent for få(ikke unike adresser i Troms) og 38,4 prosent for mange(ikke unike adresser i Finnmark) i utvalget. For aleneforeldre varierer det fra mellom 30 prosent for få(ikke unike adresser i Sør-Trøndelag) til 40,6 % for mange (ikke unike adresser i Troms) i utvalget.

Ytterpunktene med hensyn til det totale antallet Besys familier er ikke unike adresser i Finnmark(3,2 prosent for få) og ikke unike adresser i Nordland(2,3 prosent for mange).

Det er 2,3 prosent for mange kvinner i utvalget sammenliknet med populasjonen og særlig i aldersgruppene 50-59, 70-79 og 80 og eldre er overrepresentasjonen betydelig med litt over 9 % i de to første aldersgruppene og drøyt fem prosent blant de eldre.

Dersom vi ser på estimatet av totalt antall kvinner i fylkene etter unik og ikke unik adresse er det ikke unike adresser i Finnmark som har den største overrepresentasjonen i utvalget, mens det er de ikke unike adressene i Møre og Romsdal som havner i andre enden med 3,4 prosent for få kvinner.

Tabell 3.6 Prosentvis relativt avvik mellom populasjonstall og estimerte utvalgstall for aldersgrupper for menn på grunnlag av etterstratifiserte vekter

Områder	Adresser	I alt	-20	20-29	30-39	40-49	50-59	60-69	70-79	80+
Hele landet		-1,3	-1,3	1,4	-1,3	0,6	-3,9	-2,2	-12,4	2,4
Østfold	Unike	0,0	2,6	12,0	-0,9	-11,7	1,5	4,4	-7,5	-10,1
	Ikke unike	2,6	16,3	14,2	11,8	-2,8	16,7	-10,5	-41,1	-66,1
Akershus	Unike	0,6	-1,5	14,3	-1,1	6,5	-1,7	-6,2	-23,0	33,8
	Ikke unike	-2,0	-9,7	13,2	-14,1	-0,9	-1,8	-9,2	8,5	46,7
Oslo	Unike	1,7	6,9	-23,6	-5,6	5,6	-7,9	21,0	9,1	50,4
	Ikke unike	-1,1	-1,3	4,4	1,1	7,2	-1,7	-12,6	-30,0	-8,9
Hedmark	Unike	-2,4	7,2	-8,7	-15,4	18,1	7,7	-21,8	-31,4	-7,9
	Ikke unike	-4,4	-22,1	-8,9	23,9	-28,5	8,1	13,0	8,0	-7,3
Oppland	Unike	-4,6	-7,2	10,9	-21,6	2,2	-4,7	3,2	-13,0	1,0
	Ikke unike	-1,5	19,3	-5,1	-4,2	-4,8	4,7	-44,2	-31,9	29,6
Buskerud	Unike	-2,5	-2,6	-8,5	-12,0	2,2	5,5	2,7	4,8	-26,4
	Ikke unike	-2,1	-19,2	-6,2	9,3	7,0	-9,1	-1,1	20,1	13,9
Vestfold	Unike	-4,4	-0,3	-8,9	3,2	0,1	-23,3	0,3	-10,8	6,1
	Ikke unike	-8,2	1,2	-34,9	6,2	13,0	-45,0	3,3	-10,5	-1,2
Telemark	Unike	-2,8	17,5	-28,0	7,8	-11,8	-14,6	27,6	-46,6	-11,4
	Ikke unike	-2,1	-16,7	-21,5	22,2	6,5	33,4	37,7	-27,0	-84,5
Aust-Agder	Unike	-5,2	-13,7	10,7	-6,7	-18,2	-16,9	11,6	26,4	22,0
	Ikke unike	3,8	-4,2	8,1	48,7	-4,5	3,3	-51,2	9,4	-15,4
Vest-Agder	Unike	0,8	4,6	8,7	2,1	0,3	-2,9	-18,2	-23,1	39,0
	Ikke unike	11,4	7,4	-0,8	23,3	38,1	-0,3	39,0	-7,8	-28,5
Rogaland	Unike	-1,8	-5,2	4,0	-21,4	10,4	-4,2	15,1	13,0	-6,9
	Ikke unike	-3,6	-13,6	11,0	-5,9	-0,8	-25,7	17,0	16,1	0,7
Hordaland	Unike	2,2	10,6	-0,6	10,0	-1,1	-5,3	-7,9	-31,3	38,2
	Ikke unike	-4,8	-7,0	7,5	3,8	-24,4	-6,6	2,1	-10,9	-22,7
Sogn og Fjordane	Unike	-11,2	-22,5	-3,0	3,8	-13,7	-14,6	-5,5	8,1	-32,2
	Ikke unike	-3,7	-2,2	-1,4	3,1	-18,4	28,9	-19,3	-43,0	-2,0
Møre og Romsdal	Unike	-1,3	1,1	3,7	-22,2	7,8	-0,2	-21,3	11,4	21,3
	Ikke unike	2,7	16,9	4,1	-7,8	-4,1	1,0	5,6	0,9	-20,1
Sør-Trøndelag	Unike	0,5	-7,3	9,2	-4,4	3,1	2,5	23,9	-21,0	31,4
	Ikke unike	-0,2	5,2	-0,7	3,0	-7,9	1,1	-36,0	0,3	38,2
Nord-Trøndelag	Unike	-2,1	-11,3	21,8	-12,6	1,5	-11,5	13,7	9,7	0,1
	Ikke unike	-4,0	12,1	-1,7	25,4	-31,8	-12,7	-51,2	-18,7	13,7
Nordland	Unike	-0,7	2,3	1,0	-0,5	-6,0	13,0	-5,2	-30,0	2,0
	Ikke unike	-9,3	-10,3	-17,4	9,8	-1,8	-22,6	-30,8	-12,5	9,6
Troms	Unike	-1,0	5,8	-9,8	12,1	12,9	-19,7	-10,5	-14,6	-22,2
	Ikke unike	8,9	30,3	-4,2	11,7	13,9	-0,7	-5,1	-38,1	8,9
Finnmark	Unike	-2,2	14,0	-14,0	8,0	-1,3	-16,8	-1,4	-27,0	-78,5
	Ikke unike	-1,5	-31,4	21,7	4,5	20,6	-24,4	11,1	20,5	6,2
Minimum avvik	Fylkene	-11,7	-31,4	-34,9	-22,2	-31,8	-45,0	-51,2	-46,6	-84,5
Maksimum avvik	Fylkene	11,4	30,3	21,8	48,7	38,1	33,4	39,0	26,4	50,4

Skjevhetene i utvalget med hensyn til aldersfordelingen er størst for de minste fylkene. Utslaget er størst for Finnmark som hele seks ganger topper statistikken med enten for mange eller for få kvinner i årsklassene. Oppland, Telemark, Aust-Agder, Vest-Agder, Sør-Trøndelag, Nord-Trøndelag og Nordland topper ellers i en eller to årsklasser med for høyt eller for lavt antall kvinner i utvalget sammenliknet med populasjonen.

Utvalget treffer bedre for menn enn kvinner for hele landet med unntak av aldersgruppa 70-79 år der utvalget gir 12,4% for mange menn sammenliknet med populasjonen. For menn totalt får vi 1,3 prosent for høyt tall basert på utvalget.

Det er i henholdsvis Vest-Agder(for lavt tall for ikke unike adresser) og i Sogn og Fjordane(for høyt tall for unike adresser) at avvikene er størst. Jevnt over er også avvikene i fylkene lavere for menn enn for kvinner.

Dersom vi ser på avvikene mellom utvalget og populasjonen for de enkelte årsklassene er det Vestfold, Telemark, Aust-Agder, Vest-Agder og Nord-Trøndelag som har flest topper på lista med størst avvik mellom utvalget og populasjonen. Oppland, Troms og Finnmark har også en topp på avvikerlista hver.

Det små forskjeller mellom kvinner og menn med hensyn til hvor store de største avvikene er og det er i de minste fylkene at avviket jevnt over er størst. Verken Akershus, Oslo, Rogaland eller Hordaland topper noen av avvikerlistene.

I de to neste tabellene, 3.7 og 3.8, har vi sett på virkningen av kalibreringen målt ved minimum, maksimum, median, gjennomsnittet og standardavviket for den etterstratifiserte vekten, den kalibrerte vekten og forholdet mellom de to vektene.

Siden andre husholdninger som institusjoner finnes under ikke unike adresser er det opplagt at vektene kryper ned mot 1 både for etterstratifiserte og kalibrerte vekter. Vi ser videre at det er i svært mange strata at vi reduserer de kalibrerte vektene til 1/3 av de etterstratifiserte vektene, mens det kun er i Oslo(unike adresser) og Telemark(ikke unike adresser) at vi øker vektene med den maksimale verdien, dvs. at den kalibrerte vekten er tre ganger større enn den etterstratifiserte.

Medianen er størst i Oslo, unike adresser, både for etterstratifiserte og kalibrerte vekter. De minste medianene finnes i Finnmark for både etterstratifiserte (unike adresser) og kalibrerte(ikke unike adresser) vekter. Medianen til forholdet mellom kalibrert og etterstratifiserte vekter varierer mellom 0,83(ikke unike adresser i Telemark og Finnmark) og 1,04(ikke unike adresser i Møre og Romsdal).

Videre i tabell 3.8 er vektene og forholdet beskrevet ved gjennomsnittet og standardavviket. Gjennomsnittet av vektene varierer fra 69,6(67,4) for unike adresser i Finnmark til 158,9(158,4) for unike adresser i Oslo. Etterstratifiserte vekter er nevnt først med kalibrerte vekter i parentes. Standardavviket varierer fra 17,8(26,5) til 44,7(93,5) med kalibrerte vekter i parentes. Det er for unike adresser i Finnmark at vi oppnår minimum, mens størst verdi for standardavviket har vi henholdsvis i Telemark(etterstratifiserte vekter) og Vest-Agder(kalibrerte vekter), i begge tilfeller gjelder det ikke unike adresser.

