

*Øystein Amundrud og Irene Tuveng*

**Utredning av utvalgsplan for  
sentral sykefraværstatistikk**

Notater

# Innhold

<b>1 Sammen drag og konklusjon</b>	<b>2</b>
<b>2 Bakgrunn og formål</b>	<b>3</b>
<b>3 Problemstilling</b>	<b>3</b>
<b>4 Krav til nøyaktighet på estimatene</b>	<b>4</b>
<b>5 Valg av estimeringsmetode - Rate-estimatoren</b>	<b>6</b>
5.1 Stratifisert utvalgsplan	7
<b>6 Aktuell stratainn deling</b>	<b>8</b>
6.1 Stratifisering etter størrelse på bedrift	8
<b>7 Deskriptiv analyse av Bedrifts- og foretaksregisteret</b>	<b>9</b>
<b>8 Usikkerhetsberegninger</b>	<b>10</b>
8.1 Teoretiske betraktninger	10
<b>9 Simulering av sykefravær for å beregne usikkerhet ved ulike utvalgsplaner</b>	<b>12</b>
<b>10 Resultater av simuleringene</b>	<b>14</b>
<b>11 Andre kilder til usikkerhet</b>	<b>19</b>
<b>12 Rullering av utvalget</b>	<b>19</b>
<b>Æres den som æres bør</b>	<b>20</b>
<b>Litteratur</b>	<b>21</b>
<b>Vedlegg 1. Variansen til endringstall når vi ser på totalt sykefravær</b>	<b>23</b>
<b>Vedlegg 2. Næringsgrupperinger</b>	<b>25</b>
<b>Vedlegg 3. Bedriftsstørrelse og ansattprofil fordelt på næringshovedgrupper</b>	<b>27</b>
<b>De siste utgitte publikasjonene i serien Notater</b>	<b>36</b>

# 1 Sammendrag og konklusjon

Dette notatet inngår som en del i planleggingen av utvalgsundersøkelsen i prosjektet «Sentral sykefraværstatistikk».

I kapittel 2 gis det en kort oversikt over prosjektet «Sentral sykefraværstatistikk», bl.a. prosjektets organisering og hovedformål. Kapittel 3 leder leseren inn på problemstillingen vi vil se nærmere på i dette notatet. Bl.a. gis det her en beskrivelse av momenter som er vesentlige når vi skal anslå nødvendig utvalgsstørrelse.

Krav til nøyaktighet i de estimerte verdier diskuteres i kapittel 4. Vi ser her nærmere på hva en sentral statistikk skal brukes til og vurderer tallmaterialet i dagens kilder til statistikk. I kapittel 5 vurderes aktuelle stratainndelinger, før vi i kapittel 6 beskriver hvilken estimeringsmetode vi har valgt å benytte her. En deskriptiv analyse av trekkgrunlaget for denne analysen gis i kapittel 7.

I kapittel 8 starter vi med å vurdere minimumsgrensene mht utvalgsstørrelse for ulike detaljeringer på tallene og ser nærmere på varians innenfor og mellom strataene i trekkgrunlaget. Kapittel 9 beskriver hvordan vi har simulert sykefraværet innenfor hver bedrift i trekkgrunlaget gjennom å forutsette en enkel sammenheng mellom sykefravær og antall ansatte i hver bedrift.

Kapittel 10 gjengir resultatene fra simuleringene sammen med en oppsummering av viktige resultater. Kapittel 11-13 beskriver kort andre momenter som har betydning for usikkerhet i estimatene.

## Konklusjon

I samråd med brukerne velger vi å ta utgangspunkt i en utvalgsstørrelse på **10 000 bedrifter** som skal delta i undersøkelsen hvert kvartal. Vi vil benytte en stratifisert utvalgsplan og stratifiserer da etter næring og størrelse på bedrift. Usikkerheten i estimatene vil da forhåpentligvis være så liten at man kan identifisere endringer på 7 % innenfor hver næring på 2-siffernivå. Dette forutsetter at frafallet ikke blir for stort, og at en viss del av utvalget holdes fast fra gang til gang. Rullering og frafallshåndtering behandles ikke spesielt i dette notatet. Vi vil operere med totaltelling av de aller største bedriftene innenfor hver næring og de aller minste bedriftene får en trekk sannsynlighet lik null («cut-off»). Hvor vi setter grensen for totaltelling og «cut-off» varierer imidlertid for hver næring avhengig av bedriftsstrukturen i næringen. Vi har regnet med at vi kan sette «cut-off» grensen lavere enn ved 4 ansatte for enkelte strata. Enkelte små næringer med særlig få og små bedrifter må overrepresenteres i utvalget. Det bør være mulig å øke utvalget her på bekostning av større næringer.

Den endelige utvalgsplanen vil bli dokumentert i et senere notat hvor også de mer praktiske sider ved utvalgsundersøkelsen vil bli dokumentert.

## 2 Bakgrunn og formål

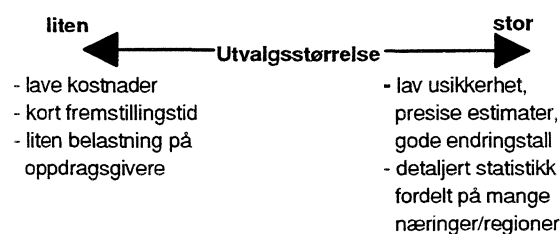
Prosjektet «Sentral sykefraværstatistikk» er et samarbeidsprosjekt mellom Rikstrygdeverket (RTV) og SSB. Prosjektet finansieres av Sosial- og helsedepartementet (SHD). I referansegruppen for prosjektet sitter foruten SHD, RTV og SSB også Direktoratet for arbeidstilsynet, Landsorganisasjonen i Norge (LO), Planleggings- og samordningsdepartementet (PSD), Næringslivets Hovedorganisasjon (NHO), Handels- og Servicenæringens Hovedorganisasjon (HSH), Kommunenes Sentralforbund (KS), Yrkesorganisasjonenes Sentralforbund (YS), Forsikringsselskapenes Arbeidsgiverforening og Akademikernes Fellesorganisasjon (AF).

Hovedformålet med prosjektet er å legge grunnlaget for en ny sentral sykefraværstatistikk som er ment å omfatte alt *sykefravær grunnet egen sykdom*. Statistikken skal bygges opp ved hjelp av RTVs registre over legemeldt sykefravær med varighet 4 dager eller mer og data over sykefravær med varighet 1-3 dager samlet inn via en *bedriftsbasert utvalgsundersøkelse*. Denne undersøkelsen skal baseres på et utvalg bedrifter som bes rapportere inn summariske tall for sykefraværet fordelt på kjønn hvert kvartal.

I dette notatet starter vi utredning av utvalgsplanen i en sentral sykefraværstatistikk. Spesielt konsentrerer vi oppmerksomheten om å anslå hvilken *utvalgsstørrelse* som er tilstrekkelig til vårt formål. Notatet er ikke ment å være noen endelig dokumentasjon av utvalgsplanen, men kun som en del av den. En endelig dokumentasjon kommer senere.

## 3 Problemstilling

Faktorer som har betydning for valg av utvalgsstørrelse er bl.a. kostnader, ønsket nøyaktighet på estimat, ønsket detaljeringsgrad på måltallene med hensyn til, f.eks. næring og region, bedriftsstørrelse og rapporteringsbelastning for den enkelte bedrift.



For en gitt utvalgsstørrelse har *estimeringsmetode* stor betydning for presisjonsgraden på estimatene. I dette notatet velger vi en estimeringsmetode ut i fra en forutsetning om at summariske absoluttmål for sykefraværet er relativt sterkt positivt korrelert med antall ansatte i bedriften. Det er dermed mulig å utnytte kjent registerkunnskap om bedrifters størrelse ved estimeringen av sykefraværsmåltall slik at nødvendig utvalgsstørrelse for gitt krav til nøyaktighet i estimatene våre reduseres.

Siden ønsket nøyaktighet på estimatet er avgjørende for valg av utvalgsstørrelse, starter vi med å angi krav til nøyaktighet som vi skal basere oss på i en sentral sykefraværstatistikk.

## 4 Krav til nøyaktighet på estimatene

I rapporten «Sentral sykefraværstatistikk», SHD 1994, beskrives formålet med en ny sentral sykefraværstatistikk. Det overordnede målet er [...] å kunne foreta en fullstendig, løpende beskrivelse og analyse av sykefraværet i Norge bl.a. som ledd i en helhetlig strategi for å bedre arbeidsmiljøet og redusere sykefraværet. ....det er forutsatt at sykefraværstatistikken skal dekke en rekke ulike formål bl.a. gi grunnlag for pålitelige utgiftsanslag, evaluering av tiltak, identifisering av risikogrupper og gi grunnlag for forskning. ...]. Det er dermed klart at i tillegg til en ren kvantifisering av sykefraværet er man ute etter gode *endringstall* (endringen i sykefraværet påfølgende kvartaler - fra år til år osv.). I tillegg til totaltall er man også interessert i estimater spesifikt for de ulike næringer, regioner og kjønn.

Mulige tema for en analyse kan f.eks. være å se på hvilke næringer som har særlig høyt sykefravær, utvikling av sykefraværet fra år til år eller fra kvartal til kvartal (sesongvariasjoner), hvordan fraværet fordeler seg på privat / offentlig sektor eller regionvise forskjeller, for å nevne noe.

For den delen av sykefraværet som baseres på en utvalgsundersøkelse skal vi begrense oss noe mht detaljeringsgrad på tallene våre. Spesielt av hensyn til rapporteringsbelastning for bedriftene vil vi ikke stille de samme kravene til statistikken for 1-3 dagers fraværet som vi vil gjøre til den øvrige sykefraværstatistikken. Vi vil ta utgangspunkt i at statistikken over 1-3 dagers fraværet skal gi gode tall med hensyn til næring, kjønn og region, og se på utvalgsstørrelser som er nødvendig alt etter hvor stor usikkerhet som er akseptabel.

Dersom vi var interessert i å se på endringer i det *totale* sykefraværet vil den delen av datamaterialet som bygger på heldekkende registerdata bidra til å redusere nødvendig utvalgsstørrelse for gitt krav til presisjon<sup>1</sup>. Imidlertid er det ønskelig å skille ut sykefravær med varighet 1-3 dager og se på dette spesielt. Det kan være interessant å studere om dette sykefraværet avviker fra sykefravær av lengre varighet og i så fall hvordan. Det er ønskelig med relativt høy presisjon i dataene, da statistikken skal brukes blant annet til å fatte tiltak og følge opp tiltak som har til hensikt å redusere sykefraværet.

Da vi er spesielt interessert i endringstallene velger vi å ta utgangspunkt i et estimat for endring når vi bestemmer oss for krav til presisjon

Hvor små endringer trenger vi å kunne observere i 1-3 dagers sykefraværet? I dagens statistikkilder opereres det med sykefraværspresenterer hvor sykefraværet måles i forhold til mulige dagsverk i perioden. I NHO's- og AD's sykefraværstatistikk presenteres endringer på ned til 0,1 prosentpoeng. Hva dette tilsvarer i prosentvise endringer avhenger selvfølgelig av nivåtallene. I dagens statistikker ligger sykefraværspresenteren (sykefravær i prosent av mulige dagsverk) for 1-3 dagers sykefraværet på rundt 1%, en endring på ett prosentpoeng tilsvarer da en endring på 10 % i realmåltallet. I «Sykefraværprosjektet, 1991-1993» LO/NHO-området, var målsetningen å redusere sykefraværet med 20 prosent. Med et nivå på 1-3 dagers sykefraværet som anslått over tilsvarer en 20 % reduksjon på rundt 0,2 prosentpoeng.

En sentral sykefraværstatistikk skal i det minste gi mulighet til å registrere de samme endringer som dagens kilder til statistikk. Hvilken utvalgsfeil det er i tallene som presenteres i dagens statistikker er uklart, men om vi nå går ut ifra at sykefraværspresenteren faktisk ligger i området

---

<sup>1</sup> I vedlegg 1 illustreres hvordan behovet for presisjon i estimatene for 1-3 dagers

0,5 - 1,5%, og setter som krav at det skal være mulig å observere endringer på 0,1 prosentpoeng med relativt stor sannsynlighet, betyr det at vi skal kunne observere endringer i antall sykefraværsgener fra periode til periode på anslagsvis 7-8%.

Det er naturlig å måle krav til nøyaktighet i % av estimatet. På den måten tar vi høyde for forskjellig utgangspunkt for estimert nivå på sykefraværet. Estimater for en endring mellom to perioder er lik differansen mellom to estimater for nivåtallene i hver av periodene.

La nå  $\tilde{t}_1$  og  $\tilde{t}_2$  betegne estimater for totalt sykefravær i to perioder man ønsker å sammenligne. Estimater for endringen er gitt ved  $\tilde{e} = \tilde{t}_2 - \tilde{t}_1$ . Variansen til endringen vil være gitt ved  $Var(\tilde{e}) = Var(\tilde{t}_2) + Var(\tilde{t}_1) - 2Cov(\tilde{t}_1, \tilde{t}_2)$ . For å forenkle analysen antar vi uavhengige totalestimater med tilnærmet lik varians, og trekking av nytt utvalg for hver periode<sup>2</sup>. Da blir variansen til endringsestimatoren  $Var(\tilde{e}) \approx 2 \cdot Var(\tilde{t})$ . Dvs variansen til endringen er ca. dobbelt så stor som variansen til hver av de to estimatene for nivåtallene. Om vi nå ønsker å identifisere en endring i 1-3 dagers sykefraværet på  $X$  % fra en periode til en annen krever vi at  $2 \cdot \sqrt{Var(\tilde{e})} / \tilde{t}_1 < X$  %<sup>3</sup>, altså at estimatet av endringen må ha et standardavvik som er mindre enn  $(X/2)$ % av totalestimatet. Dette vil igjen føre til at standardavviket til totalestimatet,  $\sqrt{Var(\tilde{t}_1)}$ , ikke må overstige  $(\frac{X}{2})\% / \sqrt{2}$  av totalestimatet,  $\tilde{t}_1$ .

