



Ole Villund

Kvalitet på yrke i registerbasert statistikk

Resultater og videre utfordringer

Notater

1 Innhold

| | | |
|-----|---|----|
| 1 | Innhold..... | 1 |
| 2 | Innledning..... | 3 |
| 2.1 | Bakgrunn | 3 |
| 2.2 | Kontroller av yrkeskoder i Arbeidstakerregisteret..... | 3 |
| 3 | Aktuelle kvalitetsmålinger..... | 4 |
| 3.1 | Presisjon | 4 |
| 3.2 | Samsvarsanalyse av yrke i Arbeidskraftundersøkelsen og Arbeidstakerregisteret..... | 5 |
| 3.3 | Datagrunnlag | 5 |
| 3.4 | Samsvar av yrkeskoder..... | 6 |
| 3.5 | Makroanalyse | 9 |
| 3.6 | Validitet | 11 |
| 3.7 | Endringer | 11 |
| 3.8 | Oppfølging..... | 13 |
| 4 | Innføringen av yrke i registerbasert sysselsettingsstatistikk..... | 13 |
| 4.1 | Bakgrunn | 13 |
| 4.2 | Oppgave..... | 13 |
| 4.3 | Gjennomføring | 14 |
| 4.4 | Kvalitet | 15 |
| 4.5 | Videre utvikling..... | 16 |
| 4.6 | Kriterier | 16 |
| 4.7 | Anvendelse | 16 |
| 4.8 | Metoder og alternativer | 16 |
| 4.9 | Evaluerings..... | 18 |
| 5 | Utvikling av metoder for kontroll av yrkeskodingen i Arbeidstakerregisteret..... | 19 |
| 5.1 | Oversikt | 19 |
| 5.2 | Status | 19 |
| 5.3 | Metodikk for store feil..... | 23 |
| 5.4 | Metodikk for små feil | 33 |
| 5.5 | Konklusjon | 38 |
| 5.6 | Videre utvikling..... | 39 |
| 6 | Vedlegg | 40 |
| 6.1 | Yrkesstatistikk | 40 |
| 6.2 | Datadokumentasjon estimeringsopplegg | 40 |
| 6.3 | Revisjonsopplegget..... | 43 |
| 7 | Henvisninger | 46 |
| | Tidligere utkommet i serien Notater..... | 47 |
| | Tabell 3-1: Frafall av yrkeskode i Arbeidstakerregisteret og estimert presisjon. 2003-2004..... | 4 |
| | Tabell 3-2: Skjevheter i frafall av yrkeskode i Arbeidstakerregisteret. 2003-2004. Absolutt og relativ forskjell..... | 5 |
| | Tabell 3-3: Samsvar av yrkes- og næringskoder pr. person i kobling av AKU og AA. 2003..... | 6 |
| | Figur 3-4: Samsvar av yrkes- og næringskoder pr. person i kobling av AKU og AA. 2003..... | 6 |
| | Tabell 3-5: Samsvar av yrkeskoder i samme næring etter kjønn, alder og yrke. Kobling av AKU og AA. 2003..... | 7 |
| | Figur 3-6: Samsvar av yrkeskoder etter kjønn og alder. Kobling av AKU og AA. 2003..... | 7 |
| | Figur 3-7: Samsvar av yrkeskoder etter yrkesfelt. Kobling av AKU og AA. 2003..... | 8 |
| | Tabell 3-8: Mikrokonsistens av yrkesfelt. Fra delutvalg av koblet AKU og AA. 2003..... | 8 |
| | Tabell 3-9: Skjevhet i delutvalg av koblet AKU og AA. 2003. Prosent og relativ skjevhet..... | 10 |
| | Tabell 3-10: Systematiske forskjeller i utvalgte yrker. Koblet AKU og AA. 2003. Prosentandeler og relative..... | 10 |
| | Tabell 3-11: Systematiske feil på 1-siffer yrke. Koblet AKU og AA. 2003. Prosentandeler og relative..... | 11 |
| | Tabell 3-12: Tilfeldige feil på 1-siffer yrke. Koblet AKU og AA. 2003. Prosentandeler og relative..... | 11 |
| | Tabell 3-13: Samsvar på yrkeskoder etter samsvar av næringskoder. Fra AKU/AA 2004..... | 12 |
| | Figur 3-14: Samsvar av yrkeskoder etter kjønn og alder. Fra AKU/AA 2003 og 2004..... | 12 |
| | Tabell 3-15: Konsistens av yrkesfelt. Delutvalg av koblet AKU og AA. Endring fra 2003 til 2004..... | 13 |
| | Tabell 5-1: Status for levering av yrkesdata i Arbeidstakerregisteret uke 31-2004..... | 19 |
| | Tabell 5-2: Fordeling av bedrift, foretak og næring etter antall ansatte. Arbeidstakerregisteret pr. uke 31-2004..... | 20 |
| | Tabell 5-3: Levert yrkesdata i prosent av ansatte, etter størrelse. Arbeidstakerregisteret pr. uke 40-2004..... | 21 |
| | Tabell 5-4: Antall bedrifter pr. foretak, etter størrelse. Arbeidstakerregisteret pr. uke 31-2004..... | 21 |
| | Tabell 5-5: Antall enheter og ansatte, etter størrelse. Arbeidstakerregisteret pr. uke 31-2004..... | 21 |
| | Tabell 5-6: Leveringsgrad av yrkesdata, etter størrelse. Arbeidstakerregisteret pr. uke 31-2004..... | 22 |

| | |
|---|----|
| Tabell 5-7: Omfanget av yrkesdatarapportering. Arbeidstakerregisteret pr. uke 31-2004 | 22 |
| Tabell 5-8: Yrkesdata i Arbeidstakerregisteret fra uke 31-35 2004, etter hovedarbeidsforhold. Antall og prosent. | 23 |
| Tabell 5-9: Antall bedrifter og ansatte i utvalgte næringer. Arbeidstakerregisteret uke 31-2004..... | 25 |
| Tabell 5-10: Oversikt over de ulike aggregeringsmetodene. | 29 |
| Tabell 5-11: Eksempel på yrkesstruktur i bedrifter og avvik fra gruppegjennomsnitt. | 29 |
| Tabell 5-12: Eksempel på yrkesstruktur i bedrifter og avvik fra gruppegjennomsnitt. | 29 |
| Tabell 5-13: Eksempel på yrkesstruktur i bedrifter og avvik fra gruppegjennomsnitt. | 30 |
| Tabell 5-14: Eksempel på yrkesstruktur i bedrifter og avvik fra gruppegjennomsnitt. | 30 |
| Tabell 5-15: Eksempel på yrkesstruktur i bedrifter og avvik fra gruppegjennomsnitt. | 31 |
| Tabell 5-16: Eksempel på yrkesstruktur i bedrifter og avvik fra gruppegjennomsnitt. | 31 |
| Tabell 5-17: Eksempel på yrkesstruktur i bedrifter og avvik fra gruppegjennomsnitt. | 31 |
| Tabell 5-18: Eksempel på yrkesstruktur i bedrifter og avvik fra gruppegjennomsnitt. | 32 |
| Tabell 5-19: Eksempel på stratum med få/tydelige avvik. $k_g=1.43$ | 33 |
| Tabell 5-20: Eksempel på stratum med mindre klar yrkesstruktur. $k_g=1.08$ | 33 |
| Tabell 5-21: Næringer hvor det forekommer mange yrkeskoder med svært små andeler. | 34 |
| Tabell 5-22: Eksempel på næringer etter yrkeshomogenitet. Fra Arbeidstakerregisteret uke 35. | 36 |
| Tabell 5-23: Eksempel på undersøkelse av mulige tilfeldige feil utfra yrkesfordeling i bedrifter..... | 37 |
| Tabell 5-24: Eksempel på undersøkelse av mulige tilfeldige feil utfra yrkesfordeling i bedrifter..... | 37 |
| Tabell 5-25: Eksempel på yrker, etter næringsspesifisitet. Fra Arbeidstakerregisteret uke 35..... | 38 |
| Tabell 6-1: Stratifisering etter næringsundergrupper og størrelsesgruppene..... | 44 |

2 Innledning

Dette notatet beskriver ulike sider ved å dokumentere og forbedre kvaliteten på yrke i registerbasert sysselsettingsstatistikk. Det omfatter kvalitetsmåling av yrke i Arbeidstakerregisteret, opplegget for å lage yrkesfordeling i den publiserte statistikken og en oppsummering om kvaliteten i den, samt forslag til metoder for revisjon av yrke i Arbeidstakerregisteret.

Notatet er utviklet i tiden fra sommeren 2004 og fram til utgivelsen. Det er derfor noe ulikheter i tidspunktet for datauttakene til de ulike undersøkelsene som er referert. Det er også slik at i løpet av 2005 skal alle foretak som benytter såkalte PAI-koder i Arbeidstakerregisteret gå over til å levere yrkeskoder. Dette vil få betydelig innvirkning på totale yrkesdata i Arbeidstakerregisteret, og vil bli dokumentert separat.

Takk til Leiv Solheim for verdifulle kommentarer på mange av de metodiske problemstillingene.

2.1 Bakgrunn

Fra og med statistikkåret 2003 var yrke med som kjennemerke i den registerbasert sysselsettingsstatistikken. Bakgrunnen for at dette var mulig var innføringen av yrkeskoder i Arbeidstakerregisteret, som omfatter de fleste sysselsatte. Alle arbeidstakerforhold i Arbeidstakerregisteret skal inneholde yrkesdata, som skal brukes i mange viktige statistikker. Arbeidsgiverne, Trygdeetaten og Statistisk sentralbyrå har lagt ned mye arbeid for å få dette på plass. Årene 2000-2002 kan man si var utviklings- og oppbyggingsfasen. De første offisielle statistikkene på bakgrunn av yrkeskoder i Arbeidstakerregisteret var sykefravær etter yrke og yrkesfordeling etter arbeidsstedskommune, m.m. Det er i den forbindelse utviklet et estimeringsopplegg for detaljert yrke på alle sysselsatte. Kvalitetsarbeidet har bestått i både å øke rapporteringen av yrkesdata og kontrollere selve yrkeskodene. Det har vært gjennomført både automatiske og manuelle kontroller og omkodning, på bakgrunn av stikkprøver og analyser av kvaliteten. For nærmere beskrivelse vises til notatene som er oppgitt i vedlegg.

2.2 Kontroller av yrkeskoder i Arbeidstakerregisteret

Her nevnes kort de viktigste metoder som har vært brukt i større eller mindre utstrekning for å vurdere og forbedre kvaliteten. Mye av kontrollene i startfasen har vært av de automatiske yrkeskodene. Metoder for korrigerings har vært både programmert (automatisk) omkodning og manuell (skjønnsmessig) koding. Dette er vel egentlig ikke revisjon, og ble ikke så godt dokumentert i starten. Man kan argumentere for at det er viktigere å dokumentere de korreksjoner som gjøres av *innrapporterte* yrkeskoder.

En enkel, men arbeidskrevende metode er å ta stikkprøver og undersøke yrkeskode på jobbnivå direkte. Tilfeldig utvalgte arbeidstakerforhold har vært listet ut, sammen med relevante tilleggsopplysninger som er koblet fra ulike kilder. Dette har gitt grunnlag for å måle kvaliteten samt gjennomføre endel korreksjoner. I noe mindre utstrekning kan det bli aktuelt som et internt kvalitetsmål.

For å undersøke den automatiske kodingen i register har en kjørt det automatiske kodesystemet på yrkestitler fra Arbeidskraftundersøkelsen (AKU), og sammenlikne med den opprinnelige (manuelt kodede) yrkeskoden. Dette ble særlig brukt i utviklingen av programmene.

Mikroanalyser hvor man undersøker yrkeskodene på person- eller jobbnivå kategori deles i to hovedtyper: Konsistens- eller samsvarsanalyser der en kobler til andre datakilder og sammenlikne yrkeskodene for samme jobb. Datakilder som yrke i Arbeidstakerregisteret har vært sammenliknet med er AKU-utvalgsdata og data fra Lønnsstatistikken. Dette er viktige uavhengige kvalitetsmål, hvor referanse som regel er AKU.

Vi har også prøvd å undersøke hvor konsekvent yrkeskodingen er, ved å sammenlikne liknende arbeidstakerforhold i Arbeidstakerregisteret med hverandre. En måler i hvilken grad arbeidstakerforhold som har samme yrkestittel/tekst, samme næringskode og utdanningskode, har like yrkeskoder. En fordel med denne metoden er at den ikke påvirkes av forskjeller i variabelinnhold i de ulike datakildene.