Tabell 3.7 Minimum, maksimum og medianen for designvektene, kalibrerte vekter og forholdet mellom disse innen hvert stratum

Områder	Adresser	Minimum			Maksimum			Median		
		Vekt	Kalvekt	forhold	Vekt	Kalvekt	forhold	Vekt	Kalvekt	forhold
Hele landet	Unike	2,1	2,4	0,33	196,8	590,3	3,00	136,4	119,9	0,98
	Ikke unike	1,2	1,1	0,33	197,1	493,3	3,00	147,6	134,6	0,99
Østfold	Unike	45,5	38,8	0,39	141,5	249,1	1,83	136,4	129,5	1,01
	Ikke unike	1,9	1,6	0,33	157,4	241,7	1,54	138,7	122,9	0,99
Akershus	Unike	51,8	49,0	0,57	160,5	353,7	2,28	155,3	144,0	0,97
	Ikke unike	2,8	3,8	0,53	185,8	386,6	2,08	153,2	139,7	0,96
Oslo	Unike	32,8	28,9	0,33	196,8	590,3	3,00	160,0	148,7	0,97
	Ikke unike	1,2	1,1	0,47	197,1	297,7	1,77	168,6	153,2	1,00
Hedmark	Unike	28,0	26,1	0,33	140,0	275,0	1,96	140,0	117,9	0,97
	Ikke unike	2,0	2,4	0,33	141,5	294,5	2,08	130,5	111,9	0,94
Oppland	Unike	3,8	4,9	0,61	140,1	221,8	1,58	140,1	127,0	0,98
	Ikke unike	1,3	1,3	0,33	142,1	278,8	1,96	142,1	119,9	1,00
Buskerud	Unike	43,0	47,9	0,56	141,3	220,6	1,56	141,3	121,3	1,00
	Ikke unike	2,0	1,7	0,50	154,4	242,7	1,75	141,9	129,7	1,01
Vestfold	Unike	6,4	6,0	0,61	139,7	293,5	2,10	139,7	123,8	0,99
	Ikke unike	2,1	2,5	0,33	167,7	286,8	1,79	160,5	139,0	0,97
Telemark	Unike	43,1	14,4	0,33	166,3	287,4	2,22	129,2	115,6	0,97
	Ikke unike	2,2	1,7	0,33	175,7	493,3	3,00	132,1	111,8	0,83
Aust-Agder	Unike	25,7	25,3	0,36	102,6	270,8	2,64	102,6	89,4	0,98
	Ikke unike	1,4	1,1	0,33	93,6	255,1	2,81	93,6	63,5	0,83
Vest-Agder	Unike	39,5	37,1	0,39	140,3	292,3	2,13	140,3	116,7	0,91
	Ikke unike	1,8	1,2	0,33	152,3	388,6	2,57	150,9	108,3	0,88
Rogaland	Unike	31,6	29,6	0,51	149,2	233,4	1,56	126,6	124,5	0,99
	Ikke unike	1,5	1,6	0,54	168,9	273,6	1,62	155,9	137,4	1,02
Hordaland	Unike	38,5	44,8	0,56	159,2	312,7	1,96	153,8	138,5	0,99
	Ikke unike	1,3	1,5	0,49	169,3	252,5	1,61	167,6	150,8	1,02
Sogn og Fjordane	Unike	3,3	3,9	0,61	103,4	175,3	1,70	103,4	87,0	0,97
	Ikke unike	2,3	2,4	0,33	106,5	225,1	2,14	106,5	80,2	0,91
Møre og Romsdal	Unike	45,5	39,0	0,55	136,4	214,7	1,57	136,4	128,4	1,00
	Ikke unike	2,7	2,6	0,36	151,5	254,8	1,74	149,4	128,7	1,04
Sør-Trøndelag	Unike	46,8	45,1	0,46	140,6	263,3	1,88	140,3	120,3	0,97
	Ikke unike	1,2	1,3	0,41	138,2	252,5	1,83	136,4	119,8	0,98
Nord-Trøndelag	Unike	23,8	30,0	0,35	126,9	271,5	2,28	119,2	96,8	0,98
	Ikke unike	3,2	3,9	0,33	119,6	214,0	2,00	119,6	99,6	1,02
Nordland	Unike	5,5	5,4	0,54	138,3	249,4	1,80	138,3	113,8	1,00
	Ikke unike	2,4	2,9	0,44	152,9	241,3	2,77	147,6	126,9	0,96
Troms	Unike	2,1	2,4	0,43	133,3	251,6	1,89	133,3	109,7	0,98
	Ikke unike	4,0	3,8	0,33	165,2	324,4	1,96	140,9	128,7	1,02
Finnmark	Unike	2,7	2,6	0,33	84,5	171,9	2,23	77,1	66,0	0,96
	Ikke unike	13,3	15,5	0,33	92,9	231,0	2,49	90,5	60,8	0,83
Minimum avvik	Fylkene	1,2	1,1	0,33	84,5	171,9	1,56	77,1	60,8	0,83
Maksimum avvik	Fylkene	51,8	49,0	0,61	92,9	590,3	3,00	168,6	153,2	1,04

Standardavviket for kalibrerte vekter øker markant sammenliknet med for etterstratifiserte vekter, særlig for ikke unike adresser i Vest-Agder og Finnmark. Den kraftige økningen av standardavviket sammenliknet med svært små endringer for gjennomsnittet viser at det er i halene av fordelingene til vektene at det skjer betydelige endringer.

Gjennomsnittsverdien for forholdet varierer fra 0,998(ikke unike adresser i Aust-Agder) til 1,013(ikke unike adresser i Telemark), mens standardavviket for forholdet varierer fra 0,149(ikke unike adresser i Oslo) til 0,585(ikke unike adresser i Vest-Agder). Dette underbygger at endringene på vektene skjer i halene av fordelingene i hver gruppe.

Tabell 3.8. Gjennomsnittet og standardavviket for designvektene, kalibrerte vekter og forholdet mellom disse innen hvert stratum

Områder	Adresser	Gjennomsnittet			Standardavviket			Standard avviket
		Vekt	Kalvekt	forhold	Vekt	Kalvekt	Forhold	Økning ved kalibrering
Hele landet	Unike	120,5	119,8	1,003	34,0	43,5	0,242	28,15
	Ikke unike	131,6	130,8	1,003	43,9	55,3	0,288	26,10
Østfold	Unike	123,6	123,1	1,003	28,0	34,2	0,178	22,00
	Ikke unike	124,7	124,2	1,005	37,5	49,2	0,263	31,22
Akershus	Unike	138,1	137,5	1,003	33,9	43,3	0,205	27,66
	Ikke unike	141,6	140,7	1,008	43,2	55,2	0,255	27,96
Oslo	Unike	158,9	158,4	1,006	44,6	65,7	0,288	47,48
	Ikke unike	148,2	147,7	1,001	43,1	47,3	0,149	9,71
Hedmark	Unike	123,7	121,4	1,004	29,7	40,6	0,285	36,60
	Ikke unike	115,0	113,7	1,003	39,1	53,9	0,305	37,93
Oppland	Unike	121,0	121,0	1,004	30,9	39,3	0,210	27,10
	Ikke unike	121,7	119,9	1,001	37,0	51,4	0,293	38,85
Buskerud	Unike	124,2	122,8	1,005	29,0	33,3	0,189	14,69
	Ikke unike	126,1	127,5	1,002	38,0	51,6	0,247	35,88
Vestfold	Unike	122,2	121,2	1,005	29,1	34,5	0,203	18,45
	Ikke unike	139,9	137,8	1,003	42,0	54,4	0,273	29,64
Telemark	Unike	120,3	119,3	1,000	28,1	50,0	0,352	78,02
	Ikke unike	129,4	128,3	1,013	44,7	76,4	0,511	70,85
Aust-Agder	Unike	91,6	90,3	1,004	20,7	30,5	0,299	47,37
	Ikke unike	72,8	73,5	0,998	27,9	50,7	0,509	82,08
Vest-Agder	Unike	126,0	124,7	1,000	25,4	44,7	0,312	75,93
	Ikke unike	132,2	133,6	1,001	39,7	93,5	0,585	135,77
Rogaland	Unike	117,2	118,4	1,001	27,3	38,1	0,190	39,48
	Ikke unike	142,9	142,1	1,003	41,1	50,3	0,226	22,57
Hordaland	Unike	139,8	139,0	1,003	31,2	40,1	0,200	28,40
	Ikke unike	147,2	145,0	1,002	42,5	46,6	0,191	9,84
Sogn og Fjordane	Unike	90,8	89,2	1,001	22,5	28,3	0,232	26,06
	Ikke unike	91,3	89,8	1,004	26,2	49,0	0,467	86,96
Møre og Romsdal	Unike	119,9	120,4	1,002	29,5	40,1	0,207	35,81
	Ikke unike	126,9	126,8	1,004	43,5	52,4	0,237	20,59
Sør-Trøndelag	Unike	119,9	120,5	1,002	32,0	48,3	0,261	51,11
	Ikke unike	116,1	116,6	1,001	35,0	46,5	0,221	33,16
Nord-Trøndelag	Unike	105,4	102,7	1,001	26,4	36,2	0,291	37,06
	Ikke unike	103,6	101,1	1,010	28,9	43,6	0,373	51,09
Nordland	Unike	117,7	116,5	1,003	31,3	38,0	0,214	21,38
	Ikke unike	128,9	124,6	1,004	39,4	45,8	0,342	16,25
Troms	Unike	111,0	110,5	1,004	30,9	40,4	0,256	30,79
	Ikke unike	136,7	136,5	1,006	36,6	61,1	0,365	66,94
Finnmark	Unike	68,6	67,0	1,004	17,8	26,5	0,353	48,75
	Ikke unike	76,0	76,5	1,003	23,0	53,6	0,575	133,40
Minimum avvik	Fylkene	68,6	67,0	0,998	17,8	26,5	0,149	9,71
Maksimum avvik	Fylkene	158,9	158,4	1,013	44,7	93,5	0,585	135,77

Ser vi på hele utvalget og vektene under ett, men fordelt på unike og ikke unike adresser, kan vi oppsummere bildet i tabell 3.9. Vi ser at for unike adresser gjelder når vi kalibrerer vil 5% av vektene øke med over 42,4 % (forholdet er større enn 1,424) og 5% av vektene reduseres med mer enn 34,2% (forholdet er mindre enn 0,658). For ikke unike adresser er de tilsvarende tallene henholdsvis 47,3% og 38,9%. Vi kan altså tilnærmet si at kalibreringen førte til at rundt 10% av vektene endret seg med mer enn 2/5 av den opprinnelige verdien. Et annet mål på endringen er spredningen målt ved standardavviket. Vi ser at standardavviket til vektene etter kalibreringen har økt fra 33,98 til 43,54 for unike adresser, mens den tilsvarende økningen for ikke unike adresser er fra 43,85 til 55,30. Vi ser også at mens gjennomsnittet av vektene har endret seg lite etter kalibreringen har både median og kvartiler beveget seg betydelig etter kalibrering. Vi ser også at både median og kvartiler til vektene for

unike adresser ligger høyere enn for ikke unike adresser både før og etter kalibrering. Vi legger også merke til at medianen synker i verdi etter kalibrering for begge grupper.

