



Jørn Ivar Hamre og Anne Vedø

Utvalgsundersøkelsen om egenmeldt sykefravær

Dokumentasjon av utvalgsplanen, utvalget for 2006 og standardfeilberegninger

Innhold

1. Innledning.....	3
2. Beskrivelse av utvalgsplan, trekking og utvalgsrulling.....	4
2.1 Utvalgsplanen.....	4
2.2 Trekkingen.....	5
2.3 Utvalgsrullingene.....	5
3. Om prinsipper, metode, analyse og trekkpopulasjonen.....	6
3.1 Prinsipper for trekking av utvalget.....	6
3.2 Dynamikk i populasjonene.....	7
3.2.1 Nye bedrifter i populasjonen siden forrige trekking.....	7
3.2.2 Bedrifter opphører.....	8
3.2.3 Bedrifter endrer stratum.....	8
3.3 Nøyaktighet og justering av trekk sannsynlighetene i 2006.....	9
3.5 Om trekkpopulasjonen for 2006.....	13
4. Dokumentasjon av oppdateringsrutinene.....	15
4.1 Tidspunkt for rulling.....	16
4.2 Datafilene, programmene og fysisk dataflytdiagram.....	16
4.2.1 Data systemet krever.....	16
4.2.2 Fysisk dataflytdiagram for utvalgstrekkingen.....	17
4.2.3 Beskrivelse av programmene.....	19
4.3 Delregister prosedyrer.....	24
4.3.1 Registrering av nytt delregister og metadata om undersøkelsen.....	24
4.3.2 Innlasting av enheter i delregister.....	24
4.3.3 Oppdatering av andre variable i delregisteret.....	26
5. Kontrolltabeller for utvalget og trekkpopulasjonen 2006.....	27
Vedlegg A.....	38
A. Variasjonskoeffisienter i ulike strata.....	38
Vedlegg B.....	39
B. Beskrivelse av anvendte formler for beregning av estimert varians.....	39
Vedlegg C.....	43
C. Utleddning av kovariansen mellom de to ratene.....	43
Vedlegg D.....	45
D. Samvariasjon i sykefraværsavviket mellom menn og kvinner.....	45
Vedlegg E.....	47
Vedlegg F.....	51
De sist utgitte publikasjonene i serien Notater.....	53

1. Innledning

Utvalgsundersøkelsen om egenmeldt sykefravær utgjør sammen med den registerbaserte statistikken over legemeldt sykefravær Sentral sykefraværstatistikk. Statistikken skal belyse utviklingen i det samlede sykefraværet. Dette skal kunne fordeles på kjønn, næring og fylke. Dette skal bl.a. danne grunnlag for å utvikle og evaluere tiltak for å redusere sykefraværet.

Statistisk sentralbyrå (SSB) henter inn summariske tall for egenmeldt sykefravær for de ansatte i utvalgte bedrifter. Et tilfeldig utvalg av rundt 10 000 bedrifter får spørreskjema hvert kvartal. Hvert år oppdateres utvalget av bedrifter som får tilsendt spørreskjemaet om egenmeldt sykefravær. Utvalgsplanen, trekkingen og oppdatering av utvalget dokumenteres i dette notatet. I tillegg dokumenterer notatet hvordan estimerte standardavvik til måltallene beregnes, samt størrelsen på relativt standardavvik.

Notatet har en **dokumentasjonsdel** som beskriver utvalgstrekkningen, samt oppdateringen 1. kvartal 2006. Denne delen er relevant for fagpersonell som jobber med statistikken og ev. ekspertbrukere av statistikken. Videre inneholder notatet en **dokumentasjon av oppdateringsrutinene** til hjelp ved kommende års utvalgssulleringer.

En versjon av utsendt informasjonsbrev til nye bedrifter og spørreskjema er trykt i vedlegg E og F.

Notatet inneholder i tillegg beregninger av relativt standardavvik, samt teknisk dokumentasjon.

Notatet er i hovedsak utarbeidet av Jørn Ivar Hamre ved Seksjon for arbeidsmarkedsstatistikk. Bidraget til Anne Vedø ved Seksjon for metoder og standarder har bestått i utledning av kovariansformlene beskrevet i vedlegg C.

2. Beskrivelse av utvalgsplan, trekking og utvalgsrullering

2.1 Utvalgsplanen

Undersøkelsen om egenmeldt sykefravær benytter et årlig roterende panelutvalg på om lag 10 000 bedrifter trukket fra Bedrifts- og foretaksregisteret.

Trekkepopulasjonen består av bedrifter i ordinær drift med ansatte fra et situasjonsuttak fra Bedrifts- og foretaksregisteret. Trekkepopulasjonen dekker i utgangspunktet alle næringer. Innenfor noen næringer er det problemer med å rapportere tall for egenmeldt sykefravær. Av kvalitetshensyn har vi derfor holdt følgende næringer utenfor trekkepopulasjonen:

<u>Næringskode:</u>	<u>Navn på næringen:</u>
01.42	Tjenester tilknyttet husdyrhold, herunder avløserlag, jordbruk
11.1	Utvinning av råolje og naturgass
61	Sjøtransport
85.334	Arbeidstrening for ordinært arbeidsmarked
85.335	Varig vernet arbeid
95	Lønnet arbeid i private husholdninger
99	Internasjonale organer og organisasjoner

Stratifisering

Trekkepopulasjonen er stratifisert etter bedriftsstørrelse og næring, ved siden av noen hjelpestratum for enklere å håndtere nye bedrifter siden forrige trekking. Stratifiseringen sikrer god representasjon av de fleste ulike 2-siffer næringene og blant store og små bedrifter.

Trekkepopulasjonen er hovedsakelig stratifisert etter 2-siffer NACE¹ **næring**. For at ingen av strataene skal bli for små har vi slått sammen enkelte 2-siffer næringer. Følgende næringskoder på samme linje er slått sammen:

<u>Næringskode:</u>	<u>Navn på næringen:</u>
01-05	Jordbruk, skogbruk og fiske
10,12-14	Bergverksdrift
15-16	Nærings- og nytelsesmiddelindustri
17-19	Tekstil- og bekledningsindustri
23-25	Oljeraffinering, kjemisk industri mv.
30-33	Elektroteknisk og optisk industri
34-35	Transportmiddelindustri
36-37	Møbelindustri og annen industri
40-41	Kraft- og vannforsyning

Trekkepopulasjonen er også stratifisert etter **bedriftsstørrelse** målt som antall ansatte i bedriftene. Vi opererer med 5 ulike størrelsesstrata for hver næring. Trekk sannsynligheten øker fra størrelsesstratum til størrelsesstratum. Størrelsesstrata 1 med de minste bedriftene har trekk sannsynlighet lik 0. Dette kalles ofte for cut-off strata. Størrelsesstrata 5 med de største bedriftene trekk sannsynlighet lik 100 prosent.

På denne måten dekker undersøkelsen rundt 15 prosent av alle ansatte, mens vi sender ut skjema til rundt 6 prosent av bedriftene. Ansattgrensene for de ulike størrelsesstrataene varierer fra næring til næring, jamfør tabell 3.5.

Vi benytter cut-off strata for at SSB ikke skal belaste de aller minste bedriftene med en stor oppgavebyrde. Ved bruk av cut-off strata gjør vi da innenfor estimeringen en antagelse om at forholdet mellom egenmeldt sykefravær blant menn (kvinner) og antall ansatte menn (kvinner) er likt i de minste og i de nest minste bedriftene innenfor hver av næringene. Dvs. at vi estimerer også sykefraværet for de minste, dvs. de er med ved beregning av oppblåsningsfaktorene.

¹ Jamfør [Standard for næringsgruppering](#), NOS C182

2.2 Trekkingen

Alle bedrifter innenfor samme stratum har **lik trekk sannsynlighet**.

Utvalget **trekkes systematisk** fra en liste for hvert stratum der bedriftene er sortert etter fylke og deretter etter antall ansatte. Dette sikrer god representasjon fra alle landets fylker, selv om vi ikke stratifiserer på fylke direkte.

Systematisk trekking er valgt fordi vi ønsker å frita bedriftene så lenge som mulig fra undersøkelsen etter at de har vært med en periode. Metoden gjør det enkelt å etablere en rotasjonsplan for bedriftene til årlige utvalgsrullinger. Valget at systematisk trekking er også begrunnet i at det gir et tilnærmet tilfeldig utvalg med god fylkesvis spredning selv om vi kun stratifiserer på næring og antall ansatte.

Kort beskrivelse av metoden systematisk trekking

Som navnet indikerer, går systematisk trekking ut på å trekke enheter på en systematisk måte, og derfor strengt tatt på en ikke-tilfeldig måte. Fordelen ved metoden er at enhetene trekkes jevnt over fra hele populasjonen. Hovedfordelen med systematisk trekking er metodens operasjonelle enkelhet når utvalget trekkes fra en liste. Metoden er også fordelaktig om en ønsker en rotasjonsplan, der enheter som har vært med får fri så lenge som mulig.

Hvis $N/n = k$, der N er populasjonsstørrelsen og n er utvalgsstørrelsen, går metoden ut på å trekke én enhet blant de første k enhetene i populasjonen, og deretter trekke hver k 'te enhet deretter. Et element av tilfeldighet introduseres ofte ved at den første enheten trekkes tilfeldig blant de første k enhetene i populasjonen. Systematisk trekking kan sees på som en tilnærming til et tilfeldig utvalg dersom rekkefølgen på enhetene i listen ikke har betydning for det vi skal måle.

2.3 Utvalgsrulleringen

Populasjonen av bedrifter endrer seg kontinuerlig. Nye enheter oppstår, andre opphører. Enheter endrer størrelse og næring. Eierforhold blir endret. Denne tidsdimensjonen må derfor tas hensyn til ved oppdatering/rullering av utvalget og i estimeringen.

Vi **oppdaterer** (rullerer) bedriftsutvalget **en gang pr. år**, slik at det reflekterer endringene som har skjedd i bedriftspopulasjonen siste året.

Systemets rotasjonsgrupper forutsetter at trekk sannsynlighetene ligger fast fra år til år.

Bedriftene i utvalget skal være **med i 4 år om gangen og være fritatt så lenge som mulig** i forhold til trekk sannsynligheten. Vi gjør unntak fra 4 års regelen for en del **nye** bedrifter i populasjonen siden forrige trekking for at de ikke skal bli underrepresentert og fordi vi vil spre nye bedrifter jevnt utover rotasjonsgruppene. Andre unntak fra 4 års regelen er bedriftene i fulltellingsstrataene.

Årsaken til at vi vil ha bedriftene med i 4 år av gangen er at kostnaden ved tilrettelegging av personalsystemene i de enkelte bedriftene anses å kunne være relativt store i forhold til å fylle ut vårt lille spørreskjema en ekstra gang. Dessuten får utvalget gode panelegenskaper, som igjen gir gode endringstall.

3. Om prinsipper, metode, analyse og trekkpopulasjonen

3.1 Prinsipper for trekking av utvalget

Følgende fire prinsipper ligger til grunn for trekkingen av utvalget²:

- Utvalget **trekkes systematisk** fra en liste der bedriftene er sortert etter fylke og deretter etter antall ansatte. Sorteringen sikrer at vi får en god spredning av utvalget geografisk, selv om vi ikke stratifiserer på fylke.
- For hvert stratum gjøres det **én tilfeldig trekking** som brukes inntil vi endrer på selve trekksansynlighetene.
- Vi rullerer **hvert år** ut minst mulig av utvalget.
- Bedriftene i utvalget skal være **med i 4 år om gangen og være fritatt så lenge som mulig** i forhold til trekksansynligheten. Vi gjør unntak fra 4 års regelen for nye bedrifter i populasjonen for at de ikke skal bli underrepresentert og fordi vi vil spre nye bedrifter jevnt utover rotasjonsgruppene.

Vi har valgt å operere med noen få ulike trekksansynligheter, lik 0 eller 1 over toerpotens, dvs. $1/2^k$, der $k \in [0,6]$. Trekksansynlighetene blir da: 1, $1/2$, $1/4$, $1/8$, $1/16$, $1/32$, $1/64$ eller 0. Disse trekksansynlighetene håndterer smidig bedrifter som havner i forskjellige strata fra år til år. Strata endringer kan skyldes endring av næring eller at bedrifter får vesentlig færre eller flere ansatte.

Siden bedriftene er med i 4 år om gangen og utvalget rulleres hvert år, vil andelen bedrifter som til en hver tid er med i undersøkelsen i et stratum være 4 ganger større enn andelen bedrifter vi rullerer/trekker inn hvert år.

- Rullerer vi hver 8. bedrift hvert år, blir andelen vi til enhver tid har med i undersøkelsen $1/2$
- Rullerer vi hver 16. bedrift hvert år, blir andelen vi til enhver tid har med i undersøkelsen $1/4$
- Rullerer vi hver 32. bedrift hvert år, blir andelen vi til enhver tid har med i undersøkelsen $1/8$
- Rullerer vi hver 64. bedrift hvert år, blir andelen vi til enhver tid har med i undersøkelsen $1/16$
- Rullerer vi hver 128. bedrift hvert år, blir andelen vi til enhver tid har med i undersøkelsen $1/32$

Det siste prinsippet gjør at dersom vi hvert år rullerer:

- hver 8. bedrift, så er bedriftene fritatt i 4 år
- hver 16. bedrift, så er bedriftene fritatt i 12 år
- hver 32. bedrift, så er bedriftene fritatt i 28 år
- hver 64. bedrift, så er bedriftene fritatt i 60 år
- hver 128. bedrift, så er bedriftene fritatt i 124 år

Det tredje og fjerde prinsippet betyr til sammen at vi rullerer 25 prosent av utvalgsdelen hvert år.

Avhengig av sannsynligheten for å være med i utvalget, deler vi populasjonen (hvert strata) inn i et vist antall puljer:

- trekksansynlighet lik $1/2$ gir 8 puljer
- trekksansynlighet lik $1/4$ gir 16 puljer
- trekksansynlighet lik $1/8$ gir 32 puljer
- trekksansynlighet lik $1/16$ gir 64 puljer
- trekksansynlighet lik $1/32$ gir 128 puljer

² Hovedprinsippene bygger på metoden presentert i Leiv Solheim sitt upubliserte notat "Strukturstatistikk 2000 - forslag til samordning av utvalgsplanen".

Tabell 3.1 Oversikt over trekksannsynligheter og steglengde

Trekksannsynlighet, andelen som til enhver tid er med i utvalget	Trekksannsynlighet, andelen som til enhver tid er med i utvalget	To'er-potens, k	Steglengde/ antall puljenummer	Årlig innrullerings prosent
100 %	1	0	0	
50 %	1/2	1	8	12,5 %
25 %	1/4	2	16	6,25 %
12,5 %	1/8	3	32	3,125 %
6,25 %	1/16	4	64	1,5625 %
3,125 %	1/32	5	128	0,78125 %
1,5625 %	1/64	6	256	0,390625 %
0 %		∞		0 %

1) To'er-potensen er det tallet en må opphøye tallet 2 i for å få trekksannsynligheten (P), dvs. $P=1/2^k$.

Når vi har sortert populasjonen slik som beskrevet i første prinsipp kan vi tildele bedriftene puljenummer ved å gi den første bedriften 1, den neste 2, den neste 3, inntil 4, 8, 16, 32, 64 eller 128, avhengig av raden i tabell 3.1, inntil alle bedriftene har fått et puljenummer. Trekking og oppstart av rotasjonsplanen skjer da som følgende:

1. Vi trekker et startpunkt i hvert stratum ved å trekke ut et **tilfeldig** lovlig puljenummer. Denne puljen skal være med KUN 1 år.
2. Bedriftene med puljenummer lik 1 + det tilfeldige startnummeret, blir med i 2 år. (Er dette nummeret ulovlig dvs. høyere enn antall puljer "trekkes" puljenummer 1.)
3. Bedriftene med puljenummer lik 1 + puljenummeret til de som skal være med i 2 år, blir med i 3 år. (Er dette nummeret ulovlig dvs. høyere enn antall puljer "trekkes" puljenummer 1.)
4. Bedriftene med puljenummer lik 1 + puljenummeret til de som skal være med i 3 år, blir med i 4 år. (Er dette nummeret ulovlig dvs. høyere enn antall puljer "trekkes" puljenummer 1.)

Selve rulleringen skjer normalt ved at puljen tas ut etter fire år, og neste puljenummer i rekken tas inn. For enkelte puljer med nye bedrifter siden forrige rullering, vil bedrifter bli rullert ut av utvalget før de har vært med i 4 år.

Strata med få bedrifter

Dersom populasjonen av bedrifter innenfor et stratum er mindre enn antall lovlige puljenummer i stratumet, gir vi hver bedrift et helt tilfeldig valgt lovlig puljenummer. En bedrift blir da med i årets undersøkelse dersom det tilfeldige puljenummeret er mellom 1 og 4.

Tilfeldig trekking av hver og en enhet i disse spesielle strataene er valgt fordi systematisk trekking her ville gitt en klynge som til tider ville blitt overrepresentert og til andre tider underrepresentert.

3.2 Dynamikk i populasjonene

Dersom populasjonen var statisk ville det opplegget som er beskrevet i foregående avsnitt kunne gjennomføres. Det er imidlertid betydelig dynamikk i populasjonen. Vi skal her **skille mellom tre typer endringer** som fører til at vi må legge til flere operasjoner for at systemet skal fungere fra år til år

- Nye bedrifter etableres
- Gamle bedrifter opphører
- Bedrifter bytter stratum

3.2.1 Nye bedrifter i populasjonen siden forrige trekking

Vi definerer indikatoren **alder**, med verdien 'NY' dersom bedriften ikke var med i trekkpopulasjonen ved forrige trekking, og med verdien 'GAMMEL' dersom bedriften var med i trekkpopulasjonen ved forrige trekking.

Siden vi har innført **alder** (NY, GAMMEL) som et kriterium for å dele bedriftene inn i strata tas dette enkelt hånd om ved at alle nye bedrifter i trekkpopulasjonen siden forrige rullering tilordnes egne strata. Vi lar strataene med nye og gamle bedrifter, gitt næring og størrelsesgruppe, ha lik trekksannsynlighet og dermed likt antall puljer. Vi

trekker bedriftene i disse strataene slik som angitt i foregående avsnitt. (Stikkordene var stratainnndeling, puljeoppdeling, nummerering i puljer og trekking av puljenumre som skal være med 1,2,3 og 4 år.) Også for de nye bedriftene er vi nødt til å "trekke" med bedrifter som bare er med 1, 2 og 3 år, for at de nye bedriftene ikke skal bli underrepresentert, i og med at de gamle bedriftene er med i 4 år av gangen³.

Bedrifter som var NYE i trekkpopulasjonen ved forrige trekking/rullering karakteriseres som GAMLE fra og med neste rullering, uavhengig om enhetene ble trukket ut eller ikke.

3.2.2 Bedrifter opphører

Dette er også rimelig greit å håndtere siden disse rett og slett forsvinner. Problemet er at de reduserer antallet enheter i stratomet og i de puljene de forsvinner fra slik at utvalget blir mindre. I gjennomsnitt skal jo like mange forsvinne fra alle puljer i et stratum slik at dette bør ikke være et stort problem.

3.2.3 Bedrifter endrer stratum

Det er flere grunner til at bedriftene endrer stratum. Alle fire dimensjoner alder (NY/GAMMEL), næring og størrelse kan være årsak til at bedriften går fra et stratum til et annet. Når en bedrift går fra et stratum til et annet må vi skille mellom tre tilfeller, beskrevet under.

Samme omløpshastighet

Samme omløpshastighet i det nye stratomet, dvs. samme trekkprosent. Da beholder bedriften puljenummeret fra forrige trekking. (Dette er tilfellet ved "aldersendring" fra ny til gammel, gitt nærings- og størrelsesstrata.)

Høyere omløpshastighet

Høyere omløpshastighet i det nye stratomet, dvs. høyere trekkprosent og færre puljer.
Da gir vi de fleste bedriftene:

Nytt puljenummer = CEIL[gammelt puljenummer / $2^{k_{gml}-k_{ny}}$],

der CEIL betyr avrunding oppover til nærmeste heltall, og

k_{gml} og k_{ny} er toerpotensene som gir trekkssannsynlighetene i stratomet ved forrige og gjeldende trekking.

Bedrifter med gammelt puljenummer lik 1,2,3 eller 4 beholder puljenummeret, slik at bedriftene som nå er inne i utvalget er med i de avtalte 4 år. Dette er viktig, men ulempen ved dette er at avvikene fra den generelle formelen skaper litt vridninger i rotasjonsplanen over tid.

For å stabilisere rotasjonsplanen/mønsteret spesialbehandler vi utvalgte puljer:

Dersom $[k_{gml} - k_{ny}] > 1$ og

$\text{CEIL}[\text{gammelt puljenummer} / 2^{k_{gml}-k_{ny}}] \in (1,2,3)$ så settes Nytt puljenummer = 1

Dersom $[k_{gml} - k_{ny}] = 1$ og gammelt puljenummer = 5 så settes Nytt puljenummer = 1

Dersom $[k_{gml} - k_{ny}] = 1$ og gammelt puljenummer = 7 så settes Nytt puljenummer = 2

Lavere omløpshastighet

Lavere omløpshastighet i det nye stratomet, dvs. lavere trekkprosent og dermed også flere puljer.