Makroanalyser tar utgangspunkt i aggregerte tall, og vi ser bort fra tilfeldige feil. Man kan si at slike analyser forteller med direkte hvordan kvaliteten er på publiserte tabeller. Sammenlikning av registerbasert yrkesstatistikk med publiserte tall fra annet hold har avdekket systematiske forskjeller. Aktuelle kilder har vært yrkesfordeling i AKU, utdanningsstatistikk, lønnsstatistikk, helsepersonellstatistikk og yrkesstatistikk fra sammenliknbare land. Videre er det sett på tall fra organisasjoner som legeföreningen, fotografforbundet, m.fl.

Ved å modellere yrke ved hjelp av tilstrekkelige forklaringsvariabler, kan en direkte identifisere jobber med usannsynlige yrkeskoder. Det er store utfordringer i detaljeringsnivået, registeregenskaper og ikke minst stor

variasjonen i de ansattes bakgrunn innen mange yrker. Endel yrker er generelle på den måten at de i liten grad kan klassifiseres utfra tilgjengelige variabler. Det ville vært interessant å undersøke om noen av metodene som ble utviklet i forbindelse med imputering og prediksjon kunne utnyttes i utviklingen av nye revisjonsmetoder.

I kapittel 5 viser vi metoder for å sammenlikne yrkesfordelingen mellom liknende bedrifter. Hensikten er å oppdage systematiske forskjeller i yrkeskodingen og identifisere enheter som bør undersøkes. Dette blir en slags mellomting mellom mikro- og makroanalysene nevnt over. Det er prinsipielt en indirekte måte, men det er enkelt å identifisere arbeidstakerforholdene når bedriften er kjent, og gjennomføre revisjon på mikronivå.

3 Aktuelle kvalitetsmålinger

3.1 Presisjon

Her benyttes metoder fra utvalgsundersøkelser på registerdata for å vurdere betydningen av manglende yrkeopplysninger. For korthets skyld brukes *frafall* om både manglende innrapportering og titler som ikke har blitt klassifisert f.eks. fordi de er for generelle til å yrkeskodes. En regner her både stillingskoder og yrkeskoder som respons. Presisjon regnes også utfra at det partielle frafallet er rent tilfeldig. Selv om dette faktisk ikke er tilfelle, brukes dette som referanse bl.a. for å vise størrelsen sammenliknet med andre kvalitetsmål. Videre er det interessant i forhold til størrelsen på de minste strata i estimeringsopplegget for yrke i registerbasert sysselsettingsstatistikk (se eget notat).

Betegner:

- (3-1) p andelen av et yrke, f.eks. 1%
 (3-2) N "populasjonen", antall i register
 (3-3) n "utvalget", antall med levert yrkesdata

Standardfeilen (standardavviket til estimatoren):

$$(3-4) \quad SE(p) = \sqrt{\frac{N-n}{N-1} \cdot \frac{1}{n} \cdot p \cdot (p-1)}$$

Sammenlikner resultat for de to siste år, og ser en viss forbedring på grunn av øket leveringsgrad.

Tabell 3-1: Frafall av yrkeskode i Arbeidstakerregisteret og estimert presisjon. 2003-2004.

| | År 2003 | År 2004 |
|--------------------------|-----------|-----------|
| Hovedarbeidstakerforhold | 1 940 086 | 1 959 429 |
| Levert yrke | 1 783 819 | 1 844 460 |
| Frafall | 8.1 % | 5.9 % |
| Eksempel p | 1 % | 1 % |
| $SE(p)$ | 2.11E-05 | 1.77E-05 |
| Standardfeilen i antall | 41.02 | 34.77 |

Relativ standardfeil er her ganske liten, omlag 0.2%.

Det partielle frafallet av yrkeskode er ikke jevnt. Vi ser kort på skjevheter som vil ha betydning for yrkesstatistikk etter viktige kjennemerker som kjønn og alder. Man kan si at fra 2003 til 2004 er det blitt mindre skjevhet etter kjønn og mer skjevhet i alder. Det er særlig de yngste og eldste som er overrepresentert i frafallsgruppen, noen som kan gi utslag i yrkesfordelingen. I estimeringsopplegget for yrke i registerbasert sysselsettingsstatistikk, stratifiseres det bl.a. etter kjønn og alder som til en viss grad kompenserer dette.

Tabell 3-2: Skjevheter i frafall av yrkeskode i Arbeidstakerregisteret. 2003-2004. Absolutt og relativ forskjell.

| | | År 2003 | | | | År 2004 | | | |
|---------|----------|-----------|-------------|-----------|----------------|-----------|-------------|-----------|----------------|
| | | I frafall | I hele reg. | Forskjell | Rel. forskjell | I frafall | I hele reg. | Forskjell | Rel. forskjell |
| | Menn | 59.31 | 51.71 | 7.60 | 15 % | 57.71 | 51.62 | 6.09 | 12 % |
| | Kvinner | 40.69 | 48.29 | -7.60 | -16 % | 42.29 | 48.38 | -6.09 | -13 % |
| I alt | 16-19 år | 3.32 | 3.29 | 0.03 | 1 % | 3.69 | 3.01 | 0.68 | 23 % |
| | 20-24 år | 7.81 | 7.76 | 0.05 | 1 % | 8.51 | 7.99 | 0.52 | 7 % |
| | 25-39 år | 36.85 | 36.19 | 0.66 | 2 % | 35.93 | 35.45 | 0.48 | 1 % |
| | 40-54 år | 35.72 | 35.61 | 0.11 | 0 % | 34.91 | 35.57 | -0.66 | -2 % |
| | 55-66 år | 15.39 | 16.30 | -0.92 | -6 % | 15.91 | 17.03 | -1.12 | -7 % |
| | 67-74 år | 0.91 | 0.84 | 0.07 | 8 % | 1.06 | 0.95 | 0.11 | 12 % |
| Menn | 16-19 år | 1.70 | 1.60 | 0.10 | 6 % | 1.85 | 1.47 | 0.38 | 26 % |
| | 20-24 år | 4.28 | 3.94 | 0.34 | 9 % | 4.36 | 4.05 | 0.31 | 8 % |
| | 25-39 år | 21.35 | 19.16 | 2.19 | 11 % | 20.26 | 18.76 | 1.50 | 8 % |
| | 40-54 år | 21.85 | 18.10 | 3.74 | 21 % | 20.66 | 18.05 | 2.61 | 14 % |
| | 55-66 år | 9.58 | 8.46 | 1.12 | 13 % | 9.93 | 8.79 | 1.14 | 13 % |
| | 67-74 år | 0.55 | 0.45 | 0.10 | 23 % | 0.65 | 0.50 | 0.15 | 30 % |
| Kvinner | 16-19 år | 1.62 | 1.69 | -0.07 | -4 % | 1.84 | 1.53 | 0.30 | 20 % |
| | 20-24 år | 3.53 | 3.82 | -0.29 | -8 % | 4.15 | 3.95 | 0.21 | 5 % |
| | 25-39 år | 15.50 | 17.02 | -1.53 | -9 % | 15.67 | 16.69 | -1.02 | -6 % |
| | 40-54 år | 13.87 | 17.51 | -3.64 | -21 % | 14.25 | 17.52 | -3.27 | -19 % |
| | 55-66 år | 5.81 | 7.85 | -2.04 | -26 % | 5.98 | 8.25 | -2.27 | -27 % |
| | 67-74 år | 0.36 | 0.40 | -0.04 | -9 % | 0.41 | 0.45 | -0.04 | -9 % |

3.2 Samsvarsanalyse av yrke i Arbeidskraftundersøkelsen og Arbeidstakerregisteret.

I kvalitetsarbeidet med yrke i Arbeidstakerregisteret (AA) har det helt fra starten av vært interessant med samsvaranalyser mellom registerdata og data fra Arbeidskraftundersøkelsen (AKU). I korte trekk går dette ut på å koble datafiler sammen på personnivå, og sammenlikne yrkeskodene.

Det er flere feilkilder ved slike analyser:

- En person kan ha flere jobber, og forskjellige yrker i dem. Kobling på personnivå kan føre til at en sammenlikner yrkeskoder for ulike arbeidsoppgaver.
- En person kan bytte jobb, eller bytte yrke i samme ansettelsesforhold. Forskyving i tid mellom de to datakildene, kan gjøre at man sammenlikner reelt forskjellige yrker.
- Til forskjell fra registerdata inneholder AKU-data også beskrivelse av arbeidsoppgaver. Det er også noen ulikheter i noen andre variabler og forskjellige metoder for selve kodingen. Tilsammen gjør dette gjør at innholdet i klassifiseringen ikke er det samme. Sagt enkelt: selv om samme jobben får to ulike yrkeskoder, kan begge være "riktige" utfra de ulike data i de to kildene.

Allikevel mener vi at sammenlikninger med AKU er viktig grunnlag for uavhengig kvalitetskontroll av yrke og andre kjennemerker i Arbeidstakerregisteret, og i registerbasert sysselsetningsstatistikk. Det er verdt å merke seg at analysen omfatter bare delpopulasjonen som har **yrkeskode** i register, hvor hovedsakelig kommuner og sykehus ikke inngår.

3.3 Datagrunnlag

Data som inngår er et delutvalg fra AKU av sysselsatte *ansatte* fra kvartalsfiler hvor det velges siste records pr. person. Resultatet er et større utvalg for å bedre grunnlaget for sammenlikning av yrker på et detaljert nivå. På grunn av det månedlige kalibreringsopplegget er da vektene i dette datasettet ikke lenger helt representative for denne "årsfilen". I samsvaranalysene her brukes kun uvektede tall fra det koblede delutvalget. Unntaket er beregning av skjevhet, der referansen er vektet årsgjennomsnitt fra årsfil med flere records pr. person.

Data fra Arbeidstakerregisteret er periodefiler med hovedarbeidstakerforhold for bosatte i alderen 16-74 år. Det innebærer at enheten er person, som gjør en kobling enklere. Yrkeskode defineres som 4 siffer fra den aktuelle 7-siffer koden direkte fra Arbeidstakerregisteret. Det betyr at kommunale stillingskoder, m.m. ikke inngår, i motsetning til noen av tidligere undersøkelser. Resultatet er yrkeskoder som er *oppgift av arbeidsgiver direkte*, eller som kodet ved Statistisk sentralbyrå og skal være *kontrollert av arbeidsgiver*.

3.4 Samsvar av yrkeskoder

3.4.1 Enkel sammenlikning av koder

For å få en enkel oversikt sammenlikner vi de to yrkeskodene pr. person. Som i tidligere forsøk ser vi hvor like kodene er ved å sammenlikne siffer for siffer (fra venstre). "yrke 0" betegner da at samme person har fått to helt forskjellige yrkeskoder. "yrke 4" betyr at for denne personen er yrkeskoden i AKU helt lik yrkeskoden i register. For en indirekte kontroll om det kan være ulike jobber, lager vi en tilsvarende sammenlikningsvariabel for næringskode. I tabellene betegner "nace 0" at personen har to helt ulike næringskoder i de to datakildene. F.eks. "nace 2" betegner at næringskodene er like i de 2 første siffer, men ikke alle 5. Vi benytter ikke organisasjonsnummer i koblingsnøkkelen denne gang, for nettopp å se på betydning av at personer har flere jobber i ulike næringer.