Med  $X = 7\%$  gir det et krav på **2,47%**. Dvs våre estimater for sykefraværet bør ikke ha større varians enn at det utgjør maks 2,47% av estimatet selv. Jo større estimat for sykefraværet, jo større kan variansen være uten at kravet til presisjon overstiges.

Kravet til varians er satt ut i fra den forutsetning at det trekkes nytt utvalg hver periode. Om vi ikke gjør det men lar hele eller deler av utvalget være det samme flere ganger, betyr det at utvalgsstørrelsen kan reduseres med samme krav til presisjon dersom det kun er endringstallene vi er opptatt av. Imidlertid er det da fare for skjevheter. Nivåtallene kan fort bli dårligere fordi frafallsprosenten over tid gjerne blir større, representativiteten til utvalget kan forringes og enhetene som er med flere ganger kan endre sin adferd nettopp fordi de er med i utvalget. Disse problemene blir mindre dersom vi bytter ut deler av utvalget hver periode - *roterende utvalg*. I sentral sykefraværstatistikk er en hovedmålsetning gode endringstall, slik at vi nok vil velge et roterende utvalg. Hvordan utvalget kan roteres kommer vi tilbake til skriver vi noe mer om i kapittel 12.

---

<sup>2</sup> Antar nytt utvalg hver periode for å forenkle analysen. Imidlertid kommer vi nok til å rullere utvalget bare delvis.

<sup>3</sup> Under antagelse om normalfordelte estimater, vil dette gi 95%-konfidensgrenser.

## 5 Valg av estimeringsmetode - Rate-estimatoren

Vi skal estimere følgende totalsummer for sykefraværet(1-3 dager) i sentral sykefraværstatistikk;

- Antall personer med sykefravær
- Antall sykefraværstilfeller
- Antall sykefraværsdager
- Antall sykefraværsdagsverk

For alle disse variablene synes det rimelig å anta positiv korrelasjon med antall ansatte i bedriften. Vi har informasjon om antall ansatte innenfor hver bedrift tilgjengelig i Bedriftsregistret. Hvis antagelsen om sterk korrelasjon er riktig så vil denne tilleggsinformasjonen kunne bidra til å gi økt nøyaktighet i estimatet for en gitt utvalgsstørrelse dersom vi velger en metode for estimering som nyttiggjør seg av informasjonen om antall ansatte. En slik metode er *rate-estimering*.

Vi skal estimere populasjonstotalen:  $t = \sum_{i=1}^N y_i$ ,  $i = 1, 2, \dots, N$ .

Dersom  $y_i$  angir antall sykefraværsdagsverk i bedrift nr.  $i$ , så er  $t$  lik populasjonstotalen for antall sykefraværsdagsverk, osv.

Vi ønsker å estimere  $t$ . Der  $t$  kan være en av de 4 totalstørrelsene nevnt over. Dette skal vi gjøre ved å trekke et utvalg enheter fra populasjonen uten tilbakelegging.

$N$  = Totalt antall bedrifter

$n$  = Antall bedrifter i utvalget.

Vi har tilleggsinformasjon om populasjonen som vi skal benytte oss av. Antall ansatte innenfor bedrift  $i$  betegnes  $x_i$ .

Totalt antall ansatte i hele populasjonen:  $x_0 = \sum_{i=1}^N x_i$ .

Rate-estimatoren er nå gitt ved: 
$$\tilde{t} = x_0 \frac{\sum_{i \in s} y_i}{\sum_{i \in s} x_i} \quad (1)$$

Denne estimatoren er forventningsrett, og en varians som følger;

$$Var(\tilde{t}) = N^2 \frac{1-f}{n} (\sigma_y^2 + R^2 \sigma_x^2 - 2R \sigma_{xy}) \quad (2)$$

$$f = \frac{n}{N}, \quad R = \frac{t}{x_0}$$

$\sigma_y^2$ : Variansen til  $y$  (sykefraværet for alle bedriftene i populasjonen).

$\sigma_x^2$ : Variansen til  $x$  (antall ansatte i bedriftene i populasjonen).

$\sigma_{xy}$ : Kovariansen mellom  $x$  og  $y$ .

$f$  angir treksannsynligheten til bedriftene. Vi ser at jo sterkere korrelert  $x$  er med  $y$  (målt ved kovarians-leddet  $\sigma_{xy}$ ), jo mer reduseres variansen til estimatoren (les: usikkerheten i estimatet).

### 5.1 Stratifisert utvalgsplan

I sykefraværstatistikken ønsker vi å publisere tall over 1-3 dagers sykefraværet fordelt på næring, region og fordelt på kjønn. Dette setter krav til størrelse på utvalget innenfor de forskjellige gruppene. For å kunne utarbeide statistikk må vi sikre oss at de enkelte gruppene sikres god representasjon i utvalget. Stratifisering reduserer usikkerheten i estimatene hvis strataene er mer homogene enn populasjonen (mindre varians). Mulige stratifiseringsvariable som vi vil vurdere er næring, størrelse på bedrift og geografisk beliggenhet.

Å stratifisere beskrives ofte som et alternativ til rateestimering. Vi skal forsøke en metode som kombinerer dette. Dvs vi estimerer verdier for de ulike totalstørrelsene innenfor hvert stratum vha rate-estimering, hvor utvalget er trukket tilfeldig og med gitt trekk sannsynlighet, og summerer opp.

Strataene betegnes med  $U_1, U_2, \dots, U_H$ , og delutvalgene betegnes med  $s_1, s_2, \dots, s_H$

$$N = \sum_{h=1}^H N_h, \quad n = \sum_{h=1}^H n_h.$$

$H$  = antall strata

$N_h$  = Populasjonen med bedrifter i stratum nr. h

$n_h$  = Utvalget med bedrifter trukket fra stratum nr. h

Metoden for estimering av  $t$  slik den er gjengitt over gjelder nå innenfor hvert stratum, slik at:

$$t = \sum_{h=1}^H t_h, \text{ hvor } t_h \text{ er } y\text{-totalen i stratum } U_h.$$

Rate-estimatoren er gitt ved;

$$\tilde{t} = \sum_{h=1}^H \tilde{t}_h = \sum_{h=1}^H \left[ x_{0h} \frac{\sum_{i \in s_h}^{N_h} y_i}{\sum_{i \in s_h}^{N_h} x_i} \right] \quad (3)$$

Denne estimatoren er fremdeles forventningsrett. Dersom vi antar at  $\tilde{t}_1, \tilde{t}_2, \dots, \tilde{t}_H$  er uavhengige kan vi skrive variansen til estimatoren som en sum av variansene innenfor hvert av strataene (her benyttes en alternativ formel for varians til den gjengitt i (2)) ;

$$Var(\tilde{t}) = \sum_{h=1}^H Var(\tilde{t}_h) = \sum_{h=1}^H \left[ N_h^2 \cdot \frac{1-f_h}{n_h} \cdot \frac{1}{N_h} \sum_{i \in U_h}^{N_h} (y_i - R_h x_i)^2 \right] \quad (4)$$

Vi ser av denne variansformelen at både antall bedrifter som er med i undersøkelsen, størrelse på bedriftene, sykefraværsandelen ( $R$ ) og antall strata har betydning for størrelsen på variansen til estimatet.

Vi vil nå se nærmere på mulige nærings- og regionsgrupperinger. Deretter vil vi se på registeret som danner trekkgrunnlaget i undersøkelsen for bl.a. å se på hvordan antall ansatte fordeler seg på bedriftene innenfor ulike næringer og regioner.



## 6 Aktuell stratainndeling

Det er spesielt ønskelig at det skal være mulig å gi næringsfordelte tall for 1-3 dagers sykefraværet fordelt på kjønn. Regionsfordeling vil også bli vurdert. Utfra usikkerhets- og kostnadsbetraktninger er det begrenset hvor fin næringsinndeling og regionsinndeling vi kan benytte. Vi skal vurdere følgende alternativer for stratainndelingen:

### Næring

<i>Grad</i>	<i>Antall nivåer</i>	<i>Beskrivelse</i>
1	1	Ingen inndeling etter næring
2	10 <sup>4</sup>	SSB næringsgrupperingsstandard, 10 nivåer
3	15	SSB næringsgrupperingsstandard, gruppe A-O
4	58	SSB næringsstandard, 2-siffernivå.

### Region

<i>Grad</i>	<i>Antall nivåer</i>	<i>Beskrivelse</i>
1	1	Hele landet
2	5	Østlandet, Sørlandet, Vestlandet, Midt-Norge, Nord-Norge
3	7 <sup>5</sup>	Oslo + Akershus, Østfold + Vestfold + Buskerud + Telemark, Hedmark + Oppland, Aust-Agder + Vest-Agder + Rogaland, Hordaland + Sogn og Fjordane + Møre og Romsdal, Sør-Trøndelag + Nord-Trøndelag, Nordland + Troms + Finnmark
4	19	Alle landets fylker

### 6.1 Stratifisering etter størrelse på bedrift

I tillegg til å stratifisere etter næring og region vi også å stratifisere med hensyn til størrelse på bedrift. Ved siden av at vi sikrer representativitet av bedrifter etter størrelsesfordeling er hensikten med denne stratifiseringen å sikre en overrepresentasjon av de største bedriftene. Ved å velge en større trekk sannsynlighet for store bedrifter kan vi nå en gitt andel arbeidstakere med et mindre utvalg enn om trekk sannsynligheten var lik for alle bedrifter. Samtidig kompliseres ikke metoden når vi holder trekk sannsynligheten lik innenfor hvert stratum. Rate-estimering innenfor hvert stratum er fremdeles forventningsrett. Størrelsesstrukturen innenfor forskjellige næringer kan være veldig forskjellig. Derfor skal vi også velge forskjellig størrelsesgrupperinger i ulike næringer.

<sup>4</sup> Se detaljert beskrivelse av de ulike næringsgrupperingsstandardene i vedlegg 1.

<sup>5</sup> Denne inndelingen følger inndelingen i REGARD (REGIONalt modellsystem for Arbeidsmarked og Demografi), utviklet ved Forskningsavdelingen i SSB, 1993. Inndelingen på grad 3 er lik den på grad 2 bortsett fra at Østlandet er oppdelt i 3 regioner

## 7 Deskriptiv analyse av Bedrifts- og foretaksregisteret

I dette notatet har vi benyttet SSB's Det sentrale bedrifts- og foretaksregisteret (BOF) som trekkgrunnlag for utvalgsundersøkelsen. Når den løpende undersøkelsen kommer i gang for fullt vil vi benytte bedriftsdelen av enhetsregisteret som trekkgrunnlag, bl.a. fordi dette registeret per i dag er mer dekkende enn BOF, og det gir opplysninger om ansatte fordelt på kjønn. Enhetsregisteret er imidlertid noe ufullstendig ennå bl.a. med hensyn til næringskode, derfor benytter vi her BOF. Ved utgangen av første halvår 1997 skal oppdateringen av Enhetsregisteret være så godt som klar, slik at vi vil kunne benytte dette ved oppstart av utvalgsundersøkelsen 3. kvartal 1997.

Vi har gjort oss kjent med BOF gjennom versjonen fra juni 1996. I dette kapitlet forsøker vi å beskrive bedriftspopulasjonen i Norge ved å skjelne litt til noen mulige stratainndelinger fra forrige kapittel.

Når man fra BOF trekker ut alle bedrifter i ordinær drift (Record-type=2, Tilstand=9) og i tillegg krever at bedriftene skal ha minst én ansatt (Sysselsetting > 0) ender man opp med en populasjon bestående av 184.235 bedrifter. Av disse mangler 122 bedrifter næringskode, 3 bedrifter er i næring P og Q, og 53 bedrifter har regionkode Svalbard, Jan Mayen eller Nordsjøen. Disse næringene og regionene har vi utelatt fra trekkgrunnlaget her, men det betyr ikke at de ikke skal inngå i det endelige opplegget for utvalgsplan. Vårt trekkgrunnlag består dermed av 184.057 bedrifter.

I vedlegg 3 foreligger en oversikt over bedriftsstørrelse og ansattprofil fordelt på næringshovedgruppe.

Grovt sett kan man si at de fleste bedriftene er små og mellomstore (bedrifter med 1-4 ansatte utgjør 70 % av den totale bedriftspopulasjonen men dekker bare 16 % av de sysselsatte), mens de fleste sysselsatte jobber i mellomstore og store bedrifter (bedrifter med mer enn 20 ansatte utgjør bare 15 % av bedriftspopulasjonen men sysselsetter hele 70 % av arbeidstakerne). Dette er en grunn til at det synes fornuftig å overrepresentere store bedrifter i utvalget.

Det er imidlertid viktig å være klar over at det er store individuelle forskjeller, spesielt mellom næringer. Som eksempel kan nevnes at Næring D (industri) består av mange mellomstore og store bedrifter, mens næring G (varehandel, reparasjoner) består av veldig mange små bedrifter. Dette har betydning for valg av størrelsesstrata og trekk sannsynligheter, som dermed må tilpasses de ulike strataene for å oppnå optimal estimering.

Det eksisterer foreløpig en del begrensninger i BOF. Bl.a. er primærnæringene (næring A og B) kun representert ved bedrifter knyttet til planteproduksjon i jordbruk og tjenester tilknyttet skogbruk samt fiskeoppdrett og klekkerier. Videre er næring L (offentlig forvaltning) ufullstendig representert, men man kan håpe at denne næringen er fullstendig dekket innen utvalgsundersøkelsen kommer i gang for fullt.