For bedrifter med gammelt puljenummer lik 1,2,3 eller 4 har vi valgt å beholde puljenummeret.

Bedrifter med andre gamle puljenummer spres tilfeldig utover, ved at de får nye puljenummer lik et tilfeldig tall i mengden av lovlige puljenummer større enn 4, siden puljenumrene 1-4 er allerede overrepresentert.

³ En alternativ metode kunne vært å ha 4 ganger så høy trekkssannsynlighet for nye bedrifter. Da kunne alle nye bedrifter som var trukket ut være med i 4 år. Vi kunne imidlertid fått en mer ubalansert fordeling av puljenumrene og rotasjon, dersom tilsiget av nye bedrifter i de enkelte næringer over tid ikke er jevn. Ved rulleringen etter ville de nye bedriftene skifte strata til "gamle bedrifter" og da få en lavere rulleringssannsynlighet, og flere puljenummer bedriftene skal spres utover på, som beskrevet i eget avsnitt under.

Med sas-kode er den formelen vi har valgt⁴ å anvende:

Nytt puljenummer = $\text{CEIL}[\text{RANDUNI}(0) * (\text{steg} - 4)] + 4$,
der $\text{RANDUNI}(0)$ returnerer et tilfeldig tall mellom 0 og 1, og der *steg* er antall ulike tillatte puljenummer i stratumet, jamfør tabell 3.1.

Endrede treksanssynligheter kan også komme av at SSB har forandret dem på bakgrunn av analyse innenfor hvert strata av utvalgsstørrelse og relativt standardavvik.

3.3 Nøyaktighet og justering av treksanssynlighetene i 2006

Når vi nøyaktighetsmålene vi satt oss?

I rapporten fra pilotprosjektet var målsetningen å kunne observere endringer fra periode til periode på omlag 7 prosent for 2-siffer næring. Det impliserer⁵ at **standardavviket til estimatet i prosent av estimatet selv ikke skal overstige 2,47 prosent**. Ut i fra det regnet en med at et utvalg på 10 000 enheter var tilstrekkelig.

Med denne utvalgsstørrelsen regnet vi også med å kunne måle signifikante endringer i sykefraværet (1-3 dager, nå egenmeldt) på 3 prosent (0,03 prosentpoeng) med en oppdeling i statistikken på 15 næringsgrupper. Det tilsvarer at standardavviket til estimatet i prosent av estimatet selv ikke skal overstige 1,06 prosent for 15 næringsgrupper.

Beregningene over bygger på en antagelse om at utvalgene i de 2 periodene vi ser på endringer mellom er uavhengige. Panelegenskapene i utvalget gir trolig noe positiv kovarians, som skal kommer til fratrekk når en beregner usikkerhet til endringstall. Dette har vi imidlertid ikke tatt hensyn til⁶.

Tabell 3.2 viser estimert standardavvik og variasjonskoeffisienter for antall tapte dagsverk pga. egenmeldt sykefravær i kvartalet for næringsgruppene som vi i publiserer på www.ssb.no/sykefratot/ (tabell 3). Formlene som er benyttet til beregning av relativt standardavvik er gjengitt i vedlegg B.

Variasjonskoeffisientene i tabell 3.2 tyder ikke på at vi når de nøyaktighetsmålene vi satt oss. Ingen av de 12 publiserte næringsgruppene har en variasjonskoeffisient mindre enn 1,06 prosent.

Vedleggstabell A.1 viser relativt standardavvik for de enkelte strata og samlet. Utvalget er delt inn i 41 forskjellige næringsstrata, mens 2-siffer næring i standard for næringsgruppering består av 60 ulike grupper. Ut ifra det bør relativt standardavvik i siste kolonne (næringsstrata summer) i vedleggstabell A.1 i alle fall være mindre enn 2,47 prosent. Vedlegg A viser at vi 3. kvartal 2005 ikke nådde nøyaktighetsmålene som vi opprinnelig satt oss. Sannsynligvis nådde vi kun målet innenfor et næringsstrata, helse- og sosialtjenester. Gjennomsnittlig relativt standardavvik 3. kvartal 2005 var beregnet til 6 prosent. Uoppgitt i tabellen i vedlegg A kan ha ført til at vi undervurderer utvalgsusikkerheten i enkelte næringer litt.

⁴ En alternativ metode ville vært å multiplisere gammelt puljenummer med 2-er potensdifferansen og så trekke fra et tilfeldig tall mellom 0 og 2-er potensdifferansen (avrundet ned). Dvs. Nytt puljenummeret = gammelt puljenummer * $(2^{k_{ny} - k_{gml}})$ - $\text{FLOOR}[\text{RANDUNI}(0) * (2^{k_{ny} - k_{gml}})]$, der FLOOR betyr avrundet ned til nærmeste heltall. Den alternative metoden bevarer rotasjonsmønsteret bedre, men gir mindre tilfeldig utvalg gitt at bedriftene med gammelt puljenummer lik 1,2,3 eller 4 skal beholde sitt puljenummer.

⁵ Jamfør Notater 1997/25: 'Utredning av utvalgsplan for sentral sykefraværstatistikk' av Øystein Amundrud og Irene Tuveng

⁶ Slike kovariansberegninger blir imidlertid trolig tilsvarende de i vedlegg C.

Tabell 3.2 Estimert standardavvik og variasjonskoeffisienter etter kjønn og næring. 3. kvartal 2005.

Næring	Estimert standardavvik			Variasjonskoeffisienten		
	Totalt	Menn	Kvinner	Totalt	Menn	Kvinner
TOTALT	6 299,9	3 767,8	4 297,1	0,98	1,20	1,31
Jordbruk, skogbruk og fiske	415,9	328,7	203,3	8,72	10,17	13,20
Industri og bergverksdrift	1 808,4	1 491,2	708,3	1,84	2,00	2,96
Kraft- og vannforsyning	290,9	255,5	123,8	6,65	7,58	12,27
Bygge- og anleggsvirksomhet	1 745,7	1 702,1	247,4	3,41	3,49	10,06
Vareh., hotell og restaurantv,	2 767,3	1 839,8	1 821,2	2,86	3,75	3,82
Transport og kommunikasjon	921,8	773,5	316,2	2,24	2,75	2,45
Finansiell tjenesteyting, forsikring	631,6	329,5	493,6	4,69	6,89	5,68
Forr, tjenesteyting, eiendomsdrift	1 843,8	1 418,8	914,8	2,77	3,83	3,09
Off, adm og forsvar, sosialforsikr,	2 206,6	1 094,6	1 543,5	4,30	5,05	5,21
Undervisning	1 530,7	668,6	1 211,5	3,78	5,73	4,21
Helse-, sosialtjenester	3 628,3	1 001,4	3 045,1	2,39	4,21	2,38
Andre sosiale og personlige tjen,	656,1	386,6	475,6	3,11	4,81	3,65

Note: Variasjonskoeffisienten er estimert standardavvik til tapte dagsverk pga. egenmeldt sykefravær i prosent av tapte dagsverk pga. egenmeldt sykefravær.

Etter at vi innførte oppgaveplikt fra og med 2. kvartal 2003 har presisjonen i utvalgsundersøkelsen blitt bedre enn da undersøkelsen var frivillig. Svarprosenten i undersøkelsen er med oppgaveplikt opp i mot 90 prosent, mot tidligere nærmere 65 prosent etter purring. Nettutvalgsstørrelsen ser likevel ikke ut til å være tilstrekkelig til å nå opprinnelige presisjonsmål.

Selv om økt utvalgsstørrelse kunne ha vært ønskelig for å nå de opprinnelige mål, har vi valgt å beholde utvalgsstørrelsen på rundt ti tusen bedrifter.

Fokus i publiseringen er hovedsakelig på nivået og endringen i alt eget sykefravær, dvs. summen av egenmeldt og legemeldt sykefravær. I og med at legemeldt sykefravær bygger på registreringer av alle sykmeldinger utfylt av leger kan en si at det ikke hefter utvalgsusikkerhet av betydning ved tallene⁷. Fokuset over gjør at vi kan redusere kravet til utvalgsusikkerheten, i og med at egenmeldte sykefraværet utgjør om lag 15 prosent av alle tapte dagsverk pga. sykefravær. Utvalgsusikkerheten til det egenmeldte sykefraværet er da naturlig å se i forhold til størrelsen på summen av egenmeldt og legemeldt sykefravær, som vist i tabell 3.3.

Tabell 3.3 Estimert standardavvik til tapte dagsverk pga. egenmeldt sykefravær i prosent av tapte dagsverk pga. egenmeldt og legemeldt sykefravær, etter kjønn og næring. 3. kvartal 2005

Næring		Variasjonskoeffisienten		
		Totalt	Menn	Kvinner
01-99	TOTALT	0,11	0,14	0,14
01-05	Jordbruk, skogbruk og fiske	0,85	0,99	1,30
10,12-37	Industri og bergverksdrift	0,23	0,26	0,32
40-41	Kraft- og vannforsyning	0,83	0,97	1,47
45	Bygge- og anleggsvirksomhet	0,40	0,42	0,89
50-55	Vareh., hotell og restaurantv.	0,28	0,41	0,33
60-64	Transport og kommunikasjon	0,19	0,23	0,22
65-67	Finansiell tjenesteyting, forsikring	0,62	0,93	0,74
70-74	Forr, tjenesteyting, eiendomsdrift	0,37	0,59	0,35
75	Off, adm og forsvar, sosialforsikr.	0,56	0,69	0,65
80	Undervisning	0,38	0,58	0,42
85	Helse-, sosialtjenester	0,26	0,51	0,25
90-99	Andre sosiale og personlige tjen.	0,31	0,45	0,38

⁷ Overgangen fra legemeldte sykefraværsdager til dagsverk går via den **berregnede** variabelen stillingsandel. På personnivå bruker vi så langt kun informasjon om arbeidsforholdet er heltid, kort- eller lang deltid fra arbeidstakerregisteret. Ved å sammenligne 37,5 timer pr. uke med hva heltid, kort- eller lang deltidsansatte fordelt på kjønn svarte i AKU at de i gjennomsnitt hadde som fast/avtalt arbeidstid pr. uke får vi stillingsandelene. Denne generaliseringen gir feil for mange enkeltpersoner og enkeltbedrifter, men gir mindre feil på aggregert nivå.

Justering og avrunding av trekksannsynlighetene i 2006

Trekksannsynlighetene avrundes i hvert stratum til nærmeste ekte brøk med nevner lik en toerpotens $1/2^K$, der $K \in [1, 7]$. I tillegg opererer vi med trekksannsynlighetene 0 og 1. Trekkesannsynlighetene kan der for være 1, $1/2$, $1/4$, $1/8$, $1/16$, $1/32$, $1/64$, $1/128$ eller 0.

Til utvalgstrekkningen for 2006 er noen av trekksannsynlighetene justert i forhold til trekksannsynlighetene for 2005. Årsaken er at trekkpopulasjonen etter størrelses- og næringsstrata har endret seg, slik at vi ville ha endt opp med for få bedrifter dersom vi hadde benyttet trekksannsynlighetene fra 2005. Enkelte stratum har derfor fått økt trekksannsynlighet fra og med trekkingen 2005. Ved oppjusteringen av trekksannsynlighetene har vi vektlagt størrelsen på relativt standardavvik i de enkelte strata, og at enkelte strata ikke skal bli for små, helst over 25 bedrifter⁸. Ved å øke trekksannsynligheten i strata med høyest relativt standardavvik (og evt. redusere trekksannsynligheten i strata med lavt relativt standardavvik) kan vi re-allokere utvalget slik at kommende observasjoner settes inn der de kaster mest av seg.

⁸ Minst 20 bedrifter i utvalget i hvert strata (der det ikke er fulltelling) gjør rate-estimatorens forventningsskjevhet neglisjerbar.

Tabell 3.4 Teoretiske trekksannsynligheter 2006, etter størrelses- og næringsgruppe.

Næringsstrata	Næringsstratanr.	Størrelsesgruppe				
		1	2	3	4	5
Jordbruk, skogbruk og fiske	1	0	12,5	50	100	100
Bergverksdrift	2	0	25	50	100	100
Olje- og gassutvinning	3	0	25	100	50	100
Nærings- og nytelsesmiddelindustri	4	0	3,125	12,5	25	100
Tekstil- og bekledningsindustri	5	0	12,5	50	100	100
Trelast- og trevareindustrien	6	0	6,25	25	50	100
Treforedling	7	0	50	50	50	100
Forlag og grafisk industri	8	0	12,5	25	50	100
Oljeraffineri, kjemisk industri mv.	9	0	25	25	50	100
Mineralproduktindustri	10	0	12,5	50	50	100
Metallindustri	11	0	50	50	50	100
Metallvareindustri	12	0	6,25	25	50	100
Maskinindustri	13	0	6,25	12,5	50	100
Elektroteknisk og optisk industri	14	0	12,5	25	100	100
Transportmiddelindustri	15	0	12,5	12,5	50	100
Møbelindustri og annen industri	16	0	12,5	25	100	100
Kraft- og vannforsyning	17	0	6,25	25	50	100
Bygge- og anleggsvirksomhet	18	0	3,125	12,5	25	100
Motorkjøretøytjenester	19	0	3,125	12,5	25	100
Agentur- og engroshandel	20	0	3,125	12,5	25	100
Detaljhandel og reparasjon av varer	21	0	1,5625	6,25	25	100
Hotell- og restaurantvirksomhet	22	0	3,125	25	50	100
Landtransport og rørtransport	23	0	6,25	50	50	100
Lufftransport	25	0	50	100	100	100
Tjenester tilknyttet transport	26	0	3,125	25	50	100
Post og telekommunikasjon	27	0	6,25	25	50	100
Finansiell tjenesteyting	28	0	6,25	12,5	100	100
Forsikring og pensjonsfond	29	0	25	50	100	100
Hjelpevirks. finansiell tj.yting	30	0	25	50	50	100
Eiendomsdrift	31	0	12,5	50	100	100
Utleievirksomhet, maskiner og utstyr	32	0	25	50	100	100
Databehandlingsvirksomhet	33	0	3,125	12,5	50	100
Forskning og utviklingsarbeid	34	0	25	25	50	100
Annen forretn. tjenesteyting	35	0	3,125	25	50	100
Off. adm. og forsvar, sosialforsikr.	36	0	3,125	6,25	50	100
Undervisning	37	0	3,125	6,25	12,5	100
Helse- og sosialtjenester	38	0	3,125	12,5	50	100
Kloakk- og renovasjonsvirksomhet	39	0	12,5	50	100	100
Interesseorganisasjoner	40	0	12,5	50	50	100
Kultur og sport	41	0	6,25	25	50	100
Annen personlig tjenesteyting	42	0	12,5	50	50	100

Note: Endrede trekksannsynligheter er markert med fet skrift. Alle endringene i trekksannsynlighetene er opp ett nivå.

3.5 Om trekkpopulasjonen for 2006

Trekkpopulasjon for 2006 er definert ut fra siste **situasjonsuttaket fra Bedrifts- og foretaksregisteret (BoF)** i desember 2005.

Seleksjoner i BoF

Enhetene i undersøkelsen er **bedrifter**, *rec_type* =2
Bedriftene skal være i **ordinær drift** *status* =B (dvs. ikke slettet)

Vi begrenser massen til registertypene:

**Enbedriftsforetak,
Flerbedriftsforetak,
Hjelpebedrifter.**

Dvs. hhv. *reg_type*⁹ IN (01,02,04)

Bedriftene skal **ikke** være **under slettede foretak** eller under foretak som ikke har utsendingsadresse.

Dvs. *status_foretak*¹⁰ ULIK 'S'

Bedriftene vil da hovedsakelig være under foretak som ikke er slettet ('B'), og i noen tilfeller under foretak som er slettet for fusjon/fisjon ('F') eller slettet som dublett('D'). Bedrifter under foretak med de to siste status typene vil også være med i trekkpopulasjonen, men vil bli ligge på vent om de trekkes ut inntil deres skjebne er avgjort.

Av kvalitetshensyn har vi, som i det opprinnelige utvalget, holdt enkelte næringsgrupper utenfor trekkgrunnlaget;

95 Lønnet arbeid i private husholdninger
99 Internasjonale organer og organisasjoner

Siden bedriftene i næringene under ikke fører statistikk over egenmeldt sykefravær som dekker alle ansatte (f.eks. sjøfolk og offshore arbeidere) dropper vi dem fra trekkpopulasjonen:

01.42 Tjenester tilknyttet husdyrhold¹¹
11.1 Utvinning av råolje og naturgass
61 Sjøtransport

Vernede bedrifter er avgrenset næringene til:

85.334 Arbeidstrening for ordinært arbeidsmarked
85.335 Varig vernet arbeid

Vernede bedrifter holdes også utenfor trekkpopulasjonen. Årsaken er at rapportert sykefravær fra vernede bedrifter tyder på at de inneholder sykefraværstall også for de handikappede. Som regel arbeider de handikappede innenfor en annen næringen enn helse og sosialtjenester, ofte med industriarbeid. Videre kan det se ut som mange av disse handikappede ikke er meldt inn i arbeidstakerregisteret. Derfor blir det også vanskelig å la de komme med under en annen næringsgruppe¹².

Kort oppsummert betyr det at vi **ikke** tar med bedrifter der

nace1 IN (00 , ' ' , 01.420 , 11.1 , 61 , 85.334 , 85.335 , 88 , 95 , 98 , 99)

Tilkobling av siste bedriftsaggregerte arbeidstakerregister

⁹ Det betyr at vi dropper AAFY (virksomhet til ikke-næringsdrivende person) *reg_type*=05 og *reg_type*=06 (trolig ubetjente bensinstasjoner).

¹⁰ Status='F' betyr slettet for fusjon/fisjon.

¹¹ Næring 01.42 omfatter Avløserlag, Dyrepensjonater, Husdyravl, servicevirks., Inseminasjon, Jordbruksavløser, Klipping, dyr, Klovklipping, Servicevirksomhet, husdyravl.

Vi kobler til et 4-ukersuttak fra **arbeidstakerregisteret** med referansedato 25/10-2005 aggregert opp på bedriftsnivå. Vi får da en optelling av antall ansatte i bedriftene pr. 25/10-2005. Det er imidlertid satt et tak på hvor mange arbeidsforhold kan ha samtidig.

Trekkpopulasjonen avgrenses videre til **bedrifter med arbeidstakere** ifølge koblingen.

Inndeling i strata

I utgangspunktet vil vi ha de samme strata inndelingene i det opprinnelige utvalget, dokumentert i Notat 98/60 tabell 2 og 3. Trekkpopulasjonen deles derfor inn i 41 **næringsgrupper** og 5 **størrelsesgrupper**. Bedriftene med færrest ansatte er i størrelsesgruppe 1, kalt **cut-off** strata. Ingen i cut-off strataene blir trukket ut til å være med i undersøkelsen. Bedriftene med flest ansatte er i størrelsesgruppe 5, også kalt **fulltellingsstrata**. Alle bedriftene i fulltellingsstrataene skal være med, hele tiden. Antall ansatte som skal til for at bedrifter havner i de ulike størrelsesgruppene avhenger av hvilken næringsgruppe bedriftene hører hjemme i, jamfør tabell 3.5.

Tabell 3.5 Inndeling av antall ansatte i ulike størrelsesgrupper etter næring.

Næring, 2-siffer NACE	Størrelsesstrata					type betegnelse
	1, "cut-off"	2	3	4	5, fulltelling	
52 , 67 , 91 , 93	1-3	4-9	10-19	20-49	50 +	A
50	1-4	5-9	10-19	20-49	50 +	B
01 , 02 , 05 , 17 - 19 , 51 , 70	1-3	4-19	20-49	50-99	100 +	C
45 , 55	1-4	5-19	20-49	50-99	100 +	D
71 , 72 , 74 , 92	1-3	4-19	20-49	50-149	150 +	E
10 , 12-16 , 20 , 22 , 26 , 28 , 29 , 36	1-4	5-19	20-49	50-149	150 +	F
37 , 40 , 41 , 60 , 63 , 80 , 90	1-4	5-19	20-49	50-149	150 +	F
23 - 25 , 30 - 33 , 61 , 66 , 73 , 75 , 85	1-4	5-19	20-99	100-299	300 +	H
21 , 27 , 34 , 35 , 64 , 65	1-4	5-19	20-99	100-399	400 +	I
11 , 62	1-4	5-49	50-99	100-999	1000 +	J

I tillegg vil vi benytte en stratifiseringsvariabel **alder**, som indikerer om bedriftene var med i trekkpopulasjonen ved forrige trekking/rullering. Bedriftene som var med i trekkpopulasjonen ved forrige ved forrige trekking/rullering defineres som "gamle", de andre som "nye".