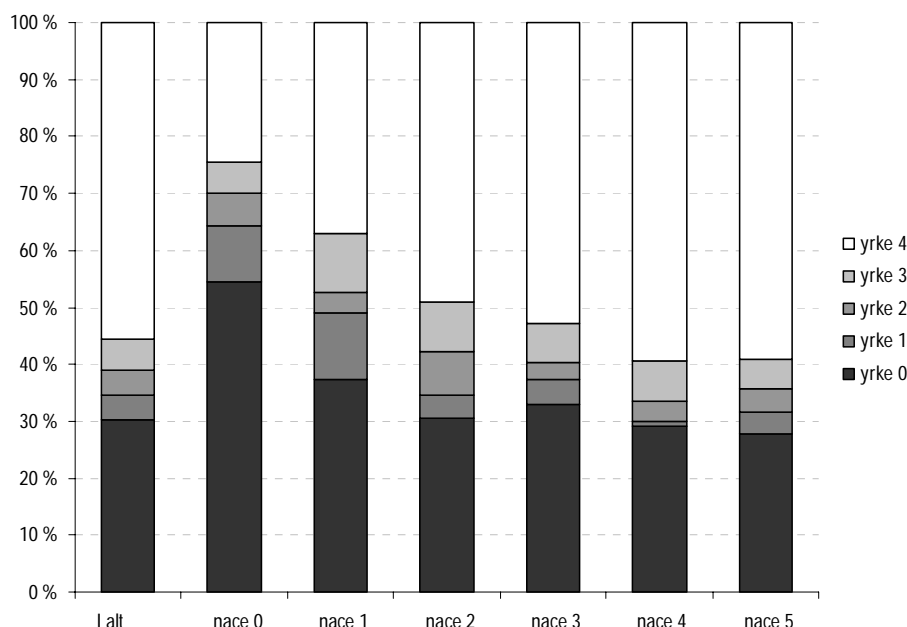
Tabell 3-3: Samsvar av yrkes- og næringskoder pr. person i kobling av AKU og AA. 2003.

| | I alt | Uoppgitt | yrke 0 | yrke 1 | yrke 2 | yrke 3 | yrke 4 |
|--------|--------|----------|--------|--------|--------|--------|--------|
| I alt | 18 096 | 6 884 | 3 404 | 488 | 471 | 621 | 6 228 |
| nace 0 | 1 229 | 429 | 437 | 78 | 46 | 43 | 196 |
| nace 1 | 397 | 159 | 89 | 28 | 8 | 25 | 88 |
| nace 2 | 1 019 | 531 | 149 | 20 | 37 | 43 | 239 |
| nace 3 | 889 | 400 | 161 | 21 | 15 | 33 | 259 |
| nace 4 | 597 | 350 | 72 | 2 | 9 | 17 | 147 |
| nace 5 | 13 965 | 5 015 | 2 496 | 339 | 356 | 460 | 5 299 |

| | 100 | 38 | 19 | 3 | 3 | 3 | 34 |
|--------|-----|----|----|---|---|---|----|
| I alt | 100 | 35 | 36 | 6 | 4 | 3 | 16 |
| nace 0 | 100 | 40 | 22 | 7 | 2 | 6 | 22 |
| nace 1 | 100 | 52 | 15 | 2 | 4 | 4 | 23 |
| nace 2 | 100 | 45 | 18 | 2 | 2 | 4 | 29 |
| nace 3 | 100 | 59 | 12 | 0 | 2 | 3 | 25 |
| nace 4 | 100 | 36 | 18 | 2 | 3 | 3 | 38 |

Uoppgitt yrke skyldes først og fremst register, ikke AKU. Vi ser bort fra uoppgitt næring i begge kilder, da dette er svært liten andel. For å få et enda klarere bilde av kvaliteten og i hvilken grad den avhenger av nærings-samsvaret, kan vi se på et diagram med noen av tallene fra tabellen. Uoppgitt er holdt utenfor, og det vises relative andeler.

Figur 3-4: Samsvar av yrkes- og næringskoder pr. person i kobling av AKU og AA. 2003.



Man ser tydelig noen hovedfunnene som ved tidligere undersøkelser:

- Det er sprik i yrkeskodene, selv når vi har grunn til å tro at det er samme type jobb.
- Yrkeskodene er oftest helt like eller helt forskjellige, det er lite gradvis likhet. Det betyr at vi ikke har så stor gevinst av å aggregere etter hierarkiet i yrkesstandarden.
- Ikke overraskende er ulikheten større jo mer forskjellig næringskoden er. Det er imidlertid endel som har samme yrke i jobber i ulike næringer, f.eks. rengjøring.

For videre analyse (også hele kap.4) bruker vi bare records som har lik næringskode, for å sammenlikne de som mest sannsynlig har samme arbeidsoppgaver. De neste tabellene har derfor en annen totalsum.

Tabell 3-5: Samsvar av yrkeskoder i samme næring etter kjønn, alder og yrke. Kobling av AKU og AA. 2003.

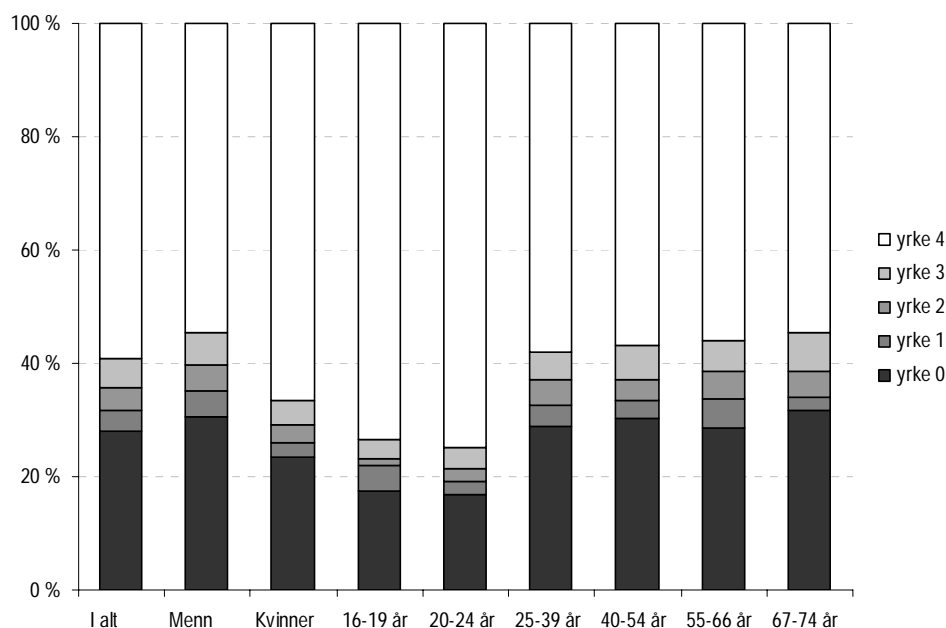
| | Antall | I alt | Uoppgitt yrke | 0 yrke | 1 yrke | 2 yrke | 3 yrke | 4 |
|----------|--------|-------|---------------|--------|--------|--------|--------|---|
| I alt | 13 965 | 5 015 | 2 496 | 339 | 356 | 460 | 5 299 | |
| Menn | 7 609 | 2 125 | 1 683 | 246 | 255 | 312 | 2 988 | |
| Kvinner | 6 356 | 2 890 | 813 | 93 | 101 | 148 | 2 311 | |
| 16-19 år | 375 | 61 | 55 | 14 | 4 | 10 | 231 | |
| 20-24 år | 921 | 199 | 121 | 18 | 15 | 27 | 541 | |
| 25-39 år | 5 021 | 1 664 | 964 | 133 | 149 | 166 | 1 945 | |
| 40-54 år | 5 139 | 2 082 | 926 | 97 | 116 | 177 | 1 741 | |
| 55-66 år | 2 447 | 991 | 416 | 76 | 70 | 77 | 817 | |
| 67-74 år | 62 | 18 | 14 | 1 | 2 | 3 | 24 | |

| Prosent | | | | | | | | |
|----------|-----|----|----|---|---|---|----|--|
| I alt | 100 | 36 | 18 | 2 | 3 | 3 | 38 | |
| Menn | 100 | 28 | 22 | 3 | 3 | 4 | 39 | |
| Kvinner | 100 | 45 | 13 | 1 | 2 | 2 | 36 | |
| 16-19 år | 100 | 16 | 15 | 4 | 1 | 3 | 62 | |
| 20-24 år | 100 | 22 | 13 | 2 | 2 | 3 | 59 | |
| 25-39 år | 100 | 33 | 19 | 3 | 3 | 3 | 39 | |
| 40-54 år | 100 | 41 | 18 | 2 | 2 | 3 | 34 | |
| 55-66 år | 100 | 40 | 17 | 3 | 3 | 3 | 33 | |
| 67-74 år | 100 | 29 | 23 | 2 | 3 | 5 | 39 | |

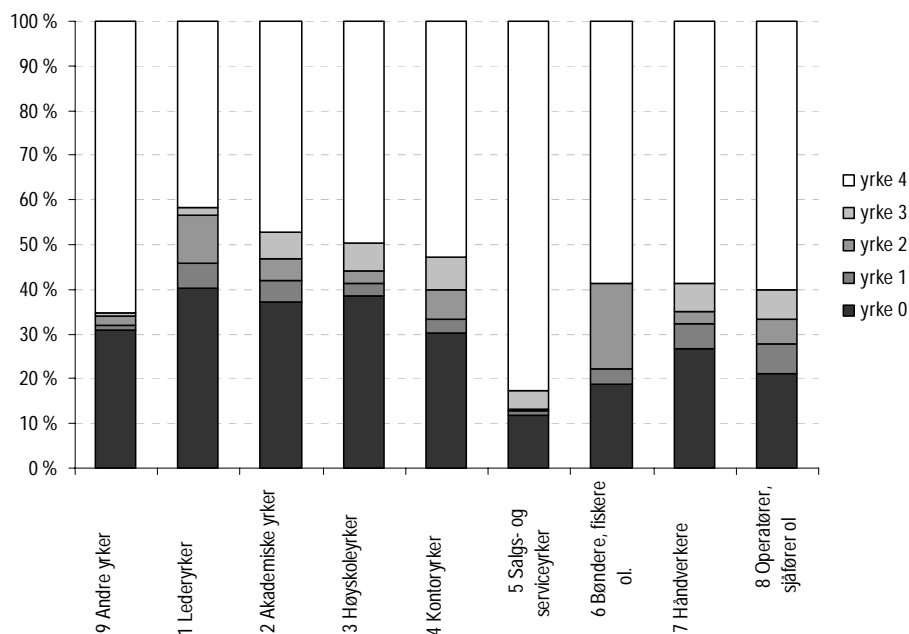
| Prosent | | | | | | | | |
|---------------------------|-----|----|----|---|----|---|----|--|
| 9 Andre yrker | 100 | 41 | 18 | 1 | 1 | 1 | 38 | |
| 1 Lederyrker | 100 | 27 | 30 | 4 | 8 | 1 | 30 | |
| 2 Akademiske yrker | 100 | 60 | 15 | 2 | 2 | 2 | 19 | |
| 3 Høyskoleyrker | 100 | 49 | 20 | 2 | 1 | 3 | 25 | |
| 4 Kontoryrker | 100 | 36 | 19 | 2 | 4 | 5 | 34 | |
| 5 Salgs- og serviceyrker | 100 | 31 | 8 | 1 | 0 | 3 | 57 | |
| 6 Bønder, fiskere ol. | 100 | 12 | 16 | 3 | 16 | 0 | 52 | |
| 7 Håndverkere | 100 | 16 | 23 | 5 | 2 | 5 | 49 | |
| 8 Operatører, sjåfører ol | 100 | 14 | 18 | 6 | 5 | 6 | 51 | |

Det går fram at kvaliteten varierer betydelig med viktige kjennemerker i statistikken, og også med yrke selv. Årsaken til den høyere andelen uoppgitt hos kvinner, skyldes nok overvekt av kommunale stillingskoder. I diagrammene er uoppgitt holdt utenfor for å gi et tydeligere bilde av den reelle kvaliteten.

Figur 3-6: Samsvar av yrkeskoder etter kjønn og alder. Kobling av AKU og AA. 2003.



Figur 3-7: Samsvar av yrkeskoder etter yrkesfelt. Kobling av AKU og AA. 2003.



3.4.2 Mikrokonsistens

For å undersøke mikrokonsistens nærmere, aggregerer vi til 1-siffer yrkesfelt (9 grupper) og tar kun med de som har oppgitt yrke i begge kilder. Tabellen viser at konsistensen ligger på mellom 60-85%. Tall for bønder og fiskere er usikre.