Tabell 3.9. En sammenlikning av vektene før og etter kalibrering

Mål	Unike adresser			Ikke unike adresser		
	Før kalibrering	Etter kalibrering	Forholdet	Før kalibrering	Etter kalibrering	Forholdet
Maksimum	196,78	590,33	3,000	197,09	493,29	3,000
99% fraktil	196,78	249,13	1,769	197,09	256,84	1,971
95% fraktil	159,96	190,80	1,424	169,28	210,45	1,473
90% fraktil	155,31	170,16	1,297	168,91	198,22	1,325
75% - øverste kvartil	140,31	147,14	1,119	168,65	170,05	1,133
50% - median	136,35	119,89	0,983	147,63	134,64	0,991
25% - nederste kvartil	91,96	86,24	0,850	92,89	90,98	0,833
10% fraktil	68,37	65,32	0,729	70,96	61,13	0,686
5% fraktil	63,30	55,85	0,658	53,27	42,16	0,611
1 % fraktil	41,15	38,77	0,510	3,23	3,08	0,333
Minimum	2,12	2,41	0,333	1,19	1,06	0,333
Gjennomsnittet	120,51	119,80	1,003	131,60	130,82	1,003
Standardavviket	33,98	43,54	0,242	43,85	55,30	0,288

4. Usikkerhet

I dette kapitlet skal vi først finne en enkel formel for å avgjøre hvilke tall som kan publiseres ut fra betingelser på nøyaktigheten. Generelt gjelder at det relative standardavviket er størst for små tall og avtar når tallene vokser, mens standardavviket øker i proporsjonalt med kvadratroten til tallet selv. Vi har valgt å følge praksis fra Folke- og bolig tellingen i 1990 og publiserte derfor tall med et relativt standardavvik mindre enn 20%. Vi satte av den grunn en minste grense for å publisere lik 5 000 husholdninger eller boliger.

I avsnitt 4.2 ser vi på en mer nøyaktig metode for å beregne usikkerheten til antallet med en bestemt egenskap for de etterstratifiserte vektene vi beregnet i avsnitt 3.2, se formel 3.9. Utrekningene bygger på at vi antar antallet husholdninger i hver stratum for utvalget er gitt. I avsnitt 4.3 argumenterer vi for at usikkerheten vi beregnet i avsnitt 4.2 også gjelder tilnærmet for de kalibrerte vektene. Til slutt i avsnitt 4.4 finner vi en enkel formel for å anslå usikkerheten til de estimerte husholdnings- og bolig tallene for fylkene og landet totalt.

4.1. En enkel formel for å bestemme hva som skal publiseres

La følgende to størrelser bety henholdsvis antallet husholdninger eller boliger i populasjonen og utvalget:

- N - antallet husholdninger eller boliger i populasjonen
- n - antallet husholdninger eller boliger i utvalget

La oss videre anta at vi skal beregne antallet husholdninger eller boliger med en bestemt egenskap, altså har vi at

$$(4.1) \quad Y_h = \begin{cases} 0 \\ 1 \end{cases}$$

der verdien 1 betyr at husholdningen eller boligen har egenskapen, mens 0 betyr at den mangler egenskapen. La oss videre innføre en betegnelse for sannsynligheten for at en husholdning har egenskapen ved

$$(4.2) \quad p = P(Y_h = 1) = E(Y_h)$$

som betyr at vi skiller ikke mellom husholdningene eller boligene, se neste avsnitt for en mer inngående diskusjon av beregningene av usikkerhet. Vi tenker oss nå at vi estimerer antallet med egenskapen ved enkel oppblåsing

$$(4.3) \quad \hat{T} = N\hat{p} = N\bar{Y}_s = N \frac{1}{n} \sum_{h \in utv} Y_h$$

Det relative standardavviket til det estimerte antallet med egenskapen er da

$$(4.4) \quad RSD(\hat{T}) = \frac{\sqrt{Var(\hat{T})}}{E(\hat{T})} = \frac{\sqrt{N^2 Var(\hat{p})}}{Np} = \frac{N \sqrt{\frac{p(1-p)}{n}}}{Np} = \sqrt{\frac{1-p}{np}}$$

Dersom vi setter et krav om at det relative standardavviket må være mindre eller lik en øvre grense, α , finner vi følgende kriterium:

$$(4.5) \quad RSD(\hat{T}) \leq \alpha \Leftrightarrow p \geq \frac{1}{n\alpha^2 + 1} \Leftrightarrow Np \geq \frac{N}{n\alpha^2 + 1} = \frac{\frac{N}{n}}{\alpha^2 + \frac{1}{n}}$$

Det betyr at vi kan lage følgende enkle regel for når et tall, \hat{T} , kan publiseres,

$$(4.6) \quad \hat{T} \geq \frac{\frac{N}{n}}{\alpha^2}$$

siden $\frac{1}{n}$ er liten og forventet verdi til \hat{T} er Np .

Dersom vi setter antall husholdninger lik 2,0 millioner og bruker at utvalget er på knappe 16 000 kan vi skrive kriteriet som publiser dersom

$$(4.7) \quad \hat{T} \geq \frac{125}{\alpha^2}$$

og vi kan lage en enkel tabell for forskjellige verdier til det relative standardavviket.

Tabell 4.1. Relativt standardavvik og antallet husholdninger eller boliger det svarer til

Det relative standardavviket - α %	Antall boliger og husholdninger
20%	3 125
10%	12 500
5%	50 000

Den nedre grensen for tall som skal publiseres ble satt til 5 000 .

Tall som ikke kan publiseres prikket, men vi behøver ikke ta hensyn til at tallet kan beregnes siden tallet ikke er følsomt, men kun for usikkert til at Statistisk sentralbyrå vil gå god for tallet. Alle tall som publiseres skal presenteres i hele tusen.

Vi kan også lage en formel for å estimere standardavviket direkte fra det publiserte antallet siden

$$(4.8) \quad SD(\hat{T}) = \sqrt{\frac{\hat{T}(N - \hat{T})}{n}}$$

- og i tabell 4.2 har vi gitt noen tall for sammenhengen mellom tallet som publiseres og usikkerheten, mens vi i figur 4.1 har vi vist hvordan standardavviket varierer med det publiserte antallet \hat{T} .

Tabell 4.2 Standardavviket som funksjon av størrelsen på det publiserte antallet og konfidensintervallene når vi antar at det er to million husholdninger i Norge og 15 000 husholdninger i utvalget

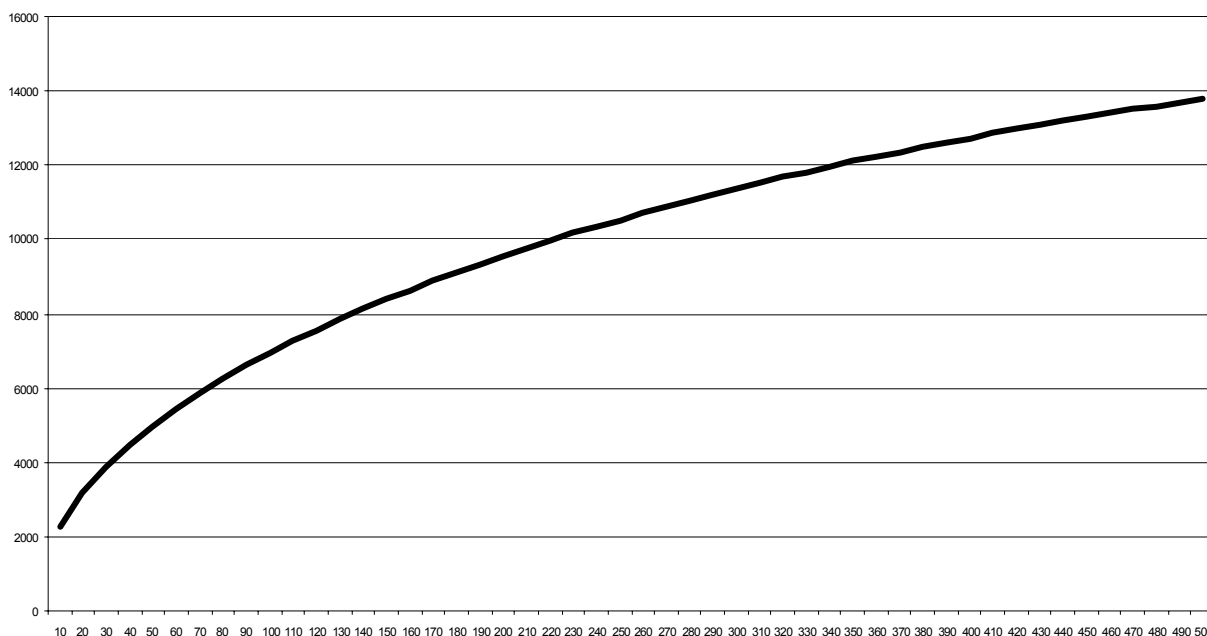
Publisert tall	Standardavviket	Relativt standardavvik	Usikkerhet = 2 standardavvik	95% Konfidensintervall
5 000	795	15,9%	1 590	[3 410,6 590]
10 000	1 123	11,2%	2 246	[7 754,12 246]
25 000	1 768	7,1%	3 537	[21 463,28 537]
50 000	2 486	5,0%	4 971	[45 029,54 971]
100 000	3 470	3,5%	6 939	[93 061,106 939]
250 000	5 265	2,1%	10 530	[239 470,260 530]
500 000	6 894	1,4%	13 787	[486 213,513 787]
1 000 000	8 960	0,9%	15 920	[984 080,1 015 920]

Vi kan da sammenfatte tabell 4.2 ved å påstå at vi har en høy sikkerhet for at dersom vi publiserer

- et foreløpig tall lik 5 000 er det endelige tallet mellom 3 500 og 6 500
- et foreløpig tall lik 10 000 er det endelige tallet mellom 7 750 og 12 250
- et foreløpig tall lik 50 000 er det endelige tallet mellom 45 000 og 55 000
- et foreløpig tall lik 100 000 er det endelige tallet mellom 93 000 og 107 000
- et foreløpig tall lik 250 000 er det endelige tallet mellom 239 500 og 260 500
- et foreløpig tall lik 500 000 er det endelige tallet mellom 486 000 og 514 000

Disse anslagene er for usikre når vi skal vurdere påliteligheten til anslaget på antallet husholdninger i alt i hvert fylke. I avsnitt 4.4 har vi vist hvordan dette kan beregnes med utgangspunkt i de inverse trekksannsynlighetene i formel (3.9).

Figur 4.1. Usikkerheten til de publiserte tallene for boliger eller husholdninger med bestemte egenskaper. Enheten på den horisontale akse er antall 1 000.



4.2. Beregne den faktiske usikkerheten ved bruk av de inverse trekksannsynlighetene

I dette avsnittet skal vi vise hvordan vi kan beregne usikkerheten til tall beregnet på grunnlag av de inverse trekksannsynlighetene, vi skal bruke vektene definert i (3.9). Dersom nå Y er en egenskap vi skal beregne det totale antallet for i et fylke eller landet i sin helhet for kan vi først beregne antallet i en adressegruppe s i fylket r ved å skrive

$$(4.9) \quad \hat{T}_{r,s} = \sum_f \sum_{h \in utv_{r,s,f}^{hush}} w_{h,r,s,f}^{hush} Y_{h,r,s,f} = \frac{M_{r,s}}{m_{r,s}^{hus}} \sum_f \frac{m_{r,s,f}^{hush}}{f} \hat{p}_{r,s,f}$$

der

$$(4.10) \quad \hat{p}_{r,s,f} = \frac{1}{m_{r,s,f}^{hush}} \sum_{h \in utv_{r,s,f}^{hush}} Y_{h,r,s,f}$$

er den estimerte andelen med egenskapen i husholdninger av adressegruppe s i fylke r som består av f antall DSF familier.