4. Dokumentasjon av oppdateringsrutinene

Rulleringen av utvalget bygger på at trekkesannsynlighetene innen hvert stratum skal være konstant. Over lang tid kan dette medføre at utvalget bli noe ubalansert. En bør derfor holde øye med utviklingen i de ulike strataene. Det er viktig at utvalgsstørrelsen ikke blir alt for liten i strata (der det er flere å ta av). Dessuten bør ikke utvalgsstørrelsen avvike alt for mye fra 10 000. (+/- 150 går sikkert greit)

Erfaring med sas, og kjennskap til undersøkelsen er forkunnskaper som forventes av brukeren. Systemet settes i gang ved å kjøre sas programmene på unix i henhold til fysisk dataflytdiagram, etter å ha fått tilgang til filene beskrevet i diagrammet.

Eventuelle usikkerhet tilknyttet løsningen

Rotasjonsplanen bygger på at trekkesannsynlighetene innen hvert strata er konstant. Endringer i trekkpopulasjonen kan medføre at enkelte strata blir alt for store eller små.

Justering av trekkesannsynlighetene rokker ved selve rotasjonsmønsteret. En bør vurdere nøye før en foretar endringer. Et av målene ved utvalgsplanen er å la bedrifter som har vært med i undersøkelsen få pause så lenge som mulig før de trekkes inn igjen. Dersom trekkesannsynligheten i et strata endres, vil også antall ulike puljenummer i stratumet forandres.

I utgangspunktet benyttes trekk sannsynlighetene fra i fjor.

Alternativt justeres trekkesannsynlighetene i enkelte strata direkte i sas-program 2 (*Les_inn_trekksannsynlighetene2006.sas*).

Ved evt. justering av trekk sannsynlighetene er det en fordel å analysere tabellen over *relativt standardavvik etter strata*. Se eget dataflytdiagram i kapittel 4.2.2 for tabellproduksjon.

Ansvarsforhold rundt systemet:

- EDB-ansvarlig og kontaktperson ved drift og systemfeil er Jørn Ivar Hamre. Vi har tidligere fått hjelp av Audun Rusti, S.410, til å oppdatere Delregisteret med andre variable, se kapittel 4.3.3.
- Prosjektleder er Kristoffer Vetvik.

Når programmene beskrevet i kapittel 4.2 er kjørt skal:

- Utvalgsfila og trekkpopulasjonen kontrolleres. Sjekk at kontrolltabellene ser fornuftig ut, samt en liten mikrodatasjekk av utvalgsfila.
- Enhetene i utvalget lastes inn i delregisteret, jamfør kapittel 4.3.1 og 4.3.2. Mer generell informasjon om delregisteret ligger på katalogen: Q:\DOK\Bof\Delregister\Info\
- Utvalgsfila legges på et fellesområde (for eksempel **\$FELLES/wk2/ham/**) slik at andre hjelpevariable til administrasjon av datafangsten kan lastes inn i delregisteret, se kapittel 4.3.3. S260 har ikke laget et datasystem for dette. Audun Rusti ved Seksjon 410 har tidligere vært behjelpelig. Send e-post til Audun Rusti og Per Erik Gjedtjernet, med kopi til Karin Synnøve Øiseth, om hjelp til oppdatering av delregisteret med andre hjelpevariable til datafangstadministrasjon.
- Flate arkivfiler (generert av sas-program nr. 6) flyttes til arkivområdet i henhold til retningslinjene i Interne dokumenter 99/6, alternativt Q:\DOK\DATADOK\INFO\Brukerbeskrivelse_online.doc. Ny metadata legges inn i DATADOK

Referanser til annen ekstern og intern dokumentasjon som kan være av interesse:

- Notat 2003/92 Undersøkelsen om egenmeldt sykefravær. Dokumentasjon av utvalgsplan, trekking og rullering for 2003
- Notat 2005/45 Estimering av fylkesfordelte og sektorfordelte tall for egenmeldt sykefravær. Dokumentasjon av metode og system, og resultatet.
- Mer informasjon om sykefraværstatistikken finns på området: <http://www.ssb.no/sykefratot/om.html>
- Q:\DOK\UTVALG\260\214H_DOK.DOC (teknisk intern dokumentasjon av utvalget trukket i 1997)

4.1 Tidspunkt for rullering

Utvalget skal rulleres hvert år i forkant av skjemautesendingene i 1. kvartal.

Nye enheter bør helst få informasjon om at de skal være med ved inngangen av kvartalet de skal svare for, slik at de har litt tid å forberede rapportuttak.

I februar hvert år sender Seksjon for bedriftsregister ut informasjonsbrev til alle foretak om forventet oppgaveplikt for kommende år. For at nytt utvalg skal være grunnlag for denne utsendingen må dette være lagt inn i delregisteret innen midten av januar.

I og med at vi har innført oppgaveplikt i undersøkelsen fra og med 2. kvartal 2003, plikter SSB å sende ut et varsel- og informasjonsbrev til alle nye bedrifter om at de er trukket ut i forkant av skjemautesendingen.

Ut i fra dette bør utvalget rulleres ferdig og legges inn i delregisteret før nyttår. Videre bør informasjonsbrevet til nye bedrifter sendes ut så tidlig som mulig i 1. kvartal. Vedlegg F inneholder utsendt informasjonsbrev til nye bedrifter.

4.2 Datafilene, programmene og fysisk dataflytdiagram

4.2.1 Data systemet krever

Systemet krever 4 inputfiler, som er beskrevet under. Filene brukes i henhold til fysisk dataflytdiagram under.

Systemet trenger siste tilgjengelige situasjonsuttak fra BoF. Ved utvalgstrekkningen 2006 ble følgende fil brukt:

\$DSB/sbbofsit/wk6/sit200512b.sas7bdat

Systemet trenger også trekkpopulasjonen fra foregående trekking. Ved utvalgstrekkningen 2006 ble trekkpopulasjonen for 2005 brukt:

\$SYKEFR/utvalg/trekking/wk24/popmednyepuljenr200412b.sas7bdat

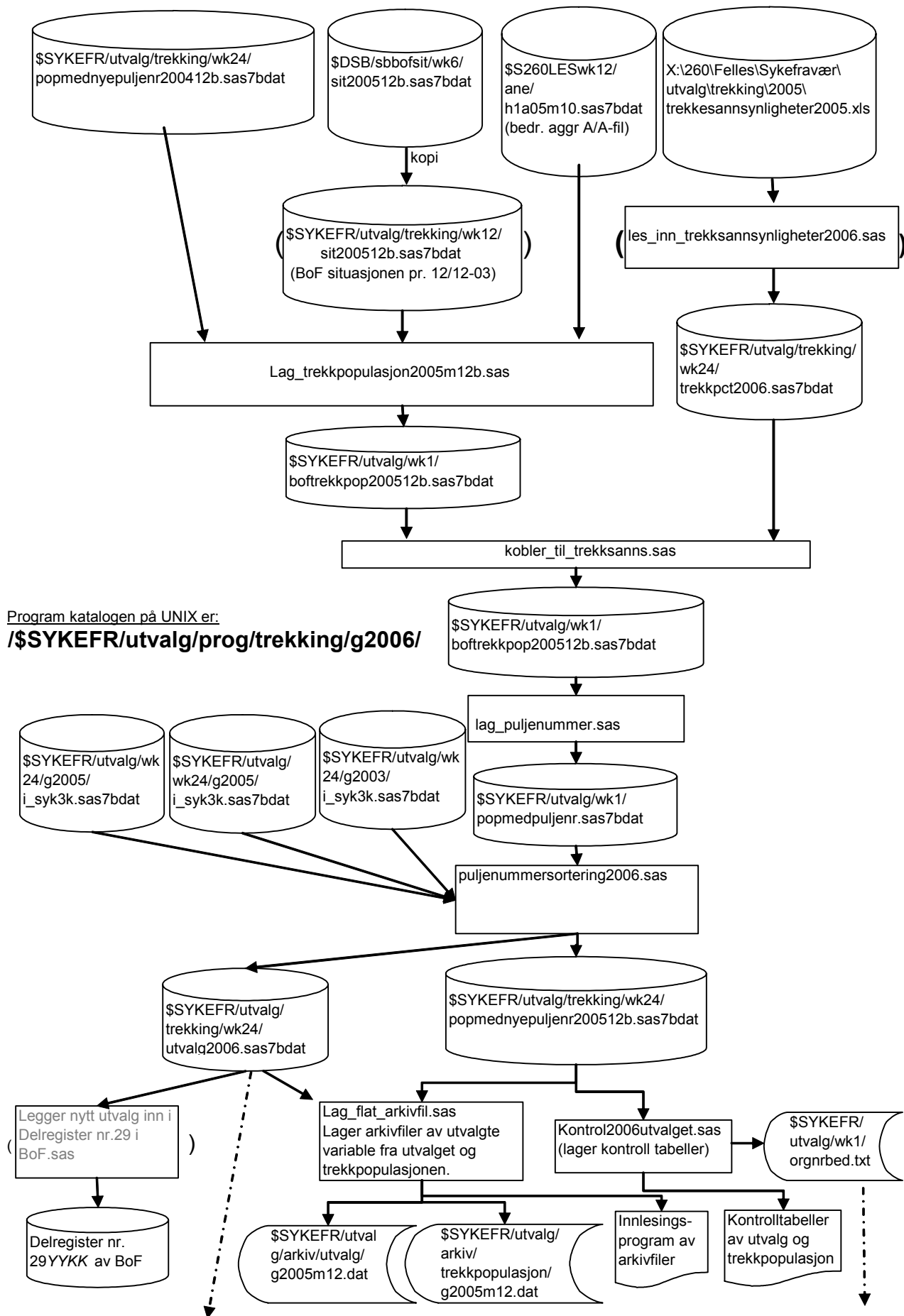
Systemet trenger også siste fil fra arbeidstakerregisteret bearbeidet for den registerbaserte sykefraværstatistikken. Ved utvalgstrekkningen 2006 ble følgende fil brukt:

\$S260LES/wk12/ane/h1a05m10.sas7bdat

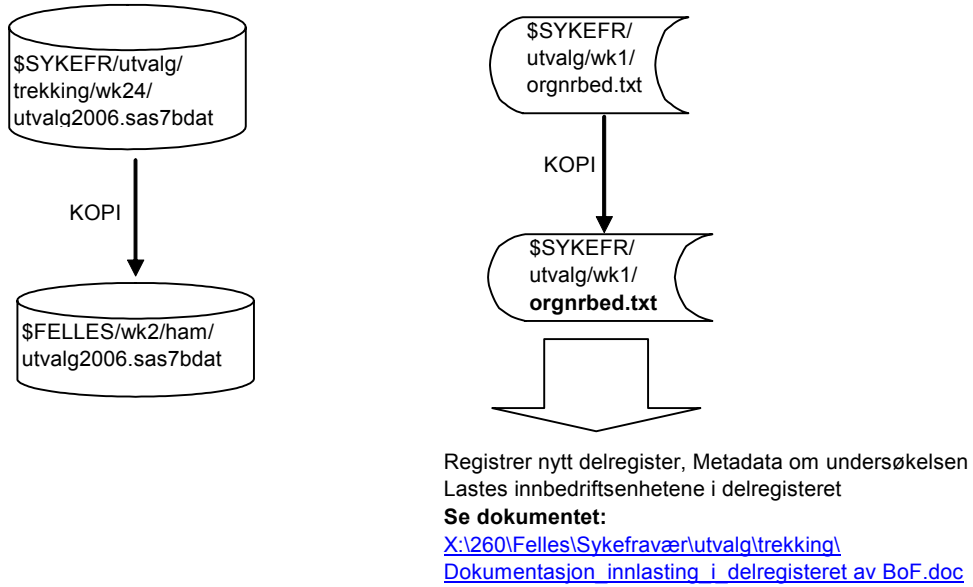
Systemet trenger også en fil med trekksannsynligheter. Filen vil være den samme fra år til år, og heter:

X:\260\Felles\Sykefravær\utvalg\trekking\2005\trekkesannsynligheter2005.xls

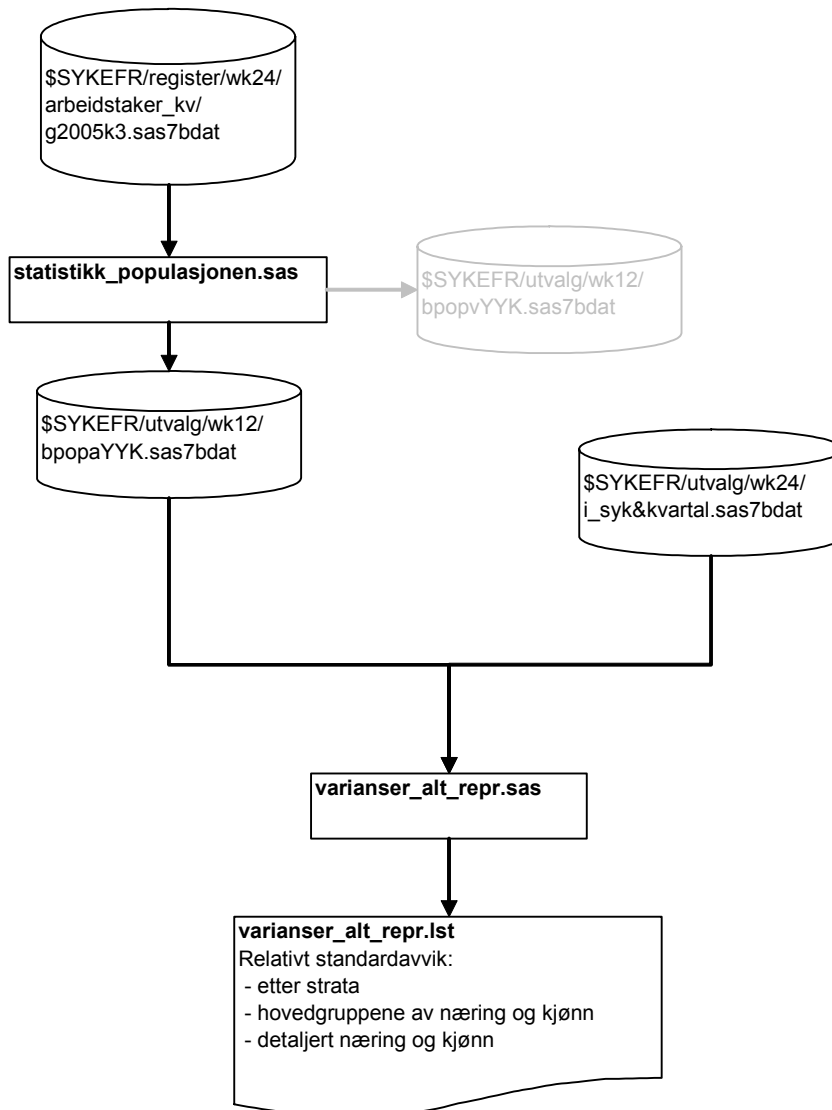
4.2.2 Fysisk dataflytdiagram for utvalgstrekkningen



Fysisk dataflytdiagram for innlasting av utvalget i delregister 290106



Fysisk dataflytdiagram for beregning av relativt standardavvik for tapte dagsverk i kvartalet grunnet egenmeldt sykefravær



4.2.3 Beskrivelse av programmene

Systemet består av syv sas-programmer, som er beskrevet under. Programmene er lagret på UNIX under katalogen:

SSYKEFR/utvalg/prog/trekking/g2006/

Programmene kjøres i SAS på unix i rekkefølgen under, og i henhold til dataflytdiagrammet over.

Program 1

```
/******  
*   Produktnummer .....: 2113                                     *  
*   Prosjekt .....: Sentral sykefraværstatistikk                 *  
*   Programkatalog .....: $SYKEFR/utvalg/prog/trekking/g2004/   *  
*   Programnavn .....: Lag_trekkpopulasjon2004m12v1.sas        *  
*   Skrevet av .....: JØRN IVAR HAMRE, S.260                    *  
*   Dato .....: 2003/02/12                                       *  
*   Versjon .....: 1.2                                           *  
*   Programmets funksjon .: Selektorer trekkpopulasjonen fra BoF. *  
*   : Kobler til Arbeidstakerregisteret for oppdatering av     *  
*   : antall ansatte.                                           *  
*   : Lager stratifiseringsvariablene: næringsgruppe (1-42),  *  
*   : størrelse (1-5)                                           *  
*   :                                                           *  
*   : Kobler til forrige trekkpopulasjon ( 2003),             *  
*   : lager stratifiseringsvariablene: bedriftsalder (NEW/OLD) *  
*   :                                                           *  
*   Anmerkninger .....:                                         *  
*   :                                                           *  
*   Filer inn .....: $SYKEFR/utvalg/wk24/bofsitv&yyyy.&mm.&bokstav. *  
*   : $S260LES/wk12/ane/h1aYYmMM                               *  
*   Filer ut .....: $SYKEFR/utvalg/wk1/boftrekkpop&yyyy.&mm.&bokstav. *  
*   :                                                           *  
*   Endret av/når .....: HAM, 2003-12-12                         *  
*   Endring .....: Utvider avgrensingen av vernede bedrifter til NACE-kode*  
*   : 85.334 og 85.335. Disse selekteres vekk fra trekkpop'n *  
*   :                                                           *  
*   Endret av/når .....: HAM, 2003-12-12                         *  
*   Endring .....: Selektorer bedr. med ansatte ut ifra siste bedrifts- *  
*   : aggregerte arbeidstakerregisterfil (4-ukersuttak).     *  
*   : Vi kobler selv til fila PAU sender til S410 for å få    *  
*   : den absolutt siste tilgjengelige fila.                 *  
*   :                                                           *  
*   Endret av/når .....: HAM, 2003-12-12                         *  
*   Årsak til endring ....:                                       *  
*   :                                                           *  
*****/
```

Program 2

```
/******  
* Produktnummer .....: 2113 *  
* Prosjekt .....: Sentral sykefraværstatistikk *  
* Programkatalog .....: $SYKEFR/utvalg/prog/trekking/g2006/ *  
* Programnavn .....: Les_inn_trekksannsynlighetene2006.sas *  
* : *  
* Skrevet av .....: JØRN IVAR HAMRE, S260 *  
* Dato .....: 16.12.2004 *  
* Versjon .....: 1.1 *  
* Programmets funksjon .: Leser inn trekksannsynlighetene til de ulike nærings- og *  
* : størrelsesstrataene. *  
* : Sannsynlighetene er utarbeidet i et *  
* : excel regneark på bakgrunn av de oppriinnelige sannsynlig- *  
* : hetene, analyse av relativt standardavvik i de ulike *  
* : næringene, og et ønske om robuste strata størrelsesmessig, *  
* : og en avrunding til  $p=1/(2*k)$  der k er et tall 1,2,..,7 *  
* : eller 0 eller uendelig, som gir  $p=0$ . *  
* : *  
* Anmerkninger .....: Disse trekksannsynlighetene BØR ligge fast en del år *  
* : med mindre nåværende %'er gir veldig skjeve, store, små utv. *  
* : (Endrede %'er forandrer rulleringssystemet) *  
* : *  
* Endret av og når : HAM 2005.12.19 *  
* : *  
* : *  
* Forandring : TREKKSANNSYNLIGHETENE I 6 celler er justert opp *  
* : *  
* Filer ut .....: $SYKEFR/utvalg/trekking/wk24/trekkpct2006 *  
* : *  
*****/
```

Program 3

```
/******  
* Produktnummer .....: 2113 *  
* Prosjekt .....: Sentral sykefraværstatistikk *  
* Programnavn .....: $SYKEFR/utvalg/prog/trekking/g2003/kobler_til_trekksanns.sas*  
* Skrevet av .....: JØRN IVAR HAMRE, S260 *  
* Dato .....: 2003-02-20 *  
* Versjon .....: 1.1 *  
* Programmets funksjon .: Kobler trekksannsynlighetene på trekkpopulasjonen *  
* : Definerer steglengde dvs. antall puljer bedriftene i *  
* : det enkelte strata deles inn i. *  
* : *  
* : Definerer også variabelen stratum. *  
* : *  
* Filer inn .....: $SYKEFR/utvalg/trekking/wk24/trekk24.trekkpct2004 (FAST) *  
* Filer inn .....: $SYKEFR/utvalg/trekking/wk1/boftrekkpop200302b.sas7bdat *  
* : *  
* Filer ut .....: $SYKEFR/utvalg/trekking/wk1/boftrekkpop200302b.sas7bdat *  
* : *  
*****/
```