Tabell 3-8: Mikrokonsistens av yrkesfelt. Fra delutvalg av koblet AKU og AA. 2003.

| | | Register | | | | | | | | |
|----------------------------|--------|---------------|--------------|--------------------|-----------------|---------------|--------------------------|-----------------------|---------------|----------------------------|
| AKU | I alt | 9 Andre yrker | 1 Lederyrker | 2 Akademiske yrker | 3 Høyskoleyrker | 4 Kontoryrker | 5 Salgs- og serviceyrker | 6 Bønder, fiskere ol. | 7 Håndverkere | 8 Operatører, sjåførere ol |
| I alt | 11 212 | 694 | 1 076 | 811 | 1 913 | 1 094 | 2 714 | 123 | 1 415 | 1 372 |
| 9 Andre yrker | 589 | 389 | 6 | 1 | 11 | 28 | 88 | 2 | 12 | 52 |
| 1 Lederyrker | 1 110 | 13 | 655 | 69 | 123 | 60 | 121 | 6 | 30 | 33 |
| 2 Akademiske yrker | 842 | 2 | 72 | 504 | 189 | 39 | 16 | - | 5 | 15 |
| 3 Høyskoleyrker | 2 235 | 19 | 173 | 195 | 1 324 | 163 | 252 | 5 | 53 | 51 |
| 4 Kontoryrker | 985 | 25 | 37 | 19 | 101 | 675 | 80 | - | 8 | 40 |
| 5 Salgs- og serviceyrker | 2 450 | 100 | 69 | 8 | 66 | 67 | 2 080 | 8 | 20 | 32 |
| 6 Bønder, fiskere ol. | 113 | 6 | 9 | - | 1 | 1 | 2 | 85 | 6 | 3 |
| 7 Håndverkere | 1 676 | 87 | 31 | 12 | 72 | 20 | 37 | 9 | 1 179 | 229 |
| 8 Operatører, sjåførere ol | 1 212 | 53 | 24 | 3 | 26 | 41 | 38 | 8 | 102 | 917 |

| | | Register | | | | | | | | |
|----------------------------|-------|---------------|--------------|--------------------|-----------------|---------------|--------------------------|-----------------------|---------------|----------------------------|
| AKU | I alt | 9 Andre yrker | 1 Lederyrker | 2 Akademiske yrker | 3 Høyskoleyrker | 4 Kontoryrker | 5 Salgs- og serviceyrker | 6 Bønder, fiskere ol. | 7 Håndverkere | 8 Operatører, sjåførere ol |
| I alt | 100 | 6 | 10 | 7 | 17 | 10 | 24 | 1 | 13 | 12 |
| 9 Andre yrker | 100 | 66 | 1 | 0 | 2 | 5 | 15 | 0 | 2 | 9 |
| 1 Lederyrker | 100 | 1 | 59 | 6 | 11 | 5 | 11 | 1 | 3 | 3 |
| 2 Akademiske yrker | 100 | 0 | 9 | 60 | 22 | 5 | 2 | - | 1 | 2 |
| 3 Høyskoleyrker | 100 | 1 | 8 | 9 | 59 | 7 | 11 | 0 | 2 | 2 |
| 4 Kontoryrker | 100 | 3 | 4 | 2 | 10 | 69 | 8 | - | 1 | 4 |
| 5 Salgs- og serviceyrker | 100 | 4 | 3 | 0 | 3 | 3 | 85 | 0 | 1 | 1 |
| 6 Bønder, fiskere ol. | 100 | 5 | 8 | - | 1 | 1 | 2 | 75 | 5 | 3 |
| 7 Håndverkere | 100 | 5 | 2 | 1 | 4 | 1 | 2 | 1 | 70 | 14 |
| 8 Operatører, sjåførere ol | 100 | 4 | 2 | 0 | 2 | 3 | 3 | 1 | 8 | 76 |

Til sammenlikning er standardfeilen i dette utvalget (ved enkel estimering) 0.1% ved 1% andel.

3.5 Makroanalyse

3.5.1 Betegnelser

3.5.1.1 Kodevariabler

(3-5) Y_R Yrke i register

(3-6) Y_A Yrke i AKU

3.5.1.2 Indikatorvariabler

(3-7) $I_{Y,R}$ Har yrke y i register

(3-8) $I_{Y,A}$ Har yrke y i AKU

3.5.1.3 Andeler

(3-9) $p_{Y,R}$ Andel yrke y i register

(3-10) $p_{Y,A}$ Andel yrke y i AKU

(3-11) $p_I = \frac{\sum_i I_{Y,R} \wedge \neg I_{Y,A}}{n}$ Andel som har yrke y i register, men ikke i AKU

(3-12) $p_{II} = \frac{\sum_i I_{Y,A} \wedge \neg I_{Y,R}}{n}$ Andel som har yrke y i AKU, men ikke i register

3.5.1.4 Systematiske feil

(3-13) $\hat{\beta} = \hat{p}_{Y,R} - \hat{p}_{Y,A}$ Systematisk målefeil

(3-14) $\beta_{relativ} = \frac{\hat{p}_{Y,R} - \hat{p}_{Y,A}}{\hat{p}_{Y,A}}$ Relativ systematisk målefeil

3.5.1.5 Tilfeldige feil

(3-15) $\tau = \sqrt{(p_I + p_{II}) - (p_I - p_{II})^2}$ Tilfeldig målefeil

(3-16) $\tau_{relativ} = \frac{\tau / \sqrt{n}}{\hat{p}_{Y,A}}$ Relativ tilfeldig målefeil

3.5.2 Skjevhet og systematiske feil

For å kunne si noe om systematiske forskjeller i yrkeskodingen, bør vi først undersøke om delutvalget er representativt med hensyn til viktige variabler. Vi vil anta at utvalget er bedre enn i noen av de tidligere undersøkelser, bl.a. fordi vi kun tar med *ansatte* i AKU, ikke alle sysselsatte (som omfatter selvstendige og familiearbeidere). Det som fortsatt kan skape skjevheter er *ansatte* som ikke meldes til Arbeidstakerregisteret (småjobber, noen ganger betegnet som LTO-sysselsatte), egenskaper ved koblingen og delutvalget (valg av 1 record pr. person, samt kravet om lik næringskode). I tillegg kommer selvsagt tilfeldigheter i det aktuelle AKU-utvalget.

Tabell 3-9: Skjevhet i delutvalg av koblet AKU og AA. 2003. Prosent og relativ skjevhet.

| | Andeler i hver kilde | | | Utvalget | | Register | |
|----------|----------------------|----------|-------|----------|---------|----------|---------|
| | Utvalget | Register | AKU | Skjevhet | Relativ | Skjevhet | Relativ |
| Menn | 59.06 | 51.71 | 51.04 | 8.03 | 16 % | 0.68 | 1 % |
| 16-19 år | 1.85 | 1.60 | 1.84 | 0.00 | 0 % | -0.24 | -13 % |
| 20-24 år | 4.73 | 3.94 | 4.02 | 0.71 | 18 % | -0.08 | -2 % |
| 25-39 år | 22.90 | 19.16 | 18.99 | 3.90 | 21 % | 0.17 | 1 % |
| 40-54 år | 19.92 | 18.10 | 17.55 | 2.36 | 13 % | 0.55 | 3 % |
| 55-66 år | 9.40 | 8.46 | 8.23 | 1.17 | 14 % | 0.23 | 3 % |
| 67-74 år | 0.28 | 0.45 | 0.40 | -0.12 | -30 % | 0.05 | 13 % |
| Kvinner | 40.94 | 48.29 | 48.96 | -8.03 | -16 % | -0.68 | -1 % |
| 16-19 år | 2.19 | 1.69 | 2.13 | 0.06 | 3 % | -0.44 | -21 % |
| 20-24 år | 4.02 | 3.82 | 4.14 | -0.11 | -3 % | -0.32 | -8 % |
| 25-39 år | 15.10 | 17.02 | 17.45 | -2.35 | -13 % | -0.42 | -2 % |
| 40-54 år | 13.29 | 17.51 | 17.17 | -3.88 | -23 % | 0.34 | 2 % |
| 55-66 år | 6.08 | 7.85 | 7.66 | -1.58 | -21 % | 0.19 | 2 % |
| 67-74 år | 0.25 | 0.40 | 0.42 | -0.17 | -41 % | -0.02 | -6 % |

Det er skjevheter i det koblede delutvalget i forhold til "populasjonen" som nå er AKU. Det er også vist skjevhet på registerfilen. Kvinner er nok underrepresentert i delutvalget pga. overvekt i kommunal sektor uten yrkeskode. Unge personer er nok overrepresentert i forhold til Arbeidstakerregisteret pga. småjobbene.

I forhold til yrke ville det sikkert vært interessant også å vurdere utdanningsnivå og –fag, samt næringsstrukturen. Det blir imidlertid lett små grupper med stor utvalgusikkerhet. For å se på effekten av de demografiske skjevhetene som vist over, forsøker vi også et kalibrert mål som vises i neste tabell.

Tabell 3-10: Systematiske forskjeller i utvalgte yrker. Koblet AKU og AA. 2003. Prosentandeler og relative.

| | Andeler | | Syst. feil | | Andel | | Syst. feil | |
|--|---------|----------|------------|---------|------------|-----------|------------|--|
| | AKU | Register | Forskjell | Relativ | Reg.kalib. | Forskjell | Relativ | |
| 1210 ADMINISTRERENDE DIREKTØRER | 1.61 | 1.95 | -0.34 | -21 % | 1.83 | -0.34 | -23 % | |
| 1222 PRODUKSJONSDIREKTØRER INNEN OLJE- OG GASSUTVINNING, BERGVERKSDRIFT, INDUSTRI, KRAFT- OG VANNFORSYNING | 1.08 | 0.97 | 0.11 | 10 % | 0.88 | 0.08 | 8 % | |
| 2130 SYSTEMUTVIKLERE OG PROGRAMMERERE | 1.68 | 1.39 | 0.29 | 17 % | 1.26 | 0.26 | 17 % | |
| 3120 DATAINGENIØRER OG -TEKNIKERE | 1.19 | 1.02 | 0.17 | 14 % | 0.94 | 0.15 | 14 % | |
| 3310 GRUNNSKOLELÆRERE | 1.39 | 1.36 | 0.04 | 3 % | 1.50 | 0.04 | 3 % | |
| 3415 TEKNISKE OG KOMMERSIELLE SALGSREPRESENTANTER | 3.09 | 1.92 | 1.17 | 38 % | 1.79 | 1.15 | 39 % | |
| 3418 BANKFUNKSJONÆRER | 1.06 | 1.25 | -0.19 | -18 % | 1.39 | -0.23 | -20 % | |
| 3432 REVISORER (IKKE STATS-AUTORISERTE) OG REGNSKAPSFØRERE | 1.38 | 0.99 | 0.39 | 28 % | 1.09 | 0.44 | 29 % | |
| 4113 SEKRETÆRER | 1.48 | 1.91 | -0.43 | -29 % | 2.32 | -0.52 | -29 % | |
| 4114 KONTORMEDARBEIDERE | 1.84 | 2.52 | -0.69 | -37 % | 2.91 | -0.78 | -37 % | |
| 4121 ØKONOMIMEDARBEIDERE OG REVISJONSASSISTENTER | 1.34 | 1.21 | 0.12 | 9 % | 1.41 | 0.21 | 13 % | |
| 4131 LAGERMEDARBEIDERE OG MATERIALFORVALTERE | 1.67 | 2.13 | -0.46 | -28 % | 1.91 | -0.42 | -28 % | |
| 5122 KOKKER | 1.17 | 0.98 | 0.19 | 16 % | 1.04 | 0.19 | 15 % | |
| 5123 HOVMESTERE, SERVITØRER OG BARKEEPERE | 1.39 | 1.47 | -0.08 | -6 % | 1.49 | -0.06 | -4 % | |
| 5131 BARNE- OG UNGDOMSARBEIDERE O.L. | 1.75 | 1.58 | 0.17 | 10 % | 1.83 | 0.19 | 9 % | |
| 5132 OMSORGSARBEIDERE OG HJELPEPLEIERE | 1.54 | 1.64 | -0.10 | -6 % | 1.94 | -0.10 | -6 % | |
| 5139 ANNET PLEIE- OG OMSORGSPERSONALE | 1.28 | 1.32 | -0.04 | -3 % | 1.45 | -0.08 | -6 % | |
| 5221 BUTIKKMEDARBEIDERE O.L. | 10.65 | 11.76 | -1.11 | -10 % | 12.55 | -1.15 | -10 % | |
| 7125 TØMRERE | 1.97 | 1.53 | 0.44 | 22 % | 1.33 | 0.38 | 22 % | |
| 7134 RØRLEGGERE OG VVS-MONTØRER | 1.02 | 0.94 | 0.08 | 8 % | 0.80 | 0.07 | 8 % | |
| 7231 BILMEKANIKERE | 1.16 | 1.10 | 0.06 | 5 % | 0.94 | 0.05 | 5 % | |
| 7241 ELEKTRIKERE, ELEKTRONIKERE OL. | 1.99 | 1.72 | 0.27 | 13 % | 1.49 | 0.24 | 14 % | |
| 8322 BUSS- OG SPORVOGNFØRERE | 1.05 | 1.06 | -0.01 | -1 % | 0.94 | -0.01 | -1 % | |
| 8323 LASTEBIL- OG VOGNTOGFØRERE | 1.53 | 1.53 | 0.00 | 0 % | 1.33 | 0.00 | 0 % | |
| 9132 RENGJØRINGSPERSONALE I BEDRIFTER OL. | 2.41 | 2.67 | -0.26 | -11 % | 3.07 | -0.26 | -9 % | |

Det viser at det er store systematiske forskjeller, selv i de aller største yrkene som er vist her. Kalibreringen har ikke særlig positiv effekt. Grunnen til at vi ikke justerer med sektor, er at det er svært få og neppe representative innen kommunale sektor som har yrkeskode i register. Aggregert yrke viser også endel forskjeller. Tilfeldige feil er relativt små sammenliknet med de systematiske, som tabellen deretter viser.