Vi finner da følgende formel for den (betingede) variansen til det estimerte antallet - betinget betyr her at vi tar antallet husholdninger i utvalget som gitt:

$$(4.11) \quad \begin{aligned} \text{Var}(\hat{T}_{r,s}) &= \text{Var}\left(\sum_f \sum_{h \in \text{utv}_{r,s,f}^{\text{hush}}} w_{h,r,s,f}^{\text{hush}} Y_{h,r,s,f}\right) = \text{Var}\left(\frac{M_{r,s}}{m_{r,s}^{\text{hus}}} \sum_f \frac{m_{r,s,f}^{\text{hush}}}{f} \hat{p}_{r,s,f}\right) \\ &= \left[\frac{M_{r,s}}{m_{r,s}^{\text{hus}}}\right]^2 \sum_f \left[\frac{m_{r,s,f}^{\text{hush}}}{f}\right]^2 \text{var}(\hat{p}_{r,s,f}) = \left[\frac{M_{r,s}}{m_{r,s}^{\text{hus}}}\right]^2 \sum_f m_{r,s,f}^{\text{hush}} \frac{p_{r,s,f}(1-p_{r,s,f})}{f^2} \end{aligned}$$

Vi kan beregne standardavviket for antallet ved å sette inn den estimerte andelen i formel (4.11):

$$(4.12) \quad SD(\hat{T}_{r,s}) = \frac{M_{r,s}}{m_{r,s}^{\text{hush}}} \sqrt{\sum_f m_{r,s,f}^{\text{hush}} \frac{\hat{p}_{r,s,f}(1-\hat{p}_{r,s,f})}{f^2}} = M_{r,s} \sqrt{\sum_f \left[\frac{m_{r,s,f}^{\text{hush}}}{m_{r,s}^{\text{hush}}}\right]^2 \frac{\hat{p}_{r,s,f}(1-\hat{p}_{r,s,f})}{f^2 m_{r,s,f}^{\text{hush}}}}$$

For et fylke r kan vi da beregne standardavviket til det publiserte tallet, $\hat{T}_r = \sum_s \hat{T}_{r,s}$, ved å summere de kvadrerte standardavvikene fra (4.12) og ta kvadratroten deretter.

$$(4.13) \quad SD(\hat{T}_r) = \sqrt{\sum_s [SD(\hat{T}_{r,s})]^2}$$

For hele landet gjør vi en tilsvarende operasjon på (4.13) og finner dermed standardavviket for det publiserte tallet, $\hat{T} = \sum_r \hat{T}_r$, lik

$$(4.14) \quad SD(\hat{T}) = \sqrt{\sum_r [SD(\hat{T}_r)]^2}$$

Opplegget bygger på følgende fire trinn:

- Vi beregner andelen husholdninger med egenskapen for et fylke r og adressegruppe s blant husholdningene som består av f familienumre, se formel (4.10)
- Vi beregner standardavviket for antallet med egenskapen i fylke r for adressegruppe s, se formel (4.12)
- Vi beregner standardavviket til antallet vi publiserer for fylket r, se formel (4.13)
- Vi beregner standardavviket til antallet vi publiserer for hele landet, se formel (4.14)

Siden de aller fleste husholdningene i utvalget består av enten 1 eller 2 DSF familier kan beregningene forenkles ytterligere ved kun å ta med de to leddene i formel (4.12). Dette gjelder kun når vi ser på tall for private husholdninger og ikke andre husholdninger siden disse husholdningene består av mange DSF familier.

4.3. Usikkerheten til de kalibrerte tallene vi publiserer

Kalibreringen som er beskrevet i avsnitt 3.5 skal forbedre treffsikkerheten til estimerte tallene for egenskaper til husholdningene og boligene i befolkningen. Når en egenskap har høy samvariasjon med de faktorene som vi kalibrerte med hensyn på er også forbedringen i forhold til de etterstratifiserte vektene som kun tar hensyn til trekkingen av utvalget og antallet DSF familier i alt størst. Det er derfor naturlig å tenke seg at usikkerheten til tall beregnet på grunnlag av de kalibrerte vektene reduseres sammenliknet med tall beregnet på grunnlag av vektene basert på de inverse trekksannsynlighetene. Dersom vi bruker standardavvikene som vi beregnet i forrige avsnitt for de kalibrerte vektene vil vi ha en øvre grense for standardavviket til de kalibrerte tallene. I hvilken grad denne

verdien ligger langt over er vanskelig å si på generell basis, men vi skal gi et argument for at forskjellen mellom standardavvikene ikke kan være veldig stor.

La oss skrive ut en sammenheng mellom de beregnede tallene med henholdsvis etterstratifiserte og kalibrerte vektorer ved å trekke fra og legge til som vist under.

$$(4.15) \quad \hat{T}_{r,s}^{kal} = \sum_f \sum_{h \in uv_{r,s,f}^{hush}} w_{h,r,s,f}^{kal} Y_{h,r,s,f} = \sum_f \sum_{h \in uv_{r,s,f}^{hush}} (w_{h,r,s,f}^{kal} - w_{h,r,s,f}) Y_{h,r,s,f} + \hat{T}_{r,s}$$

Vi kan derfor forklare forskjellen mellom det kalibrerte og det etterstratifiserte tallet ved å studere

$$\bullet \sum_f \sum_{h \in uv_{r,s,f}^{hush}} (w_{h,r,s,f}^{kal} - w_{h,r,s,f}) Y_{h,r,s,f}$$

mens for å kunne antyde om forskjellen mellom standardavviket til det kalibrerte og etterstratifiserte tallet må vi se på

$$(4.16) \quad Var\left[\sum_f \sum_{h \in uv_{r,s,f}^{hush}} (w_{h,r,s,f}^{kal} - w_{h,r,s,f}) Y_{h,r,s,f}\right] \leq \sum_f \sum_{h \in uv_{r,s,f}^{hush}} (w_{h,r,s,f}^{kal} - w_{h,r,s,f})^2 Var(Y_{h,r,s,f})$$

Vi konkluderte på bakgrunn av tabell 3.9 at for 90% av vektene er avviket mellom den kalibrerte og etterstratifiserte vektene mindre enn 2/5 av verdien til den etterstratifiserte vekten og avviket er alltid mindre enn 2 siden forholdet mellom den kalibrerte og etterstratifiserte vekten skal ligge mellom 1/3 og 3. Vi vil derfor i gjennomsnitt ha for 90% av vektene i (4.16) at

$$\bullet (w_{h,r,s,f}^{kal} - w_{h,r,s,f})^2 \leq \frac{4}{25} [w_{h,r,s,f}]^2$$

mens for de siste 10% av vektene har vi at

$$\bullet (w_{h,r,s,f}^{kal} - w_{h,r,s,f})^2 \leq 4 [w_{h,r,s,f}]^2$$

Setter vi inn disse skrankene så vil vi ha en forventet øvre grense for variansen gitt ved

$$(4.17) \quad Var\left[\sum_f \sum_{h \in uv_{r,s,f}^{hush}} (w_{h,r,s,f}^{kal} - w_{h,r,s,f}) Y_{h,r,s,f}\right] \leq 0,56 \cdot Var(\hat{T}_{r,s})$$

Dette viser at størrelsesorden på standardavviket til den kalibrerte og etterstratifiserte estimatoren for antallet med en egenskap er av samme orden.

Vi ser derfor ikke noen grunn til å gå videre med opplegget for å beskrive kvaliteten til de foreløpige tallene utover de beregningene som er beskrevet i foregående avsnitt. Dessuten er målefeil også en betydelig feilkilde som vi ikke har tatt hensyn til. Det er derfor ingen grunn til å tro at beregningene i foregående avsnitt overvurderer den totale feilen i tallene i forhold til den ukjente sannheten om husholdninger og boliger som vi søker etter gjennom bolig tellingen.

4.4. Usikkerheten til antallet husholdninger i hvert fylke og for hele landet

Det estimerte antallet husholdninger i fylket r av adrestype s kan vi beregne ved summere (3.10) over antall DSF familier i husholdningene:

$$(4.18) \quad \hat{M}_{r,s}^{hush} = \sum_f w_{r,s,f}^{hush} \cdot m_{r,s,f}^{AKU+}$$

der

- $m_{r,s,f}^{AKU+} = m_{r,s,f}^{hush}$

er antall uttrukne DSF familier som hører til husholdninger som består av f antall DSF familier. Usikkerheten til estimatet i (4.18) vil være knyttet til fordelingen av DSF familiene i AKU+ utvalget med hensyn på hvor mange DSF familier det er husholdningene som disse uttrukne DSF familiene hører til. Siden vi kan anta at antallet husholdninger gruppert etter antallet DSF familier i husholdningene er tilnærmet multinomisk fordelt kan vi derfor skrive opp følgende estimat for standardavviket til antallet husholdninger i et fylke innen hver adressegruppe lik

$$(4.19) \quad SD(\hat{M}_{r,s}^{hush}) = \sqrt{\sum_f [w_{r,s,f}^{hush}]^2 m_{r,s,f}^{AKU+} \left(1 - \frac{m_{r,s,f}^{AKU+}}{m_{r,s}^{AKU+}}\right) - \sum_{f \neq g} w_{r,s,f}^{hush} w_{r,s,g}^{hush} m_{r,s}^{AKU+} \frac{m_{r,s,f}^{AKU+}}{m_{r,s}^{AKU+}} \frac{m_{r,s,g}^{AKU+}}{m_{r,s}^{AKU+}}}$$

Denne formelen kan vi bruke videre til å beregne standardavviket til antallet husholdninger for fylkene og for heler landet:

$$(4.20) \quad SD(\hat{M}_r^{hush}) = \sqrt{\sum_s [SD(\hat{M}_{r,s}^{hush})]^2}$$

$$(4.21) \quad SD(\hat{M}^{hush}) = \sqrt{\sum_s [SD(\hat{M}_r^{hush})]^2}$$

For utvalget har vi tilnærmet at i underkant av 80% av husholdningene består av en DSF familie og rundt 20% av husholdningene består av to DSF familier, mens kun et par prosent av husholdningene i utvalget inneholder flere enn 2 DSF familier, se vedlegg E for en fordeling av utvalget etter fylker. Dersom vi ser bort fra denne lille andelen som ikke består av 1 eller 2 DSF familier finner vi at standardavviket til antallet husholdninger for et stratum, formel (4.19), blir lik

$$(4.22) \quad SD(\hat{M}_{r,s}^{hush}) = \frac{2}{5} \frac{M_{r,s}}{m_{r,s}^{hush}} \sqrt{m_{r,s}^{hush}}$$