Program 4

```
/******  
*   Produktnummer .....: 2113                                     *  
*   Prosjekt .....: Sentral sykefraværstatistikk                   *  
*   Programnavn .....: $SYKEFR/utvalg/prog/trekking/g2004/lag_puljenummer.sas *  
*   Skrevet av .....: JØRN IVAR HAMRE, S260                       *  
*   Dato .....: 2003/02/25                                        *  
*   Versjon .....: 1.3                                           *  
*   Programmets funksjon .: A) Programmet tildeler alle bedrifter i populasjonen et *  
*   : puljenummer som avgjør rullering.                             *  
*   :                                                                     *  
*   : Vi deler opp populasjonen i 6 deler som behandles hver     *  
*   : seg:                                                           *  
*   *                                                               *  
*   1. BEDRIFTENE SOM VAR MED I POPULASJONEN VED FORRIGE RULLERING *  
*   DER TREKKSANNSYNLIGHETEN og steglengden ER UENDRET.          *  
*   (De beholder sitt gamle puljenummer.)                         *  
*   *                                                               *  
*   2. BEDRIFTENE SOM VAR MED I POPULASJONEN VED FORRIGE RULLERING *  
*   DER TREKKSANNSYNLIGHETEN HAR ØKT dvs. STEGLENGDEN er REDUSERT *  
*   pga. endret stratum                                           *  
*   *                                                               *  
*   3. BEDRIFTENE SOM VAR MED I POPULASJONEN VED FORRIGE RULLERING *  
*   DER TREKKSANNSYNLIGHETEN HAR FALT dvs. STEGLENGDEN er ØKT    *  
*   pga. endret stratum                                           *  
*   *                                                               *  
*   4. "NYE" BEDRIFTER I POPULASJONEN TILKOMMET ETTER FORRIGE RULLERING *  
*   og "gamle" bedrifter fra fulltelling- eller cut-off strataene *  
*   Dersom de enkelte delpopulasjonene er større enn steglengden yy>=1 *  
*   (Bedriftene, som i forkant er sortert etter fylke og deretter ant.ansatte, *  
*   innenfor hvert strata får løpende puljenummer mellom 1 og steglengden. *  
*   Startverdien (første obs.) er tilfeldig trukket mellom 1 og steglengden, *  
*   som avhenger av strataets faste trekksannsynlighet.)          *  
*   Enheter som tilfeldigvis får puljenummer 5 skal være med i 4 år *  
*   Enheter som tilfeldigvis får puljenummer 4 skal være med i 3 år *  
*   Enheter som tilfeldigvis får puljenummer 3 skal være med i 2 år *  
*   Enheter som tilfeldigvis får puljenummer 2 skal være med i 1 år *  
*   Enheter som tilfeldigvis får puljenummer 1 får fri lengst mulig *  
*   Enheter som tilfeldigvis får puljenummer 6 rullereds inn neste år *  
*   Enheter som tilfeldigvis får puljenummer 7 rullereds inn om 2 år *  
*   osv... (jamfør puljenummersortering.sas)                       *  
*   *                                                               *  
*   5. "NYE" BEDRIFTER I POPULASJON TILKOMMET ETTER FORRIGE RULLERING *  
*   og "gamle" bedrifter fra fulltelling- eller cut-off strataene *  
*   DERSOM DELpopulasjonene IKKE ER STØRRE ENN STEGLENGDEN, yy<1 ev. yy2aar<1 *  
*   (Disse gis puljenummer HELT TILFELDIG mellom 1 og steglengden) *  
*   *                                                               *  
*   6. BEDRIFTENE i CUT-OFF STRATAENE og FULLTELLING STRATAENE    *  
*   *                                                               *  
*   Endret av :                                                     *  
*   Endret når :                                                   *  
*   Endringen består i :                                         *  
*   :                                                             *  
*   Filer inn .....: $SYKEFR/utvalg/wk1/boftrekpop&yyyy.&mm.&bokstav. *  
*   Filer ut .....: $SYKEFR/utvalg/wk1/popmedpuljenummer&yyyy.&mm.&bokstav. *  
*   :                                                             *  
*****/
```

Program 5

```
/******  
* Produktnummer .....: 2113 *  
* Prosjekt .....: Sentral sykefraværstatistikk *  
* Programnavn .....: $SYKEFR/utvalg/prog/trekking/g2006 *  
* Programnavn .....: puljenummersortering2006.sas *  
* : *  
* Skrevet av .....: JØRN IVAR HAMRE, S260 *  
* Dato .....: 2003/02/25 *  
* Versjon .....: 1.2 *  
* Programmets funksjon .: Programmet justerer alle puljenummer bortsett fra *  
* : i fulltelling- og cut-off strataene. *  
* : *  
* : De med puljenummer 1 får NYTT puljenummer lik *  
* : steglengden. *  
* : *  
* : Alle andre puljenummer får NYTT puljenummer *  
* : lik puljenummer minus 1. *  
* : *  
* : Dermed fortsetter følgende å stemme: *  
* : *  
* : Nyttpuljenummer=' 1' er med i 1 år *  
* : Nyttpuljenummer=' 2' er med i 2 år *  
* : Nyttpuljenummer=' 3' er med i 3 år *  
* : Nyttpuljenummer=' 4' er med i 4 år *  
* : Nyttpuljenummer=' 5' IKKE med i år, rulleres inn om 1 år*  
* : Nyttpuljenummer=' 6' IKKE med i år, rulleres inn om 2 år*  
* : Nyttpuljenummer=' 7' IKKE med i år, rulleres inn om 3 år*  
* : *  
* : Programmet lager også en variabel: utvalgAAAA *  
* : der AAAA står for året trekkpopulasjonen skal gjelde for *  
* : *  
* : utvalgAAAA indikerer om bedriften er med i utvalget i år.*  
* : utvalgAAAA=0 ==> IKKE MED I ÅR *  
* : utvalgAAAA=1 ==> ER MED I ÅR *  
* : utvalgAAAA=2 ==> ER MED I ÅR OG 1 ÅR TIL *  
* : utvalgAAAA=3 ==> ER MED I ÅR OG 2 ÅR TIL *  
* : utvalgAAAA=4 ==> ER MED I ÅR OG 3 ÅR TIL *  
* : utvalgAAAA=5 ==> ER MED I ÅR OG FREMOVER *  
* : (dvs. fulltelling) *  
* : *  
*****  
* HUSK Å OPPDATERE MED %LET UTV MED EN VERDI høyere enn til nå tildelte *  
* puljenummer men KUN 5 SIFFER OG LEGG TIL EN NY I_SYK-FIL HVERT ÅR *  
* *  
*****  
* Filer inn .....: $SYKEFR/utvalg/trekking/wk1/ *  
* : popmedpuljenr&yyyy.&mm.&bokstav..sas7bdat *  
* : *  
* : $SYKEFR/utvalg/wk24/g2001/i_syk3k.sas7bdat *  
* : $SYKEFR/utvalg/wk24/g2002/i_syk3k.sas7bdat *  
* : $SYKEFR/utvalg/wk24/g2003/i_syk3k.sas7bdat *  
* : $SYKEFR/utvalg/wk24/g2004/i_syk3k.sas7bdat *  
* : $SYKEFR/utvalg/wk24/g2005/i_syk3k.sas7bdat *  
* : ...EN NY FOR HVERRT ÅR !!!! *  
* : *  
* Filer ut .....: $SYKEFR/utvalg/trekking/wk24/ : *  
* : popmednyepuljenr&yyyy.&mm.&bokstav..sas7bdat *  
* : utvalg&aaaa..sas7bdat *  
*****/
```

Program 6

```
*****;
* Prosjekt .....: Sentral sykefraværstatistikk;
* Program navn .....: lag_flat_arkivfil.sas ;
* katalog .....: $SYKEFR/utvalg/prog/trekking;
* Skrevet av .....: Jørn Ivar Hamre (HAM);
* Dato .....: mai 2003;
* Funksjon .....: lager FLAT fil og;
* .....: tilrettelegger metadata for Datadok;
* .....:;
* Filer inn .....: trekk24.popmednyepuljenr&yyyy.&mm.&bokstav.;
* .....: trekk24.utvalg&yyyy.;
* .....:;
* Filer ut .....: g&yyyy.m&mm..dat (FLYTTES TIL ARKIV OMRÅDET);
* .....: g&yyyy.m&mm..dat (FLYTTES TIL ARKIV OMRÅDET);
* .....:;
* Metadata til datadok : trekkpopulasjon&yyyy..ddok,;
* .....: utvalg&yyyy.m&mm..ddok,;
*****;
* OBS HUSK Å FLYTTE DEN FØRSTE FLATFILA TIL ARKIVOMRÅDET FØR NESTE LAGES ;
* .....;
*****;
```

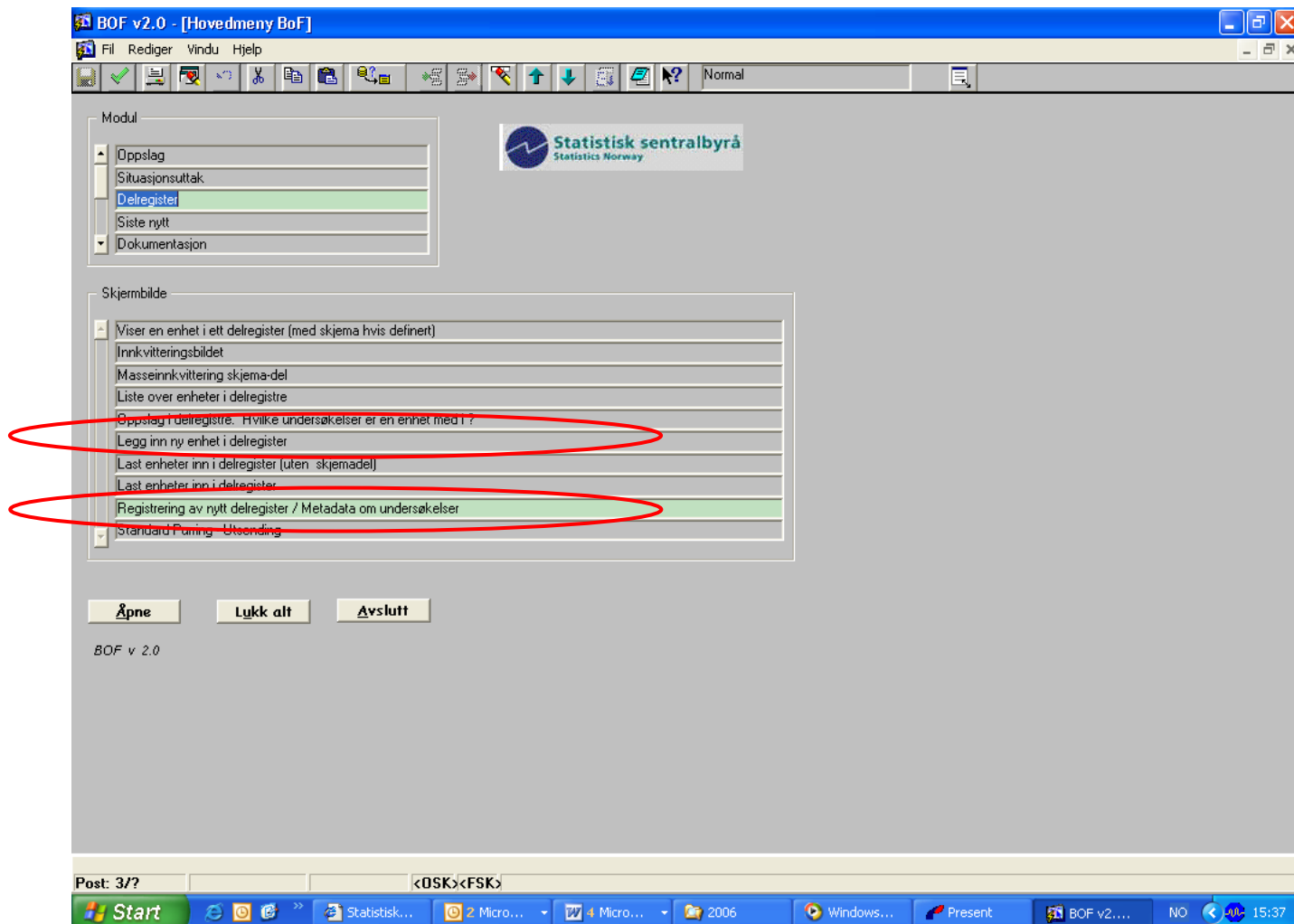
Program 7

```
/******
* Produktnummer .....: 2113 *
* Prosjekt .....: Sentral sykefraværstatistikk *
* Programnavn .....: $SYKEFR/utvalg/prog/trekking/ *
* Programnavn .....: kontroll2006utvalget.sas *
* .....: *
* Skrevet av .....: JØRN IVAR HAMRE, S260 *
* Dato .....: 2003/03/03 *
* Versjon .....: 1.2 *
* Programmets funksjon .: Lager kontrolltabeller for utvalget. *
* .....: *
* Filer inn .....: $SYKEFR/utvalg/trekking/wk24/ : *
* .....: popmednyepuljenr&yyyy.&mm.&bokstav..sas7bdat *
* .....: utvalg&aaaa..sas7bdat *
* Filer ut .....: *
* .....: *
*****/
```


4.3 Delregister prosedyrer

4.3.1 Registrering av nytt delregister og metadata om undersøkelsen

I eget skjermbilde av delregistermodulen må først nytt delregister registreres, og metadata om undersøkelsen legges inn.



4.3.2 Innlasting av enheter i delregister.

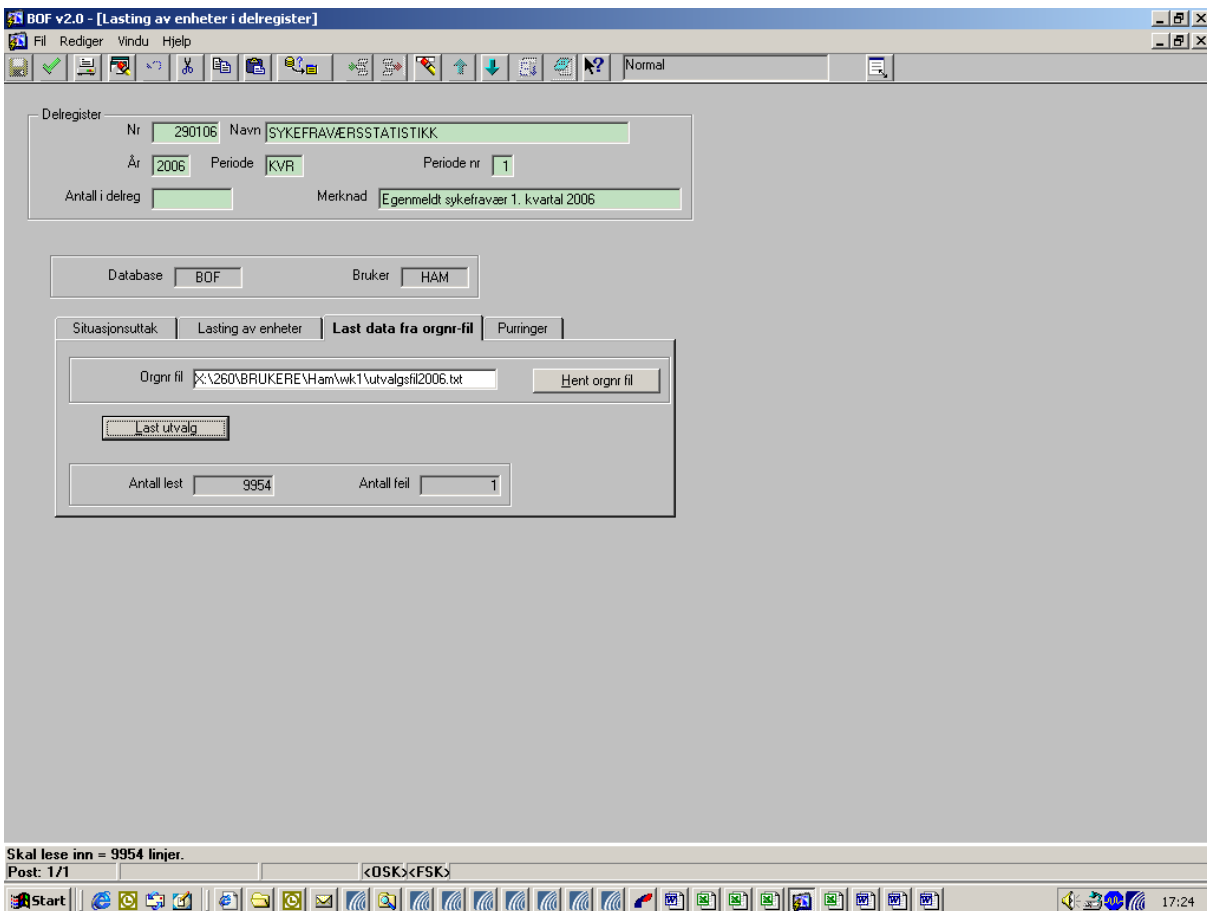
Deretter lastes enhetene inn i delregisteret i eget skjermbilde i delregister-modulen.

Variabelen **Organisasjonsnummer, bedrift** fra den nye utvalgsfila er nøkkelen. Variabelen gjøres klar på en fil som lagres i Editor (**WinEdit**). Organisasjonsnummer, bedrift limes inn i Editoren, og lagres som tekstfil, se illustrasjonen under.

- **Unngå blanke poster (topp, bunn og midt i)!**
- **Det er viktig at nummeret er i posisjon 1-9 uten tusenskille.**

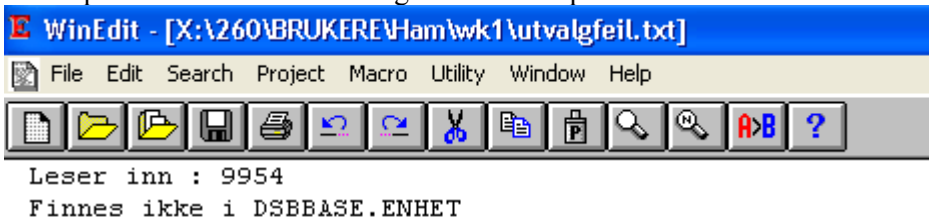


Gå nå til arkfanen 'Last data fra orgnr-fil', og hent fila lagret i Editoren.



Dersom du har en blanke poster på lista di eller prøver å laste inn enheter som ikke finnes i BOF-basen får du alle feil i en fil (utvalgsfeil.txt) som legges på samme område som du lagret tekstfilen din.

Eksemplet under viser feilmelding dersom siste post er blank.



4.3.3 Oppdatering av andre variable i delregisteret

Delregisteret blir automatisk oppdatert med en del BoF kjennemerker.

Likevel gjenstår det noe som i forhold til å oppdatere med andre kjennemerker for nye bedrifter og oppdatere kjennemerker fra tidligere delregistre for de "gamle" bedriftene.

Utvalget bygger på situasjonsuttaket 200512b fra BOF, men er ikke oppdatert med informasjon fra S820 sin oracle-base KPR1 schema sykefravær.

Vi kan benytte samme "oppskrift" for innlasting av andre kjennemerker fra utvalget som i fjor.

Delregisteret suppleres med:

- **data fra delregisteret fra forrige kvartal (290405):** Kontaktpersoner, kommentarer etc., og
- **nye variable fra utvalg2006-fila** som vi ønsker å ha med:

utvalg2006 Char 1, med verdier fra 1-5

(Variabelen, utvalg2006, gir info om hvor mange år den enkelte bedrift skal være med i undersøkelsen i denne omgang, der verdien 5 betyr at de alltid skal være med. Variabelen kan være nyttig info for K820 og for S.410 når informasjonsbrev om forventet oppgaveplikt skal sendes ut til neste år)

Andre variable som K820 trenger fra utvalg2006-fila heter:

bedralder Char 1 Bedriftsalder binær, gamle=1 og nye=2

lopenr Char 5 Løpenummer

Variable fra Delregister 290405 som vi ønsker å ha med:

Skjematype (FORE, UFO, PBED)

kvitteringstype

Undertype

Under arkfanen kommentarer ønsker vi også å ha med informasjonen under (slik det vanligvis gjøres fra kvartal til kvartal)

kontaktperson, navn

e-post

telefon

faks

interne merknader

eksterne merknader

For de **gamle bedriftene** ønsker vi dessuten at adresseopplysningene tas fra **Delregister 290405**, mens for de **nye** tar vi det fra utvalgsfila/BoF.

Som tidligere, må noen av verdiene fra Delregister 290405 imidlertid blankes ut i Delregister 290106:

kvitteringstype='K' (Kvittert) blankes ut

kvitteringstype='S' (Utsettelse) blankes ut

kvitteringstype='D' (Tatt ut) blankes ut dersom undertype='K5' (Holdes midlertidig utenfor)

5. Kontrolltabeller for utvalget og trekkpopulasjonen 2006

Trekkpopulasjonen 2006

status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
B (IKKE-SLETTET)	169987	100.00	169987	100.00

Trekkpopulasjonen 2006 etter hvorvidt bedriften var en del av trekkpopulasjonen ved forrige utvalgstrekkning.

Bedriftsalder	Frequency	Percent	Cumulative Frequency	Cumulative Percent
NY	16799	9.88	16799	9.88
GAMMEL	153188	90.12	169987	100.00

Trekkpopulasjonen 2006 etter størrelsesstrata.

Størrelsesstrata	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1 Cut-off	82569	48.57	82569	48.57
2	60351	35.50	142920	84.08
3	19711	11.60	162631	95.67
4	6076	3.57	168707	99.25
5 Full-telling	1280	0.75	169987	100.00

Trekkpopulasjonen og utvalget for 2006 etter størrelsesstrata. Prosentfordeling av antall bedrifter.