Tabell 3-11: Systematiske feil på 1-siffer yrke. Koblet AKU og AA. 2003. Prosentandeler og relative.

| | Andeler | | Syst. feil | | Andel | | Syst. feil | |
|----------------------------|---------|----------|------------|---------|------------|-----------|------------|--|
| | AKU | Register | Forskjell | Relativ | Reg.kalib. | Forskjell | Relativ | |
| 9 Andre yrker | 5.25 | 5.71 | 6.19 | -18 % | 6.57 | -0.86 | -15 % | |
| 1 Lederyrker | 9.90 | 9.50 | 9.60 | 3 % | 9.20 | 0.30 | 3 % | |
| 2 Akademiske yrker | 7.51 | 7.34 | 7.23 | 4 % | 7.03 | 0.32 | 4 % | |
| 3 Høyskoleyrker | 19.93 | 20.18 | 17.06 | 14 % | 17.30 | 2.88 | 14 % | |
| 4 Kontoryrker | 8.79 | 9.66 | 9.76 | -11 % | 10.64 | -0.98 | -10 % | |
| 5 Salgs- og serviceyrker | 21.85 | 23.53 | 24.21 | -11 % | 25.89 | -2.36 | -10 % | |
| 6 Bønder, fiskere ol. | 1.01 | 0.96 | 1.10 | -9 % | 1.05 | -0.08 | -8 % | |
| 7 Håndverkere | 14.95 | 13.17 | 12.62 | 16 % | 11.06 | 2.11 | 16 % | |
| 8 Operatører, sjåførere ol | 10.81 | 9.94 | 12.24 | -13 % | 11.27 | -1.33 | -13 % | |

Tabell 3-12: Tilfeldige feil på 1-siffer yrke. Koblet AKU og AA. 2003. Prosentandeler og relative.

| | Andeler | | Systematiske feil | | Tilfeldige feil | |
|----------------------------|---------|----------|-------------------|---------|-----------------|---------|
| | AKU | Register | Feil | Relativ | Feil | Relativ |
| 9 Andre yrker | 4.83 | 5.67 | -0.84 | -17 % | 2.63 | 1 % |
| 1 Lederyrker | 9.01 | 9.44 | -0.43 | -5 % | 4.95 | 0 % |
| 2 Akademiske yrker | 7.42 | 7.50 | -0.08 | -1 % | 4.37 | 0 % |
| 3 Høyskoleyrker | 19.75 | 16.39 | 3.36 | 17 % | - | - |
| 4 Kontoryrker | 8.91 | 9.79 | -0.88 | -10 % | 4.43 | 0 % |
| 5 Salgs- og serviceyrker | 22.84 | 24.51 | -1.66 | -7 % | 3.68 | 0 % |
| 6 Bønder, fiskere ol. | 0.72 | 0.88 | -0.16 | -22 % | 1.44 | 2 % |
| 7 Håndverkere | 15.73 | 13.40 | 2.32 | 15 % | - | - |
| 8 Operatører, sjåførere ol | 10.78 | 12.42 | -1.63 | -15 % | 3.17 | 0 % |

3.6 Validitet

Validering av yrkeskode må prinsipielt baseres på informasjon om konkrete arbeidsoppgaver. I 2002, da store deler av arbeidstakerforholdene hadde levert yrkestittel som tekst, kunne dette operasjonalisere ved tittel og f.eks. nærings- og utdanningskode. For tiden leverer de fleste kode, og en mindre andel nye forhold leveres med tekst. Det er derfor i enda mindre grad mulig å gjøre en direkte validering av klassifisering i forhold til arbeidsoppgaver. Forsøk på validering av yrke i register må skje ved hjelp av en rekke indirekte variabler. I AKU har man tekst med både yrkestittel og arbeidsoppgaver, men det er så langt ikke gjort en uavhengig måling av dette. Noen koherenstester er referert i notat 2004/46.

3.7 Endringer

For å vurdere endringer i kvaliteten kjøres akkurat den samme analysen på data fra årene 2003 og 2004. Dette istedenfor å sammenlikne med tidligere publiserte kontroller, hvor det er brukt litt ulike utgangspunkt. Data som inngår er fra AKU 1.-3. kvartal 2003, det gir 41 832 records for ansatte, hvor utvelgelsen av siste record gir 19 095 personer.

Fra Arbeidstakerregisteret brukes periodefil fra uke 44-2004, som har 1 959 429 gyldige hovedarbeidstakerforhold for bosatte 16-74 år. Den koblede fila inneholder 16 402 eller omlag 85%, som er omtrent som forventet utfra andel *ansatte* som defineres som sysselsatte tilknyttet Arbeidstakerregisteret.

Tabell 3-13: Samsvar på yrkeskoder etter samsvar av næringskoder. Fra AKU/AA 2004.

| Antall 2004 | I alt | Uoppgitt | yrke 0 | yrke 1 | yrke 2 | yrke 3 | yrke 4 |
|------------------|--------|----------|--------|--------|--------|--------|--------|
| I alt | 16 402 | 6 286 | 3 001 | 439 | 423 | 567 | 5 686 |
| nace 0 | 1 082 | 400 | 378 | 70 | 36 | 40 | 158 |
| nace 1 | 323 | 128 | 95 | 25 | 5 | 9 | 61 |
| nace 2 | 786 | 459 | 93 | 7 | 17 | 27 | 183 |
| nace 3 | 439 | 234 | 62 | 12 | 3 | 21 | 107 |
| nace 4 | 288 | 191 | 34 | 3 | 5 | 1 | 54 |
| nace 5 | 13 484 | 4 874 | 2 339 | 322 | 357 | 469 | 5 123 |
| Andel 2004 | I alt | Uoppgitt | yrke 0 | yrke 1 | yrke 2 | yrke 3 | yrke 4 |
| I alt | 100 | 38 | 18 | 3 | 3 | 3 | 35 |
| nace 0 | 100 | 37 | 35 | 6 | 3 | 4 | 15 |
| nace 1 | 100 | 40 | 29 | 8 | 2 | 3 | 19 |
| nace 2 | 100 | 58 | 12 | 1 | 2 | 3 | 23 |
| nace 3 | 100 | 53 | 14 | 3 | 1 | 5 | 24 |
| nace 4 | 100 | 66 | 12 | 1 | 2 | 0 | 19 |
| nace 5 | 100 | 36 | 17 | 2 | 3 | 3 | 38 |
| Endring fra 2003 | I alt | Uoppgitt | yrke 0 | yrke 1 | yrke 2 | yrke 3 | yrke 4 |
| I alt | 0 | 0 | -1 | 0 | 0 | 0 | 1 |
| nace 0 | 0 | 2 | -1 | 0 | -1 | 1 | -1 |
| nace 1 | 0 | 0 | 7 | 1 | 0 | -3 | -3 |
| nace 2 | 0 | 6 | -3 | -1 | -2 | -1 | 0 |
| nace 3 | 0 | 8 | -4 | 1 | -1 | 1 | -5 |
| nace 4 | 0 | 7 | 0 | 1 | 0 | -3 | -6 |
| nace 5 | 0 | 0 | -1 | 0 | 0 | 0 | 0 |

Figur 3-14: Samsvar av yrkeskoder etter kjønn og alder. Fra AKU/AA 2003 og 2004.

| Antall 2004 | I alt | Uoppgitt | yrke 0 | yrke 1 | yrke 2 | yrke 3 | yrke 4 |
|------------------|--------|----------|--------|--------|--------|--------|--------|
| I alt | 13 484 | 4 874 | 2 339 | 322 | 357 | 469 | 5 123 |
| Menn | 7 189 | 1 949 | 1 550 | 237 | 252 | 316 | 2 885 |
| Kvinner | 6 295 | 2 925 | 789 | 85 | 105 | 153 | 2 238 |
| 16-19 år | 323 | 48 | 48 | 13 | 6 | 7 | 201 |
| 20-24 år | 936 | 204 | 117 | 13 | 17 | 32 | 553 |
| 25-39 år | 4 651 | 1 500 | 879 | 120 | 131 | 157 | 1 864 |
| 40-54 år | 5 023 | 2 014 | 882 | 117 | 134 | 187 | 1 689 |
| 55-66 år | 2 468 | 1 080 | 392 | 59 | 68 | 83 | 786 |
| 67-74 år | 83 | 28 | 21 | 0 | 1 | 3 | 30 |
| Andel 2004 | I alt | Uoppgitt | yrke 0 | yrke 1 | yrke 2 | yrke 3 | yrke 4 |
| I alt | 100 | 36 | 17 | 2 | 3 | 3 | 38 |
| Menn | 100 | 27 | 22 | 3 | 4 | 4 | 40 |
| Kvinner | 100 | 46 | 13 | 1 | 2 | 2 | 36 |
| 16-19 år | 100 | 15 | 15 | 4 | 2 | 2 | 62 |
| 20-24 år | 100 | 22 | 13 | 1 | 2 | 3 | 59 |
| 25-39 år | 100 | 32 | 19 | 3 | 3 | 3 | 40 |
| 40-54 år | 100 | 40 | 18 | 2 | 3 | 4 | 34 |
| 55-66 år | 100 | 44 | 16 | 2 | 3 | 3 | 32 |
| 67-74 år | 100 | 34 | 25 | 0 | 1 | 4 | 36 |
| Endring fra 2003 | I alt | Uoppgitt | yrke 0 | yrke 1 | yrke 2 | yrke 3 | yrke 4 |
| I alt | 0 | 0 | -1 | 0 | 0 | 0 | 0 |
| Menn | 0 | -1 | 0 | 0 | 1 | 0 | 1 |
| Kvinner | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 16-19 år | 0 | -1 | 0 | 0 | 1 | -1 | 0 |
| 20-24 år | 0 | 0 | 0 | -1 | 0 | 0 | 0 |
| 25-39 år | 0 | -1 | 0 | 0 | 0 | 0 | 1 |
| 40-54 år | 0 | -1 | 0 | 0 | 1 | 1 | 0 |
| 55-66 år | 0 | 4 | -1 | -1 | 0 | 0 | -1 |
| 67-74 år | 0 | 5 | 2 | -2 | -2 | -1 | -3 |

En må kunne si utfra disse undersøkelsene at hverken data eller kvalitet har endret seg stort, og at forskjellen i kvalitet mellom viktige kjennemerker er større. Også endringene i mikrokonsistens er små.