Tenker vi oss videre at denne fordelingen mellom antallet husholdninger som består av henholdsvis 1 eller 2 DSF familier gjelder for fylkene og Norge totalt i tillegg kan vi bruke formel (4.22) for alle totaltall for husholdninger eller boliger vi publiserer. Dersom vi regner at dette forholdet mellom DSF familier i populasjonen og husholdninger i utvalget er lik 125 så har vi følgende enkle formel for å finne standardavviket til det estimerte husholdningstallet for enten fylker eller hele landet:

$$(4.23) \quad SD(\hat{M}^{hush}) = 50 \sqrt{m^{hush}}$$

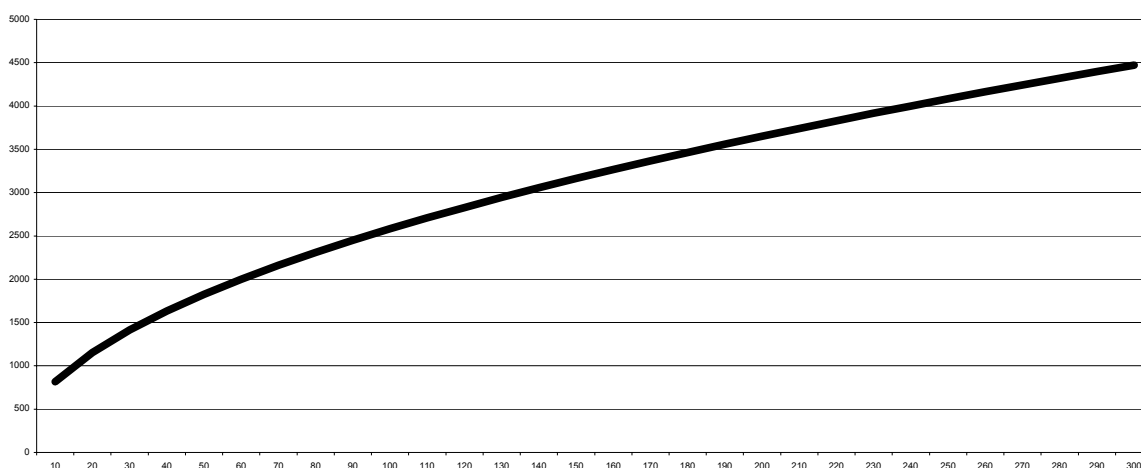
I tabell 4.3 har vi gitt noen tall som eksempler. Vi ser at for det minste fylket som er Finmark med rundt 40 000 husholdninger bør vi ikke bomme stort mer enn med 1 500 husholdninger, mens for de andre fylkene med mellom 75 000 til 300 000 husholdninger bør vi ikke bomme med mer enn fra

2 250 til 4 500 husholdninger og for hele landet bør det foreløpige antallet husholdninger ligge innenfor en feilmargin på høyst 12 500 husholdninger. Figur 4.2 viser kurven for usikkerheten målt på dette viset.

Tabell 4.3. Standardavviket til antall husholdninger

Antall husholdninger	Standardavviket	2 standardavvik	95 % konfidensintervall
37 500	800	1 600	[35 900,39 100]
75 000	1 125	2 250	[72 750,77 250]
150 000	1 600	3 200	[146 800, 153 200]
300 000	2 250	4 500	[295 500, 304 500]
2 000 000	6 250	12 500	[1 987 500, 2 012 500]

Figur 4.2 Usikkerheten målt ved to standardavvik til det totale antallet husholdninger eller boliger i fylkene. Tallene er i antall 1 000 på horisontal akse.



5. Imputering

I avsnitt 2.4 konkluderte vi med at de fire adressegruppene er karakterisert ved følgende egenskaper med hensyn på at vi mangler svar, altså vi har ikke mottatt skjema fra noen av Besys familiene i husholdningen:

- For boliger/husholdninger med unik adresse og stedfortreder, typisk eksempel er rekkehus, kan vi med stor sannsynlighet finne fram korrekt formell husholdning og med stor sannsynlighet imputere korrekte boligopplysninger
- For boliger med unik adresse uten stedfortreder, typisk eksempel er enebolig uten sokkelleilighet og våningshus, kan vi finne fram korrekt husholdning, men med liten sannsynlighet imputere korrekte boligopplysninger
- For boliger uten unik adresse med stedfortreder, typisk eksempel er blokker, har vi liten mulighet til å finne fram korrekt husholdning, men stor sannsynlighet for å imputere korrekte boligopplysninger
- For boliger uten unik adresse uten stedfortreder, typisk eksempel er enebolig med sokkelleilighet, har vi høy sannsynlighet for å finne fram til korrekt husholdning, men svært lav sannsynlighet til imputere korrekte boligopplysninger.

Dersom vi summerer dette ser vi at det er langt lettere å finne fram til korrekt husholdning på de skjemaene vi mangler enn å imputere boligopplysninger for disse husholdningenes boliger.

Det kan imidlertid gjennomføres en viktig kontroll for å undersøke frafallet. Vi kan nemlig sette inn det vi tror er den korrekte husholdningen for frafallet og vi kan da kjøre ut husholdningstall både basert på totalutvalget og nettoutvalget. Dersom vi ser liten forskjell mellom disse to tabellene med husholdningstall vil det gi større trygghet for at også boligtallene er lite påvirket av frafallet.

Vi ser ovenfor at det kun for boliger uten unik adresse og med stedfortreder, det typiske er blokker, at vi har reell mulighet til å imputere boligopplysninger av høy kvalitet. Dette kommer av at for boliger med unik adresse og stedfortreder, det typiske er rekkehus, mangler vi datagrunnlaget siden de andre skjemaene til disse rekkehusene ikke er bearbeidet. Altså dersom vi har frafall finnes det ingen annen bolig i samme rekkehus som vi kan hente opplysninger fra. Det betyr at i den gruppa (blokkleilighetene) der vi faktisk har reell mulighet til å imputere riktige boligopplysninger er det samtidig vanskeligst å finne fram til den riktige husholdningen.

Dette betyr at det er lite å hente på å imputere boligopplysningene på de skjemaene vi mangler i utvalget vårt. Vi konkluderer derfor med at vi ikke imputerer direkte for enhetsfracfallet, men bruker etterstratifisering etter fylker og adressegrupper etterfulgt av kalibrering med hensyn på de fem BESYS familietyper og de åtte aldersgruppene for henholdsvis menn og kvinner.

6. Referanser

- | | |
|---|---|
| Fosen Johan, Anne Gro Hustoft og Bengt Oscar Lagerstrøm | Ny spørresekvens for å identifisere husholdninger i utvalgsundersøkelser. Notater 2001/25 |
| Heldal, Johan(1992) | A Method for Calibration of Weights in Sample Surveys. Arbeidsnotat fra Avdeling for personstatistikk, 3/1992 |
| Heldal, Johan(2000) | Kalibrering av AKU. Notater 2000/7 |
| Noack, Turid, Aslaug Hurlen Foss og Ingvild Hauge Byberg(2001) | Gjete kongens harer - rapport med arbeidet for å få samboere mer innpasset i statistikken. Rapporter 2001/40. |
| Solheim, Leiv, Johan Fosen, Aslaug Hurlen Foss og Li Chun Zhang(2002) | Folke- og bolig tellingen 2001: Hvor nøyaktig blir statistikken. Kommer i serien Notater |
| Solheim, Leiv og Li-Chun Zhang(2001) | AKU: Utvalgsplanen, utvalget og husholdninger. Upublisert notat. Lso, lcz 25. juni 2001. |
| SSB(2001) | Standard for gruppering av familier og husholdninger. NOS C 677, 2001 |
| Zhang, Li-Chun(2000) | Post-stratification and calibration - A synthesis. The American Statistician, vol. 54, 174-184 |
| Zhang, Li-Chun(2001) | Estimeringsmetode for familie-/husholdningsfordeling. Notater 2001/50 |

Vedlegg A Beregne vektor på person- og husholdningsnivå - en sammenlikning

A.1. Innledning

Vi antar at vi har en husholdningsegenskap eller verdi

- Y_h

som vi på grunnlag av observasjoner fra et utvalg ønsker å predikere for hele populasjonen, dvs. vi ønsker å finne

$$(A.1) \quad T_Y = \sum_{h \in \text{utv}} Y_h + \sum_{h \notin \text{utv}} Y_h$$

der den første summen blir observert fra utvalget, mens den andre summen ikke observeres og må predikeres. Dersom vi predikerer de enkelte Y -ene utenfor utvalget

$$(A.2) \quad \hat{Y}_h$$

ved hjelp av utvalget får vi en predikert total

$$(A.3) \quad \hat{T}_Y = \sum_{h \in \text{utv}} Y_h + \sum_{h \notin \text{utv}} \hat{Y}_h$$

Vi skal i dette vedlegget se hvordan vi kan beregne (A.3) dersom vi stiller som krav at når vi setter inn antall menn i husholdningen som Y verdi, dvs.

$$(A.4) \quad Y_h = x_h$$

der x_h er antall menn i husholdningen h , gir prediksjonen i (A.3) totalt antall menn i husholdningene. Vi skal undersøke to metoder for å beregne (A.3), nemlig direkte på husholdningene og indirekte på personene i husholdningene før vi til slutt sammenlikner metodene.

La oss imidlertid før vi går videre, føre inn betegnelser på antallet husholdninger i utvalget og populasjonen:

- M
- m

I neste avsnitt skal vi anta at det totale antallet husholdninger, M , er kjent, mens vi i avsnitt A.3 skal kun anta at antallet personer i alt og antallet menn i alt er kjent.