Størrelses- strata	Trekk- populasjon	Utvalg
Total	100,0	100,0
1 Cut-off	51,6	0,0
2	35,6	33,2
3	10,4	31,7
4	2,4	22,2
5 Full-telling	0,0	12,9

Utvalget for 2006, etter størrelsesstrata og hvor mange år bedriftene skal være med.

Frequency Percent Row Pct Col Pct	Størrelsesstrata				Total
	2	3	4	5 Full- telling	
1 Med i 1 år	1055 10.60 43.11 31.89	877 8.81 35.84 27.83	515 5.17 21.05 23.26	0 0.00 0.00 0.00	2447 24.59
2 Med i 2 år	788 7.92 37.35 23.82	832 8.36 39.43 26.40	490 4.92 23.22 22.13	0 0.00 0.00 0.00	2110 21.20
3 Med i 3 år	756 7.60 38.36 22.85	724 7.27 36.73 22.98	491 4.93 24.91 22.18	0 0.00 0.00 0.00	1971 19.80
4 Med i 4 år	709 7.12 37.83 21.43	690 6.93 36.82 21.90	475 4.77 25.35 21.45	0 0.00 0.00 0.00	1874 18.83
5 Alltid med	0 0.00 0.00 0.00	28 0.28 1.81 0.89	243 2.44 15.67 10.98	1280 12.86 82.53 100.00	1551 15.58
Total	3308 33.24	3151 31.66	2214 22.24	1280 12.86	9953 100.00

Utvalget for 2006 etter hvorvidt bedriften var en del av trekkpopulasjonen ved forrige utvalgstreking.

Bedriftsalder	Frequency	Percent	Cumulative Frequency	Cumulative Percent
NY	358	3.60	358	3.60
GAMMEL	9595	96.40	9953	100.00

Utvalget for 2006 etter hvorvidt bedriften var med i fjor eller er ny i utvalget og størrelsesstrata.

bybedrift	Størrelsesstrata				Total
	2	3	4	5 full- tellings	
Nye	1090 10.95 37.55 32.95	1018 10.23 35.07 32.31	674 6.77 23.22 30.44	121 1.22 4.17 9.45	2903 29.17
Gamle	2218 22.28 31.46 67.05	2133 21.43 30.26 67.69	1540 15.47 21.84 69.56	1159 11.64 16.44 90.55	7050 70.83
Total	3308 33.24	3151 31.66	2214 22.24	1280 12.86	9953 100.00

Utvalget for 2006 etter hvorvidt bedriften var med i det opprinnelige utvalget i 1997 og hvor mange år bedriftene skal være med. Frekvens og prosenter

Utvalget 2006 | **utvalgt97**
 | (De utvalgte 10 000 bedriftene i 1997)

Frequency			Total
Percent			
Row Pct			
Col Pct	0 NEI	1 JA	
Med 1 år	2041	52	2093
	24.26	0.62	24.88
	97.52	2.48	
	27.88	4.76	
Med 2 år	1685	54	1739
	20.03	0.64	20.67
	96.89	3.11	
	23.02	4.95	
Med 3 år	1572	66	1638
	18.69	0.78	19.47
	95.97	4.03	
	21.48	6.04	
Med 4 år	1447	109	1556
	17.20	1.30	18.50
	92.99	7.01	
	19.77	9.98	
Alltid med i utvalget	575	811	1386
	6.84	9.64	16.48
	41.49	58.51	
	7.86	74.27	
Total	7320	1092	8412
	87.02	12.98	100.00

Frequency Missing = 1541

Trekkpopulasjonen 2006 til sentral sykefraværstatistikk
Antall bedrifter etter nærings- og størrelsesstrata.

Næringsstrata, verdier fra 1 til 42	I alt	Størrelsesstrata					5 Fulltelling
		1 Cut-off	2	3	4		
I alt	169987	82569	60351	19711	6076	1280	
01 Jordbruk, skogbruk og fiske	7323	5776	1435	92	15	5	
02 Bergverksdrift	369	170	158	30	9	2	
03 Olje- og gassutvinning	206	80	76	16	33	1	
04 Nærings- og nytelsesmiddelindustri	1795	467	756	336	181	55	
05 Tekstil- og bekledningsindustri	510	247	191	53	15	4	
06 Trelast- og trevareindustrien	1050	477	386	129	50	8	
07 Treforedling	104	31	24	29	16	4	
08 Forlag og grafisk industri	1866	1031	616	126	74	19	
09 Oljeraffinering, kjemisk industri mv.	584	233	167	137	35	12	
10 Mineralproduktindustri	622	271	223	75	42	11	
11 Metallindustri	151	40	45	31	29	6	
12 Metallvareindustri	1408	545	594	194	68	7	
13 Maskinindustri	1363	674	479	119	76	15	
14 Elektroteknisk og optisk industri	929	429	326	135	32	7	
15 Transportmiddelindustri	841	303	278	173	67	20	
16 Møbelindustri og annen industri	886	453	299	93	35	6	
17 Kraft- og vannforsyning	881	321	345	152	54	9	
18 Bygge- og anleggsvirksomhet	15347	8557	5341	1083	263	103	
19 Motorkjøretøytjenester	6213	2556	1721	1384	492	60	
20 Agentur- og engroshandel	11723	5765	4948	736	179	95	
21 Detaljhandel og reparasjon av varer	24129	8539	10398	3731	1239	222	
22 Hotell- og restaurantvirksomhet	7200	2984	3255	760	143	58	
23 Landtransport og rørtransport	7148	5101	1652	254	104	37	
25 Lufttransport	119	24	63	12	18	2	
26 Tjenester tilknyttet transport	2315	1126	893	198	74	24	
27 Post og telekommunikasjon	1223	342	488	315	69	9	
28 Finansiell tjenesteyting	1402	476	653	231	35	7	
29 Forsikring og pensjonsfond	380	196	114	53	11	6	
30 Hjelpevirks. finansiell tj.yting	522	216	185	54	47	20	
31 Eiendomsdrift	6253	4687	1429	102	26	9	
32 Utleievirksomhet, maskiner og utstyr	847	419	390	30	8	0	
33 Databehandlingsvirksomhet	3269	1845	1103	216	80	25	
34 Forskning og utviklingsarbeid	310	97	88	93	28	4	
35 Annen forretn. tjenesteyting	17275	10549	5359	868	408	91	
36 Off. adm. og forsvar, sosialforsikr.	5348	1614	2163	1281	236	54	
37 Undervisning	6463	1471	2085	1984	834	89	
38 Helse- og sosialtjenester	20725	8453	8056	3547	580	89	
39 Kloakk- og renovasjonsvirksomhet	629	262	263	79	23	2	
40 Interesseorganisasjoner	3538	2171	884	288	144	51	
41 Kultur og sport	3462	1820	1360	207	58	17	
42 Annen personlig tjenesteyting	3259	1751	1062	285	146	15	

Utvalget 2006 til sentral sykefraværstatistikk. Antall bedrifter etter nærings- og størrelsesstrata.

Næringsstrata, verdier fra 1 til 42	I alt	Størrelsesstrata				
		1 Cut-off	2	3	4	5 Fulltelling
I alt	9953	0	3308	3151	2214	1280
01 Jordbruk, skogbruk og fiske	242	0	180	42	15	5
02 Bergverksdrift	70	0	42	17	9	2
03 Olje- og gassutvinning	58	0	26	16	15	1
04 Nærings- og nytelsesmiddelindustri	165	0	25	40	45	55
05 Tekstil- og bekledningsindustri	70	0	23	28	15	4
06 Trelast- og trevareindustrien	84	0	25	28	23	8
07 Treforedling	41	0	10	11	16	4
08 Forlag og grafisk industri	175	0	86	35	35	19
09 Oljeraffinering, kjemisk industri mv.	102	0	44	28	18	12
10 Mineralproduktindustri	100	0	30	37	22	11
11 Metallindustri	53	0	17	17	13	6
12 Metallvareindustri	127	0	40	49	31	7
13 Maskinindustri	96	0	24	19	38	15
14 Elektroteknisk og optisk industri	107	0	36	32	32	7
15 Transportmiddelindustri	117	0	40	23	34	20
16 Møbelindustri og annen industri	104	0	39	24	35	6
17 Kraft- og vannforsyning	90	0	26	34	21	9
18 Bygge- og anleggsvirksomhet	491	0	197	138	53	103
19 Motorkjøretøytjenester	435	0	67	187	121	60
20 Agentur- og engroshandel	403	0	164	96	48	95
21 Detaljhandel og reparasjon av varer	1023	0	215	285	301	222
22 Hotell- og restaurantvirksomhet	473	0	147	197	71	58
23 Landtransport og rørtransport	327	0	118	114	58	37
25 Lufttransport	61	0	29	12	18	2
26 Tjenester tilknyttet transport	153	0	37	55	37	24
27 Post og telekommunikasjon	148	0	32	76	31	9
28 Finansiell tjenesteyting	104	0	31	31	35	7
29 Forsikring og pensjonsfond	72	0	30	25	11	6
30 Hjelpevirks. finansiell tj.yting	125	0	50	30	25	20
31 Eiendomsdrift	279	0	199	45	26	9
32 Utleievirksomhet, maskiner og utstyr	125	0	105	12	8	0
33 Databehandlingsvirksomhet	136	0	39	25	47	25
34 Forskning og utviklingsarbeid	67	0	21	27	15	4
35 Annen forretn. tjenesteyting	770	0	227	266	186	91
36 Off. adm. og forsvar, sosialforsikr.	327	0	72	88	113	54
37 Undervisning	403	0	72	129	113	89
38 Helse- og sosialtjenester	1175	0	345	449	292	89
39 Kloakk- og renovasjonsvirksomhet	105	0	35	45	23	2
40 Interesseorganisasjoner	389	0	125	146	67	51
41 Kultur og sport	191	0	99	46	29	17
42 Annen personlig tjenesteyting	370	0	139	147	69	15

Faktiske trekksannsynligheter for 2006 til sentral sykefraværstatistikk. Andelen bedrifter i utvalget etter nærings- og størrelsesstrata.

Næringsstrata, verdier fra 1 til 42		Størrelsesstrata					5 Full- telling
		I alt	1 Cut-off	2	3	4	
01-99	I alt	5.9	0.0	5.5	16.0	36.4	100.0
01	Jordbruk, skogbruk og fiske	3.3	0.0	12.5	45.7	100.0	100.0
02	Bergverksdrift	19.0	0.0	26.6	56.7	100.0	100.0
03	Olje- og gassutvinning	28.2	0.0	34.2	100.0	45.5	100.0
04	Nærings- og nytelsesmiddelindustri	9.2	0.0	3.3	11.9	24.9	100.0
05	Tekstil- og bekledningsindustri	13.7	0.0	12.0	52.8	100.0	100.0
06	Trelast- og trevareindustrien	8.0	0.0	6.5	21.7	46.0	100.0
07	Treforedling	39.4	0.0	41.7	37.9	100.0	100.0
08	Forlag og grafisk industri	9.4	0.0	14.0	27.8	47.3	100.0
09	Oljeraffinering, kjemisk industri mv.	17.5	0.0	26.3	20.4	51.4	100.0
10	Mineralproduktindustri	16.1	0.0	13.5	49.3	52.4	100.0
11	Metallindustri	35.1	0.0	37.8	54.8	44.8	100.0
12	Metallvareindustri	9.0	0.0	6.7	25.3	45.6	100.0
13	Maskinindustri	7.0	0.0	5.0	16.0	50.0	100.0
14	Elektroteknisk og optisk industri	11.5	0.0	11.0	23.7	100.0	100.0
15	Transportmiddelindustri	13.9	0.0	14.4	13.3	50.7	100.0
16	Møbelindustri og annen industri	11.7	0.0	13.0	25.8	100.0	100.0
17	Kraft- og vannforsyning	10.2	0.0	7.5	22.4	38.9	100.0
18	Bygge- og anleggsvirksomhet	3.2	0.0	3.7	12.7	20.2	100.0
19	Motorkjøretøytjenester	7.0	0.0	3.9	13.5	24.6	100.0
20	Agentur- og engroshandel	3.4	0.0	3.3	13.0	26.8	100.0
21	Detaljhandel og reparasjon av varer	4.2	0.0	2.1	7.6	24.3	100.0
22	Hotell- og restaurantvirksomhet	6.6	0.0	4.5	25.9	49.7	100.0
23	Landtransport og rørtransport	4.6	0.0	7.1	44.9	55.8	100.0
25	Lufttransport	51.3	0.0	46.0	100.0	100.0	100.0
26	Tjenester tilknyttet transport	6.6	0.0	4.1	27.8	50.0	100.0
27	Post og telekommunikasjon	12.1	0.0	6.6	24.1	44.9	100.0
28	Finansiell tjenesteyting	7.4	0.0	4.7	13.4	100.0	100.0
29	Forsikring og pensjonsfond	18.9	0.0	26.3	47.2	100.0	100.0
30	Hjelpevirks. finansiell tj.yting	23.9	0.0	27.0	55.6	53.2	100.0
31	Eiendomsdrift	4.5	0.0	13.9	44.1	100.0	100.0
32	Utleievirksomhet, maskiner og utstyr	14.8	0.0	26.9	40.0	100.0	-
33	Databehandlingsvirksomhet	4.2	0.0	3.5	11.6	58.8	100.0
34	Forskning og utviklingsarbeid	21.6	0.0	23.9	29.0	53.6	100.0
35	Annen forretn. tjenesteyting	4.5	0.0	4.2	30.6	45.6	100.0
36	Off. adm. og forsvar, sosialforsikr.	6.1	0.0	3.3	6.9	47.9	100.0
37	Undervisning	6.2	0.0	3.5	6.5	13.5	100.0
38	Helse- og sosialtjenester	5.7	0.0	4.3	12.7	50.3	100.0
39	Kloakk- og renovasjonsvirksomhet	16.7	0.0	13.3	57.0	100.0	100.0
40	Interesseorganisasjoner	11.0	0.0	14.1	50.7	46.5	100.0
41	Kultur og sport	5.5	0.0	7.3	22.2	50.0	100.0
42	Annen personlig tjenesteyting	11.4	0.0	13.1	51.6	47.3	100.0

Prosentpoengsdifferanse mellom faktiske- og teoretiske trekk sannsynligheter 2006 etter næring- og størrelsesstrata. De teoretiske trekk sannsynlighetene er satt lik ekte brøker med 2er potens ($1/2^{**k}$, der $k=0-7$)

Næringsstrata, verdier fra 1 til 42		Størrelsesstrata				
		1 Cut-off	2	3	4	5 Full-telling
01	Jordbruk, skogbruk og fiske	0.0	0.0	-4.3	0.0	0.0
02	Bergverksdrift	0.0	1.6	6.7	0.0	0.0
03	Olje- og gassutvinning	0.0	9.2	0.0	-4.5	0.0
04	Nærings- og nytelsesmiddelindustri	0.0	0.2	-0.6	-0.1	0.0
05	Tekstil- og bekledningsindustri	0.0	-0.5	2.8	0.0	0.0
06	Trelast- og trevareindustrien	0.0	0.2	-3.3	-4.0	0.0
07	Treforedling	0.0	-8.3	-12.1	0.0	0.0
08	Forlag og grafisk industri	0.0	1.5	2.8	-2.7	0.0
09	Oljeraffinerer, kjemisk industri mv.	0.0	1.3	-4.6	1.4	0.0
10	Mineralproduktindustri	0.0	1.0	-0.7	2.4	0.0
11	Metallindustri	0.0	-12.2	4.8	-5.2	0.0
12	Metallvareindustri	0.0	0.5	0.3	-4.4	0.0
13	Maskinindustri	0.0	-1.2	3.5	0.0	0.0
14	Elektroteknisk og optisk industri	0.0	-1.5	-1.3	0.0	0.0
15	Transportmiddelindustri	0.0	1.9	0.8	0.7	0.0
16	Møbelindustri og annen industri	0.0	0.5	0.8	0.0	0.0
17	Kraft- og vannforsyning	0.0	1.3	-2.6	-11.1	0.0
18	Bygge- og anleggsvirksomhet	0.0	0.6	0.2	-4.8	0.0
19	Motorkjøretøytjenester	0.0	0.8	1.0	-0.4	0.0
20	Agentur- og engroshandel	0.0	0.2	0.5	1.8	0.0
21	Detaljhandel og reparasjon av varer	0.0	0.5	1.4	-0.7	0.0
22	Hotell- og restaurantvirksomhet	0.0	1.4	0.9	-0.3	0.0
23	Landtransport og rørtransport	0.0	0.9	-5.1	5.8	0.0
25	Lufttransport	0.0	-4.0	0.0	0.0	0.0
26	Tjenester tilknyttet transport	0.0	1.0	2.8	0.0	0.0
27	Post og telekommunikasjon	0.0	0.3	-0.9	-5.1	0.0
28	Finansiell tjenesteyting	0.0	-1.5	0.9	0.0	0.0
29	Forsikring og pensjonsfond	0.0	1.3	-2.8	0.0	0.0
30	Hjelpevirks. finansiell tj.yting	0.0	2.0	5.6	3.2	0.0
31	Eiendomsdrift	0.0	1.4	-5.9	0.0	0.0
32	Utleievirksomhet, maskiner og utstyr	0.0	1.9	-10.0	0.0	-
33	Databehandlingsvirksomhet	0.0	0.4	-0.9	8.8	0.0
34	Forskning og utviklingsarbeid	0.0	-1.1	4.0	3.6	0.0
35	Annen forretn. tjenesteyting	0.0	1.1	5.6	-4.4	0.0
36	Off. adm. og forsvar, sosialforsikr.	0.0	0.2	0.6	-2.1	0.0
37	Undervisning	0.0	0.3	0.3	1.0	0.0
38	Helse- og sosialtjenester	0.0	1.2	0.2	0.3	0.0
39	Kloakk- og renovasjonsvirksomhet	0.0	0.8	7.0	0.0	0.0
40	Interesseorganisasjoner	0.0	1.6	0.7	-3.5	0.0
41	Kultur og sport	0.0	1.0	-2.8	0.0	0.0
42	Annen personlig tjenesteyting	0.0	0.6	1.6	-2.7	0.0

Trekkpopulasjonen og utvalg for 2006 til sentral sykefraværstatistikk
 Antall bedrifter og ansatte, etter næringsstrata. Absolutte tall og andel i utvalget

Næringsstrata	Bedrifter			Ansatte		
	Antall	Popula-	Andelen	Antall	Populas-	Andelen
	Utvalg-	sjonen	i	Utvalget	jonen	ansatte i
	et		utvalget			utvalget
I alt	9953	169987	5.9	718650	2091491	34.4
Jordbruk, skogbruk og fiske	242	7323	3.3	4319	23090	18.7
Bergverksdrift	70	369	19.0	2252	4072	55.3
Olje og gassutvinning	58	206	28.2	6404	13177	48.6
Nærings og nytelsesmiddelindustri	165	1795	9.2	21249	50462	42.1
Tekstil og bekledingsindustri	70	510	13.7	2534	5006	50.6
Trelast og trevareindustrien	84	1050	8.0	4852	14726	32.9
Treforedling	41	104	39.4	6127	7206	85.0
Forlag og grafisk industri	175	1866	9.4	12016	25204	47.7
Oljeraffinering, kjemisk industri mv.	102	584	17.5	11491	20558	55.9
Mineralproduktindustri	100	622	16.1	5201	10210	50.9
Metallindustri	53	151	35.1	7331	11853	61.8
Metallvareindustri	127	1408	9.0	6059	19914	30.4
Maskinindustri	96	1363	7.0	9360	21251	44.0
Elektroteknisk og optisk industri	107	929	11.5	10081	18195	55.4
Transportmiddelindustri	117	841	13.9	19460	36118	53.9
Møbelindustri og annen industri	104	886	11.7	5542	11121	49.8
Kraft og vannforsyning	90	881	10.2	5059	15185	33.3
Bygge og anleggsvirksomhet	491	15347	3.2	33126	139871	23.7
Motorkjøretøytjenester	435	6213	7.0	10457	54227	19.3
Agentur og engroshandel	403	11723	3.4	24092	100047	24.1
Detaljhandel og reparasjon av varer	1023	24129	4.2	33132	180781	18.3
Hotell og restaurantvirksomhet	473	7200	6.6	21321	79058	27.0
Landtransport og rørtransport	327	7148	4.6	20728	51639	40.1
Lufttransport	61	119	51.3	9318	10110	92.2
Tjenester tilknyttet transport	153	2315	6.6	10080	27708	36.4
Post og telekommunikasjon	148	1223	12.1	16372	38195	42.9
Finansiell tjenesteyting	104	1402	7.4	14963	29015	51.6
Forsikring og pensjonsfond	72	380	18.9	6177	8603	71.8
Hjelpevirks. finansiell tj.yting	125	522	23.9	4074	6234	65.4
Eiendomsdrift	279	6253	4.5	5838	23264	25.1
Utleievirksomhet, maskiner og utstyr	125	847	14.8	1692	5105	33.1
Databehandlingsvirksomhet	136	3269	4.2	12787	33278	38.4
Forskning og utviklingsarbeid	67	310	21.6	5805	11915	48.7
Annen forretn. tjenesteyting	770	17275	4.5	53794	147163	36.6
Off. adm. og forsvar, sosialforsikr.	327	5348	6.1	52812	144863	36.5
Undervisning	403	6463	6.2	48499	185900	26.1
Helse og sosialtjenester	1175	20725	5.7	164078	428223	38.3
Kloakk og renovasjonsvirksomhet	105	629	16.7	3724	7520	49.5
Interesseorganisasjoner	389	3538	11.0	9740	21714	44.9
Kultur og sport	191	3462	5.5	10965	31502	34.8
Annen personlig tjenesteyting	370	3259	11.4	5739	18208	31.5