Man må også være klar over at tallene i bedriftsregisteret baserer seg på opptil 2 år gamle opplysninger. Dermed vil registeret mangle de aller nyeste bedriftene, og samtidig inneholde bedrifter som har «gått inn». De opplyste sysselsettingstall vil også kunne forventes å være noe unøyaktige i forhold til dagens situasjon. Dette vil imidlertid være bedre i Enhetsregisteret.

## 8 Usikkerhetsberegninger

Som en grov tommelfingerregel kan man si at man *minst* behøver 25 utvalgte enheter innen hvert stratum for å gi anslag for stratumstotalene. Med bruk av 5 størrelsesgrupper (hvorav én faller fra p.g.a cut-off) får man følgende formel for *absolutt minste* mulige utvalgsstørrelse for en gitt stratifisering:  $n = r \cdot a \cdot 4 \cdot 25$ ,  $r$ : antall regioner,  $a$ : antall næringer.

Antall næringer	Antall regioner	Minimum utvalgsstørrelse
58	1	5.800
15	19	28.500
15	7	10.500
15	5	7.500
15	1	1.500
10	19	19.000
10	7	7.000
10	5	5.000
10	1	1.000
1	19	1.900
1	7	700
1	5	500
1	1	100

Tabellen ovenfor gir nedre grense på utvalgsstørrelsen for de aktuelle stratainnodelinger. Siden største realistiske utvalgsstørrelse synes å være 20.000 kan vi allerede nå slå fast at det er uaktuelt å satse på å estimere sykefravær for 19 regioner (fylker) fordelt på 15 næringer. Vi kan også kutte ut alternativet med 19 regioner fordelt på 7 næringer, siden realistisk utvalgsstørrelse vil ligge en god del høyere enn de tabulerte verdier. I tillegg til de rene utvalgsfeil må vi også ta høyde for andre feilkilder som frafall, feilrapporteringer etc.

### 8.1 Teoretiske betraktninger

Vi benytter formelen for rate-estimatoren, ligning (4), og formel for variansen til denne, ligning (5).

Variansen til stratumtotal-estimatoren  $\tilde{t}_h$  estimeres med formelen;

$$\tilde{Var}(\tilde{t}_h) = \left( \frac{\bar{x}_h}{\bar{x}_{s_h}} \right) \cdot N_h^2 \cdot \frac{1-f_h}{n_h} \cdot \frac{1}{n_h-1} \sum_{i \in s_h} (y_i - \tilde{R}_h x_i)^2 \quad (5)$$

hvor  $\tilde{R}_h = \frac{\bar{y}_{s_h}}{\bar{x}_{s_h}}$ .

Vi lar nå  $h$  også løpe over *størrelsesstrataene* (4 strata for hver kombinasjon av næring og region). For å estimere totalt sykefravær for en gitt nærings- og regionkombinasjon må vi summere estimatene for hvert størrelsesstratum.

La  $\tilde{t}_{r,a}$  betegne estimatoren for totalt sykefravær i region  $r$  og næring  $a$ . For enkelhets skyld sløyfer vi region- og næringsindeksene for estimatene innen hvert enkelt størrelsesstratum, og vi betegner disse  $\tilde{t}_2$ ,  $\tilde{t}_3$ ,  $\tilde{t}_4$  og  $\tilde{t}_5$ . Disse størrelsene estimeres med formel (4). Her angir laveste indeks estimatoren for de minste bedriftene. I tillegg kommer det ukjente sykefraværet i cut-off-stratumet. Vi lar  $\tilde{t}_1$  betegne estimatoren for cut-off-fraværet. Totalt fravær for region  $r$  og næring  $n$  er dermed gitt ved  $\tilde{t}_{r,a} = \sum_{j=1}^5 \tilde{t}_j$ . Vi velger å estimere cut-off-fraværet utfra estimert fravær i stratumet for de minste

bedriftene:  $\tilde{t}_1 = \frac{x_{01}}{x_{02}} \tilde{t}_2$ , hvor  $x_{01}$  er totalt antall ansatte i cut-off-stratumet, og  $x_{02}$  er totalt antall ansatte i størrelsesstratum 2. Dette gjøres fordi vi antar at sykefraværprosenten for cut-off-bedriftene er sterkest korrelert med sykefraværprosenten i stratumet med de minste bedriftene.

Summa, summarum får vi at totalt sykefravær innenfor region  $r$  og næring  $n$  estimeres med

$$\tilde{t}_{r,a} = (1 + x_{01}/x_{02}) \cdot \tilde{t}_2 + \sum_{j=3}^5 \tilde{t}_j \quad (6)$$

Når det gjelder variansen til  $\tilde{t}_{r,a}$ , så estimeres den som summen av variansene til hvert ledd i høyresiden i ligning (6). Vi antar m.a.o. at sykefraværet i små bedrifter er uavhengig av sykefraværet i store bedrifter.

$$\tilde{Var}(\tilde{t}_{r,a}) = \tilde{Var}\left((1 + x_{01}/x_{02}) \cdot \tilde{t}_2\right) + \sum_{j=3}^5 \tilde{Var}(\tilde{t}_j) = (1 + x_{01}/x_{02})^2 \cdot \tilde{Var}(\tilde{t}_2) + \tilde{Var}(\tilde{t}_3) + \tilde{Var}(\tilde{t}_4) \quad (7)$$

Legg merke til at estimert varians for  $\tilde{t}_5$  blir 0, siden vi har fulltelling for de største bedriftene, og dermed ingen utvalgsfeil for dette stratumet.

## 9 Simulering av sykefravær for å beregne usikkerhet ved ulike utvalgsplaner

Vi ønsker nå å gi anslag for hvor store usikkerheter som er knyttet til de ulike utvalgsplaner. For å kunne si noe om hvilken usikkerhet som knytter seg til estimatene ved å velge forskjellig utvalgsstørrelse, stratifisering og trekk sannsynligheter trenger vi å vite noe om hvordan sykefraværet faktisk er i bedriftene. Nå har vi lite eller ingen tilgjengelig informasjon om hvordan sykefravær med varighet 1-3 dager ser ut. Før vi har gjennomført i alle fall én undersøkelse har vi heller ikke noen tidligere data som dekker alle sektorer. Så hvordan kan vi da si noe om nødvendig utvalgsstørrelse før vi har noe empirisk materiale å støtte oss på?

Det vi gjør er at vi antar noe om hvordan vi forventer at sykefraværet varierer bedrifter i mellom. Videre benytter vi de dataene vi har over nivå på sykefraværet for å anslå omfanget.

Vi har valgt å gjennomføre såkalte *Monte-Carlo-simuleringer*, hvor det antas at antall sykefraværsdager i en bedrift er en stokastisk variabel som er proporsjonal med antall sysselsatte i bedriften. Vi konsentrerer oss her om å simulere antall sykefraværsdagsverk, slik at resultatene gjelder for denne størrelsen spesielt.

Eksisterende statistikker for sykefravær i bedrifter viser at korttidssykefraværet utgjør omlag 1% av antall *mulige dagsverk*. Sysselsettingsvariabelen i bedriftsregisteret angir antall *ansatte i bedriften*, og i våre simuleringer antar vi at gjennomsnittlig sykefravær med varighet 1-3 dager er 1% denne variabelen. Med 1 årsverk = 300 dagsverk<sup>6</sup> får vi så at *gjennomsnittlig korttidsfravær i dagsverk = 3 · antall ansatte*. Vi antar videre at spredningen (les: standardavviket) på sykefraværet i bedriftspopulasjonen er proporsjonal med antall ansatte.

Vi har benyttet følgende 2 modeller i våre simuleringer:

Modell 1.  $Y \sim N(3x, x^2/4)$  , sykefraværet er normalfordelt med forventning =  $3 \cdot \text{antall ansatte}$  og varians =  $(\text{antall ansatte})^2/4$  .

Modell 2.  $Y \sim N(3x, x^2)$  , samme som for modell 1, med varians =  $(\text{antall ansatte})^2$  .

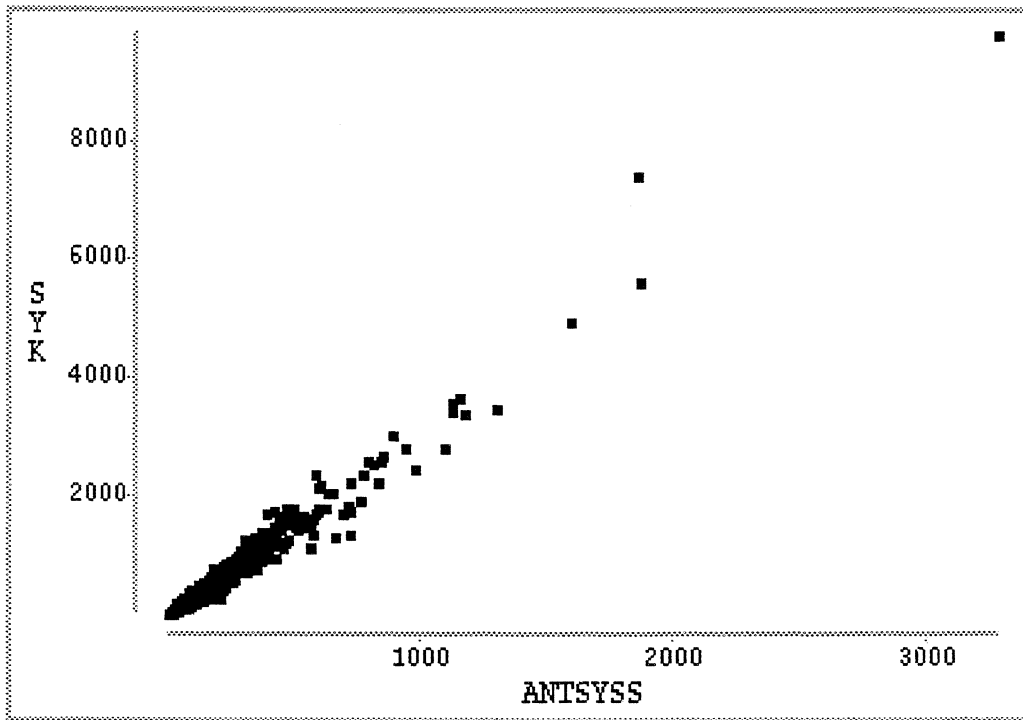
Nedenfor er vist 95% konfidensgrenser for disse modellene for noen bedriftsstørrelser:

95 % konfidensintervall for simulert sykefravær (i dager) for noen utvalgte bedriftsstørrelser				
Antall ansatte	1	10	100	500
Modell 1	[2 , 4]	[20 , 40]	[200 , 400]	[1000 , 2000]
Modell 2	[1 , 5]	[10 , 50]	[100 , 500]	[500 , 2500]

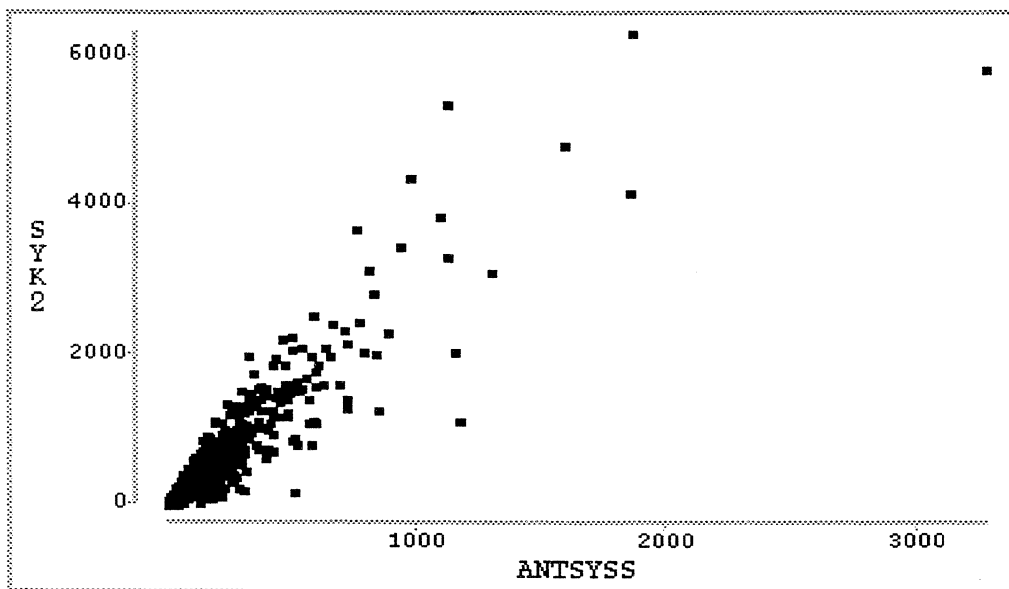
For å illustrere hvordan de simulerte sykefraværsverdier varierer med størrelse på bedrift har vi for de to modellene plottet simulerte verdier for næring D (Industri).

<sup>6</sup> Tallet 300 ble hentet ut fra løse luften. I virkeligheten utgjør 1 årsverk ca. 254 dagsverk. Denne forskjellen har imidlertid ingenting å si for de estimerte usikkerheter og konklusjonene i dette notatet.

**Fig. 1. Simulerte verdier for sykefravær i næring D. Modell 1**



**Fig. 2. Simulerte verdier for sykefravær næring D. Modell 2**



## 10 Resultater av simuleringene

Vi har simulert sykefraværsgangsverk for 4 ulike stratifiseringer etter næring og region, og delt inn i naturlige størrelsesstrata<sup>7</sup>. Vi har lagt spesiell vekt på stratifisering etter næring da det prioriteres foran region av brukerne. Vi har variert utvalgsstørrelse, dekningsgrad og trekksannsynlighet (se tabell 1), og estimert sykefraværet og tilhørende standardavvik ut i fra de simulerte verdier (se tabell 2). Disse verdiene er estimert ved hjelp av ligningene (3), (5), (6) og (7).