A.2. Husholdningsberegninger

Vi antar at følgende modell mellom husholdningsegenskapen og antall menn gjelder for populasjonen:

$$(A.5) \quad Y_h = \alpha + \beta x_h + \varepsilon_h$$

der

$$(A.6) \quad \begin{aligned} E(\varepsilon_h) &= 0 \\ \text{Var}(\varepsilon_h) &= \sigma^2 \end{aligned}$$

Minste kvadraters metode gir da følgende estimater for de to ukjente parametrene:

$$(A.7) \quad \begin{aligned} \hat{\alpha} &= \bar{Y}_s - \hat{\beta}\bar{x}_s \\ \hat{\beta} &= \frac{\sum_{h \in utv} (x_h - \bar{x}_s) Y_i}{\sum_{h \in utv} (x_h - \bar{x}_s)^2} \end{aligned}$$

der

- $\bar{Y}_s = \frac{1}{m} \sum_{h \in utv} Y_h$
- $\bar{x}_s = \frac{1}{m} \sum_{h \in utv} x_h$

er gjennomsnittet av henholdsvis Y-ene og x-ene i utvalget. Bruker vi disse parameterestimaterne til å predikere de ikke observerte Y-ene

$$(A.8) \quad \hat{Y}_h = \hat{\alpha} + \hat{\beta}x_h = \bar{Y}_s - \hat{\beta}\bar{x}_s + \hat{\beta}x_h = \bar{Y}_s + \hat{\beta}(x_h - \bar{x}_s)$$

finner vi følgende form på prediksjonen i (A.3):

$$(A.9) \quad \begin{aligned} \hat{T}_Y &= \sum_{h \in utv} Y_h + \sum_{h \notin utv} [(\bar{Y}_s + \hat{\beta}(x_h - \bar{x}_s))] = m\bar{Y}_s + (M - m)\bar{Y}_s + \hat{\beta}[\sum_{h \in utv} x_h - (M - m)\bar{x}_s] \\ &= M\bar{Y}_s + \hat{\beta}[M\bar{x} - m\bar{x}_s - (M - m)\bar{x}_s] = M\bar{Y}_s + M(\bar{x} - \bar{x}_s)\hat{\beta} \\ &= M \frac{1}{m} \sum_{h \in utv} Y_h + M(\bar{x} - \bar{x}_s) \frac{\sum_{h \in utv} (x_h - \bar{x}_s) Y_h}{\sum_{h \in utv} (x_h - \bar{x}_s)} = \sum_{h \in utv} \left\{ \frac{M}{m} \left[1 - \frac{(\bar{x} - \bar{x}_s)(\bar{x}_s - x_h)}{\hat{\sigma}_x^2} \right] Y_h \right\} \end{aligned}$$

der

- $\sum_{h \notin utv} x_h = \sum_h x_h - \sum_{h \in utv} x_h = M\bar{x} - m\bar{x}_s$
- $\hat{\sigma}_x^2 = \frac{1}{m} \sum_{h \in utv} (x_h - \bar{x}_s)^2$

La oss betegne uttrykket inne i parentesen i den siste summen i (A.9), utenom Y, som en vekt:

$$(A.10) \quad w_h = \frac{M}{m} \left[1 - \frac{(\bar{x} - \bar{x}_s)(\bar{x}_s - x_h)}{\hat{\sigma}_x^2} \right]$$

Vi kan da undersøke egenskapene til disse vektene for husholdningene definert i (A.10). For det første finner vi at summen av vektene er lik antall husholdninger:

$$\begin{aligned}
(A.11) \quad \sum_{h \in utv} w_h &= \sum_{h \in utv} \frac{M}{m} \left[1 - \frac{(\bar{x} - \bar{x}_s)(\bar{x}_s - x_h)}{\hat{\sigma}_x^2} \right] = \sum_{h \in utv} \frac{M}{m} - \sum_{h \in utv} \left[\frac{M}{m} \frac{(\bar{x} - \bar{x}_s)(\bar{x}_s - x_h)}{\hat{\sigma}_x^2} \right] \\
&= \frac{M}{m} m - \frac{M(\bar{x} - \bar{x}_s)}{\hat{\sigma}_x^2} \sum_h (\bar{x}_s - x_h) = M
\end{aligned}$$

Den siste likheten følger av at

- $\sum_{h \in utv} (\bar{x}_s - x_h) = m\bar{x}_s - m\bar{x}_s = 0$

For det andre finner vi at vektene multiplisert med antallet menn i husholdningene summerer til det totale antallet menn i alle husholdninger i populasjonen:

$$\begin{aligned}
(A.12) \quad \sum_{h \in utv} w_h x_h &= \sum_{h \in utv} \frac{M}{m} \left[1 - \frac{(\bar{x} - \bar{x}_s)(\bar{x}_s - x_h)}{\hat{\sigma}_x^2} \right] x_h = \sum_{h \in utv} \frac{M}{m} x_h - \sum_{h \in utv} \left[\frac{M}{m} \frac{(\bar{x} - \bar{x}_s)(\bar{x}_s - x_h)}{\hat{\sigma}_x^2} \right] x_h \\
&= \frac{M}{m} m\bar{x}_s - \frac{M(\bar{x} - \bar{x}_s)}{m\hat{\sigma}_x^2} \sum_h (\bar{x}_s - x_h) x_h = M\bar{x}_s - \frac{M(\bar{x} - \bar{x}_s)}{m\hat{\sigma}_x^2} \sum_h (\bar{x}_s - x_h)^2 \\
&= M\bar{x}_s - \frac{M(\bar{x} - \bar{x}_s)}{m\hat{\sigma}_x^2} m\hat{\sigma}_x^2 = M\bar{x}_s - M(\bar{x} - \bar{x}_s) = M\bar{x} = \sum_h x_h
\end{aligned}$$

siden

- $\begin{aligned} \sum_{h \in utv} (\bar{x}_s - x_h) x_h &= \sum_{h \in utv} (\bar{x}_s - x_h)(x_h - \bar{x}_s + \bar{x}_s) = \sum_{h \in utv} (\bar{x}_s - x_h)^2 - \sum_{h \in utv} (\bar{x}_s - x_h) \bar{x}_s \\ &= \sum_{h \in utv} (\bar{x}_s - x_h)^2 - \bar{x}_s \sum_{h \in utv} (\bar{x}_s - x_h) = \sum_{h \in utv} (\bar{x}_s - x_h)^2 - \bar{x}_s \cdot 0 = \sum_{h \in utv} (\bar{x}_s - x_h)^2 = m\hat{\sigma}_x^2 \end{aligned}$

Vi har da vist at med modellen gitt i (A.5-6) finner vi vektene definert i (A.10) har egenskapene at de summerer til antallet husholdninger i populasjonen, dvs. når $Y=1$, og dessuten gir riktig antall menn i populasjonen, dvs. når $Y =$ antall menn i husholdningen.

Setter vi inn antallet personer i husholdningene som husholdningsegenskap finner vi

$$(A.13) \quad \hat{N} = \sum_h w_h n_h = M \frac{n}{m} + M \frac{(\bar{x} - \bar{x}_s)}{\hat{\sigma}_s^2} \frac{1}{m} \sum_h (x_h - \bar{x}_s)(n_h - \bar{n}_s) = M\bar{n}_s + M \frac{(\bar{x} - \bar{x}_s)}{\hat{\sigma}_s^2} \hat{\sigma}_{x,n}$$

der

- n_h - antall personer i husholdning h i utvalget
- n - antall personer i alt i utvalget
- $\bar{n}_s = \frac{1}{m} \sum_h n_h = \frac{n}{m}$ - gjennomsnittlig antall personer i husholdningene i utvalget
- $\hat{\sigma}_{x,n} = \frac{1}{m} \sum_h (x_h - \bar{x}_s)(n_h - \bar{n}_s)$ - estimert kovariasjon mellom antall menn og antall personer

Siden det er en positiv samvariasjon mellom antall menn og antall personer i husholdningene ser vi fra (A.13) at vi reduserer antall personer fra den enkle oppblåsingen, antall husholdninger multiplisert med antall personer per husholdning, når vi har flere menn per husholdning i utvalget enn i populasjonen og omvendt øker antallet personer i populasjonen når antallet menn i utvalgshusholdningene er mindre enn i husholdningene totalt i populasjonen.

Dette opplegget kunne vi brukt dersom vi hadde en full oversikt over husholdningene i populasjonen eller i hvert fall kjente antallet husholdninger i populasjonen.

A. 3. Personberegninger

I dette avsnittet antar vi at antallet husholdninger er ukjent, mens antallet personer i alt og antallet menn i alt er kjent. Vi skal nå gjennomføre en tilsvarende øvelse på personnivå som i foregående avsnitt. La oss først innføre noen betegnelser:

- n_h - antall personer i husholdning h
- n - antall personer i utvalget
- N - antall personer i populasjonen

Da har vi selvsagt følgende enkle identiteter når vi summerer over henholdsvis utvalget og hele populasjonen:

$$(A.14) \quad \begin{aligned} n &= \sum_{h \in \text{utv}} n_h \\ N &= \sum_h n_h \end{aligned}$$

La videre for hver person i husholdningen

$$(A.15) \quad z_{i,h} = \begin{cases} 0 \\ 1 \end{cases}$$

der verdien 1 betyr at personen er en mann, mens 0 betyr kvinne. Det er da slik at summen av z-ene for personene i en husholdning er lik antall menn i husholdningen - altså

$$(A.16) \quad \sum_{i=1}^{n_h} z_{i,h} = x_h$$

Definer videre en "predikert" verdi for personene i hver husholdning som er lik gjennomsnittet av z-ene:

$$(A.17) \quad \hat{z}_{i,h} = \bar{z}_{.,h} = \frac{1}{n_h} \sum_{i=1}^{n_h} z_{i,h}$$

Summen av de "predikerte" z-ene i (A.17) er selvsagt også lik antallet menn i husholdningen. Videre definerer vi også gjennomsnittlig antall menn per person eller som ser forholdet mellom antallet menn og antallet personer:

$$(A.18) \quad \begin{aligned} \bar{z}_s &= \frac{1}{n} \sum_{h \in utv} \sum_{i=1}^{n_h} \hat{z}_{i,h} = \frac{m}{n} \frac{1}{m} \sum_{h \in utv} x_h = \frac{m}{n} \bar{x}_s \\ \bar{z} &= \frac{1}{N} \sum_h \sum_{i=1}^{n_h} \hat{z}_{i,h} = \frac{M}{N} \frac{1}{M} \sum_h x_h = \frac{M}{N} \bar{x} \end{aligned}$$

Definer videre "Y-verdien" til hver person i husholdningen som personens andel av Y-verdien

$$(A.19) \quad U_{i,h} = \frac{Y_h}{n_h}$$

slik alle personer i en husholdning har samme "Y-verdi", dvs. U-verdien er den samme for alle personene i husholdningen. Vi ser også at summen av U-verdiene over personene i en husholdning er lik Y-verdien til husholdningen siden

$$(A.20) \quad \sum_{i=1}^{n_h} U_{i,h} = \sum_{i=1}^{n_h} \frac{Y_h}{n_h} = n_h \frac{Y_h}{n_h} = Y_h$$

La oss videre beregne gjennomsnittet av U-verdiene over utvalget og hele populasjonen:

$$(A.21) \quad \begin{aligned} \bar{U}_s &= \frac{1}{n} \sum_{h \in utv} \sum_{i=1}^{n_h} U_{i,h} = \frac{m}{n} \frac{1}{m} \sum_{h \in utv} Y_h = \frac{m}{n} \bar{Y}_s \\ \bar{U} &= \frac{1}{N} \sum_h \sum_{i=1}^{n_h} U_{i,h} = \frac{M}{N} \frac{1}{M} \sum_h Y_h = \frac{M}{N} \bar{Y} \end{aligned}$$