Trekkpopulasjonen og utvalg for 2006 til sentral sykefraværstatistikk
 Antall bedrifter og ansatte, etter næringsstrata. Absolutte tall og i prosent

Næringsstrata	Bedrifter				Ansatte			
	Antall		I % av totalen		Antall		I % av totalen	
	Popula- sjonen	Utval- get	Popula- sjonen	Utvalg- et	Popula- sjonen	Utvalg- et	Popula- sjonen	Utvalg- et
I alt	169987	9953	100.00	100.00	2091491	718650	100.00	100.00
Jordbruk, skogbruk og fiske	1547	242	0.91	2.43	14229	4319	0.68	0.60
Bergverksdrift	369	70	0.22	0.70	4072	2252	0.19	0.31
Olje og gassutvinning	206	58	0.12	0.58	13177	6404	0.63	0.89
Nærings og nytelsesmiddelindustri	1795	165	1.06	1.66	50462	21249	2.41	2.96
Tekstil og bekledningsindustri	510	70	0.30	0.70	5006	2534	0.24	0.35
Trelast og trevareindustrien	1050	84	0.62	0.84	14726	4852	0.70	0.68
Treforedling	104	41	0.06	0.41	7206	6127	0.34	0.85
Forlag og grafisk industri	1866	175	1.10	1.76	25204	12016	1.21	1.67
Oljeraffinering, kjemisk industri mv.	584	102	0.34	1.02	20558	11491	0.98	1.60
Mineralproduktindustri	622	100	0.37	1.00	10210	5201	0.49	0.72
Metallindustri	151	53	0.09	0.53	11853	7331	0.57	1.02
Metallvareindustri	1408	127	0.83	1.28	19914	6059	0.95	0.84
Maskinindustri	1363	96	0.80	0.96	21251	9360	1.02	1.30
Elektroteknisk og optisk industri	929	107	0.55	1.08	18195	10081	0.87	1.40
Transportmiddelindustri	841	117	0.49	1.18	36118	19460	1.73	2.71
Møbelindustri og annen industri	886	104	0.52	1.04	11121	5542	0.53	0.77
Kraft og vannforsyning	881	90	0.52	0.90	15185	5059	0.73	0.70
Bygge og anleggsvirksomhet	15347	491	9.03	4.93	139871	33126	6.69	4.61
Motorkjøretøytjenester	6213	435	3.65	4.37	54227	10457	2.59	1.46
Agentur og engroshandel	11723	403	6.90	4.05	100047	24092	4.78	3.35
Detaljhandel og reparasjon av varer	24129	1023	14.19	10.28	180781	33132	8.64	4.61
Hotell og restaurantvirksomhet	7200	473	4.24	4.75	79058	21321	3.78	2.97
Landtransport og rørtransport	7148	327	4.21	3.29	51639	20728	2.47	2.88
Lufttransport	119	61	0.07	0.61	10110	9318	0.48	1.30
Tjenester tilknyttet transport	2315	153	1.36	1.54	27708	10080	1.32	1.40
Post og telekommunikasjon	1223	148	0.72	1.49	38195	16372	1.83	2.28
Finansiell tjenesteyting	1402	104	0.82	1.04	29015	14963	1.39	2.08
Forsikring og pensjonsfond	380	72	0.22	0.72	8603	6177	0.41	0.86
Hjelpevirks. finansiell tj.yting	522	125	0.31	1.26	6234	4074	0.30	0.57
Eiendomsdrift	6253	279	3.68	2.80	23264	5838	1.11	0.81
Utleievirksomhet, maskiner og utstyr	847	125	0.50	1.26	5105	1692	0.24	0.24
Databehandlingsvirksomhet	3269	136	1.92	1.37	33278	12787	1.59	1.78
Forskning og utviklingsarbeid	310	67	0.18	0.67	11915	5805	0.57	0.81
Annen forretn. tjenesteyting	17275	770	10.16	7.74	147163	53794	7.04	7.49
Off. adm. og forsvar, sosialforsikr.	5348	327	3.15	3.29	144863	52812	6.93	7.35
Undervisning	6463	403	3.80	4.05	185900	48499	8.89	6.75
Helse og sosialtjenester	20725	1175	12.19	11.81	428223	164078	20.47	22.83
Kloakk og renovasjonsvirksomhet	629	105	0.37	1.05	7520	3724	0.36	0.52
Interesseorganisasjoner	3538	389	2.08	3.91	21714	9740	1.04	1.36
Kultur og sport	3462	191	2.04	1.92	31502	10965	1.51	1.53
Annen personlig tjenesteyting	9035	370	5.32	3.72	27069	5739	1.29	0.80

Trekkpopulasjonen og utvalg for 2006 til sentral sykefraværstatistikk
 Antall bedrifter og ansatte, etter foretakets forretningsadresse. Absolutte tall og i prosent

Fylke (foretn.adr. bedrift)	Bedrifter				Ansatte			
	Antall		I % av totalen		Antall		I % av totalen	
	Populasjonen	Utvalget	Populasjonen	Utvalget	Populasjonen	Utvalget	Populasjonen	Utvalget
Hele landet	169987	9953	100.00	100.00	2091491	718650	100.00	100.00
Østfold	8720	470	5.13	4.72	101225	31401	4.84	4.37
Akershus	16264	1015	9.57	10.20	209690	82235	10.03	11.44
Oslo	23296	1770	13.70	17.78	371763	185013	17.78	25.74
Hedmark	6863	327	4.04	3.29	75013	20227	3.59	2.81
Oppland	6782	357	3.99	3.59	72684	19700	3.48	2.74
Buskerud	9263	503	5.45	5.05	104475	32722	5.00	4.55
Vestfold	7946	448	4.67	4.50	88396	23975	4.23	3.34
Telemark	5890	310	3.46	3.11	65790	16901	3.15	2.35
Aust-Agder	3716	197	2.19	1.98	38605	10087	1.85	1.40
Vest-Agder	5862	317	3.45	3.18	73194	20817	3.50	2.90
Rogaland	13431	854	7.90	8.58	172630	56889	8.25	7.92
Hordaland	14741	936	8.67	9.40	197222	74130	9.43	10.32
Sogn og Fjordane	4672	196	2.75	1.97	44313	8831	2.12	1.23
Møre og Romsdal	9423	491	5.54	4.93	99041	25853	4.74	3.60
Sør-Trøndelag	9561	627	5.62	6.30	125118	43648	5.98	6.07
Nord-Trøndelag	5000	220	2.94	2.21	49099	11777	2.35	1.64
Nordland	8796	471	5.17	4.73	93886	23968	4.49	3.34
Troms	6079	260	3.58	2.61	66694	18079	3.19	2.52
Finnmark	3367	146	1.98	1.47	29566	5075	1.41	0.71
Svalbard	111	14	0.07	0.14	1337	614	0.06	0.09
Jan Mayen	1	0	0.00	0.00	1	0	0.00	0.00
Kontinentalsokkelen	40	21	0.02	0.21	10940	6674	0.52	0.93
Missing	163	3	0.10	0.03	809	34	0.04	0.00

Trekkpopulasjonen og utvalg for 2006 til sentral sykefraværstatistikk
 Antall bedrifter og ansatte, etter fylke. Absolutte tall og andel i utvalget

Fylke (foretn.adr. bedrift)	Bedrifter			Ansatte		
	Antall		Andelen i utvalget	Antall		Andelen ansatte i utvalget
	Utvalget	Populasjonen		Utvalget	Populasjonen	
Hele landet	9953	169987	5.9	718650	2091491	34.4
Østfold	470	8720	5.4	31401	101225	31.0
Akershus	1015	16264	6.2	82235	209690	39.2
Oslo	1770	23296	7.6	185013	371763	49.8
Hedmark	327	6863	4.8	20227	75013	27.0
Oppland	357	6782	5.3	19700	72684	27.1
Buskerud	503	9263	5.4	32722	104475	31.3
Vestfold	448	7946	5.6	23975	88396	27.1
Telemark	310	5890	5.3	16901	65790	25.7
Aust-Agder	197	3716	5.3	10087	38605	26.1
Vest-Agder	317	5862	5.4	20817	73194	28.4
Rogaland	854	13431	6.4	56889	172630	33.0
Hordaland	936	14741	6.3	74130	197222	37.6
Sogn og Fjordane	196	4672	4.2	8831	44313	19.9
Møre og Romsdal	491	9423	5.2	25853	99041	26.1
Sør-Trøndelag	627	9561	6.6	43648	125118	34.9
Nord-Trøndelag	220	5000	4.4	11777	49099	24.0
Nordland	471	8796	5.4	23968	93886	25.5
Troms	260	6079	4.3	18079	66694	27.1
Finnmark	146	3367	4.3	5075	29566	17.2
Svalbard	14	111	12.6	614	1337	45.9
Jan Mayen	0	1	0.0	0	1	0.0
Kontinentalsokkelen	21	40	52.5	6674	10940	61.0
Missing	3	163	1.8	34	809	4.2

Trekpopulasjonen 2006 til sentral sykefraværstatistikk
 Antall bedrifter etter størrelsesstrata og fylke, (bedriftens forretningssted)

bed_f_fylke	Størrelsesstrata					5 Full-telling
	I alt	1 Cut-off	2	3	4	
I alt	169987	82569	60351	19711	6076	1280
Østfold	8720	4241	3100	1012	314	53
Akershus	16264	8251	5351	1850	635	177
Oslo	23296	11420	7821	2557	1099	399
Hedmark	6863	3411	2396	824	201	31
Oppland	6782	3303	2436	806	215	22
Buskerud	9263	4597	3239	1054	320	53
Vestfold	7946	3944	2755	924	278	45
Telemark	5890	2826	2107	752	187	18
Aust-Agder	3716	1887	1281	410	122	16
Vest-Agder	5862	2769	2091	726	237	39
Rogaland	13431	6185	4964	1623	552	107
Hordaland	14741	7041	5189	1770	621	120
Sogn og Fjordane	4672	2352	1717	498	93	12
Møre og Romsdal	9423	4390	3674	1077	253	29
Sør-Trøndelag	9561	4481	3438	1209	357	76
Nord-Trøndelag	5000	2499	1846	529	108	18
Nordland	8796	4144	3392	993	235	32
Troms	6079	2996	2185	716	162	20
Finnmark	3367	1672	1269	360	62	4
Svalbard	111	50	46	11	2	2
Jan Mayen	1	1	0	0	0	0
Kontinentalsokkelen	40	1	4	5	23	7
Missing	163	108	50	5	0	0

Utvalget 2006 til sentral sykefraværstatistikk
 Antall bedrifter etter størrelsesstrata og fylke, (bedriftens forretningssted)

bed_f_fylke	Størrelsesstrata					5 Full-telling
	I alt	1 Cut-off	2	3	4	
I alt	9953	0	3308	3151	2214	1280
Østfold	470	0	152	153	112	53
Akershus	1015	0	318	284	236	177
Oslo	1770	0	459	462	450	399
Hedmark	327	0	103	125	68	31
Oppland	357	0	127	121	87	22
Buskerud	503	0	173	164	113	53
Vestfold	448	0	169	140	94	45
Telemark	310	0	108	113	71	18
Aust-Agder	197	0	72	70	39	16
Vest-Agder	317	0	94	103	81	39
Rogaland	854	0	280	285	182	107
Hordaland	936	0	326	275	215	120
Sogn og Fjordane	196	0	80	76	28	12
Møre og Romsdal	491	0	187	164	111	29
Sør-Trøndelag	627	0	202	221	128	76
Nord-Trøndelag	220	0	94	71	37	18
Nordland	471	0	193	168	78	32
Troms	260	0	92	97	51	20
Finnmark	146	0	68	52	22	4
Svalbard	14	0	8	3	1	2
Jan Mayen	0	0	0	0	0	0
Kontinentalsokkelen	21	0	1	3	10	7
Missing	3	0	2	1	0	0

A. Variasjonskoeffisienter i ulike strata

Tabell A.1 Estimert standardavvik i prosent av estimatene, etter størrelses- og næringsstrata. 3. kvartal 2005

Næringsstrata	Størrelsestrata					5 Full-telling
	I alt	1 Cut-off	2	3	4	
I alt	1,0	7,2	2,8	1,9	1,4	0,9
Jordbruk, skogbruk og fiske	8,7	19,0	10,7	22,6	15,9	13,8
Bergverksdrift	14,4	130,0	45,3	14,9	5,3	-
Nærings og nytelsesmiddelindus	4,0	74,2	18,8	13,8	7,0	2,6
Tekstil og bekledningsindustri	5,3	53,5	18,7	8,2	3,5	-
Trelast og trevareindustrien	6,9	44,6	15,3	10,8	11,1	-
Treforedling	1,6	39,8	15,2	9,1	-	-
Forlag og grafisk industri	5,9	36,3	14,2	11,1	15,0	4,6
Oljeraffinering, kjemisk indus	5,9	69,5	25,8	15,8	9,0	3,8
Mineralproduktindustri	4,8	32,4	12,0	10,2	6,8	10,2
Metallindustri	7,0	120,7	39,6	14,2	13,0	1,7
Metallvareindustri	7,1	54,7	17,8	8,0	9,3	-
Maskinindustri	4,9	66,7	23,9	9,7	6,9	-
Elektroteknisk og optisk indus	7,4	98,7	35,9	11,9	9,2	-
Transportmiddelindustri	7,0	62,5	20,8	15,2	10,2	13,0
Møbelindustri og annen industr	7,8	57,6	22,7	10,3	11,1	7,9
Kraft og vannforsyning	6,7	76,3	22,7	11,6	9,3	6,2
Bygge og anleggsvirksomhet	3,4	15,9	6,3	6,3	5,8	4,5
Motorkjøretøytjenester	6,5	42,2	19,6	9,7	7,4	3,0
Agentur og engroshandel	5,4	34,2	11,5	7,7	8,1	2,2
Detaljhandel og reparasjon av	4,9	26,9	9,7	9,1	5,9	2,9
Hotell og restaurantvirksomhet	6,2	51,5	16,2	8,6	6,9	4,4
Landtransport og rørtransport	5,3	31,4	16,4	9,0	7,3	4,0
Lufttransport	1,7	94,7	14,4	7,7	-	-
Tjenester tilknyttet transport	5,7	51,3	18,9	9,0	6,0	2,8
Post og telekommunikasjon	2,7	25,7	6,1	4,5	4,9	6,5
Finansiell tjenesteyting	5,7	67,9	18,6	10,3	9,6	-
Forsikring og pensjonsfond	4,0	48,4	21,1	6,0	13,3	-
Hjelpevirks. finansiell tj.yti	17,1	103,1	42,3	22,0	14,3	23,3
Eiendomsdrift	7,4	21,5	11,2	9,4	4,8	9,6
Utleievirksomhet, maskiner og	10,8	39,1	13,2	25,4	-	-
Databehandlingsvirksomhet	6,3	42,1	15,9	11,4	9,1	-
Forskning og utviklingsarbeid	5,4	54,8	16,9	10,5	10,7	-
Annen forretn. tjenesteyting	3,6	19,9	8,0	4,8	4,4	5,0
Off. adm. og forsvar, sosialfo	4,3	47,7	14,0	10,0	3,7	2,7
Undervisning	3,8	68,6	16,3	7,0	5,6	2,0
Helse og sosialtjenester	2,4	24,0	7,7	3,3	4,2	1,6
Kloakk og renovasjonsvirksomhe	7,2	55,7	17,2	9,9	3,7	18,0
Interesseorganisasjoner	6,1	29,7	15,5	7,8	7,4	4,2
Kultur og sport	5,9	40,0	15,0	15,6	10,9	4,2
Annen personlig tjenesteyting	5,2	20,0	9,1	7,6	8,8	9,8

Note: Blanke celler er forårsaket av missing i telleren.

B. Beskrivelse av anvendte formler for beregning av estimert varians

I sentral sykefraværstatistikk ønsker vi å estimere følgende summer for tapte dagsverk grunnet egenmeldt sykefravær i (del)populasjonen:

grov næring (h) for menn (m) og kvinner (k) hver for seg:

$$t_{mh} = \sum_{i=1}^{N_h} y_{mhi} \quad \text{og} \quad t_{kh} = \sum_{i=1}^{N_h} y_{khi}$$

næring totalt:

$$t_h = t_{kh} + t_{mh} = \sum_{i=1}^{N_h} y_{mhi} + \sum_{i=1}^{N_h} y_{khi}$$

hele populasjonen:

$$t = \sum_{h=1}^H t_h = \sum_{h=1}^H (t_{mh} + t_{kh}) = \sum_{h=1}^H \left(\sum_{i=1}^{N_h} y_{mhi} + \sum_{i=1}^{N_h} y_{khi} \right)$$

menn totalt og for kvinner totalt:

$$t_m = \sum_{h=1}^H t_{mh} = \sum_{h=1}^H \left(\sum_{i=1}^{N_h} y_{mhi} \right) \quad \text{og} \quad t_k = \sum_{h=1}^H t_{kh} = \sum_{h=1}^H \left(\sum_{i=1}^{N_h} y_{khi} \right)$$

N_h er antall bedrifter i statistikkpopulasjonen i stratum h

y_{mhi} angir antall tapte dagsverk grunnet egenmeldt sykefravær for menn i bedrift i i stratum h.

y_{khi} angir antall tapte dagsverk grunnet egenmeldt sykefravær for kvinner i bedrift i i stratum h.

H er antall strata populasjonen deles inn i.

Delpopulasjonene betegnes U_1, U_2, \dots, U_H , og delutvalgene betegnes S_1, S_2, \dots, S_H .

La n_h er antall bedrifter i nettoutvalget i stratum nr. h.

Vi har da at antall bedrifter i populasjonen og utvalget hhv. er :

$$N = \sum_{h=1}^H N_h \quad \text{og} \quad n = \sum_{h=1}^H n_h$$

La x_{hi} angi antall ansatte menn i bedrift nr. i.

La x_{ki} angi antall ansatte kvinner i bedrift nr. i.

La x_{0mh} angi antall ansatte menn i populasjonen i stratum nr. h

La x_{0kh} angi antall ansatte kvinner i populasjonen i stratum nr. h

Det er rimelig å anta en positiv korrelasjon mellom størrelsen på det egenmeldte sykefraværet i bedriftene og antall ansatte i bedriftene. I og med at vi har informasjon om antall ansatte menn og antall ansatte kvinner i hver eneste bedrift i populasjonen, kan tilleggsinformasjonen bidra til mer nøyaktige estimater, gitt utvalgsstørrelse, dersom vi foretar rateestimering.

Måltallene er derfor estimert med en rate-estimator innenfor hvert stratum bortsett fra cut-off strataene.

Formelen er som følger:
$$\tilde{t}_h = \tilde{t}_{mh} + \tilde{t}_{kh} = x_{0mh} \cdot \frac{\sum_{i \in S_h} y_{mi}}{\sum_{i \in S_h} x_{mi}} + x_{0kh} \cdot \frac{\sum_{i \in S_h} y_{ki}}{\sum_{i \in S_h} x_{ki}}$$

Siden estimatoren \tilde{t}_h er tilnærmet forventningsrett for stor n_h og fordi det er vanskelig å beregne utvalgsusikkerheten for en rateestimator gitt systematisk utvalgstrekkning, beregner vi en tilnærmet varians for rateestimatoren gitt tilfeldig utvalg som mål for usikkerheten til vår rate-estimator.

Vi antar at sykefraværet i ulike bedrifter er uavhengig, men at det kan være avhengighet mellom menn og kvinner sitt sykefravær innad i samme bedrift. Våre undersøkelser tyder på noe positiv samvariasjon i sykefraværet mellom menn og kvinner i samme bedrift, jmfør vedlegg D. Variansformelen kan da skrives som følgende: Variansen til estimatorne for hvert av strataene bortsett fra cut-off strataene er da gitt ved formelen $Var(\tilde{t}_h) = Var(\tilde{t}_{mh} + \tilde{t}_{kh}) = Var(\tilde{t}_{mh}) + Var(\tilde{t}_{kh}) + 2 \cdot COV(\tilde{t}_{mh}, \tilde{t}_{kh})$

$$Var(\tilde{t}_h) \approx N_h^2 \cdot \frac{1 - f_h}{n_h} \cdot \frac{1}{N_h} \cdot \left[\sum_{i=1}^{N_h} (y_{mi} - R_{mh} x_{mi})^2 + \sum_{i=1}^{N_h} (y_{ki} - R_{kh} x_{ki})^2 \right] + 2 \cdot COV(\tilde{t}_{mh}, \tilde{t}_{kh})$$

der:

y_{mi} angir antall tapte dagsverk grunnet egenmeldt sykefravær for menn i bedrift i.

y_{ki} angir antall tapte dagsverk grunnet egenmeldt sykefravær for kvinner i bedrift i.