Med dette kan vi si noe om hvilken utvalgsstørrelse som er tilstrekkelig for å tilfredsstillere kravet vi har satt til presisjon, og evt hvilken presisjon vi kan håpe på for gitte utvalgsstørrelser og detaljeringsgrader.

Det må imidlertid presiseres at relevansen i resultatene avhenger fullt og helt av at forutsetningen om at det er relativt sterk positiv korrelasjon mellom sykefraværsgangs i en bedrift og antall ansatte i samme bedrift. Derfor er denne analysen kun ment å gi en første indikasjon på nødvendig utvalgsstørrelse. Når utvalgsundersøkelsen er satt i gang vil vi etterhvert kunne teste relevansen i denne forutsetningen og da evt. gjøre nødvendige endringer i utvalgsplanen.

**Tabell 1**

Stratifisering etter hovednæringsområde i alt 15 næringsgrupper (A-O)											
Antall regioner	Antall næringer	Observasjonsenhet	Totalt antall bedrifter	Totalt antall ansatte	Strata-inndeling	Antall bedrifter i stratum	Antall ansatte i stratum	Trekksannsynlighet	Antall uttrukne bedrifter	Dekningsgrad, bedrifter, %	Dekningsgrad, ansatte, %
1	15	D - Industri	14180	267808	1-4 5-19 20-49 50-149 150+	7616 4072 1415 773 304	14352 38446 43427 64679 106904	0 0,09 0,25 0,5 1	0 366 354 387 304	10,0	57,33
1	15	D - Industri	14180	267808	1-4 5-19 20-49 50-149 150+	7616 4072 1415 773 304	14352 38446 43427 64679 106904	0 0,05 0,1 0,4 1	0 203 141 309 304	6,75	51,85
1	15	G - Varehandel, reparasjon av kjøretøy og hush.artikler	51703	278679	1-2 3-4 5-19 20-49 50+	25407 9794 14378 1694 430	32833 33602 124304 48559 39381	0 0,1 0,2 0,5 1	0 979 2875 847 430	10,0	32,86
1	15	G - Varehandel, reparasjon av kjøretøy og hush.artikler	51703	278679	1-2 3-4 5-19 20-49 50+	25407 9794 14378 1694 430	32833 33602 124304 48559 39381	0 0 0,1 0,3 1	0 0 1438 508 430	4,6	23,83
1	15	A - Jordbruk og skogbruk	5301	13489	1-4 5-19 20-49 50-149 150+	4098 796 351 40 16	5743 2614 2688 1157 1287	0 0,15 0,3 0,8 1	0 119 105 32 16	5,1	25,4

<sup>7</sup> Størrelsesstratumsgrenser er her valgt ut i fra størrelsesfordelingen i næringen. Målet er å få til en noenlunde lik dekningsgrad for antall ansatte i utvalget samtidig som det sikres representativitet mht størrelsesfordelingen i næringen og evt. regionen.

Forts. tabell 1

Stratifikering etter detaljert næring, 58 næringsgrupper (2-siffer i SSBs næringsstandard)											
Antall regioner	Antall næringer	Observasjonsenhet	Totalt antall bedrifter	Totalt antall ansatte	Strata-inndeling	Antall bedrifter i stratum	Antall ansatte i stratum	Trekk sannsynlighet	Antall uttrukne bedrifter	Dekningsgrad, bedrifter, %	Dekningsgrad, ansatte, %
1	58	24-Prod. av kjemikalier og kjemiske produkter	221	13752	1-4 5-19 20-149 150-300 300+	66 58 70 18 9	138 591 4538 4008 4477	0 0,05 0,1 0,3 1	0 6 7 6 9	12,7	45,7
1	58	24-Prod. av kjemikalier og kjemiske produkter	221	13752	1-4 5-19 20-149 150-300 300+	66 58 70 18 9	138 591 4538 4008 4477	0 0,03 0,04 0,1 1	0 2 3 2 9	7,2	38,3
1	58	52 - Detaljhandel	29760	139443	1-4 5-19 20-49 50-149 150+	20887 7996 750 119 8	40529 67170 21306 8703 1735	0 0,3 0,4 0,5 1	0 2398 300 60 8	9,3	25,2
1	58	52 - Detaljhandel	29760	139443	1-4 5-19 20-49 50-149 150+	20887 7996 750 119 8	40529 67170 21306 8703 1735	0 0,11 0,2 0,25 1	0 900 150 30 8	3,7	11,2
1	58	66-Forsikring og pensjonsfond	379	4223	1-4 5-19 20-49 50-149 150+	283 66 15 11 4	567 565 493 903 1695	0 0,3 0,4 0,7 1	0 20 6 8 4	10,0	65,0
1	58	66-Forsikring og pensjonsfond	379	4223	1-4 5-19 20-49 50-149 150+	283 66 15 11 4	567 565 493 903 1695	0 0,1 0,2 0,36 1	0 7 3 4 4	4,75	51,8
Stratifikering etter næringshovedområde og region, 10 næringsgrupper og 7 regioner											
Antall regioner	Antall næringer	Observasjonsenhet	Totalt antall bedrifter	Totalt antall ansatte	Strata-inndeling	Antall bedrifter i stratum	Antall ansatte i stratum	Trekk sannsynlighet	Antall uttrukne bedrifter	Dekningsgrad, bedrifter, %	Dekningsgrad, ansatte, %
7	10	G (varehandel), Hordaland, Sogn og Fjordane, Møre og Romsdal	8362	41373	1-2 3-5 6-12 13-24 25+	4037 2216 1470 419 220	5309 8424 11755 6811 9074	0 0,09 0,14 0,5 1	0 201 210 209 220	10,0	35,97
7	10	A (jord og skog), Nord-Norge	300	577	3 4+	21 22	63 174	0,33 1	7 22	9,7	33,79
7	10	C (bergverk og utvinning), Hedmark, Oppland	50	289	5-19 20+	11 3	102 112	0,2 1	2 3	10,0	45,81
Stratifikering etter næringshovedområde og fylke, 15 næringer og 19 fylker											
Antall regioner	Antall næringer	Observasjonsenhet	Totalt antall bedrifter	Totalt antall ansatte	Strata-inndeling	Antall bedrifter i stratum	Antall ansatte i stratum	Trekk sannsynlighet	Antall uttrukne bedrifter	Dekningsgrad, bedrifter, %	Dekningsgrad, ansatte, %
19	15	H (hotell og restaurant), Finnmark	169	1463	1-3 4-9 10-19 20-54 55+	90 36 26 13 4	146 210 342 399 366	0 0,11 0,15 0,31 1	0 4 4 4 4	9,4	38,56



**Tabell 2**

<b>Stratifikering etter hovednæringsområde. I alt 15 næringsgrupper (A-O)</b>							
<i>Stratum</i>	<i>Simuleringsvarians</i>	<i>Simulert sann verdi</i>	<i>Estimat</i>	<i>Avvik</i>	<i>Avvik %</i>	<i>Estimert std.avvik</i>	<i>Std.avvik i % av estimat</i>
D - 10% utvalg	$\chi^2/4$	799647,3	799754,6	107,319	0,013	2302,229	<b>0,29</b>
D - 10% utvalg	$\chi^2$	797903,0	793840,2	4062,855	0,51	4315,846	<b>0,54</b>
D - 6,75% utvalg	$\chi^2/4$	799647,3	797048,9	2598,393	0,32	3092,306	<b>0,39</b>
D - 6,75% utvalg	$\chi^2$	797903,0	788664,8	9238,248	1,16	5860,053	<b>0,74</b>
G - 10% utvalg	$\chi^2/4$	835829,0	835562,2	266,7755	0,032	1514,532	<b>0,18</b>
G - 10% utvalg	$\chi^2$	831862,8	827954,7	3908,156	0,4698	3019,262	<b>0,36</b>
G - 4,6% utvalg	$\chi^2/4$	835829,0	839803,2	-3974,2	-0,475	1790,832	<b>0,21</b>
G - 4,6% utvalg	$\chi^2$	831862,8	832232,3	369,496	0,044	3622,8	<b>0,44</b>
A - 5,1% utvalg	$\chi^2/4$	40248,4	39726,12	522,2752	1,3	568,9827	<b>1,43</b>
A - 5,1% utvalg	$\chi^2$	40348,95	41106,32	-757,374	-1,88	1185,633	<b>2,88</b>
<b>Stratifikering etter detaljert næring. 58 næringsgrupper (2-siffer i SSBs næringsstandard)</b>							
<i>Stratum</i>	<i>Simuleringsvarians</i>	<i>Simulert sann verdi</i>	<i>Estimat</i>	<i>Avvik</i>	<i>Avvik %</i>	<i>Estimert std.avvik</i>	<i>Std.avvik i % av estimat</i>
24-12,7% utvalg	$\chi^2/4$	40852,92	41688,9	-835,977	-2,05	615,7746	<b>1,48</b>
24-12,7% utvalg	$\chi^2$	40232,33	44911,55	-4679,22	-11,63	1844,528	<b>4,11</b>
24 -7,2% utvalg	$\chi^2/4$	40852,92	38971,38	1881,537	4,606	1360,25	<b>3,5</b>
24 -7,2% utvalg	$\chi^2$	40232,33	37600,12	2632,208	6,543	1114,979	<b>3,0</b>
52 - 9,3% utvalg	$\chi^2/4$	419105,1	419935,3	-830,188	-0,2	1155,183	<b>0,28</b>
52 - 9,3% utvalg	$\chi^2$	417428,5	416504,1	924,4804	0,22	2374,906	<b>0,57</b>
52 - 3,7% utvalg	$\chi^2/4$	419105,1	418043,1	1062,028	0,253	2125,269	<b>0,51</b>
52 - 3,7% utvalg	$\chi^2$	417428,5	413891,4	3537,134	0,847	4336,579	<b>1,05</b>
66- 10% utvalg	$\chi^2/4$	13399,92	13355,46	44,46	0,33	143,3766	<b>1,07</b>
66- 10% utvalg	$\chi^2$	11836,78	11425,75	411,031	3,47	276,9818	<b>2,42</b>
66-4,75% utvalg	$\chi^2/4$	13399,92	13027,48	372,4438	2,779	273,7301	<b>2,1</b>
66-4,75% utvalg	$\chi^2$	11836,78	11972,5	-135,724	0,115	512,8473	<b>4,3</b>
<b>Stratifikering etter næringshovedområde og region. 10 næringsgrupper og 7 regioner.</b>							
<i>Stratum</i>	<i>Simuleringsvarians</i>	<i>Simulert sann verdi</i>	<i>Estimat</i>	<i>Avvik</i>	<i>Avvik %</i>	<i>Estimert std.avvik</i>	<i>Std.avvik i % av estimat</i>
G,Vestlandet 10% utvalg	$\chi^2/4$	125.402,3	125.849,2	-446,9	0,36	570,58	<b>0,45</b>
G,Vestlandet 10% utvalg	$\chi^2$	126.685,6	126.408,9	276,7	0,22	1.141,17	<b>0,92</b>
A,Nord-Norge 9,7% utvalg	$\chi^2/4$	1743,7	1830,6	-86,9	5,0	52,87	<b>2,89</b>
A,Nord-Norge 9,7% utvalg	$\chi^2$	1756,4	1930,1	-173,7	9,9	105,74	<b>5,48</b>
C, Innlandet 10% utvalg	$\chi^2/4$	939,9	870,4	69,5	7,40	24,26	<b>2,79</b>
C, Innlandet 10% utvalg	$\chi^2$	1012,8	873,7	139,1	13,7	48,5	<b>5,55</b>
<b>Stratifikering etter næringshovedområde og fylke. 15 næringer og 19 fylker.</b>							
<i>Stratum</i>	<i>Simuleringsvarians</i>	<i>Simulert sann verdi</i>	<i>Estimat</i>	<i>Avvik</i>	<i>Avvik %</i>	<i>Estimert std.avvik</i>	<i>Std.avvik i % av estimat</i>
H, Finnmark	$\chi^2/4$	4.690,4	4.750,5	-60,1	1,28	200,74	<b>4,23</b>
H, Finnmark	$\chi^2$	4.991,8	4.111,9	-120,1	2,41	401,46	<b>7,85</b>

Selv om disse simuleringene ikke er av det mest omfattende slaget, kan vi allikevel trekke noen viktige slutninger:

- Med et utvalg på ca. 10 000 bedrifter og en detaljering på 15 næringer ser det ut som at en usikkerhetsgrense på 7 % (dvs at standardavviket til estimatet ikke overstiger 2,47% av estimatet selv) kan tilfredsstilles med god margin så lenge ikke strataene blir for små, i alle fall for de undersøkte strata. Simuleringene som er gjort for næring G tyder på at selv for næringer med en stor andel små bedrifter kan vi ha en cut-off grense på 4 ansatte.
- For å studere mulighetene innenfor mer detaljert næring (58 grupper) har vi sett på tre til dels ulike næringsgrupper i våre simuleringer. Næring 24 er en relativt liten næring, men har en stor andel «store» bedrifter. Vi har satt grensen for bedrifter som totaltelles til 300 ansatte og mer. Andel bedrifter med i utvalget er 12,7 %, og dekningsgraden hele 45,7%. Resultatet av simuleringen her er relativt stor usikkerhet i tallene. Trolig skyldes dette at næringen er liten, kun 221 bedrifter i hele næringen. Detaljhandel kom meget godt ut, selv med en lavere andel bedrifter i utvalget (9,3%) og en mye lavere dekningsgrad. Dette skyldes høyst sannsynlig at denne næringen er mye større, i underkant av 30 000 bedrifter. Næring 66 (Forsikring og pensjonsfond) er en mellomstor næring. Den har mange «små», men også mange «store» bedrifter, slik at dekningsgraden blir relativt høy når vi totalteller bedrifter med 150 ansatte og mer. For en mer detaljert næringsinndeling viser våre simuleringer at med et utvalg på 20 000 bedrifter vil vi sannsynligvis holde oss innenfor grensen på 2,47 %. Dersom vi reduserer utvalget til 10 000 bedrifter antyder våre simuleringer at kravet om å kunne observere endringer på 7% kan holde hardt. Ved å variere trekkssannsynlighet og «cut-off»-grensene noe etter strukturen i strataene f.eks. gjennom å totaltelle de aller minste strataene, kan det allikevel være mulig å nå kravet til presisjon med 10 000 bedrifter i utvalget. Når vi i tillegg vil holde deler av utvalget fast mellom hver undersøkelse vil en utvalgsstørrelse på 10 000 bedrifter nok kunne forsvares .
- Små strata med hovedsakelig små bedrifter kommer relativt dårlig ut i simuleringen. De aller minste strataene må evt. slås sammen med andre strata, eller totaltelles.
- Når det gjelder muligheter for regionsinndeling har vi sett på en grov inndeling (7 regioner) og en fylkesvis inndeling. Med 10 næringsgrupper og 7 regioner gir våre simuleringer som resultat at med et utvalg på ca 20 000 bedrifter og dersom modell 1 gjelder for sammenhengen mellom sykefravær og antall ansatte, så kan vi kanskje nå kravet vi har satt til presisjon. Om imidlertid modell 2 gjelder, overskrides grensen relativt klart i alle fall i de undersøkte strata og spesielt da for de små strataene. For en regioninndeling på fylkesnivå i tillegg til en næringsinndeling på 15, overskrides 1,76 %-grensen klart, selv om dekningsprosenten i antall sysselsatte er god. Dette antas hovedsakelig å skyldes at vi har såpass få som 4 bedrifter i mange strata (<25 som var «tommefingerregel»).
- Vi må være klar over at næring D og G er noen av de største næringene i Norge, og innenfor disse næringene ser 5% utvalg ut til å være tilstrekkelig, selv for 7 regioner. For næringer med få bedrifter og en overvekt av små bedrifter er et 5% utvalg bedrifter kanskje ikke tilstrekkelig med de krav vi har satt til presisjon. Ved å øke utvalgsstørrelsen i små næringer på bekostning av store og variere cut-off grensen kan vi trolig allikevel holde oss innenfor grensen satt til presisjon med et utvalg på 10 000 bedrifter for 10 næringer og 7 regioner.
- Hvor gode estimater vi får, avhenger tydelig av spredningen på dataene, jfr. simulert varians  $= x^2/4$  og  $x^2$ . Vi må anta at våre resultater er i snilleste laget, siden vi har sterk lineær

sammenheng mellom sykefravær og antall sysselsatte i vårt datasett. (Korrelasjonskoeffisient på mellom 0,93 og 0,96 i de simulerte dataene).

- Det er nødvendig med individuell tilpasning av størrelsesgrenser avhengig av hvilken næring og region man befinner seg i.
- Man kan forvente å tjene noe med en mer *optimal allokering* på størrelsesstrata. Det finnes metoder som fordeler antall enheter til hvert stratum slik at variansen totalt blir minst mulig. Dette kan redusere usikkerheten i estimatene noe. Man kan velge trekk sannsynlighetene proporsjonale med standardavviket på antall ansatte i hvert størrelsesstratum, og også her håpe på en liten usikkerhetsgevinst.
- Vi har ikke sagt noe om estimater fordelt på kjønn. På det nåværende tidspunkt er det heller ikke mulig å si noe sikkert, siden vi ikke har kapasitet til å knytte kjønnsopplysninger til bedriftsregisteret. Tall fordelt på kjønn vil kanskje ha større usikkerhet enn stratumstotalene. Vi antar imidlertid at usikkerheten for de kjønns spesifikke estimatene vil være i samme størrelsesorden som for stratumstotalene. Man må derimot være klar over at innen enkelte næringer / regioner kan kjønnsfordelingen være såvidt skjev at presise estimater er umulig uten en overrepresentering av antall bedrifter. Kjønns spesifikke endringstall innen enkelte strata vil kunne ha en større usikkerhet knyttet til seg enn det kravet vi har satt til presisjon på totaltallene.
- Det er viktig å være klar over at alle beregninger og konklusjoner på bakgrunn av disse er foretatt utfra dagens bedriftsstruktur, i alle fall så langt dette gjenspeiles i Bedrifts- og foretaksregisteret. Disse forholdene kan endres over tid, og det er nødvendig med en løpende kvalitetsvurdering av utvalgsplan og estimeringsprosedyre.

## 11 Andre kilder til usikkerhet

De usikkerheten vi har omtalt, har kun vært tilknyttet utvalgsplanen, altså det som kalles utvalgsfeil. I tillegg har vi usikkerhetskilder tilknyttet registerfeil (registeret inneholder 2 år gamle opplysninger), målefeil (bedriftene fører spørreskjemaene galt, enten p.g.a. misforståelse eller bevisst feilrapportering) og frafall (sannsynligvis vil en del bedrifter være trege / unnlate å levere).

Det er vanskelig å si noe om hvordan disse kildene vil virke inn på den totale usikkerheten i estimatene, men det er uansett viktig å redusere disse usikkerhetskildene så langt det er mulig. At bedriftsregisteret inneholder unøyaktige opplysninger for antall ansatte, kan rettes opp ved at spørreskjemaet inneholder en rubrikk over antall sysselsatte/årsverk/dagsverk. Målefeil kan til en viss grad motvirkes ved at bedriftene får tilstrekkelig veiledning og motivasjon til utfylling av spørreskjemaet.

## 12 Rullering av utvalget

Full utskiftning av utvalget fra gang til gang gir større varians for endringsestimatet enn rene panelundersøkelser (dvs. fast utvalg fra gang til gang). Panelundersøkelser vil imidlertid kunne gi noe misvisende totalestimater hvis utvalget er skjevt. En gylden middelvei kan være å benytte rullerende utvalg, der en del av utvalget (av de bedriftene som har trekk sannsynlighet  $< 1$ ) er felles mellom to etterfølgende undersøkelser. På denne måten sikrer man seg lavere usikkerhet i endringstallene enn i tilfellet med full utskiftning. Med et rent panelutvalg for ett år ad gangen, og en viss rullering kun én gang i året gir bedriftene tid til å bli kjent med skjema slik at feilrapporteringer reduseres. Med årlige faste utvalg sparer vi kostnader som f.eks. informering til nye bedrifter i utvalget, identifisering av bedriftene som trekkes ut og endringstallene pr. kvartal blir bedre.

## **Æres den som æres bør**

En stor takk rettes til Anne Sofie Abrahamsen som har bidratt med verdifulle innspill og kommentarer til utvalgsplanen.

## Litteratur

- (1) Arbeidsgruppe med representanter fra RTV, SSB, NHO, LO, KS, AD og SHD, *Sentral sykefraværstatistikk*, Sosial- og helsedepartementet, 1994.
- (2) Dalgard og Tuveng, *Sentral sykefraværstatistikk, utkast til rapport fra Pilotprosjektet*, februar 1997. Upublisert notat.
- (3) Jan F. Bjørnstad, *Generell utvalgsteori*, Byråskolen, Internt kompendium, Statistisk sentralbyrå, April 1996.
- (4) Klaus Mohn, *Industriysselssetting og produksjonsteknologi i norske regioner*, SSB rapport 93/22, Statistisk sentralbyrå, 1993.
- (5) Irene Tuveng, *Utvalgsplan*, Internt notat, Seksjon for arbeidsmarkedsstatistikk, SSB, 04.juni 1996.
- (6) Helge Næsheim, *Kostnadsanslag ny sykefraværstatistikk*, Internt notat, Seksjon for arbeidsmarkedsstatistikk, SSB, 5 august 1996.
- (7) *Statistisk årbok 1995*, Statistisk sentralbyrå, 1995.
- (8) *Lønns- og fraværstatistikk, 1.kvartal 1995*, Næringslivets hovedorganisasjon, 1995.
- (9) *Fraværstatistikk 94-95*, Kommunenes sentralforbund, 1996.
- (10) Ib Thomsen, *Prinsipper og metoder for Statistisk sentralbyrås utvalgsundersøkelser*, Samfunnsøkonomiske studier nr. 33, Statistisk sentralbyrå 1991.
- (11) Särndal, C-E., Swensson, B., Wretman, J., *Model Assisted Survey Sampling*, Springer-Verlag, 1992
- (12) Scheaffer, Mendenhall og Ott, *Elementary survey sampling*, fourth edition.
- (13) Anne Sofie Abrahamsen, Arnfinn Schjalm, Anders Falnes, Egil Heilund, Randi Jule, Leiv Solheim, *Oversikt over innhold, utvalgsplaner og metoder i statistikker basert på bedrifter og foretak*, SSB rapport 95/32, Statistisk sentralbyrå, 1995.



## Vedlegg 1. Variansen til endringstall når vi ser på totalt sykefravær

Dersom vi kun er interessert i endringer i *totalt sykefravær* kan vi redusere kravene til korttidsstatistikken betydelig. Vi illustrerer med et eksempel. Anta vi ønsker å kunne observere en endring i *totalt sykefravær* på 7% . Anta videre en proporsjonal økning slik at andel korttidssykefravær er den samme i periodene vi ser på. Totalt sykefravær ( $t$ ) = Korttidssykefravær ( $x$ ) + langtidssykefravær ( $y$ ).

$$\tilde{t}_1 = x_1 + y_1 \quad \text{Var}(\tilde{t}_1) = \text{Var}(x_1) + \text{Var}(y_1) = \text{Var}(x_1)$$

$$\tilde{t}_2 = x_2 + y_2 \quad \text{Var}(\tilde{t}_2) = \text{Var}(x_2) + \text{Var}(y_2) = \text{Var}(x_2)$$

$$\text{Var}(\tilde{e}) = \text{Var}(\tilde{t}_2 - \tilde{t}_1) = \text{Var}(x_2) + \text{Var}(x_1) = 2\text{Var}(x)$$

$$2 \cdot \text{Std}\left(\frac{\tilde{t}_2 - \tilde{t}_1}{t_1}\right) = \frac{2\sqrt{2}}{x_1 + y_1} \text{Std}(x) \leq 7\%$$

Vi antar standardavviket til korttidssykefraværet øker med nivået på sykefraværet,  $\text{Std}(x) = ax$  og at korttidssykefraværet utgjør omlag 17% av totalt antall sykedager.

$$2 \cdot \text{Std}\left(\frac{\tilde{t}_2 - \tilde{t}_1}{\tilde{t}_1}\right) = \frac{x_1}{x_1 + y_1} a 2\sqrt{2} = 0,17 \cdot 2\sqrt{2}a \leq 0,07$$

$$\Rightarrow a = 14,6\%$$

Dvs. at vi nå kan ha et standardavvik i % av estimatet på 14,6% i motsetning til 2,47% . Dersom det kun er endringer i det totale nivået på sykefraværet vi ønsker å kunne registrere betyr dette at utvalgsstørrelsen kan reduseres betraktelig for uendret krav til presisjon.





## Vedlegg 2. Næringsgrupperinger

### SSB næringsgruppering (SN94), 15 nivåer, gruppe A-O

Gruppe	Beskrivelse	2-sifret NACE-kode
A <sup>8</sup>	Jordbruk og skogbruk	01-02
B <sup>9</sup>	Fiske	05
C	Bergverksdrift og utvinning	10-14
D	Industri	15-37
E	Kraft- og vannforsyning	40-41
F	Bygge- og anleggsvirksomhet	45
G <sup>10</sup>	Varehandel, reparasjon av kjøretøyer og husholdningsapparater	50-52
H	Hotell- og restaurantvirksomhet	55
I <sup>11</sup>	Transport og kommunikasjon	60-64
J	Finansiell tjenesteyting og forsikring	65-67
K	Eiendomsdrift, forretningsmessig tjenesteyting, utleievirksomhet	70-74
L <sup>12</sup>	Offentlig forvaltning	75
M	Undervisning	80
N	Helse- og sosialtjenester	85
O	Andre sosiale og personlige tjenester	90-93

I tillegg inneholder SN94-standardene områdene P: Lønnet husarbeid (95) og Q: Internasjonale organer og organisasjoner (99). Disse gruppene har imidlertid svært ufullstendige registeropplysninger og er hold utenfor i denne utredningen. I utgangspunktet går vi imidlertid inn for at alle næringshovedområder representeres i trekkgrunnlaget og evt. også i utvalget. Vi vil vurdere kvaliteten på Enhetsregisteret med tanke på dette når vi skal i gang med utvalgsundersøkelsen.

### SSB Næringsgrupperingsstandard, 10 nivåer

Gruppe	Beskrivelse	2-sifret NACE-kode
1	Jordbruk, skogbruk og fiske	01-05
2	Utvinning av råolje og naturgass	11
3	Bergverksdrift og industri	10, 12-37
4	Kraft- og vannforsyning	40-41
5	Bygge- og anleggsvirksomhet	45
6	Varehandel, vedlikehold og reparasjon, hotell- restaurant	50-55
7	Transport og kommunikasjon	60-64
8	Finansiell tjenesteyting og forsikring	65-67
9	Eiendomsdrift, forretningsmessig tjenesteyting, utleievirksomhet	70-74
10 <sup>13</sup>	Off. forvaltning, undervisning, helse- og sosial, andre sosiale og personlige tjenester, lønnet husarbeid, internasjonale organer	75-99

<sup>8</sup> Inneholder kun tjenester tilknyttet planteproduksjon i jordbruk og tjenester tilknyttet skogbruk

<sup>9</sup> Inneholder kun fiskeoppdrett og klekkerier

<sup>10</sup> Unntatt 50.2 Vedlikehold og reparasjon av motorkjøretøyer og 52.7 Reparasjon av husholdningsvarer og varer til personlig bruk

<sup>11</sup> Unntatt 60.3 Rørtransport og 61.101/2 Utenriks sjøfart og Kysttrafikk i Europa.