Vi skal nå anta at vi har følgende modell mellom U-ene og z-ene:

$$(A.22) \quad U_{i,h} = \gamma + \lambda \hat{z}_{i,h} + \eta_{i,h} \quad \text{der} \quad \begin{aligned} E(\eta_{i,h}) &= 0 \\ \text{Var}(\eta_{i,h}) &= \tau^2 \end{aligned}$$

Bruker vi minste kvadraters metode og ser bort fra at feilleddene innen en husholdning ikke er uavhengige finner vi som i foregående avsnitt at

$$(A.23) \quad \hat{\gamma} = \bar{U}_s - \hat{\lambda} \bar{z}_s = \frac{m}{n} (\bar{Y}_s - \hat{\lambda} \bar{x}_s)$$

$$(A.24) \quad \hat{\lambda} = \frac{\sum_{h \in utv} \sum_{i=1}^{n_h} (\hat{z}_{i,h} - \bar{z}_s) U_{i,h}}{\sum_{h \in utv} \sum_{i=1}^{n_h} (\hat{z}_{i,h} - \bar{z}_s)^2} = \frac{\sum_{h \in utv} (\bar{z}_h - \bar{z}_s) Y_h}{\sum_{h \in utv} n_h (\bar{z}_h - \bar{z}_s)^2}$$

Vi bruker estimatene i (A.23-A.24) til å predikere en ikke observervert U-verdi og finner da

$$(A.25) \quad \hat{U}_{i,h} = \hat{\gamma} + \hat{\lambda} \hat{z}_{i,h}$$

som vi setter inn i prediksjonslikningen:

$$\begin{aligned}
\hat{T}_U &= \sum_{h \in utv} \sum_{i=1}^{n_h} U_{i,h} + \sum_{h \in utv} \sum_{i=1}^{n_h} \hat{U}_{i,h} = n\bar{U}_s + \sum_{h \in utv} \sum_{i=1}^{n_h} [\bar{U}_s - \hat{\lambda}\bar{z}_s + \hat{\lambda}\hat{z}_{i,h}] \\
\text{(A.26)} \quad &= n\bar{U}_s + (N-n)(\bar{U}_s - \hat{\lambda}\bar{z}_s) + \hat{\lambda}(N\bar{z} - n\bar{z}_s) = N\bar{U}_s + N(\bar{z} - \bar{z}_s)\hat{\lambda} \\
&= \sum_{h \in utv} \sum_{i=1}^{n_h} \left\{ \frac{N}{n} \left[1 - \frac{(\bar{z} - \bar{z}_s)(\bar{z}_s - \bar{z}_h)}{\hat{\tau}_z^2} \right] U_{i,h} \right\} = \sum_{h \in utv} \left\{ \frac{N}{n} \left[1 - \frac{(\bar{z} - \bar{z}_s)(\bar{z}_s - \bar{z}_h)}{\hat{\tau}_z^2} \right] Y_h \right\}
\end{aligned}$$

Vi kan da skrive opp vektene basert på z-ene og finner da

$$\text{(A.27)} \quad w_{i,h} = w_{h,z} = \frac{N}{n} \left[1 - \frac{(\bar{z} - \bar{z}_s)(\bar{z}_s - \bar{z}_h)}{\hat{\tau}_z^2} \right]$$

for vektene til den enkelte person. Vi ser at vektene er like for alle personer i samme husholdning. Vi finner videre hvilke egenskaper vektene har. Først estimerer vi antall husholdninger i populasjonen:

$$\text{(A.28)} \quad \sum_{h \in utv} \sum_{i=1}^n w_{i,h} \frac{1}{n_h} = \sum_{h \in utv} w_{h,z} = \sum_{h \in utv} \frac{N}{n} \left[1 - \frac{(\bar{z} - \bar{z}_s)(\bar{z}_s - \bar{z}_h)}{\hat{\tau}_s^2} \right] = \frac{N}{n} m = N \frac{m}{n} = \frac{N}{n/m} = \hat{M}$$

Vi estimerer antall husholdninger ved multiplisere antallet personer i populasjonen med det gjennomsnittlige antallet husholdninger per person i utvalget eller beregnet forholdet mellom antallet personer i populasjonen antall personer per husholdning i utvalget. Estimaten av antall husholdninger er med andre ord uforandret fra enkel oppblåsing.

Videre kan vi vise at vi får riktig antall personer ved å summere vektene for personer:

$$\begin{aligned}
\text{(A.29)} \quad &\sum_{h \in utv} \sum_{i=1}^n w_{i,h} = \sum_{h \in utv} w_{h,z} n_h = \sum_{h \in utv} \frac{N}{n} \left[1 - \frac{(\bar{z} - \bar{z}_s)(\bar{z}_s - \bar{z}_h)}{\hat{\tau}_s^2} \right] n_h \\
&= \frac{N}{n} \sum_{h \in utv} n_h - \frac{N(\bar{z} - \bar{z}_s)}{n \hat{\tau}_s^2} \sum_{h \in utv} (\bar{z}_s - \bar{z}_h) n_h = \frac{N}{n} n - \frac{N(\bar{z} - \bar{z}_s)}{n \hat{\tau}_s^2} [n\bar{z}_s - n\bar{z}_s] = N
\end{aligned}$$

Til slutt gjenstår at vi finner igjen antallet menn i populasjonen ved å summere på riktig måte over utvalget - vi setter inn antall menn for Y-verdien:

$$\begin{aligned}
\text{(A.30)} \quad &\sum_{h \in utv} \sum_{i=1}^n w_{i,h} \frac{x_h}{n_h} = \sum_{h \in utv} w_{h,z} n_h \bar{z}_h = \sum_{h \in utv} \frac{N}{n} \left[1 - \frac{(\bar{z} - \bar{z}_s)(\bar{z}_s - \bar{z}_h)}{\hat{\tau}_s^2} \right] \bar{z}_h \\
&= \frac{N}{n} \sum_{h \in utv} n_h \bar{z}_h - \frac{N(\bar{z} - \bar{z}_s)}{n \hat{\tau}_s^2} \sum_{h \in utv} (\bar{z}_s - \bar{z}_h) n_h \bar{z}_h = \frac{N}{n} n\bar{z}_s + \frac{N(\bar{z} - \bar{z}_s)}{n \hat{\tau}_s^2} \sum_{h \in utv} n_h (\bar{z}_h - \bar{z}_s)^2 \\
&= N\bar{z}_s + N(\bar{z} - \bar{z}_s) = N\bar{z} = \sum_h x_h
\end{aligned}$$

Vi har da vist at vektene i (A.27) gir oss det riktige antallet menn i populasjonen i likhet med hva vektene i (A.10) viste. Dessuten følger det av vektene i (A.27) at vi får riktig antall personer i tillegg, mens vi fra vektene i (A.10) får det rette antallet husholdninger.

A.4. Sammenlikning

Vi har i avsnitt A.2 antatt at vi kjente antallet husholdninger i populasjonen og ut fra dette vist hvordan vi ved en enkel modell for husholdningsegenskapen der en antar denne er avhengig av antall menn i husholdningen vist hvordan vi kan utlede vekter slik at vi når vi setter inn antall menn som husholdningsegenskapen får vi det korrekte antallet menn i populasjonen ved å bruke disse vektene når vi summerer over utvalget.

I avsnitt A.3 har vi dessuten vist ved en liknende statistisk modell for husholdningsegenskapens andel på den enkelte person forklart ved andelen menn i husholdningen per person at vi finner vekter som er konsistente med antall personer i alt og antall menn i alt i populasjonen. Vi kan i dette tilfellet ikke kjenne annet enn det totale antallet menn og personer i populasjonen og antallet menn og personer i husholdningene i utvalget.

I praksis er det opplegget i A.3 som er det naturlige siden vi ikke kjenner gjennomsnittet per husholdning i populasjonen, ellers var det ikke nødvendig med undersøkelsen, mens vi kjenner selvsagt den gjennomsnittlige verdien per person i populasjonen, og det er dette som er poenget siden antallet personer er kjent i populasjonen.

Vedlegg B Enkle person- og familievekter

Dersom vi nå definerer to korreksjoner av vekten i (3.31) lik henholdsvis

$$(B.1) \quad \begin{aligned} W_{i,r,s,f}^P &= \frac{N_{r,s}}{f \cdot m_{r,s}^{hush} n_{i,r,s,f}} \\ W_{i,r,s,f}^F &= \frac{M_{r,s}}{f \cdot m_{r,s}^{hush} n_{i,r,s,f}} \end{aligned}$$

for personer og familier finner vi ved summere vektene over alle personene i utvalget at

$$(B.2) \quad \sum_i W_{i,r,s,f}^P = \sum_{hush} \sum_{fam} \sum_{pers} \frac{N_{r,s}}{f \cdot m_{r,s}^{hush} \cdot n_{i,r,s,f}} = \sum_{hush} \sum_{fam} \frac{N_{r,s}}{f \cdot m_{r,s}^{hush}} = \sum_{hush} \frac{N_{r,s}}{m_{r,s}^{hush}} = \frac{N_{r,s}}{m_{r,s}^{hush}} m_{r,s}^{hush} = N_{r,s}$$

$$(B.3) \quad \sum_i W_{i,r,s,f}^F = \sum_{hush} \sum_{fam} \sum_{pers} \frac{M_{r,s}}{f \cdot m_{r,s}^{hush} \cdot n_{i,r,s,f}} = \sum_{hush} \sum_{fam} \frac{M_{r,s}}{f \cdot m_{r,s}^{hush}} = \sum_{hush} \frac{M_{r,s}}{m_{r,s}^{hush}} = \frac{M_{r,s}}{m_{r,s}^{hush}} m_{r,s}^{hush} = M_{r,s}$$

Vi kunne for øvrig også ha definert vektene gjennom å telle antallet personer i husholdningen direkte og ville da fått

$$(B.4) \quad \begin{aligned} \tilde{W}_{i,r,s,f}^P &= \frac{N_{r,s}}{m_{r,s}^{hush} n_{i,r,s}} \\ \tilde{W}_{i,r,s,f}^F &= \frac{M_{r,s}}{m_{r,s}^{hush} n_{i,r,s}} \end{aligned}$$

og ser at disse vektene er like for alle personer i husholdningen. Dersom vi bruker vektene i (B.4) finner vi at disse også gir det korrekte antallet personer og familier i populasjonen.