$R_{mh} = \frac{t_{mh}}{x_{0mh}}$ og $R_{kh} = \frac{t_{kh}}{x_{0kh}}$ hhv. er sykefraværsandelen for menn og for kvinner i delpopulasjonen h,

$f = \frac{n_h}{N_h}$ er trekkssannsynligheten i stratum h.

Variansen $Var(\tilde{t}_h)$ over kan estimeres ved formelen under når en tar hensyn til kovariansen i sykefraværet mellom menn og kvinner i samme bedrift:

$$\hat{Var}(\tilde{t}_h) = \left(\frac{\bar{x}_{0mh}}{\bar{x}_{mh}} \right)^2 \cdot N_h^2 \cdot \frac{1 - f_h}{n_h} \cdot \frac{1}{n_h - 1} \cdot \sum_{i \in S_h} (y_{mi} - \hat{R}_{mh} x_{mi})^2 + \left(\frac{\bar{x}_{0kh}}{\bar{x}_{kh}} \right)^2 \cdot N_h^2 \cdot \frac{1 - f_h}{n_h} \cdot \frac{1}{n_h - 1} \cdot \sum_{i \in S_h} (y_{ki} - \hat{R}_{kh} x_{ki})^2 + 2 \cdot \hat{COV}(\tilde{t}_{mh}, \tilde{t}_{kh})$$

der $\hat{R}_{mh} = \frac{\bar{y}_{mh}}{\bar{x}_{mh}}$ og $\hat{R}_{kh} = \frac{\bar{y}_{kh}}{\bar{x}_{kh}}$ hhv. er sykefraværsandelen for menn og for kvinner i utvalget innenfor stratum h.

Videre er

\bar{x}_{0mh} gjennomsnittlig antall ansatte menn i delpopulasjon h, mens

\bar{x}_{mh} er gjennomsnittlig antall ansatte menn i utvalget innenfor stratum h.

\bar{x}_{0kh} gjennomsnittlig antall ansatte kvinner i delpopulasjon h, mens

\bar{x}_{kh} er gjennomsnittlig antall ansatte kvinner i utvalget innenfor stratum h.

Den estimerte kovariansen beregnes ved å anvende formelen under, jmfør utledningen i vedlegg B.

$$\hat{COV}(\tilde{t}_{mh}, \tilde{t}_{kh}) = \left(\frac{\bar{x}_{0mh}}{\bar{x}_{mh}} \right) \cdot \left(\frac{\bar{x}_{0kh}}{\bar{x}_{kh}} \right) \cdot N_h^2 \cdot \frac{1 - f_h}{n_h} \cdot \frac{1}{n_h - 1} \cdot \sum_{i \in S_h} (y_{mi} - \hat{R}_{mh} x_{mi}) \cdot (y_{ki} - \hat{R}_{kh} x_{ki})$$

Variansen i strata h kan da estimeres ved følgende formel:

$$\begin{aligned} \hat{Var}(\tilde{t}_h) &= \left(\frac{\bar{x}_{0mh}}{\bar{x}_{mh}} \right)^2 N_h^2 \frac{1-f_h}{n_h} \frac{1}{n_h-1} \sum_{i \in S_h} (y_{mi} - \hat{R}_{mh} x_{mi})^2 \\ &\quad + \left(\frac{\bar{x}_{0kh}}{\bar{x}_{kh}} \right)^2 N_h^2 \frac{1-f_h}{n_h} \frac{1}{n_h-1} \sum_{i \in S_h} (y_{ki} - \hat{R}_{kh} x_{ki})^2 \\ &\quad + 2 \left(\frac{\bar{x}_{0mh}}{\bar{x}_{mh}} \right) \left(\frac{\bar{x}_{0kh}}{\bar{x}_{kh}} \right) N_h^2 \frac{1-f_h}{n_h} \frac{1}{n_h-1} \sum_{i \in S_h} (y_{mi} - \hat{R}_{mh} x_{mi}) \cdot (y_{ki} - \hat{R}_{kh} x_{ki}) \end{aligned}$$

Cut-off strata

Siden utvalgsstørrelsen er null innenfor cut-off strataene med de minste bedriftene, innenfor hver næring blir sykefraværet for menn og kvinner hhv. estimert med formelen:

$$\tilde{t}_{m1} = \frac{x_{0m1}}{x_{0m2}} \cdot \tilde{t}_{m2} \quad \text{og} \quad \tilde{t}_{k1} = \frac{x_{0k1}}{x_{0k2}} \cdot \tilde{t}_{k2},$$

der variablene med fotskrift 2 representerer størrelsesstratum 2, med de nest minste bedriftene.

Variansen for estimatoren i cut-off stratomet (innenfor hver næring) for hhv. menn og for kvinner er **gitt**

forutsetningene da: $Var(\tilde{t}_{m1}) = \left(\frac{x_{0m1}}{x_{0m2}} \right)^2 \cdot var(\tilde{t}_{m2})$ og $Var(\tilde{t}_{k1}) = \left(\frac{x_{0k1}}{x_{0k2}} \right)^2 \cdot var(\tilde{t}_{k2})$

Variansformelen for sykefraværestimatet for **menn** (summen av de 5 størrelsesstrataene) blir da:

$$\begin{aligned} Var(\tilde{t}_m) &= Var\left(\left[1 + \frac{x_{0m1}}{x_{0m2}} \right] \cdot \tilde{t}_{m2} + \sum_{j=3}^5 \tilde{t}_{mj} \right) \\ &= \left[1 + \frac{x_{0m1}}{x_{0m2}} \right]^2 \cdot Var(\tilde{t}_{m2}) + \sum_{j=3}^5 Var(\tilde{t}_{mj}) \end{aligned}$$

Denne variansen kan estimeres med formelen:

$$\hat{Var}(\tilde{t}_m) = \left(1 + \frac{x_{0m1}}{x_{0m2}} \right)^2 \cdot \hat{Var}(\tilde{t}_{m2}) + \sum_{j=3}^5 \hat{Var}(\tilde{t}_{mj})$$

der vi for fotskrift h = 2, 3, 4 og 5 bruker estimatoren:

Variansen for sykefraværestimatet for **kvinner** blir helt tilsvarende.

Estimert varians for summen av de 5 størrelsesstrataene (inkl. cut-off-strataet):

Fotskrift for å indikere ulike næringsstrata er droppet for enkelhets skyld. Formlene gjelder for hver av næringsstrataene.

$$\begin{aligned}
\text{Var}\left(\sum_{j=1}^5 \tilde{t}_j\right) &= \text{Var}\left(\tilde{t}_{m1} + \tilde{t}_{k1} + \tilde{t}_{m2} + \tilde{t}_{k2} + \sum_{j=3}^5 (\tilde{t}_{mj} + \tilde{t}_{kj})\right) \\
&= \text{Var}\left(\frac{x_{0m1}}{x_{0m2}} \tilde{t}_{m2} + \frac{x_{0k1}}{x_{0k2}} \tilde{t}_{k2} + \tilde{t}_{m2} + \tilde{t}_{k2} + \sum_{j=3}^5 (\tilde{t}_{mj} + \tilde{t}_{kj})\right) \\
&= \text{Var}\left(\left(1 + \frac{x_{0m1}}{x_{0m2}}\right) \tilde{t}_{m2} + \left(1 + \frac{x_{0k1}}{x_{0k2}}\right) \tilde{t}_{k2} + \sum_{j=3}^5 (\tilde{t}_{mj} + \tilde{t}_{kj})\right) \\
&= \left(1 + \frac{x_{0m1}}{x_{0m2}}\right)^2 \text{Var}(\tilde{t}_{m2}) + \left(1 + \frac{x_{0k1}}{x_{0k2}}\right)^2 \text{Var}(\tilde{t}_{k2}) + 2\left(1 + \frac{x_{0m1}}{x_{0m2}}\right)\left(1 + \frac{x_{0k1}}{x_{0k2}}\right) \text{Cov}(\tilde{t}_{k2}, \tilde{t}_{m2}) \\
&\quad + \sum_{j=3}^5 (\text{Var}(\tilde{t}_{mj}) + \text{Var}(\tilde{t}_{kj}) + 2 \cdot \text{Cov}(\tilde{t}_{kj}, \tilde{t}_{mj}))
\end{aligned}$$

Denne variansen estimeres som følgende:

$$\begin{aligned}
\hat{\text{Var}}\left(\sum_{j=1}^5 \tilde{t}_j\right) &= \left(1 + \frac{x_{0m1}}{x_{0m2}}\right)^2 \hat{\text{Var}}(\tilde{t}_{m2}) + \left(1 + \frac{x_{0k1}}{x_{0k2}}\right)^2 \hat{\text{Var}}(\tilde{t}_{k2}) \\
&\quad + 2 \cdot \left(1 + \frac{x_{0m1}}{x_{0m2}}\right)\left(1 + \frac{x_{0k1}}{x_{0k2}}\right) \hat{\text{Cov}}(\tilde{t}_{k2}, \tilde{t}_{m2}) \\
&\quad + \sum_{j=3}^5 (\hat{\text{Var}}(\tilde{t}_{mj}) + \hat{\text{Var}}(\tilde{t}_{kj}) + 2 \cdot \hat{\text{Cov}}(\tilde{t}_{kj}, \tilde{t}_{mj})) \\
&= \frac{x_{0m1}}{x_{0m2}} \left(2 + \frac{x_{0m1}}{x_{0m2}}\right) \hat{\text{Var}}(\tilde{t}_{m2}) + \frac{x_{0k1}}{x_{0k2}} \left(2 + \frac{x_{0k1}}{x_{0k2}}\right) \hat{\text{Var}}(\tilde{t}_{k2}) \\
&\quad + 2 \cdot \left(\frac{x_{0m1}}{x_{0m2}} + \frac{x_{0k1}}{x_{0k2}} + \frac{x_{0m1}}{x_{0m2}} \frac{x_{0k1}}{x_{0k2}}\right) \hat{\text{Cov}}(\tilde{t}_{k2}, \tilde{t}_{m2}) \\
&\quad + \sum_{j=2}^5 (\hat{\text{Var}}(\tilde{t}_{mj}) + \hat{\text{Var}}(\tilde{t}_{kj}) + 2 \cdot \hat{\text{Cov}}(\tilde{t}_{kj}, \tilde{t}_{mj}))
\end{aligned}$$

der alle de estimerte variansene og kovariansene som inngår er beskrevet tidligere.

Etter siste likhetstegn i uttrykket over representerer de to første linjene cut-off strataet, mens størrelsesstratum 2 til 5 er representert ved respektive deler i summen i siste linje.

Ved å ta kvadratrotten av variansen får vi **standardavviket** (std) til estimatoren, som er et uttrykk for variasjonen i estimatoren målt i samme enhet som estimatoren:

$$\text{Std}(\tilde{t}_a) = \sqrt{\text{Var}(\tilde{t}_a)} \quad , \quad \text{som estimeres med:} \quad \hat{\text{Std}}(\tilde{t}_a) = \sqrt{\hat{\text{Var}}(\tilde{t}_a)}$$

Vi får **variasjonskoeffisienten**, dvs. relativt standardavvik, ved å dividere standardavviket på estimatoren:

$$\text{CV}(\tilde{t}_a) = 100 \cdot \text{Std}(\tilde{t}_a) / \tilde{t}_a \quad , \quad \text{som estimeres ved:} \quad \hat{\text{CV}}(\tilde{t}_a) = 100 \cdot \hat{\text{Std}}(\tilde{t}_a) / \tilde{t}_a$$

Forfatter: Anne Vedo

C. Utledning av kovariansen mellom de to ratene

Generelt er $Cov(aX, bY) = abCov(X, Y)$, så

$$Cov(\tilde{t}_{mh}, \tilde{t}_{kh}) = Cov\left(x_{0mh} \cdot \frac{\sum_{i \in S_h} y_{mi}}{\sum_{i \in S_h} x_{mi}}, x_{0kh} \cdot \frac{\sum_{i \in S_h} y_{ki}}{\sum_{i \in S_h} x_{ki}}\right) = x_{0mh} \cdot x_{0kh} \cdot Cov(\hat{R}_{mh}, \hat{R}_{kh})$$

På side 181 i Cochran (se [2]) er det en formel for å estimere kovariansen mellom to rater. Når vi setter inn ratene får vi:

$$\hat{C}ov(\hat{R}_{mh}, \hat{R}_{kh}) = \frac{1}{n_h(n_h - 1)\bar{x}_{mh}\bar{x}_{kh}} \sum_{i=1}^{n_h} (y_{mi}y_{ki} - \hat{R}_{mh}y_{ki}x_{mi} - \hat{R}_{kh}y_{mi}x_{ki} + \hat{R}_{mh}\hat{R}_{kh}x_{mi}x_{ki})$$

Forklaring:

For enkelhets skyld holder vi heretter stratumet fast, slik at vi kan droppe indeks h . Vi gjør en lineær tilnærming:

$$\hat{R}_m - R_m \approx \frac{\bar{y}_m - R_m\bar{x}_m}{\bar{X}_m}, \quad \hat{R}_k - R_k \approx \frac{\bar{y}_k - R_k\bar{x}_k}{\bar{X}_k}$$

Her betegner små x -er og y -er utvalgsgjennomsnitt, mens store X -er betegner populasjonsgjennomsnitt. R -ene er de sanne ratene, dvs. konstanter.

$$\begin{aligned} Cov(\hat{R}_m, \hat{R}_k) &= \frac{1}{\bar{X}_m\bar{X}_k} Cov(\bar{y}_m - R_m\bar{x}_m, \bar{y}_k - R_k\bar{x}_k) \\ &= \frac{1}{\bar{X}_m\bar{X}_k} [Cov(\bar{y}_m, \bar{y}_k) - Cov(\bar{y}_m, R_k\bar{x}_k) - Cov(R_m\bar{x}_m, \bar{y}_k) + Cov(R_m\bar{x}_m, R_k\bar{x}_k)] \end{aligned}$$

Fra Särndal et al (se [1]) side 211 oppgave 5.22, får vi at kovariansen mellom to gjennomsnitt ved enkelt tilfeldig utvalg kan estimeres forventningsrett med:

$$\hat{C}ov(\bar{y}_m, \bar{y}_k) = \frac{1-f}{n} \frac{1}{n-1} \sum_s (y_{mi} - \bar{y}_m)(y_{ki} - \bar{y}_k)$$

Når vi setter inn estimatoren for de 4 kovariansene får vi følgende estimator:

$$\hat{C}ov(\hat{R}_m, \hat{R}_k) = \frac{1-f}{\bar{X}_m\bar{X}_k n(n-1)} \cdot \left[\begin{aligned} &\sum_{i \in S} (y_{mi} - \bar{y}_m)(y_{ki} - \bar{y}_k) - \sum_{i \in S} (y_{mi} - \bar{y}_m)(R_k x_{ki} - R_k \bar{x}_k) \\ &- \sum_{i \in S} (R_m x_{mi} - R_m \bar{x}_m)(y_{ki} - \bar{y}_k) + \sum_{i \in S} (R_m x_{mi} - R_m \bar{x}_m)(R_k x_{ki} - R_k \bar{x}_k) \end{aligned} \right]$$

Dette er ikke noen ordentlig estimator før vi har byttet ut de ukjente R -ene med \hat{R} -er:

$$C\hat{ov}(\hat{R}_m, \hat{R}_k) = \frac{1-f}{\bar{X}_m \bar{X}_k n(n-1)} \cdot \left[\begin{aligned} & \sum_{i \in S} (y_{mi} - \bar{y}_m)(y_{ki} - \bar{y}_k) - \sum_{i \in S} (y_{mi} - \bar{y}_m)(\hat{R}_k x_{ki} - \hat{R}_k \bar{x}_k) \\ & - \sum_{i \in S} (\hat{R}_m x_{mi} - \hat{R}_m \bar{x}_m)(y_{ki} - \bar{y}_k) + \sum_{i \in S} (\hat{R}_m x_{mi} - \hat{R}_m \bar{x}_m)(\hat{R}_k x_{ki} - \hat{R}_k \bar{x}_k) \end{aligned} \right]$$

Hvis vi skriver ut alle leddene og utnytter at $\hat{R}_k \bar{x}_k = \bar{y}_k$, $\hat{R}_m \bar{x}_m = \bar{y}_m$, sitter vi igjen med de samme 4 leddene som Cochran har i sin sum.

Så forskjellen er at Cochran ikke har tatt med endelig korreksjonsfaktor $(1-f)$ i brøken foran summen, og at han deler på utvalgsgjennomsnittene i stedet for populasjonsgjennomsnittene. Når det gjelder endelig korreksjonsfaktor må det lønne seg å ta den med hvis utvalgsandelen er stor. Når det gjelder bruk av utvalgsgjennomsnitt kontra populasjonsgjennomsnitt, er det ikke så opplagt hva som er best. I Cochran er ikke populasjonsgjennomsnittene nødvendigvis kjente, så han har ikke noe valg.

Hvis vi bruker populasjonsgjennomsnitt, kan de forkortes mot populasjonstotalene, slik at

$$\begin{aligned} C\hat{ov}(\tilde{t}_m, \tilde{t}_k) &= N^2 \frac{1-f}{n(n-1)} \cdot \sum_{i=1}^n (y_{mi} y_{ki} - \hat{R}_m y_{ki} x_{mi} - \hat{R}_k y_{mi} x_{ki} + \hat{R}_m \hat{R}_k x_{mi} x_{ki}) \\ &= N^2 \frac{1-f}{n(n-1)} \cdot \sum_{i=1}^n (y_{mi} - \hat{R}_m x_{mi})(y_{ki} - \hat{R}_k x_{ki}) \end{aligned}$$

Hvis vi bruker utvalgsgjennomsnitt, får vi inn to faktorer $\frac{\bar{X}_m}{\bar{x}_m}$ og $\frac{\bar{X}_k}{\bar{x}_k}$ i tillegg:

$$C\hat{ov}(\tilde{t}_m, \tilde{t}_k) = N^2 \frac{1-f}{n(n-1)} \frac{\bar{X}_m}{\bar{x}_m} \frac{\bar{X}_k}{\bar{x}_k} \cdot \sum_{i=1}^n (y_{mi} - \hat{R}_m x_{mi})(y_{ki} - \hat{R}_k x_{ki})$$

I så fall får vi en kovariansestimator som er helt analog med Särndals variansestimator for rateestimatoren på side 236. Dette er kanskje det mest naturlige valget her, særlig siden Särndals variansestimater er brukt i notatet.

Referanser:

[1] Särndal, C-E., Swensson, B. & Wretman, J. (1992): *Modell Assisted Survey Sampling*. Springer.

[2] Cochran, W. G. (1977): *Sampling techniques*. 3rd ed. New York: Wiley.

D. Samvariasjon i sykefraværsavviket mellom menn og kvinner

Her er samvariasjon i sykefraværsavviket mellom menn og kvinner i utvalgte strata undersøkt.

Hvert punkt i figurene under gjelder avvikene:

$$plot_m = (y_{mi} - \hat{R}_{mh} x_{mi}) \text{ og } plot_k = (y_{ki} - \hat{R}_{kh} x_{ki}) \text{ for en og samme bedrift,}$$

der $\hat{R}_{mh} = \frac{\bar{y}_{mh}}{\bar{x}_{mh}}$ og $\hat{R}_{kh} = \frac{\bar{y}_{kh}}{\bar{x}_{kh}}$ er gjennomsnittlig sykefraværsandel i utvalget innenfor stratum h hhv. blant menn og kvinner.

Følgende notasjon er benyttet:

y_{mi} angir antall tapte dagsverk grunnet egenmeldt sykefravær for menn i bedrift i.

y_{ki} angir antall tapte dagsverk grunnet egenmeldt sykefravær for kvinner i bedrift i.

\bar{y}_{mh} angir gjennomsnittlig antall tapte dagsverk grunnet egenmeldt sykefravær for menn i stratum h.

\bar{y}_{kh} angir antall tapte dagsverk grunnet egenmeldt sykefravær for kvinner i stratum h.

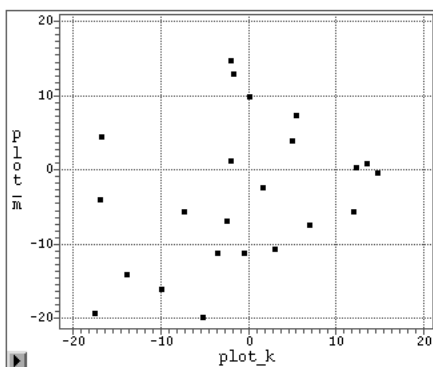
x_{mi} er antall ansatte menn i utvalget i bedrift i.

x_{ki} er antall ansatte kvinner i utvalget i bedrift i.