<sup>12</sup> Ufullstendige registeropplysninger

<sup>13</sup> Unntatt 99.9 Uoppgitt næringskode

Vi ser at de to standardene til en viss grad inneholder de samme gruppene, mens andre grupper overlapper. Av det som ble sagt i foregående avsnitt synes det som om gruppe 1 og gruppe 10 vil bli noe redusert / falle fra

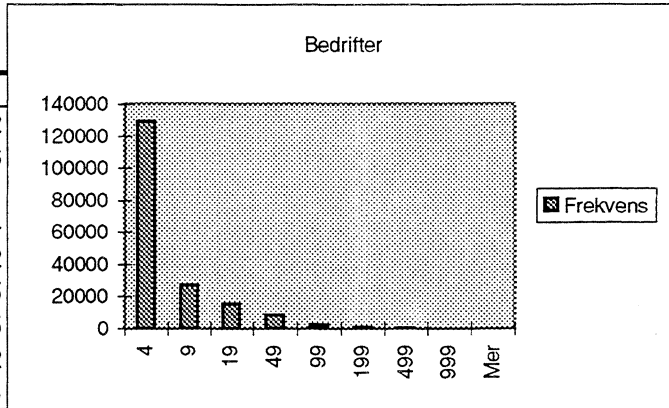
### **Vedlegg 3. Bedriftsstørrelse og ansattprofil fordelt på næringshovedgrupper**

Det sentrale bedrifts- og foretaksregisteret per juni 1996

Samtlige næringer

Hele landet

<b>Bedrifter</b>	
Intervall	Frekvens
4	129172
9	26916
19	15461
49	8624
99	2342
199	975
499	475
999	102
Mer	43



Antall bed: 184110

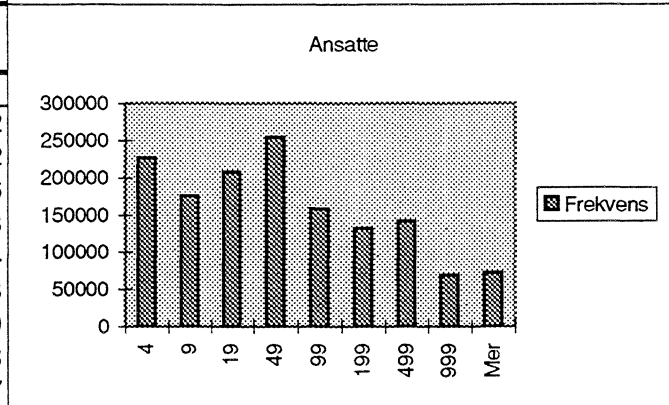
Gj.snitt: 7,812938

Std.avvik: 38,53927

Max: 3532

Antall ans. 1438460

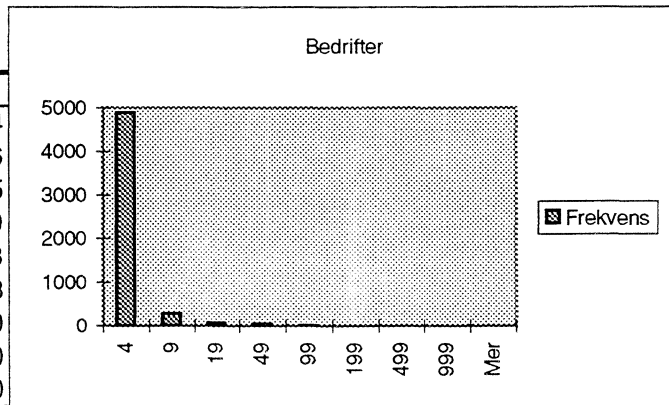
<b>Ansatte</b>	
Intervall	Frekvens
4	226502
9	175852
19	207086
49	254543
99	158354
199	132143
499	142200
999	69143
Mer	72637



Hele landet

Næring A

<b>Bedrifter</b>	
Intervall	Frekvens
4	4894
9	276
19	75
49	40
99	13
199	3
499	0
999	0
Mer	0



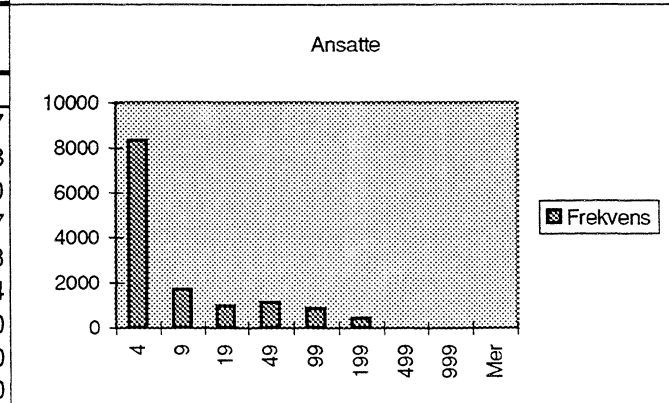
Antall bed: 5301

Gj.snitt: 2,544614

Std.avvik: 5,61868

Max: 180

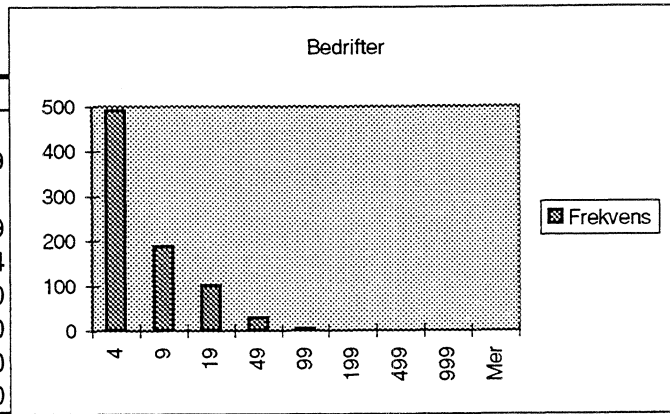
<b>Ansatte</b>	
Intervall	Frekvens
4	8357
9	1708
19	980
49	1157
99	853
199	434
499	0
999	0
Mer	0



Antall ans. 13489

Hele landet  
Næring B

<b>Bedrifter</b>	
Intervall	Frekvens
4	491
9	189
19	101
49	29
99	4
199	0
499	0
999	0
Mer	0



Antall bed. 814

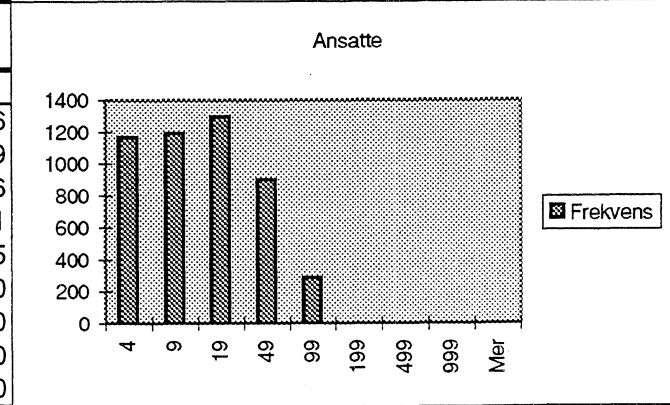
Gj.snitt: 5,94226

Std.avvik: 7,887297

Max: 99

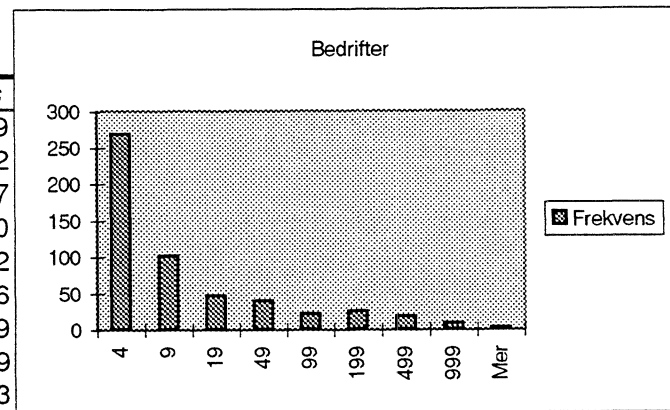
Antall ans. 4837

<b>Ansatte</b>	
Intervall	Frekvens
4	1166
9	1189
19	1296
49	901
99	285
199	0
499	0
999	0
Mer	0



Hele landet  
Næring C

<b>Bedrifter</b>	
Intervall	Frekvens
4	269
9	102
19	47
49	40
99	22
199	26
499	19
999	9
Mer	3



Antall bed: 537

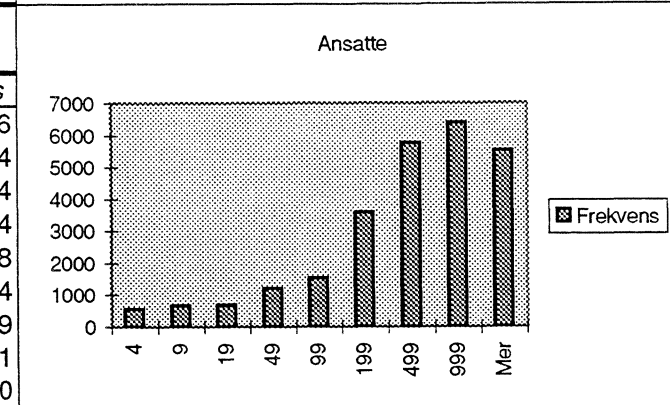
Gj.snitt: 48,28678

Std.avvik: 182,7646

Max: 2906

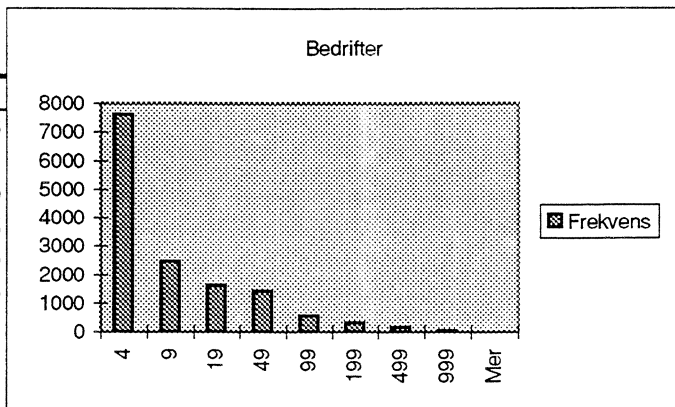
Antall ans. 25930

<b>Ansatte</b>	
Intervall	Frekvens
4	566
9	664
19	684
49	1204
99	1538
199	3594
499	5769
999	6391
Mer	5520



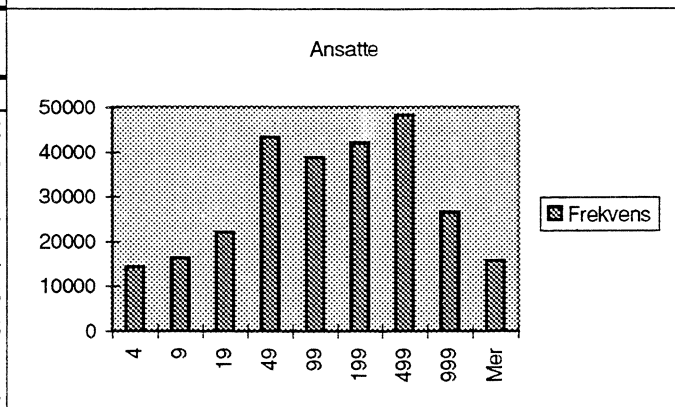
Hele landet  
Næring D

<b>Bedrifter</b>	
Intervall	Frekvens
4	7616
9	2453
19	1619
49	1415
99	559
199	309
499	159
999	40
Mer	10



Antall bed. 14180  
Gj.snitt: 18,88632  
Std.avvik: 69,66543

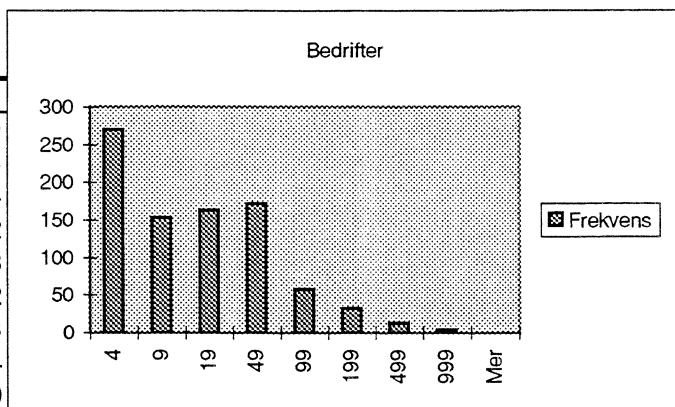
<b>Ansatte</b>	
Intervall	Frekvens
4	14352
9	16376
19	22070
49	43427
99	38834
199	42026
499	48336
999	26643
Mer	15744



Max: 3310  
Antall ans. 267808

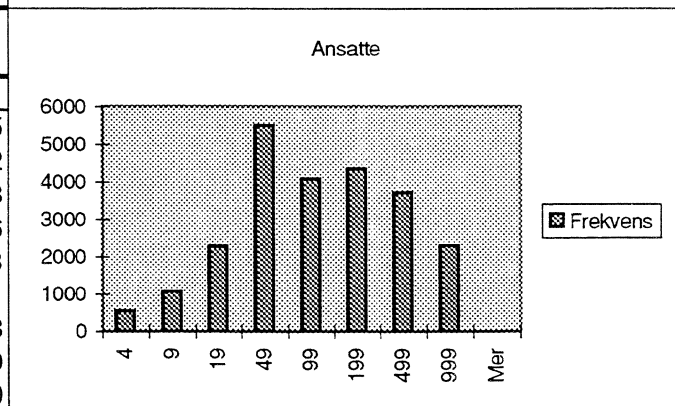
Hele landet  
Næring E

<b>Bedrifter</b>	
Intervall	Frekvens
4	270
9	153
19	164
49	172
99	58
199	32
499	13
999	4
Mer	0