Tabell 3-15: Konsistens av yrkesfelt. Delutvalg av koblet AKU og AA. Endring fra 2003 til 2004.

| AKU | Register | | | | | | | | | | |
|---------------------------|--------------------|-------------|--------------|--------------------|-----------------|---------------|--------------------------|-----------------------|---------------|---------------------------|----|
| | I alt ⁹ | Andre yrker | 1 Lederyrker | 2 Akademiske yrker | 3 Høyskoleyrker | 4 Kontoryrker | 5 Salgs- og serviceyrker | 6 Bønder, fiskere ol. | 7 Håndverkere | 8 Operatører, sjåfører ol | |
| I alt | 0 | 0 | -1 | 1 | -1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 9 Andre yrker | 0 | 0 | 0 | 0 | -1 | -1 | -2 | 1 | 3 | 0 | 0 |
| 1 Lederyrker | 0 | 0 | 3 | 0 | -1 | 1 | -1 | -1 | -1 | -1 | -1 |
| 2 Akademiske yrker | 0 | 0 | 1 | 1 | 0 | -1 | -1 | 0 | 0 | 0 | 0 |
| 3 Høyskoleyrker | 0 | 0 | 0 | 1 | -1 | 1 | -1 | 0 | 0 | 0 | 0 |
| 4 Kontoryrker | 0 | 1 | -1 | 0 | 0 | -3 | 2 | 0 | 0 | 0 | 0 |
| 5 Salgs- og serviceyrker | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 Bønder, fiskere ol. | 0 | 2 | -4 | 3 | 0 | 0 | -1 | -2 | 3 | -2 | -2 |
| 7 Håndverkere | 0 | -1 | 0 | 0 | 0 | 0 | 0 | -1 | 2 | -1 | -1 |
| 8 Operatører, sjåfører ol | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | -1 | 3 |

3.8 Oppfølging

Som i tidligere undersøkelser finner vi at det er tildels betydelige forskjeller mellom yrkeskodingen i register og AKU. Det er også vanskelig å skille ut hva som skyldes inkonsistente data og rene feilkodinger. Det er allikevel nyttig å foreta samme type analyse med jevne mellomrom, for å vurdere endringer i kvaliteten. Dette fordi man kan betrakte AKUs yrkeskoding som konstant, fordi kodingen her skjer på samme måte og i stor grad de samme personer ved Seksjon for databehandling. I Arbeidstakerregisteret derimot skjer det flere endringer. Det utføres ulike tiltak for å bedre kvaliteten, samt at det skjer stadige endringer i datagrunnlaget ved at flere bruker koder istedenfor tekst, og flere leverer maskinelt istedenfor papirskjema. En særlig stor endring vil være når et stort antall arbeidstakerforhold med nåværende *stillingskoder* får oppgitt *yrkeskode*.

4 Innføringen av yrke i registerbasert sysselsettingsstatistikk

4.1 Bakgrunn

Dette skal gi en kort evaluering av arbeidet med å lage yrkesfordelingene i den publiserte statistikken og en oppsummering om kvaliteten. Yrke er nytt kjennemerke i den registerbasert sysselsettingsstatistikken fra og med statistikkåret 2003. Bakgrunnen for dette var innføringen av yrke i Arbeidstakerregisteret, som omfatter de fleste sysselsatte. Arbeidskraftundersøkelsen (AKU) publiserer yrke for alle sysselsatte og har gjort det i mange år. Dette er en meget stor spørreundersøkelse som gir detaljerte yrkesfordeling etter kjønn, og aggregerte yrkesfordelinger på f.eks. arbeidstid, fylke og alder. Allikevel er det behov for enda mer detaljert yrkesstatistikk. I tillegg til sykefraværstatistikk etter yrke, er det også etterspørsel etter nivå tall for flere detaljerte yrker og yrkesfordeling på kommunenivå. Det var derfor flere grunner til å få yrke med i registerbasert sysselsettingsstatistikk. Først i 2000 begynte innrapportering av yrke til Arbeidstakerregisteret så smått, og det har tatt sin tid å få tilstrekkelig med kvalitetssikrede yrkeskoder på plass. Det var meningen at Folketellingen 2001 skulle inneholde yrke, i likhet med tidligere folketellinger. Det lot seg ikke gjøre, men når yrkesdata i register er etablert vil dette inngå i framtidige tellinger.

4.2 Oppgave

Målet var å få med yrkesfordelinger i den registerbasert sysselsettingsstatistikken pr. 1. november 2003, som var planlagt publisert 11. juni 2004. Det måtte utvikles metoder for å gi slike fordelinger på alle registerbaserte sysselsatte. De omfatter mer enn bare yrkeskodene som er levert til Arbeidstakerregisteret: partielt frafall, altså sysselsatte fra Arbeidstakerregisterdata uten yrkeskode, og sysselsatte definert fra andre kilder uten yrkesdata.

Det ble etterhvert klart at sysselsatte i kommune/fylke¹ og helseforetak² ikke skulle være med i statistikkgrunnlaget. Dette skyldes at de har innrapporterte *stillingskoder* og mange av disse er for generelle til å kunne konverteres til yrkeskoder. Delpopulasjonen i statistikken er da privat sektor og staten unntatt helseforetakene (sykehus, m.m.).

¹ utfra kode for institusjonell sektor. Delmengden "de som bruker PAI-kode" er ikke helt den samme.

² utfra organisasjonsnummer, liste fra helsepersonellstatistikken.

4.3 Gjennomføring

Forutsetningene for opplegget:

- Det er ikke noe mål å lage yrkesstatistikk som allerede dekkes av AKU.
- Yrkesfordeling for alle sysselsatte lages direkte til mest detaljert nivå (4-siffer yrkeskode)
- På aggregert nivå er det aktuelt å lage kommunetall og eventuelt andre variabler som ikke gis i AKU.
- De aggregerte tallene skulle være konsistente med de detaljerte, altså ikke estimert direkte.
- Det brukes ikke kalibrering f.eks. med AKU.

Metodikken som ble utviklet er i hovedsak beskrevet i notat 2004/46, del 5. En kort oppsummering:

- Statistikken er antall sysselsatte.
- Estimator er gruppegjennomsnitt.
- Både imputering og prediksjon er deterministisk ikke-informativ SHG-modell.

Den praktiske gjennomføringen innebærer:

- Dataenheten er fødselsnummer, altså antall personer.
- Hver sysselsatt i delpopulasjonene får en komplett yrkesfordeling i form av en sannsynlighetsvektor. Hver record har 352 numeriske variabler, *ikke* en kategorivariabel med 352 mulige yrkeskoder.
- Yrkesfordelingen pr. person velges blant inntil 8 modeller innen hver delpopulasjon. Den fordelingen som velges er gruppegjennomsnittet fra den matchende gruppe som bedømmes som best. Dette skjer uavhengig av hvilken modell gruppa tilhører, selv om det selvsagt er en sammenheng mellom kvalitet og aggregeringsnivået for gruppa. Det er altså *ikke* en predeterminert modell og det brukes *ikke* regresjon.

En nærmere drøfting av det metodiske i opplegget finnes tilslutt i dette kapittelet.

4.3.1 Teknisk arbeid og publisering

Alle programmer er skrevet i SAS base v.8.02. Behandlingen av store registerfiler gjør at alt må kjøres på Unix (OSF1). Begrensede maskinresurser gjorde det nødvendig å skrive om endel prosesser til SQL istedenfor tradisjonelle "SSB-prosedyrer". For framtidige endringer er det verdt å merke seg at SQL-kompetanse er mer standardisert og utbredt enn det proprietære programmeringsspråket i SAS. I forbindelse med kontrollarbeidet ble det gjort videre analyse i Excel som antakelig bør gjøres i SAS for mer effektive rutiner.

Alle metoder og programmer ble testet på 2002-data, for å komme i gang med kontrollarbeidet i tide. Ulike estimater ble sammenliknet med bl.a. følgende kilder:

- Publisert AKU-statistikk.
- Egne estimater fra AKU for sammenliknbar delpopulasjon.
- Publisert helsepersonellstatistikk.
- Tabeller over ansatte etter stillingsbetegnelse i Statens sentrale tjenestemannsregister (SST).
- Tall fra visse organisasjoner, f.eks. Legeforeningen.
- Publiserte tabeller over sysselsatte etter næring, fra register og næringsstatistikken.

Når 2003-data var klare var det relativt kort tid igjen til ytterligere kontroller. På bakgrunn av dette anbefales at man som denne gang tester alt på fjorårets data, særlig ved endringer av metodene. Av hensyn til den øvrige statistikken kan man ikke regne med lengre tid fra data blir klare til publisering.

Standard for yrkesklassifisering har 4 nivåer med tallkoder. Den 7-sifrede koden for yrkestittel består av 4-siffer yrkeskode + et løpenummer for de ulike titlene innen hvert yrke. Det var ikke aktuelt å publisere 7-siffer yrkestittel. Det ble besluttet å aggregere noen 4-siffer yrkeskoder slik at publiseringsnivået er 289 grupper av 352 yrker. Begrunnelsen for dette er vurdering av klassifiseringen av enkelte yrker, og er nærmere forklart i et Web-skriv som ble publisert samtidig med statistikken.

Publiserte tabeller er *antall sysselsatte* i hele tall. Til sammenlikning bruker AKU 1000 og prosent. Se vedlegg for nærmere om kjennemerker for krysstabeller.

Ved sending til statistikkbanken måtte 1-siffer yrkesfelt gjøres om til 2-sifret tall, utfra deres spesifikasjoner. På oppfordring fra statistikkbanken ble også laget en kodeliste for de 289 yrkesgruppene. I publisering benevnes de imidlertid med yrkeskodene som inngår og ikke med denne koden.

4.4 Kvalitet

Vi forsøker her å oppsummere kvaliteten utfra et sett av vanlige kriterier for statistikkpublisering. Ikke alle kriterier har vært like nøye undersøkt.

4.4.1 Relevans

Yrke er etterspurt av mange brukere som offentlige etater, organisasjoner og media. Yrke skal avspeile konkrete arbeidsoppgaver, og har stor interesse for analyse av sykefravær, utdanning, lønn, helse og mange andre forhold. Yrke i register er særlig relevant for detaljert statistikk og levering av mikrodata til forskning.

4.4.2 Nøyaktighet

For hele delpopulasjonen regner vi med 10% tilfeldige feil utfra metoden alene. Fra undersøkelse av validiteten i Arbeidstakerregisteret har vi beregnet at omlag 7% er feilkodet. Den statistiske usikkerhet pga. partielt frafall her er ca. 2%. Presisjon være særlig lav for sysselsatte som er definert utfra LTO uten kobling til Arbeidstakerregisteret. Denne gruppa utgjør omlag 10% av alle sysselsatte.

Det kan være mulig med noe forbedring hvis en estimerer 1-siffer yrkesfelt direkte, men man mister konsistens med 4-siffer. Det kunne muligens også ved å estimere til de "noe aggregerte" 4-siffer yrkene (de 289 gruppene), men jeg tror ikke effekten står i forhold til arbeidet.

4.4.3 Aktualitet og punktlighet

Tabeller med yrkesfordeling og kommentarer ble publisert som en integrert del av registerbasert sysselsettingsstatistikk på det planlagte publiseringstidspunktet kl.10.00, 11.juni 2004. Referansetiden er første uke i november 2003, noe som gir en produksjonstid på over sju måneder. En er avhengig av eksterne etater og andre seksjoner for å kunne motta, koble, bearbeide og kontrollere datafiler fra diverse administrative registre.

Utfra AKU kan man vurdere årlige endringstall og vi kan si at yrkesstrukturen ikke endrer seg særlig raskt. Allikevel er det minst to forhold, som det kunne være interesse for, men som ikke er med i denne publiseringen:

- Sesongmessige endringer i visse yrker.
- Strømninger innen yrker og mellom yrker.

Analysen av dette kunne gjøres fra Arbeidstakerregisterdata som både har kortere produksjonstid og hvor det lages flere periodefiler i året.

4.4.4 Tilgjengelighet og klarhet

Publisert materiale er gratis tilgjengelig på Internett. En går utfra at de fleste brukere har tilgang til dette på en enkel måte. Det må bemerkes at det ligger mange statistikker under emne 06.01, så det er ikke så oversiktlig som man skulle ønske seg. Flere av statistikkene har også mange tabeller listet opp. Den største tabellen med yrke utgjør omtrent 6 A4-sider, og er nok i lengste laget for en webside. De mest detaljerte tabellene legges kun i statistikkbanken.

De fleste navn på yrkeskategoriene er forkortet eller skrevet om i forhold til betegnelsene i Standard for yrkesklassifisering (STYRK 1998). Allikevel må vi nok anta det er endel tilfeller der det fortsatt er vanskelig å forstå avgrensningen av arbeidsoppgavene utfra den valgte betegnelsen. Generelt kan vi vel si at mangfoldigheten og endringstakten i arbeidslivet gjør det vanskelig å lage yrkeskategorier som både er presise og intuitive. Det er viktig å fortsette arbeidet med å gjøre standarden bedre kjent, og jobbe for at viktige aktører benytter den.