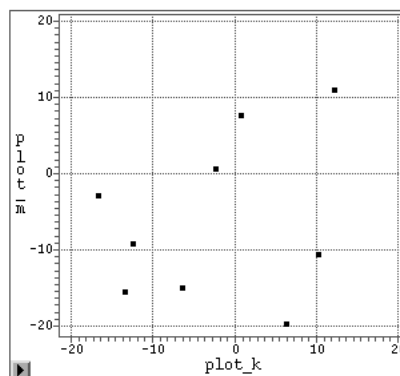
\bar{x}_{mh} er gjennomsnittlig antall ansatte menn i utvalget innenfor stratum h.

\bar{x}_{kh} er gjennomsnittlig antall ansatte kvinner i utvalget innenfor stratum h.

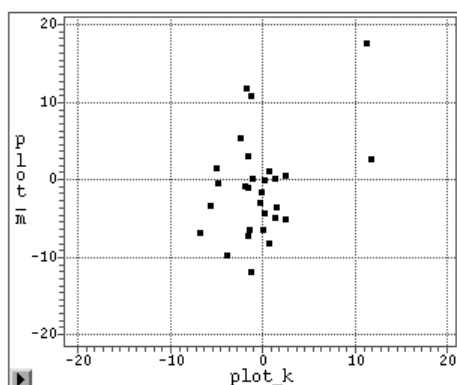
```
nar_str2 = 04
strat = 4
```



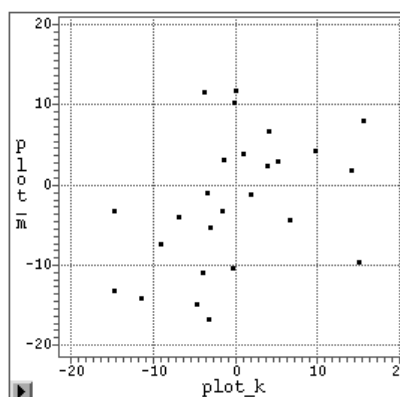
```
nar_str2 = 04
strat = 5
```



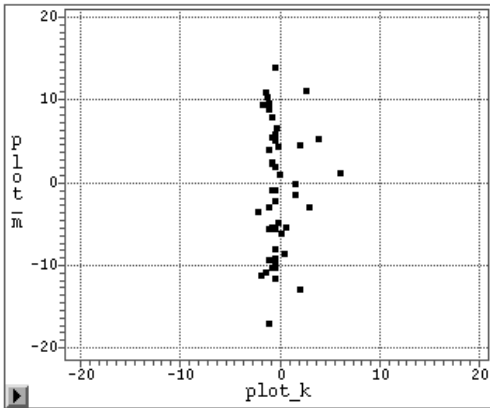
```
nar_str2 = 08
strat = 3
```



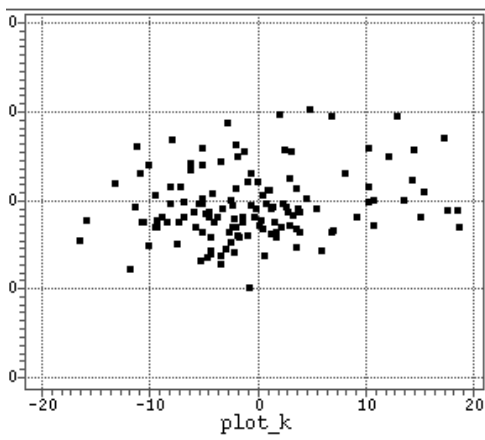
```
nar_str2 = 08
strat = 4
```



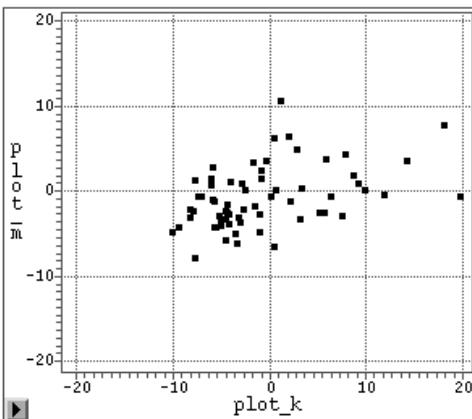
```
nar_str2 = 12
strat = 3
```



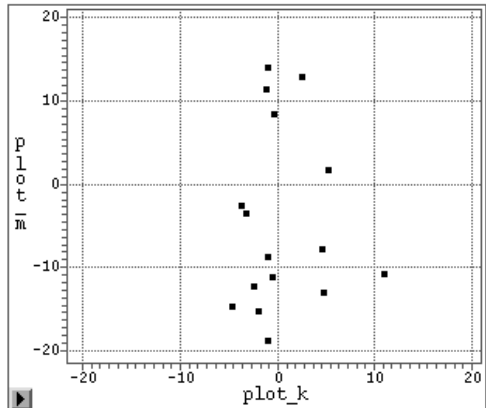
```
nar_str2 = 37
strat = 3
```



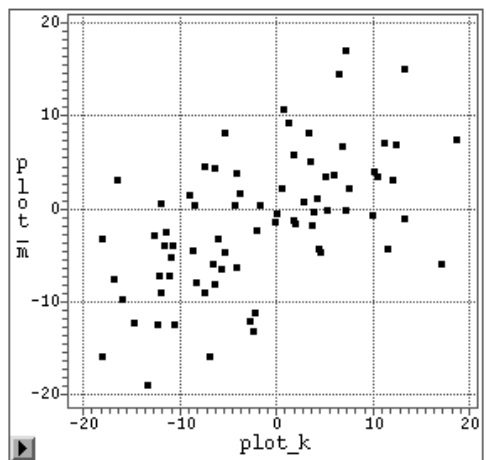
```
nar_str2 = 40
strat = 4
```



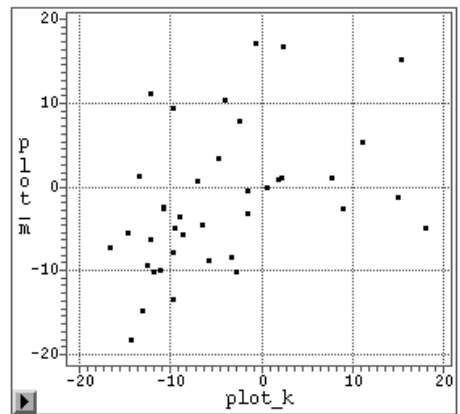
```
nar_str2 = 12
strat = 4
```



```
nar_str2 = 37
strat = 4
```



```
nar_str2 = 40
strat = 5
```



Konklusjon

De strataindelte figurene over tyder på at det er noe samvariasjon i sykefraværet til menn og kvinner innenfor samme bedrift.



Skjema for føring av egenmeldt sykefravær

Vedlegg E



Hvis endring, før opp korrekt
antall ansatte her:

Viktig: Blanketten skal leses maskinelt, så det er derfor viktig at utfyllingen blir utført nøyaktig. **Bruk blå eller svart penn.**

Sett kryss slik: og ikke slik: Hvis kryss i feil rute:

Skriv tall slik:

1	2	3	4	5	6	7	8	9	0
---	---	---	---	---	---	---	---	---	---

Opplysninger om bedriften:

Kontakt-
person:

E-post:

Telefon:

Telefaks:

Fyll ut her hvis opplysninger mangler eller er endret:



Skjemaet gjelder kun den utvalgte bedriften og ikke hele foretaket dersom bedriften er en del av et større foretak. Se bedriftens organisasjonsnummer og antall ansatte trykt ovenfor. Både faste og midlertidig ansatte som mottar lønn under sykdom, skal tas med. Er det svært vanskelig å oppgi tall for bedrifter enkeltvis, ta kontakt slik at vi kan finne andre løsninger.

Skjemaet gjelder kun egenmeldt sykefravær. Fravær pga. sykt barn holdes utenfor. Dersom bedriften ikke har hatt egenmeldt sykefravær i det aktuelle kvartalet, skal skjemaet likevel sendes inn. Før da opp "0" i "I alt" kolonnen.

Eventuelle spørsmål om utfyllingen kan rettes til:

Telefon: 62 88 50 65

E-postadresse: sykefra@ssb.no

Telefaks: 62 88 55 51



1 Har bedriften gitt anledning til utvidet egenmeldingsperiode utover 3 sammenhengende kalenderdager i

⊥

Ja → Hvor mange dager kan en enkel egenmelding vare?

Nei

Antall kalenderdager

2 Egenmeldt sykefravær for bedrifter.

	Menn	Kvinner	I alt
Antall <u>personer</u> med egenmeldt sykefravær	<input type="text"/>	<input type="text"/>	<input type="text"/>
Antall <u>egenmeldinger</u>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Antall egenmeldte sykefraværs <u>dager</u>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Antall <u>dagsverk</u> med egenmeldt sykefravær med én desimal	<input type="text"/>	<input type="text"/>	<input type="text"/>

⊥

Eventuell annen informasjon:

Hvordan går man fram ved utfyllingen av sykefraværstallene?

Antall personer med egenmeldt sykefravær:

Her føres opp det samlede antall personer som har hatt ett eller flere egenmeldte sykefravær i løpet av kvartalet. Personer som har hatt flere slike sykefraværstilfeller telles bare én gang.

Antall egenmeldinger:

Her føres opp det totale antall egenmeldinger i kvartalet. En egenmelding defineres som én eller flere sammenhengende kalenderdager med sykdom. Har en ansatt flere egenmeldinger i kvartalet, føres følgelig alle egenmeldingene opp. Deltidsarbeidende som ikke jobber daglig, skal regnes som en sammenhengende egenmelding fra den første sykefraværsdagen til vedkommende er tilbake på jobb igjen.

Antall egenmeldte sykefraværsdager:

Her oppgis det samlede antall sykefraværsdager som alle egenmeldinger utgjør i perioden. Det er kun normalarbeidsdager som her skal regnes med som sykefraværsdager. D.v.s. at generelle fridager og helgedager i sykefraværperioden skal holdes utenfor. Har man vært fraværende under en halv dag, telles ikke denne dagen med.

Antall dagsverk med egenmeldte sykefravær:

Her føres opp det samlede antall dagsverk som bortfaller pga. egenmeldt sykefravær. Dagsverk med sykefravær beregnes ved å ta utgangspunkt i antall sykefraværsdager og korrigerer for stillingsandel dersom det dreier seg om deltidsansatte. (Hvis stillingsandelen er varierende for enkelte ansatte, kan det brukes et anslag for gjennomsnittlig stillingsandel). For heltidsansatte blir antall sykefraværsdagsverk og sykefraværsdager det samme.

Formelen for dagsverk med sykefravær blir som følger:

Antall sykefraværsdager * Stillingsandel = Dagsverk med sykefravær.

Eks. på beregning av dagsverk med sykefravær:

- En arbeidstaker som jobber 4 timer daglig, er syk i tre dager. Dette gir følgende dagsverk med sykefravær (vi antar her at det er 7,5 timer i en normal hel arbeidsdag):

$$3 * \frac{4}{7,5} = 1,6$$

- En arbeidstaker i en 60% stilling er syk én dag. Dette gir følgende dagsverk med sykefravær:

$$1 * 60/100 = 0,6$$

- En ansatt som arbeider i $\frac{3}{4}$ stilling og er syk i to dager, får følgende dagsverk med sykefravær:

$$2 * \frac{3}{4} = 1,5$$

- En ansatt som jobber 25 timer i uken og har to dagers sykefravær, får dagsverk med sykefravær som følger:

$$2 * \frac{25}{37,5} = 1,3$$

(Vi antar her at full arbeidsuke utgjør 37,5 timer.)



Tilbakerapportering av sykefraværsprosenter:

Sykefraværsprosenten er beregnet som dagsverk med sykefravær i prosent av avtalte dagsverk.



Dagsverk med egenmeldt sykefravær bygger på tallene som bedriften rapporterte til SSB på skjema for de aktuelle kvartalene.

Dagsverk med legemeldt sykefravær bygger på data fra et register i Rikstrygdeverket over sykemeldinger fylt ut av leger. Tallene gjelder bare bosatte arbeidstakere i Norge 16-69 år.

Avtalte dagsverk er beregnet ut fra data i Rikstrygdeverket sitt arbeidstakerregister. Her ligger data som arbeidsgiverne sender om inntak og opphør av arbeidsforhold og den arbeidstid pr. uke som er knyttet til disse arbeidsforholdene. I beregningene er det gjort fratrukk for feriedagsverk og det er gjort beregninger av arbeidstid som i begge tilfeller bygger på nasjonale gjennomsnitt. For små bedrifter kan dette gi en viss feil i sykefraværsprosenten. Tallene gjelder bare bosatte arbeidstakere i Norge 16-69 år.

På SSB sin webside: <http://www.ssb.no/sykefra/> får du mer informasjon, og tall for sykefravær.

Spørsmål eller kommentar om tilbakerapporterte tall kan rettes til:

Christoffer Berge, tlf: 21 09 46 44, e-postadresse: sykefra@ssb.no eller faksnr: 21 09 49 96.



Takk for innsatsen!

«Navn_»
«Adr_»
«Pnr_» «Psted_»

Org.nr.: «Orgnr_f_»

Oslo, 12.01.2006
Deres ref.: , Vår ref.: 06/12
Saksbehandler: Christoffer Berge
Seksjon for arbeidsmarkedsstatistikk

Din bedrift er trukket ut til undersøkelsen om egenmeldt sykefravær

Kvartalsvis statistikk over sykefraværet utarbeides av Statistisk sentralbyrå (SSB) på grunnlag av bedriftenes tall for egenmeldt sykefravær og registerbaserte tall for legemeldt sykefravær. Statistikken viser hvordan sykefraværet utvikler seg over tid i ulike næringer og er et nyttig redskap i arbeidet med å redusere sykefraværet.

SSB fornyer utvalget en gang i året. Din bedrift er nå med i utvalget. Små og mellomstore bedrifter må regne med å delta i ca. fire år, mens store bedrifter vil være med hele tiden. Bedriften vil få tilsendt et skjema for føring av egenmeldt sykefravær per kvartal. Se vedlagt eksempel på skjema. Vi gjør oppmerksom på at dette skjemaet ikke skal sendes inn nå. Skjema for 1. kvartal 06 blir sendt ut i uke 13.

Opplysningene som gis er underlagt *taushetsplikt* etter statistikklovens § 2-4, og vil bare bli brukt til å utarbeide offisiell statistikk, og til statistisk bruk for forskning og offentlig planlegging.

Svarfrist

Første skjema for utfylling blir sendt ut i uke 13 og svarfristen for 1. kvartal er 4. mai 2006.

Som takk for hjelpen får svarbedriftene etter hvert tilbakerapportert tall for *alt* sykefravær i kvartalet, inkludert det som er legemeldt, både for bedriften selv, næringsgruppen bedriften tilhører og for hele landet.

Lurer dere på noe?

Spørsmål kan rettes til Christoffer Berge på tlf. 21 09 46 44 eller Jørn Ivar Hamre på tlf. 21 09 49 29.
E-post: sykefra@ssb.no

Opplysningsplikt

Opplysningene om det egenmeldte sykefraværet hentes nå inn med hjemmel i statistikkloven § 2-2. Dette er et enkeltvedtak dere kan klage på etter forvaltningslovens kapittel VI dersom dere mener dere ikke har plikt eller lovlig adgang til å gi opplysningene. Etter statistikkloven § 2-3, har Statistisk sentralbyrå myndighet til å ilegge dem som ikke gir opplysninger innen en oppgitt frist tvangsmulkt. Merk atplikten til å gi opplysninger ikke faller bort selv om mulkten blir betalt. Nærmere informasjon om klageretten er vedlagt.

Med vennlig hilsen

Helge N. Næsheim
seksjonssjef

Klage på vedtak om opplysningsplikt

Hjemmelsgrunnlag

Oppgavene innhentes med hjemmel i lov om offisiell statistikk og Statistisk sentralbyrå (statistikkloven) av 16.06.1989 nr 54 § 2-2, jf. kgl.res. av 16.06.1989 nr 387 og delegasjonen fra Finansdepartementet av 13.02.1990 nr 1228.

Klageadgang

Man kan klage på vedtaket om opplysningsplikt etter reglene i forvaltningsloven kapittel VI. Dette gjelder også senere vedtak om avvisning av klage. Følgende regler gjelder dersom man skal klage:

- Vedtaket påklages til Finansdepartementet som klageinstans, men klagen skal fremsettes skriftlig for, det vil si sendes til, **Statistisk sentralbyrå, Seksjon for arbeidsmarkedsstatistikk**.
- Fristen for å klage er 3 - tre - uker fra det tidspunkt foretaket har mottatt vedtaket. Det er tilstrekkelig at klagen er postlagt innen klagefristens utløp.

Etter § 29 i forvaltningsloven kan klagefristen i særlige tilfeller forlenges. Etter § 31 kan en klage som er sendt inn for sent allikevel bli behandlet dersom den opplysningspliktige i sin klage sannsynliggjør at han eller hun ikke kan lastet for overskridelsen av tidsfristen.

Få med i klagen hvilken statistisk undersøkelse det gjelder og hvem den opplysningspliktige er.

Videre må dere begrunne klagen, og dere bør også ta med andre opplysninger som kan ha betydning for vurderingen av klagen.

Klagen må være undertegnet av den opplysningspliktige eller en annen person med fullmakt. Oppgi nummeret som Statistisk sentralbyrå bruker på foretaket. Det står ved siden av postadressen på forsiden av arket.

De sist utgitte publikasjonene i serien Notater

- 2006/32 E.Cometa Rauan: Undersøking om foreldrebetaling i barnehagar, januar 2006. 46s.
- 2006/33 T. Skarðhamar: Kriminalitet gjennom ungdomstiden blant nordmenn og ikke-vestilige innvandrere. En analyse av fødselskullet 1977. 36s.
- 2006/34 N. Hagesæther og L.-C. Zhang: Om arbeidsledighet i AKU og Arena. 19s.
- 2006/35 T. Hægeland, Lars J. Kirkebøen og Oddbjørn Raaum: Skoleresultater 2005. En kartlegging av karakterer fra grunnskoler og videregående skoler i Norge. 83s.
- 2006/36 S. Skaare: Undersøkelse om «Utbrenthet i enkelte yrker» 2005. Dokumentasjonsrapport. 68s.
- 2006/37 O.F. Vaage: Barn og unges idrettsdeltakelse og foreldres inntekt. Analyse med data fra Levekårsundersøkelsen 2004. 31s.
- 2006/38 A.Vedø og L. Solheim: En praktisk innføring i utvalgsplanlegging. 40s.
- 2006/39 H.C. Hougen: Samordnet levekårsundersøkelse 2005 - tverrsnittundersøkelsen. Dokumentasjonsrapport. 156s.
- 2006/40 T. Nøtnæs, S. Bytingsvik og B. Hole: Resultater fra brukertesting av ssb.no. 34s.
- 2006/41 KOSTRA. Arbeidsgrupperapporter 2006. 169s.
- 2006/42 T. Gulbrandsen: Levekårsundersøkelse blant studenter. Dokumentasjonsrapport. 66s.
- 2006/43 A-G. Jørstad: Overvåkingssystemet for bedrifter i Bof. 19s.
- 2006/44 M. Høstmark og B.O. Lagerstrøm: Undersøkelse om Arbeidsmiljø: Destruktiv atferd i arbeidslivet. Dokumentasjonsrapport. 43s.
- 2006/45 T.K. Schjerven og K.Å. Wass: Faglig modell og rammeverk i StatRes. 67s.
- 2006/47 K. Henriksen: Utvalgsplan til konsumprisindeksens nye matvareindeks - Basert på strekkodedata. 23s.
- 2006/48 A.B. Thorud, D. Rafat, S. Ferstad og E. Vinju: Tverrgående revisjon i KOSTRA - Bedring av påliteligheten i nøkkeltallene. 65s.
- 2006/49 T. Granseth: Grensehandel. En analyse av kvaliteten av data. 48s.
- 2006/50 E. Engelién, H. Høie og M. Steinnes: Bygging i strandsona. Metode og resultater. 18s.
- 2006/51 A. Akselsen, K.I. Bøe og Ø. Sivertstøl: FD - Trygd. Dokumentasjonsrapport. Arbeidssøkere, 1.1.1992-30.4.2001. 75s.
- 2006/53 G. Claus: Inntekts- og formuesundersøkelsen for personlig næringsdrivende 2004. Dokumentasjon. 28s.
- 2006/52 L. Østby: Bruk av velferdsordninger blant nyankomne innvandrere fra de nye EØS-landene i 2005. 34s.
- 2006/54 J. Heldal: Logistisk regresjon - kurskompendium i byråskolens kurs SM507. 51s.
- 2006/55 L.H. Thingstad: Varehandelsstatistikk 2002 - omsetning etter varegruppe. 59s.
- 2006/56 H.Kull Brofoss og A. Barstad: Internasjonale erfaringer med områderettede tiltak i storbyer. En litteraturstudie. 101s.
- 2006/57 B. Bye og I. Ringdal: Disaggregering av helse-, omsorg- og utdanningstjenester i MSG6-modellen. 39s.
- 2006/59 Leiemarkedsundersøkelsen 2006. Dokumentasjonsrapport. 43s.