Antall bed. 866  
Gj.snitt: 27,58891  
Std.avvik: 58,58207

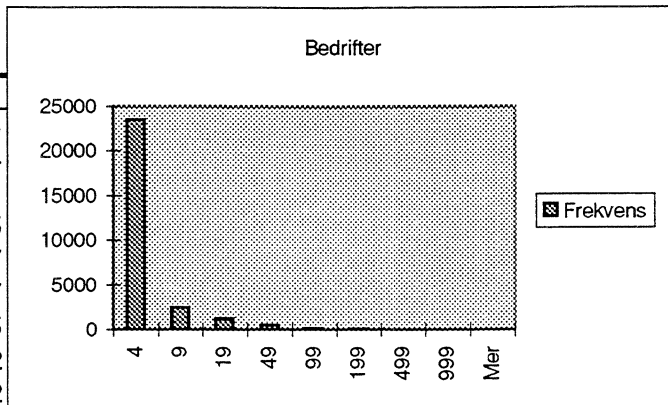
<b>Ansatte</b>	
Intervall	Frekvens
4	545
9	1062
19	2298
49	5506
99	4083
199	4361
499	3728
999	2309
Mer	0



Max: 643  
Antall ans. 23892

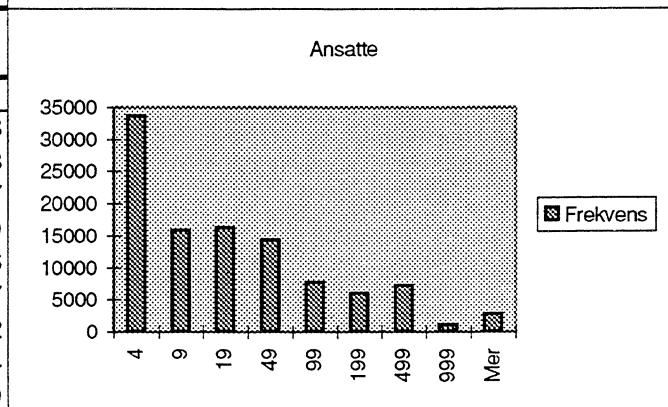
Hele landet  
Næring F

<b>Bedrifter</b>	
Intervall	Frekvens
4	23513
9	2434
19	1231
49	495
99	117
199	47
499	25
999	2
Mer	2



Antall bed. 27866  
Gj.snitt: 3,770617  
Std.avvik: 17,66408

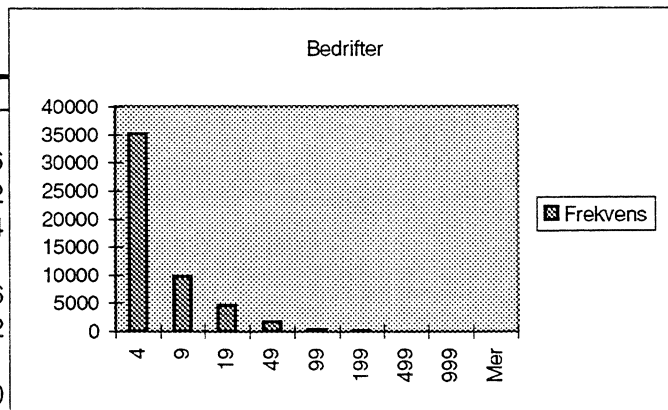
<b>Ansatte</b>	
Intervall	Frekvens
4	33688
9	15858
19	16247
49	14370
99	7756
199	6077
499	7212
999	1094
Mer	2790



Max: 1694  
Antall ans. 105092

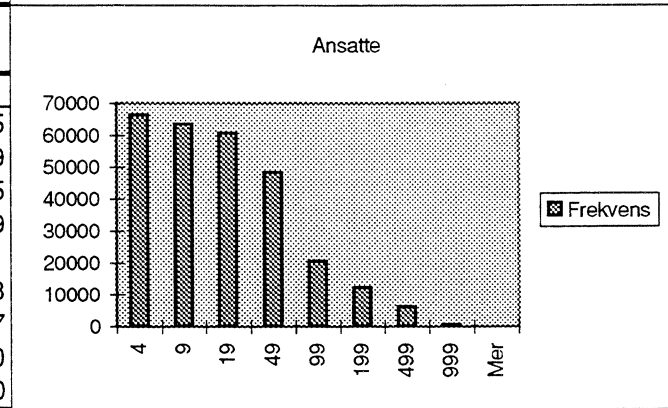
Hele landet  
Næring G

<b>Bedrifter</b>	
Intervall	Frekvens
4	35201
9	9756
19	4622
49	1694
99	311
199	96
499	22
999	1
Mer	0



Antall bed. 51703  
Gj.snitt: 5,389997  
Std.avvik: 11,23944

<b>Ansatte</b>	
Intervall	Frekvens
4	66435
9	63579
19	60725
49	48559
99	20491
199	12223
499	6157
999	510
Mer	0

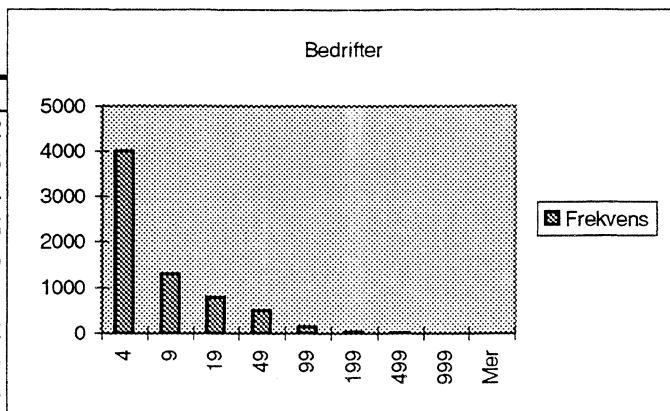


Max: 510  
Antall ans. 278679



Hele landet  
Næring H

<b>Bedrifter</b>	
Intervall	Frekvens
4	4012
9	1286
19	784
49	495
99	149
199	30
499	12
999	0
Mer	0



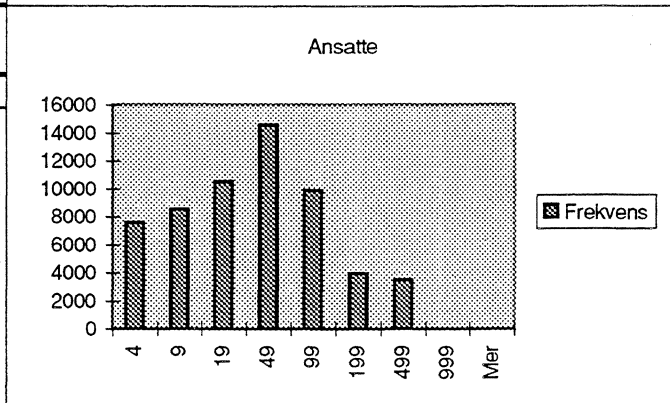
Antall bed. 6768

Gj.snitt: 8,651891

Std.avvik: 19,12478

Max: 358

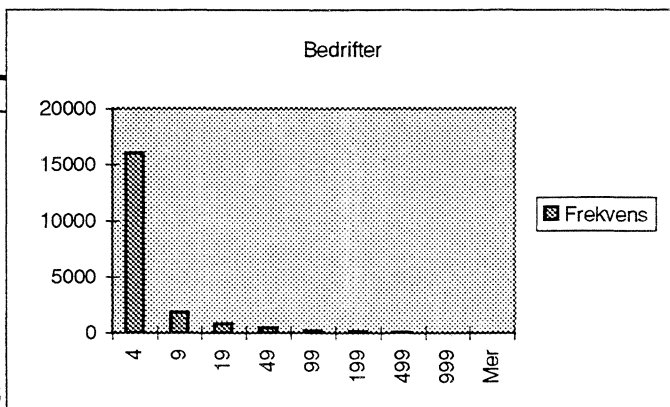
<b>Ansatte</b>	
Intervall	Frekvens
4	7581
9	8529
19	10525
49	14559
99	9904
199	3949
499	3509
999	0
Mer	0



Antall ans. 58556

Hele landet  
Næring I

<b>Bedrifter</b>	
Intervall	Frekvens
4	16036
9	1829
19	819
49	449
99	164
199	93
499	60
999	16
Mer	6



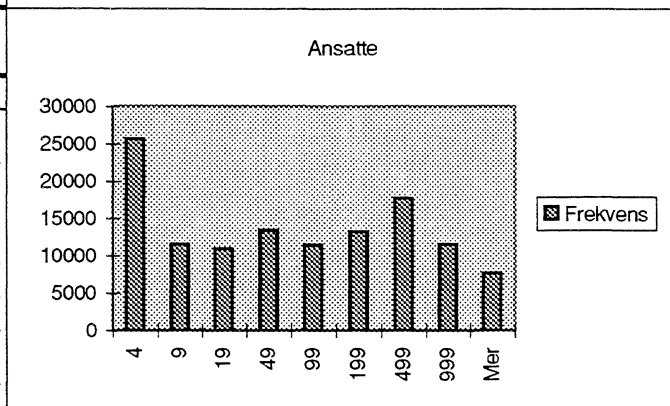
Antall bed. 19472

Gj.snitt: 6,340078

Std.avvik: 37,37571

Max: 1428

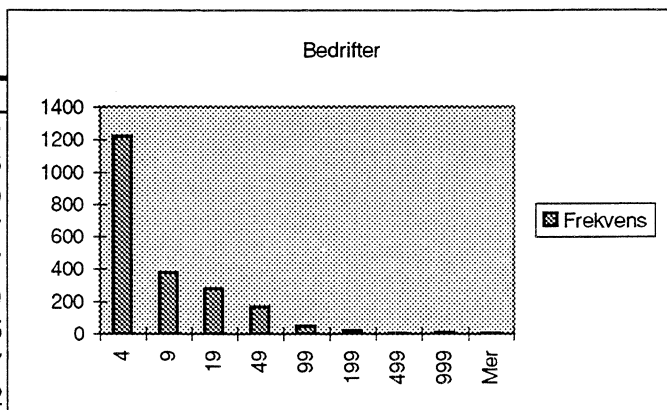
<b>Ansatte</b>	
Intervall	Frekvens
4	25729
9	11594
19	10963
49	13465
99	11465
199	13244
499	17724
999	11573
Mer	7697



Antall ans. 123454

Hele landet  
Næring J

<b>Bedrifter</b>	
Intervall	Frekvens
4	1220
9	376
19	279
49	167
99	47
199	19
499	5
999	7
Mer	2



Antall bed. 2122

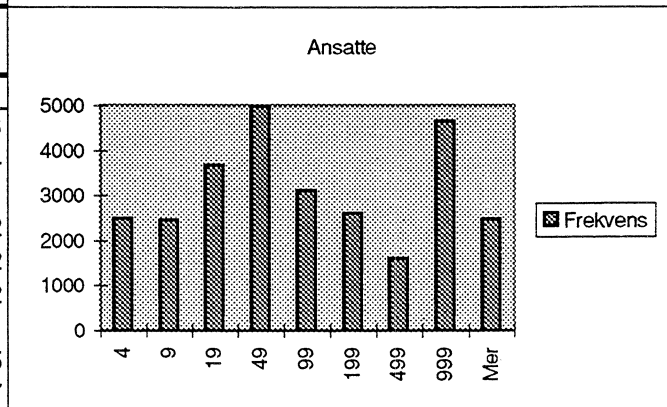
Gj.snitt: 13,24081

Std.avvik: 59,01224

Max: 1341

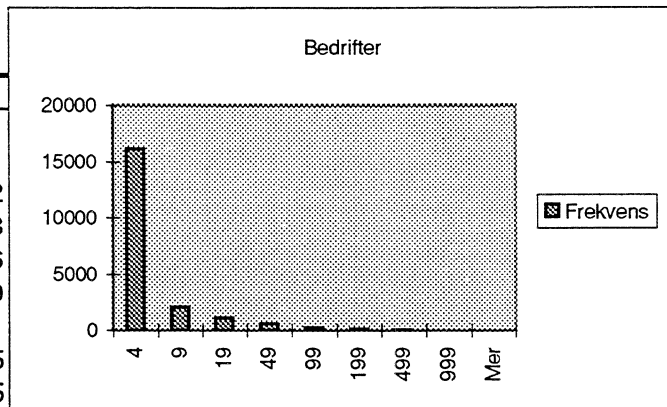
Antall ans. 28097

<b>Ansatte</b>	
Intervall	Frekvens
4	2503
9	2454
19	3681
49	4982
99	3112
199	2612
499	1611
999	4655
Mer	2487



Hele landet  
Næring K

<b>Bedrifter</b>	
Intervall	Frekvens
4	16091
9	2041
19	1082
49	598
99	206
199	99
499	61
999	5
Mer	5