4.4.5 Sammenliknbarhet

Siden yrkesfordelingen i registerbasert sysselsettingsstatistikk for en delpopulasjon er den ikke sammenliknbar med de publiserte AKU-tallene for alle sysselsatte. Fra og med 3-siffer nivå er yrkesfordelinger fra AKU sammenliknbare med nordiske og de fleste europeiske land. En registerbasert statistikk for samme delpopulasjon ville nok vært sammenliknbar på samme nivå. Siden 2003 var det første året, er ikke sammenlikning bakover mulig. For kontrollformål lages yrkesfordeling for samme delpopulasjon i 2004, altså en sammenlikning over tid.

4.4.6 Sammenheng

AKU er den viktigste kilden til offisiell yrkesstatistikk, og det er viktig å belyse sammenhengen mellom denne og registerbasert sysselsettingsstatistikk. Samsvarsanalyser med deler av registerdata og AKU-data som vist tidligere i notatet gir ca. 70% samsvar på det mest detaljerte nivået. Selv om det er samme definisjon av yrke og samme standard for klassifisering, er det noe ulikt innhold. Dette skyldes bl.a. ulikheter i:

- Data generelt: det er informasjon om endel sysselsatte som hentes fra diverse registre som gjør at de ikke er helt konsistent med opplysninger AKU.

- Yrkesdata: i AKU oppgir sysselsatte selv sin yrkestittel og arbeidsoppgaver, i register rapporteres kun yrkestittel og den føres av arbeidsgiver.
- Klassifisering – det er noe ulikheter i hvilke variabler som er tilgjengelige for koderne. Særlig gjelder dette konkrete arbeidsoppgaver. Det spørres uttrykkelig etter arbeidsoppgaver i AKU. I register består teksten som regel bare av yrkestittel, eventuelt med en ganske kort tilleggstekst.
- Koding – i AKU skjer all koding manuelt, mens for register er det overveiende koder og noe automatisk og manuell koding av tekst.

Samsvarsanalyser med data fra lønnsstatistikken gir omlag samme konsistens. Det er her samme kilde nemlig arbeidsgiver, men noe ulikheter i leverte data og tilgjengelige støttevariabler. Hvis det på sikt blir slik at lønnsstatistikken henter yrkeskodene fra Arbeidstakerregisteret, vil disse to statistikker få et konsistent kjennemerke for yrke. Analyser med AKU-data har vært gjennomført bare for noen få næringer.

I sykefraværstatistikken publiseres sykefraværspersent og dagsverk fordelt etter det samme detaljeringsnivå med data fra Arbeidstakerregisteret. På bakgrunn av bl.a. sykepengeregler kan en anta at særlig gruppa selvstendig næringsdrivende ha en annen karakteristikk. Så en kan si at yrke i sykefraværstatistikken er konsistent fordi det har samme kilde som for sysselsatte definert utfra Arbeidstakerregisteret, mens statistikkvariabelen ikke kan ekstrapoleres til alle sysselsatte.

4.5 Videre utvikling

Siden dette er et helt nytt opplegg er det muligheter for forbedringer selv innenfor de eksisterende forutsetninger og begrensninger. Det er absolutt på sin plass med en diskusjon av fordeler og ulemper med opplegget, ikke minst med tanke på forbedringer som kan bli gjennomført i god tid innen arbeidet med neste publisering starter. Diskusjonen tar utgangspunkt i at data og tilgjengelige resurser er omtrent de samme.

4.6 Kriterier

Man kan si at opplegget virket uvant på alle som kom i kontakt med det. Det er også slik at det har kommet en rekke forslag før, under og etter utviklingsarbeidet. For å vurdere opplegget og eventuelle endringer i en helhet, kan det være greit å klarlegge ulike typer av kriterier. Praktiske kriterier må ikke undervurderes, som datatekniske (data, programmer, maskiner). Arbeidstekniske (kunnskaper, ferdigheter, arbeidsoppgaver, rutiner) er burde være mer tilgjengelig for forandring, det samme kan man si om vaner og tradisjoner. Når det gjelder publisering av statistikk kan vi skille mellom direkte kvantifiserbare mål, som presisjon, konsistens, osv. og de mer ikke-kvantifiserbare mål, som relevans, klarhet osv. Videre kan en vurdere en teori utfra slike ting som sammenheng med ytre rammer og andre teorier, indre sammenheng og logikk, intuisjon, forklaringskraft, osv.

4.7 Anvendelse

Den nåværende anvendelsen av yrke i registerbasert sysselsettingsstatistikk brukes til bl.a.:

- Publisering av nivååll i den rutinemessige registerbasert sysselsettingsstatistikken.
- Oppdragsbasert statistikk.
- Mikrodata til forskning.
- Interne analyser.

Det er skrevet dataprogrammer som gir krysstabeller eller mikrodata etter behov, for å minimalisere praktisk ulemper med opplegget i forhold til dagens anvendelser. En kan selvsagt ikke se bort fra at framtidige anvendelser kan by på nye praktiske problemer.

4.8 Metoder og alternativer

I diskusjon om alternative estimeringsopplegg kan man drøfte metoder, modeller og data litt uavhengig av hverandre, selv om valgene selvsagt vil påvirke hverandre. Vi kan skille mellom noen hovedtyper av metoder:

- Imputering
 - Regelbasert
 - Modellbasert
 - Deterministisk
 - Stokastisk
- Prediksjon
 - Deterministisk
 - Stokastisk

Ved modellbasert imputering har man i prinsippet valget mellom "hot-deck" (samtidig fra samme kilde) og "cold-deck" (fra annen kilde eller samme enhet tidligere). Prediksjon er nødvendigvis "cold-deck" og modellbasert. Vi kommer nærmere tilbake til dette under *datakilder*.

Regelbasert imputering er deterministisk og krever en katalogisering som bygger på f.eks. skjønn, jus eller logikk. Ulempen med en skjønnsmessig modell er i første rekke resurser og til en viss grad validiteten i valgene. I yrkesopplegget brukes dette bare på vernepliktige i førstegangstjeneste.

Deterministisk betyr å gi en bestemt verdi, altså samme verdi ved å gjenta prosessen. Stokastisk innebærer at man velger en tilfeldig verdi innenfor et gitt sett av sannsynlige. Hvis du kjører opplegget flere ganger, får noen flere ulike koder. Det gjør at estimatet har en varians, altså en tilleggsusikkerhet. Fordelen med en stokastisk metode er at det kan gis en realistisk³ verdi. En modellbasert deterministisk kan benytte urealistiske⁴ verdier, som i det nåværende opplegget. Man kunne også dele ut realistiske verdier, men estimatet kan da bli ikke forventningsrett. I valget mellom de to sistnevnte typene må vi altså avveie forholdet mellom større varians eller større skjevhet.

Opplegget derimot gir koherente tall på alle nivåer og kryssende variabler. Det er nok uvant å operere med desimaltall, men tallene avrundes i tabellene. Det kan gi diskrepans hvis brukere av tabellene aggregerer videre fra flere tabeller. Utover dette er det også andre rene metodiske ulemper som omtales under *modeller*. Både metodene realistisk-deterministisk og stokastisk kan som sagt gi større metodiske ulemper som skjevhet og varians.

4.8.1 Modeller

Opplegget tar utgangspunkt i SHG⁵-modellen for partielt frafall. Det benyttes 24 ulike modeller med mange tusen grupper. En gruppe skal ha en mest mulig homogen yrkesstruktur, helst bare 1 yrke. For at ikke tilfeldige feil skal forsterkes, er det fastsatt en minimumstørrelse på gruppene. Det fører til at store yrker overestimeres, og at noen grupper blir heterogene.

Omfanget av heterogenitet, både antallet grupper og spredning av yrke, var en av de viktigste årsakene til å bruke frekvensdistribusjon istedenfor stokastisk eller regelbasert imputering med 1 kode pr. person. Avveining av hvilken modell som gir størst feil vil avhenge av bruken (detaljeringsgraden, kryssende variabler, osv.)

Heuristisk modell har blitt forslått bl.a. for selvstendig næringsdrivende. Det vil kreve betydelige resurser og høyt innslag av skjønn å lage en komplett katalog. F.eks. omkodningen av stillingskoder til yrkeskoder er en slik modell. Danmarks statistikk bruker dette i noen utstrekning på yrke i registerbasert sysselsettingsstatistikk.

Enda et forslag er regresjonsmodell, som vel ikke er mulig med dagens maskinkapasitet i forhold til det ønskede detaljeringsnivå.

Det er også grunn til å diskutere det forhold at det benyttes ikke-informativ modell, da frafallet godt kan være informativt. Årsaken er at enheten for levering er arbeidsgiver, og at yrkesdistribusjonen vil være spesifikk med hensyn til arbeidsgivers aktivitet.

4.8.2 Data

Nedenfor oppsummeres inndata i dagens opplegg. Inndeling i delpopulasjonene baseres på kilde til sysselsettingsstatus, som kan betraktes som 1. tilleggsvariabel. Det benyttes ikke lønn eller andre numeriske variabler (alder er gruppert).

| Mottaker | Datakilde | Variabler |
|----------------------------|------------------------|---------------------------------|
| Arbeidstakerregister (84%) | Arbeidstakerregisteret | Næring, Utdanning, Kjønn, Alder |
| LTO (9%) | AKU-ansatte | Næring, Utdanning, Kjønn, Alder |
| Selvstendige (7%) | AKU-selvstendige | Næring, Utdanning, Kjønn, Alder |
| Vernepliktige | Vernepliktsregister | Kilde til sysselsettingsstatus |

I tillegg til å ekskludere flere på mottagersiden, er det mange alternativer på giversiden. Det er klart at ved å ekskludere LTO-gruppa, ville den totale kvaliteten bli bedre. Hvis vi forutsetter at målet fortsatt er kompletthet, kan vi begrense diskusjonen til alternative givere. Det er flere giverdatakilder som har vært foreslått eller kunne tenkes.

³ En donorverdi, altså verdien fra en annen enhet

⁴ f.eks. en gjennomsnittsverdi fra en gruppe

⁵ Svarhomogene grupper, altså egentlig enheter med samme frafallsmodell. Her er det *personer med samme yrke*.

| Mulige data | Aktuelle variabler |
|--|---|
| Offentlige lønnsregister: PAL, SST, STS, MAR | Stillingskoder, hvorav mange uspesifikke. Tekst. |
| Statistisk sentralbyrås lønnstatistikk | Stillingskoder, yrkeskoder, andre koder, tekst. Lønn. ⁶ |
| Selvangivelsesregister | Lønn- og andre inntektsdata. |
| LTO | Lønn. Det er også koblet på diverse variabler fra tilleggsregistre. |
| Arbeidstakerregisteret | I tillegg til de som brukes: Bedriftsstørrelse, foretakstype, mm. |
| AKU | Andre variabler? |

| Datainnhenting | Forutsetninger |
|---|---|
| Direkte imputering ved kobling på mikronivå | <ol style="list-style-type: none"> Må koble samme arbeidsoppgaver.⁷ Mikroyrkesdata må være av bedre kvalitet. |
| Modellering fra andre datakilder | <ol style="list-style-type: none"> Giverpopulasjon må være korresponderende. Yrkesdata på gruppenivå må være av bedre kvalitet. |

AKU-data er benyttet som en skalerbar delpopulasjon av givere, altså modellering og ikke direkte imputering. Bruken av spørreundersøkellesdata har flere ulemper som utvalgsusikkerhet og andre feilkilder, og at mange små yrker ikke kommer med. På den annen side er AKU den eneste kilden med data om konkrete arbeidsoppgaver, og den eneste kilden med yrkesdata for selvstendig næringsdrivende.

Dataegenskapene og forutsetningene for å anvende nye kildedata som nevnt ovenfor kan sammenfattes slik:

- Giverkilden må inneholde yrkesdata eller nøkkel som kan kobles til yrkesdata.
- Nye yrkesdata må være bedre og/eller mer omfattende.
- Koblingen må skje på et relevant nivå. Krever en entydig nøkkel.

Det er viktig at kunnskap om og erfaring med de ulike registrene utnyttes i vurderingen av kildedata.

4.8.3 Kalibrering

Totaltallene for ansatte og selvstendig næringsdrivende kalibreres med AKU. Det er ikke aktuelt å kalibrere på mindre grupper, f.eks. yrkesfelt, selv om det foreligger betydelig skjevhet. Dette fordi det vil innebære å endre faktisk leverte yrkeskoder og skape større feil på mikronivå (frafallet er ikke stort nok). Det ble derfor valgt en annen strategi med hovedpunkter:

- Aggregere noen yrker som har overlappende koding, og endel småyrker. Det ble publisert på 289 grupper. Til sammenlikning publiseres det 105 yrker (av 352) i AKU.
- Publisere informasjon om ulikheter i innholdet.
- Ikke publisere så mange krysstabeller.