Dersom vi i tillegg forsøker å beregne antall husholdninger for de to vektsettene i (B.1) vil det gi forskjellig resultat og det er vanskelig å finne noen god begrunnelse hvorfor vi skal velge den ene framfor den andre. Vi må for personvektene la husholdningsverdien være det inverse av antall personer i husholdningen, mens vi for familievektene bruker det inverse av antall familier. Vi finner da at

$$(B.5) \quad \begin{aligned} \sum_i W_{i,r,s,f}^P \frac{1}{n_{i,r,s}} &= \sum_{hush} \sum_{fam} \sum_{pers} \frac{N_{r,s}}{f \cdot m_{r,s}^{hush} \cdot n_{i,r,s,f}} \frac{1}{n_{i,r,s}} = \sum_{hush} \sum_{fam} \frac{N_{r,s}}{f \cdot m_{r,s}^{hush}} \frac{1}{n_{i,r,s}} \\ &= \sum_{hush} \frac{N_{r,s}}{m_{r,s}^{hush}} \frac{1}{n_{i,r,s}} = N_{r,s} \frac{1}{m_{r,s}^{hush}} \sum_{hush} \frac{1}{n_{i,r,s}} = N_{r,s} \frac{1}{\bar{n}_{r,s}^{Harmonisk}} \end{aligned}$$

$$(B.6) \quad \begin{aligned} \sum_i W_{i,r,s,f}^F \frac{1}{f} &= \sum_{hush} \sum_{fam} \sum_{pers} \frac{M_{r,s}}{f \cdot m_{r,s}^{hush} \cdot n_{i,r,s,f}} \frac{1}{f} = \sum_{hush} \sum_{fam} \frac{M_{r,s}}{f \cdot m_{r,s}^{hush}} \frac{1}{f} \\ &= \sum_{hush} \frac{M_{r,s}}{m_{r,s}^{hush}} \frac{1}{f} = M_{r,s} \frac{1}{m_{r,s}^{hush}} \sum_{hush} \frac{1}{f} = M_{r,s} \frac{1}{\bar{f}_{r,s}^{Harmonisk}} \end{aligned}$$

Vi har altså funnet at

- Dersom vi bruker personvektene beregner vi antallet husholdninger lik forholdet mellom antallet personer i populasjonen og det harmoniske gjennomsnittet av antallet personer i husholdningene, mens når vi bruker familievektene beregner vi antallet husholdninger lik forholdet mellom antallet familier og det harmoniske gjennomsnittet av antallet familier i husholdningene

Bruker vi i stedet vektene definert i (B.4) finner vi akkurat de samme formlene for antallet husholdninger som når vi brukte (B.1).

Vedlegg C Detaljert fordeling av DSF familiene bosatt i boligbygg

Bygningstype	Totalt	Unik adresse	Ikke unik adresse
Alle boligbygg	1 995 379	1 285 354	710 025
111 Enebolig	926 923	917 718	9 205
112 Enebolig med m/hybel/sokkelleilighet	123 698	3 232	120 466
113 Våningshus	89 682	88 616	1 066
121 Del av tomannsbolig - vertikal	94 519	80 723	13 796
122 Tomannsbolig, horisontaldelt	81 447	3 789	77 658
123 Del av våningshus. tomannsb./vert.	3 276	832	2 444
124 Våningsh. tomannsb./horisont	4 118	72	4 046
131 Del av rekkehus m/3-4 boliger	62 552	58 619	3 933
132 Del av rekkehus m/5 boliger e. fl.	64 459	58 963	5 496
133 Del av kjede/atr. hus innt. 4 boliger	10 126	9 262	864
134 Del av kjede/atr. hus med 5 bol. el. fl.	5 243	4 674	569
135 Terrassehus	18 891	7 758	11 133
136 Andre småhus med 3-4 boliger	66 495	17 827	48 668
141 Boligblokk på 2 etasjer	20 628	8 018	12 610
142 Boligblokk på 3 og 4 etasjer	237 476	9 127	228 349
143 Boligblokk på 5 etasjer eller mer	126 301	1 200	125 101
144	5 625	1 194	4 431
145	14 609	853	13 756
146	2 613	14	2 599
151 Trygdeb., aldershjem, HVPU-bolig	9 131	2 127	7 004
152 Studenthjem/studentboliger	1 892	141	1 751
153 Annen bygning for fellesskap/l.	2 438	522	1 916
16. Fritidsbolig	3 228	3 178	50
17. Koie - seterhus - liknende	22	17	5
18. Garasje og uthus til bolig	341	262	79
19. Annen boligbygning	19 646	6 616	13 030

**Vedlegg D Beregnete befolkningstall basert på AKU+
utvalget sammenliknet med befolknings-
tallene per 3. november 2001**

Fylker	Befolkningstall	AKU+ utvalget	Avvik	Avvik, %
Hele landet	4 520 900	4 524 823	4 923	0,09
01 Østfold	252 500	248 234	-4 266	-1,69
02 Akershus	473 700	466 734	-6 966	-1,47
03 Oslo	514 800	512 629	-2 171	-0,42
04 Hedmark	187 900	197 551	9 651	5,14
05 Oppland	183 300	187 862	4 562	2,49
06 Buskerud	239 600	235 000	-4 600	-1,92
07 Vestfold	216 300	218 688	2 388	1,10
08 Telemark	165 700	169 469	3 769	2,27
09 Aust-Agder	102 800	105 160	2 360	2,30
10 Vest-Agder	157 700	154 190	-3 510	-2,23
11 Rogaland	377 600	374 137	-3 463	-0,92
12 Hordaland	441 100	441 346	246	0,06
14 Sogn og Fjordane	107 300	113 034	5 734	5,34
15 Møre og Romsdal	243 900	236 589	-7 311	-3,00
16 Sør-Trøndelag	266 100	270 150	4 050	1,52
17 Nord-Trøndelag	127 400	131 551	4 151	3,26
18 Nordland	237 600	241 866	4 266	1,80
19 Troms	151 600	145 779	-5 821	-3,84
20 Finmark	73 900	74 856	956	1,29

Vedlegg E Fordeling av husholdningene i utvalget etter antall DSF familier og fylker

Fylker	Antall DSF familier i husholdningen				
	I alt	En	To	3-9	10-
Hele landet	15 782	12 149	3 290	148	195
01 Østfold	907	720	164	13	10
02 Akershus	1 407	1 064	322	11	10
03 Oslo	1 774	1 327	397	21	29
04 Hedmark	686	524	144	8	10
05 Oppland	662	513	132	7	10
06 Buskerud	843	651	179	5	8
07 Vestfold	746	581	155	4	6
08 Telemark	595	476	107	5	7
09 Aust-Agder	494	397	80	6	11
10 Vest-Agder	515	433	71	5	6
11 Rogaland	1 209	969	211	17	12
12 Hordaland	1 314	1 042	246	7	19
14 Sogn og Fjordane	476	374	91	5	6
15 Møre og Romsdal	803	608	168	8	19
16 Sør-Trøndelag	974	692	269	2	11
17 Nord-Trøndelag	511	390	111	6	4
18 Nordland	850	631	202	6	11
19 Troms	557	419	129	5	4
20 Finmark	459	338	112	7	2

De sist utgitte publikasjonene i serien Notater

- 2002/15 A. Rolland: Søkelys på det gode liv. 37s.
- 2002/17 D. Rønningen og D. Fredriksen: Beskatningen av pensjonister. 41s.
- 2002/18 D. Rønningen: Overganger fra arbeidsmarkedet til trygd. En litteraturoversikt. 34s.
- 2002/19 F. Gundersen og L. Solheim: Regionalisering av FoU-statistikken. 43s.
- 2002/20 L. Vågane: Omnibusundersøkelsen november/desember 2001. Dokumentasjonsrapport. 56s.
- 2002/21 G. Claus, O. Haugen, P.M. Holt og E. Knutsen: Regnskapsstatistikk. Næringsoppgaver for ikke-finansielle aksjeselskaper, 1999. Dokumentasjon. 34s.
- 2002/22 M. Takle: Befolkningsstatistikk på rute-nett. Dokumentasjon. 35s.
- 2002/23 D. Roll-Hansen, S. Ferstad, M. Stålnacke, P. Tuhus og R. Nøtnæs: En spørreskjemametodisk gjennomgang av datainnsamling gjennom Grunnskolens informasjonssystem (GSI). 109s.
- 2002/24 T.P. Bøe og I. Håland: Dokumentasjon av arbeidskraftundersøkelsen (AKU). 85s.
- 2002/25 A. Akselsen og T. Sandnes: FD - Trygd: Dokumentasjonsrapport. Stønader til enslig forsørger. 1992-2000. 46s.
- 2002/26 E. Rønning: Statistisk sentralbyrås tidsbruksundersøkelse 2000/01. Dokumentasjon og resultater fra intervjuet. 125s.
- 2002/27 S. Myro og C. Torp: Stedsfesting av bedrifter i Bedrifts- og foretaksregisteret. Hovedprosjekt. 37s.
- 2002/28 C. Nordseth og T. Sandnes: FD - Trygd: Dokumentasjonsrapport. Foreløpig uførestønad. 1992-2000. 37s.
- 2002/29 S. Derakhshanfar og T. Sandnes: FD - Trygd: Dokumentasjonsrapport. Økonomisk sosialhjelp. 1992-2000. 36s.
- 2002/30 I. Johansen: Undersøking om foreldrebetaling i barnehagar, januar 2002. 42s.
- 2002/31 T.M. Køber, H. Moafi, E. Rønning og Ø. Sivertstøl: Bruk av forløpsdatabaser i Statistisk sentralbyrå. 60s.
- 2002/32 T.M. Normann: Omnibusundersøkelsen februar/mars 2002. Dokumentasjonsrapport. 37s.
- 2002/33 S. Reid: Bosettingskriteriene i inntektssystemet til kommunene. Erfaringer med overgang til ny beregningsmåte og nye bosettingskriterier, 2002. 43s.
- 2002/34 K.E. Engebretsen, P.E. Gjedtjernet, S. Kristoffersen, P.G. Larssen og J.H. Wang: Mottak og tilrettelegging av SLN-data. 49s.
- 2002/35 D. Rafat: Analyse av sammenheng mellom ektefellers sysselsetting i en familie. 27s.
- 2002/36 A. Bruvoll og T. Bye: En vurdering av avfallspolitikkens bidrag til løsning av miljø - og ressursproblemer. 31s.
- 2002/37 K.I. Bøe: B.R. Joneid: KOSTRA revisjonssystem. Malverk for generelt revisjonssystem - KOSTRA-data. Revidert utgave. 66s.
- 2002/38 N. Arnesen, G. Daugstad, O.E. Hallingstad, E. Skretting Lunde og B.Vold: Kvalitetssikring i KOSTRA. Forslag til dokumentasjonsrutiner med erfaring fra FylkesKOSTRA-helsetjenester, somatikk. 54s.
- 2002/39 H. Moafi: Omlegging av folkehøgskolestatistikk. Overgang til elektronisk rapportering. 31s.
- 2002/40 Ø. Kleven: Mediebrukundersøkelsen 2001. Dokumentasjonsrapport. 43s.
- 2002/41 Ø. Kleven: Samordnet levekårsundersøkelse 2000 - panelundersøkelsen. Dokumentasjonsrapport. 129s.