Antall bed. 20188

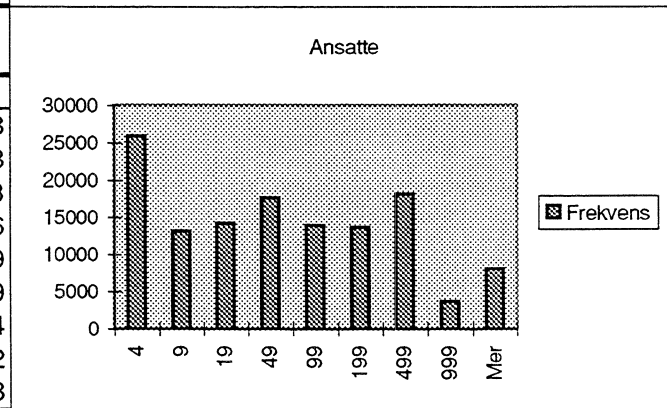
Gj.snitt: 6,370963

Std.avvik: 38,16964

Max: 3532

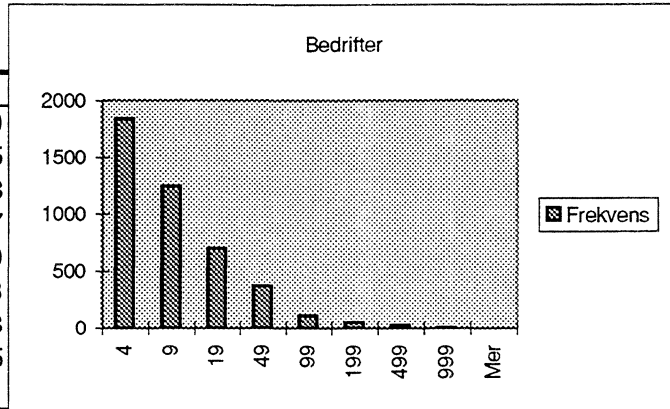
Antall ans. 128617

<b>Ansatte</b>	
Intervall	Frekvens
4	25938
9	13148
19	14233
49	17676
99	13949
199	13679
499	18204
999	3672
Mer	8118



Hele landet  
Næring L

<b>Bedrifter</b>	
Intervall	Frekvens
4	1840
9	1245
19	698
49	367
99	110
199	53
499	28
999	5
Mer	1



Antall bed. 4347

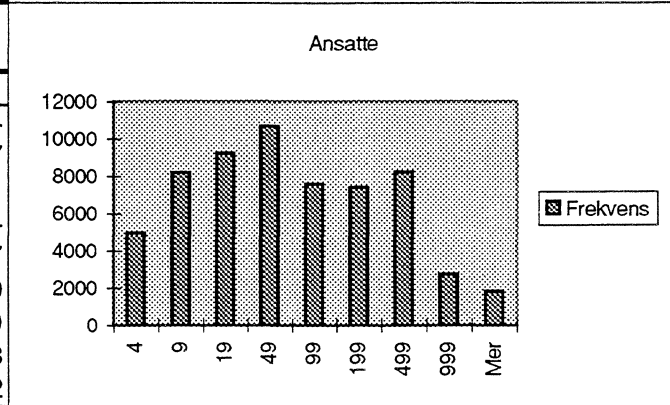
Gj.snitt: 14,02162

Std.avvik: 44,944

Max: 1832

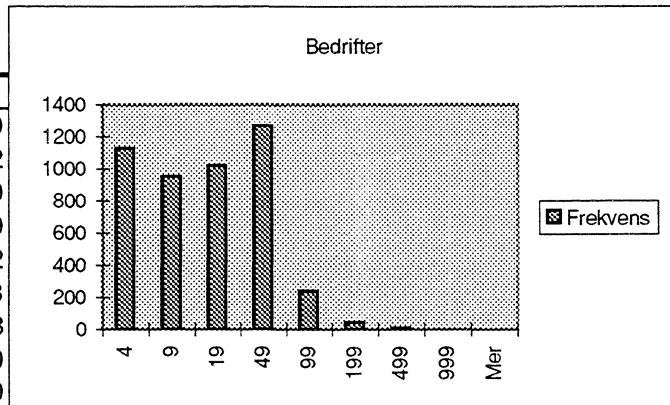
Antall ans. 60952

<b>Ansatte</b>	
Intervall	Frekvens
4	4974
9	8217
19	9251
49	10704
99	7577
199	7419
499	8230
999	2748
Mer	1832



Hele landet  
Næring M

<b>Bedrifter</b>	
Intervall	Frekvens
4	1129
9	952
19	1019
49	1269
99	242
199	43
499	8
999	0
Mer	0



Antall bed. 4662

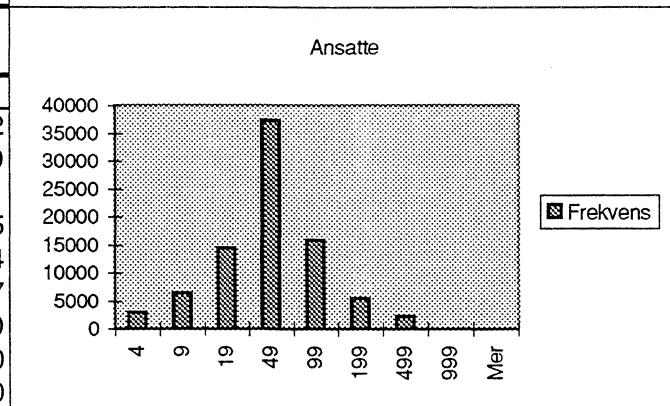
Gj.snitt: 18,25161

Std.avvik: 22,92092

Max: 446

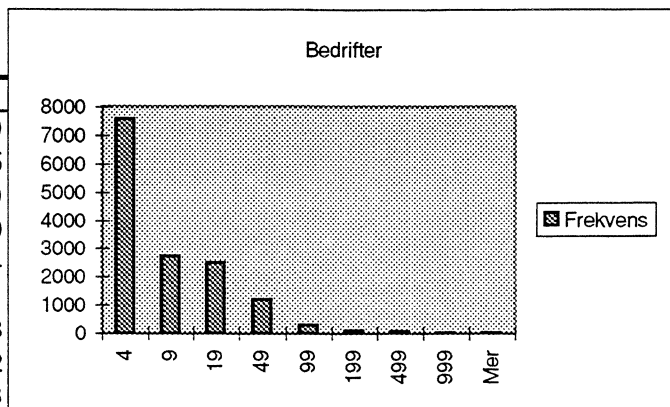
Antall ans. 85089

<b>Ansatte</b>	
Intervall	Frekvens
4	2942
9	6430
19	14531
49	37405
99	15884
199	5557
499	2340
999	0
Mer	0



Hele landet  
Næring N

<b>Bedrifter</b>	
Intervall	Frekvens
4	7600
9	2725
19	2510
49	1199
99	294
199	101
499	58
999	12
Mer	13



Antall bed. 14512

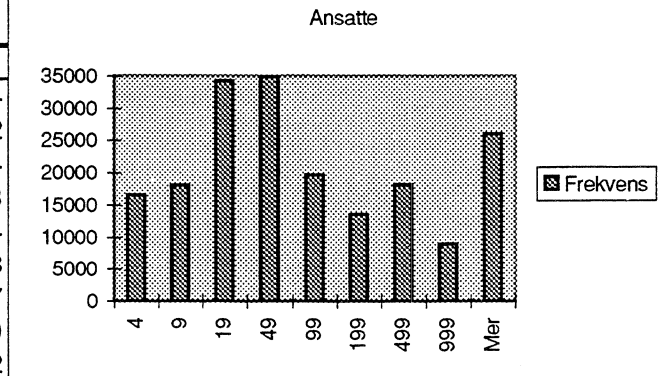
Gj.snitt: 13,12114

Std.avvik: 73,4655

Max: 3424

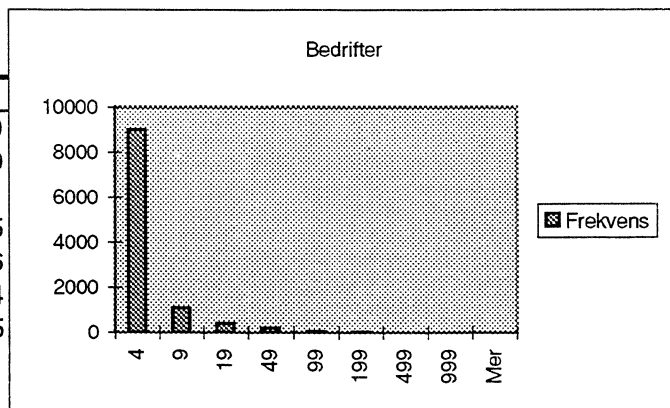
Antall ans. 190414

<b>Ansatte</b>	
Intervall	Frekvens
4	16614
9	18082
19	34324
49	34863
99	19674
199	13578
499	18207
999	8990
Mer	26082



Hele landet  
Næring O

<b>Bedrifter</b>	
Intervall	Frekvens
4	8990
9	1099
19	411
49	195
99	46
199	24
499	5
999	1
Mer	1



Antall bed. 10772

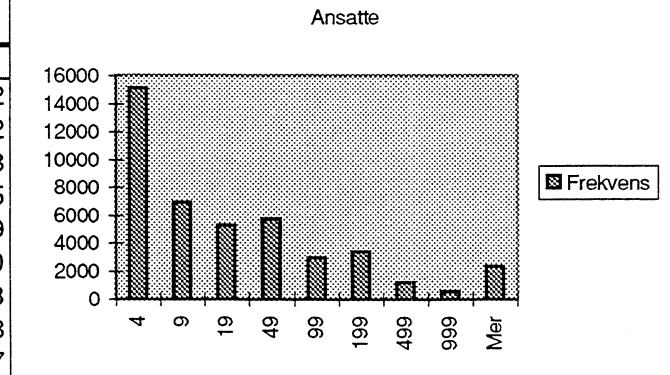
Gj.snitt: 4,04326

Std.avvik: 25,58178

Max: 2367

Antall ans. 43554

<b>Ansatte</b>	
Intervall	Frekvens
4	15112
9	6962
19	5278
49	5765
99	2949
199	3390
499	1173
999	558
Mer	2367



## De sist utgitte publikasjonene i serien Notater

- 96/53 F.R. Aune: Konsekvenser av en nordisk avgiftsharmonisering på elektrisitetsområdet. 22s.
- 96/54 M.V. Dysterud og P. Schøning: SSB-AVLØP. 187s.
- 96/55 E. Vassnes og I. Tuveng: Datagrunnlag for analyse av personers overgang fra utdanning til arbeid: Dokumentasjon. 58s.
- 96/56 K. Flugsrud, O.K. Hunnes og E. Lasson: Metode for beregning av energivarebruk og utslipp på grunnkretser: Beregninger for 1992 og 1993 for kommunene Oslo, Drammen, Bergen og Trondheim. 61s.
- 96/57 T. Kalve: Bedre barnevernsdata på edb-lesbart medium. 42s.
- 96/58 E. Midtlyng og A.A. Ritland: Leseferdigheter i den voksne befolkningen i Norge: Pilotundersøkelse: Dokumentasjonsrapport. 53s.
- 96/59 A. Sundvoll og L. Solheim: Undersøkelse om kopiering på universiteter og høyskoler: Pilotundersøkelse: Dokumentasjonsrapport. 48s.
- 96/60 A. Sundvoll: Undersøkelse om levekår og nærmiljø i Bergen: Dokumentasjonsrapport. 53s.
- 96/61 A. Bråten: Populasjon og utvalg - konsumprisindeksen. 58s.
- 96/62 M. Kjelsrud og A. Torstensen: Innvandreres tilknytning til arbeidsmarkedet. Situasjonen i november 1994. Bruttoendringer mellom november 1993 og november 1994: Dokumentasjon og analyse. 170s.
- 96/63 H.M. Teigum: Samordnet levekårsundersøkelse 1996 - tverrsnittsundersøkelsen: Dokumentasjonsrapport. 57s.
- 96/64 Å. Kaurin: Emballasjestatistikk: Utprøving av metoder og forslag til metode for innhenting av data til en nasjonal statistikk over emballasjeavfall. 46s.
- 97/1 S. Opdahl: Levekårsundersøkelse blant mottakere av grunnstønad: Dokumentasjonsrapport med tabeller. 138s.
- 97/2 E. Berg og K. Rypdal: Historisk utvikling og fremskrivning av forbruket av noen miljøskadelige produkter. 23s.
- 97/3 A. Sundvoll: Undersøkelse om velferdsstatens gleder og byrder: Dokumentasjonsrapport. 88s.
- 97/4 M.S. Bjerkseth: Evaluering av ny metode for utarbeidelse av strukturstatistikk ved Seksjon 460. 145s.
- 97/6 E. Gulløy, S. Blom og A.A. Ritland: Levekår blant innvandrere 1996: Dokumentasjonsrapport med tabeller. 205s.
- 97/7 S. Blom og A.A. Ritland: Levekår blant innvandrere 1996: Del 2: Tabeller for nordmenn. 1997. 222s.
- 97/8 T.C. Mykkelbost: Resultater fra brukerundersøkelse i forbindelse med NOS 306: Utslipp til luft i norske kommuner 1993. 21s.
- 97/9 H.M. Teigum: Omnibusundersøkelsene 1996: Dokumentasjonsrapport. 136s.
- 97/10 P.O. Lande og T. Heimdal: GERIX START: System- og brukardokumentasjon. 49s.
- 97/11 A. Barstad: Frihetens århundre? Levekår i Norge i et 100-årsperspektiv. 37s.
- 97/12 G. Sparby: Inntekts- og formuesundersøkelsen 1992: Dokumentasjon. 101s.
- 97/13 V. Pedersen: Inntekts- og formuesundersøkelsen 1993: Dokumentasjon. 94s.
- 97/14 V. Pedersen: Inntekts- og formuesundersøkelsen 1994: Dokumentasjon. 93s.
- 97/15 Metodevalg og kostnader ved etablering og drift av et boligregister. 29s.
- 97/16 K. Vassenden: Innvandrerstatistikkprosjektet: Styringsgruppas evaluering. 34s.
- 97/19 H.M. Teigum: Verdiundersøkelsen 1996: Dokumentasjonsrapport. 84s.
- 97/20 T. Ouren og T. Vik: Prosjektrapport: Voksenopplæringsprosjektet 1995-1996. 24s.

Statistisk sentralbyrå

*Oslo:*  
Postboks 8131 Dep.  
0033 Oslo

Telefon: 22 86 45 00  
Telefaks: 22 86 49 73

*Kongsvinger:*  
Postboks 1260  
2201 Kongsvinger

Telefon: 62 88 50 00  
Telefaks: 62 88 50 30

ISSN 0806-3745