Det ble bestemt at de aggregerte tallene skulle være konsistente med de detaljerte, altså ikke estimert direkte. Hvis målet hadde vært bedre aggregerte tall, ville de med den aktuelle metoden vært estimert direkte.

4.9 Evaluering

For å gjennomføre endring av estimeringsmetoden må utbytte stå i forhold til resursinnsatsen. Det er flere utfordringer i å fastsette kriterier for f.eks. kvalitetsforbedring. Uansett metode er det selvsagt viktig at også alle ikke-kvantitative kvalitetskriterier vurderes. Det avgjørende er forholdet mellom total kvalitetsøkning og total resursbruk. Igjen er kunnskap om de ulike registrenes styrke og svakheter viktig i tillegg til rene beregninger.

Eksempler på kvantitative undersøkelser og forutsetninger:

Mikrokonsistens forutsetter minst to forhold:

1. Det eksisterer en uavhengig kilde med kjent yrkesvaliditet.
2. Denne kilden kan kobles på arbeidsoppgavenivå.

Ingen av disse kan oppfylles direkte, men operasjonaliseres utfra tilgjengelig variabler. Målingen selv blir da følsom for ikke-målte egenskaper. Eks. er kobling med AKU-data.

Makrosammenlikninger er enklere, men krever også at:

1. Delpopulasjonen er den samme (eller kjent og ikke-informativ skjevhet).
2. Innholdet i de kryssende variablene er likt.

Eks. sammenlikning med Helsepersonellstatistikken.

⁶ Utvalget inneholder omlag 640 000 yrkeskoder. Delpopulasjonen er privat sektor unntatt varehandel.

⁷ En vanlig operasjonalisering er forsøk på å identifisere jobben (sysselsettingsforholdet).

Interne statistiske målinger som presisjon og koherens er enklere i forhold til særegenheter i ulike datakilder, registeregenskaper, m.m.

5 Utvikling av metoder for kontroll av yrkeskodingen i Arbeidstakerregisteret

5.1 Oversikt

Kapittelet beskriver begynnelsen på kontroll- og revisjonsopplegg for yrke i Arbeidstakerregisteret. Dette er ment å danne grunnlag for videre kvalitetsarbeid, bl.a. et opplegg der Seksjon der Databearbeiding skal begynne med kontroll og revisjon, i tillegg til manuell yrkeskoding.

Hovedsakelig baserer kontrollmetodene seg på å sammenlikne yrkesfordelingen i bedrifter. En sammenlikner da frekvensfordelinger ved diverse statistiske metoder, uten å implisere validiteten i selve yrkeskodene. Premissene for dette er at det er mulig å karakterisere bedrifter med samme fordeling av arbeidsoppgaver. En slik karakteristikk må operasjonaliseres ved tilgjengelige variabler og vil derfor være følsom for andre forhold enn bare yrkeskodingen, f.eks. relasjonen mellom næringskoder og reell aktivitet. Det stiller derfor store krav til skjønn i kontrollene, og de som utfører dette har lang erfaring i koding av yrke og inngående kjennskap til yrkes- og næringsstandardene.

Innledningsvis gis endel nøkkeltall om status for yrke i Arbeidstakerregisteret og referanser om bedriftsstruktur mv. Tilslutt gis en rekke konkrete eksempler på beregninger for enkeltbedrifter fra aktuelle data med ulike metoder.

Noen av emnene kan eventuelt også være interessant for andre registerdata eller andre kilder med mange enheter og høyt detaljeringsnivå.

5.2 Status

5.2.1 Levering av yrkesdata

Arbeidstakerregisteret i uke 31-2004 inneholdt 2 413 008 gyldige arbeidstakerforhold for bosatte. Da mange har flere jobber merker vi oss at det utgjør 2 073 612 personer. Disse er ansatt ved 123 935 ulike foretak, som består av 163 947 ulike bedrifter. De er fordelt på 687 forskjellige næringskoder. Det partielle frafallet av yrke er like under 7%, som er nærmere en halvering fra 2003. Antall personer er beregnet ved å velge et arbeidstakerforhold pr. person, som regel hovedarbeidstakerforholdet. Andelen 'stillingskode' er for en stor del kommunale og andre som benytter såkalte PAI-koder. Det arbeides for at denne gruppa skal gå over til vanlige yrkeskoder. For statlige og maritime arbeidstakerforhold er det mulig at disse vil fortsette å levere egne stillingskoder. Vi kan konkludere med at for tiden har omlag tre av fem standard yrkeskode i Arbeidstakerregisteret.

Tabell 5-1: Status for levering av yrkesdata i Arbeidstakerregisteret uke 31-2004

| | Antall arbeidstakerforhold | % | Antall personer | % |
|---------------|----------------------------|-------|-----------------|-------|
| Frafall | 166 217 | 6.89 | 134 292 | 6.48 |
| Stillingskode | 763 824 | 31.65 | 636 607 | 30.70 |
| Yrkeskode | 1 482 967 | 61.46 | 1 302 713 | 62.82 |

5.2.2 Struktur

Både for å analysere frafall og yrkesfordelingen i bedrift og foretak, er det interessant å studere organisering og størrelsesforhold på enhetene. Næringskode og størrelse i form av antall ansatte er aktuelle variabler for sammenlikning av bedrifter. Andre forhold som utdanning og demografi blir ikke behandlet her. For å få en bakgrunn vises her endel aktuelle størrelsesfordelinger. "Ansatte" er beregnet ved å velge 1 record pr. person, som regel hovedarbeidstakerforhold.

Tabell 5-2: Fordeling av bedrift, foretak og næring etter antall ansatte. Arbeidstakerregisteret pr. uke 31-2004

| Bedrift | Antall | | Ansatte | |
|-----------|---------|-------|-----------|-------|
| I alt | 163 947 | 100 % | 2 073 612 | 100 % |
| 1-9 | 121 546 | 74 % | 403 914 | 19 % |
| 10-99 | 39 699 | 24 % | 995 063 | 48 % |
| 100-999 | 2 642 | 2 % | 548 113 | 26 % |
| 1000-9999 | 60 | 0 % | 126 522 | 6 % |

| Foretak | | | | |
|-----------|---------|-------|-----------|-------|
| I alt | 123 935 | 100 % | 2 073 612 | 100 % |
| 1-9 | 98 207 | 79 % | 305 407 | 15 % |
| 10-99 | 23 005 | 19 % | 549 674 | 27 % |
| 100-999 | 2 497 | 2 % | 657 442 | 32 % |
| 1000-9999 | 222 | 0 % | 496 495 | 24 % |
| 10000- | 4 | 0 % | 64 594 | 3 % |

| Næring | | | | |
|-----------|-----|-------|-----------|-------|
| I alt | 687 | 100 % | 2 073 612 | 100 % |
| 1-9 | 32 | 5 % | 131 | 0 % |
| 10-99 | 88 | 13 % | 3 806 | 0 % |
| 100-999 | 254 | 37 % | 110 646 | 5 % |
| 1000-9999 | 270 | 39 % | 867 174 | 42 % |
| 10000- | 43 | 6 % | 1 091 855 | 53 % |

De fleste bedrifter er små, men de fleste arbeidstakere jobber i mellomstore og litt større bedrifter. Mengden foretak tilsier at de fleste består av kun en bedrift. Når det gjelder detaljert gruppering av næring så har standarden 693 koder på dette nivå. De fleste arbeidstakerforholdene samles i de største næringene. Det er hele 120 koder hvor det nesten ikke inngår arbeidstakerforhold.

5.2.3 Partiell frafall av yrkeskode

For korthets skyld brukes **frafall** om både manglende innrapportering og innrapportering som ikke har blitt klassifisert (f.eks. yrkestitler som er for generelle til å yrkeskodes). Den neste tabellen viser i hvor stor grad bedrifter og foretak rapporterer en eller annen form for yrkesdata på sine ansatte, som yrkes- eller stillingskode, fordelt etter størrelsen på enheten.

Tabell 5-3: Lvert yrkesdata i prosent av ansatte, etter størrelse. Arbeidstakerregisteret pr. uke 40-2004

| Ansatte i bedrift | Antall bedrifter | Gj.sn. | Q1 | Median |
|-------------------|------------------|--------|-------|--------|
| I alt | 166 282 | 92 % | 100 % | 100 % |
| 0 - 9 | 123 489 | 91 % | 100 % | 100 % |
| 10 - 99 | 40 143 | 95 % | 95 % | 100 % |
| 100 - 999 | 2 595 | 94 % | 95 % | 99 % |
| 1 000 - 9 999 | 55 | 89 % | 98 % | 100 % |
| Ansatte i foretak | Antall foretak | Gj.sn. | Q1 | Median |
| I alt | 125 587 | 92 % | 100 % | 100 % |
| 0 - 9 | 99 746 | 91 % | 100 % | 100 % |
| 10 - 99 | 23 157 | 94 % | 93 % | 100 % |
| 100 - 999 | 2 456 | 95 % | 95 % | 99 % |
| 1 000 - 9 999 | 224 | 94 % | 95 % | 99 % |
| 10 000 - | 4 | 74 % | 49 % | 99 % |

5.2.4 Struktur

Med tanke på både rent frafall og analyser av yrkesfordelinger kan det være greit å vise noen organisasjonsmessige forhold som referanse. Selv om de aller fleste bedrifter er små, er det slik at de fleste arbeidstakerforhold befinner seg i middels til store enheter. Kun 15% av arbeidstakerforholdene er i foretak med færre enn 10 ansatte. Nesten tre av fire foretak består av kun en bedrift.

Tabell 5-4: Antall bedrifter pr. foretak, etter størrelse. Arbeidstakerregisteret pr. uke 31-2004

| Ansatte i foretak | I alt | 1 bedr. | 2-10 bedr. | 11-100 bedr. | over 101 bedr. |
|-------------------|---------|---------|------------|--------------|----------------|
| I alt | 165 163 | 120 244 | 18 811 | 22 833 | 3 275 |
| 0 - 9 | 101 378 | 100 050 | 1 328 | . | . |
| 10 - 99 | 31 781 | 19 540 | 11 142 | 1 099 | . |
| 100 - 999 | 22 564 | 648 | 6 114 | 15 410 | 392 |
| 1 000 - 9 999 | 8 509 | 6 | 227 | 6 211 | 2 065 |
| 10 000 - 99 999 | 931 | . | . | 113 | 818 |
| Ansatte i bedrift | I alt | 1 bedr. | 2-10 bedr. | 11-100 bedr. | over 101 bedr. |
| I alt | 165 163 | 120 244 | 18 811 | 22 833 | 3 275 |
| 0 - 9 | 122 602 | 100 050 | 10 178 | 11 062 | 1 312 |
| 10 - 99 | 39 912 | 19 540 | 7 817 | 10 701 | 1 854 |
| 100 - 999 | 2 593 | 648 | 806 | 1 033 | 106 |
| 1 000 - 9 999 | 56 | 6 | 10 | 37 | 3 |

Tabell 5-5: Antall enheter og ansatte, etter størrelse. Arbeidstakerregisteret pr. uke 31-2004

| Foretaksstørrelse | Antall foretak | Ansatte |
|-------------------|------------------|-----------|
| I alt | 126 377 | 2 065 772 |
| 0 - 9 | 100 663 | 313 854 |
| 10 - 99 | 23 017 | 546 001 |
| 100 - 999 | 2 473 | 650 453 |
| 1 000 - 9 999 | 220 | 492 301 |
| 10 000 - 99 999 | 4 | 63 163 |
| Bedriftsstørrelse | Antall bedrifter | Ansatte |
| I alt | 165 163 | 2 065 772 |
| 0 - 9 | 122 602 | 410 843 |
| 10 - 99 | 39 912 | 997 479 |
| 100 - 999 | 2 593 | 536 641 |
| 1 000 - 9 999 | 56 | 120 809 |

For å se nærmere på omfanget av frafall etter den administrative enheten foretak, kan vi vekte den gjennomsnittlige leveringsgraden med antall ansatte. Vi ser da at det totale frafallet er på under 7% og at det er enkelte foretak som står

for det største frafallet. Særlig tydelig er dette i de største foretakene, men også endel i de små. Vi kan godt si at dette er negativt med hensyn til statistikken (skjevhet og systematiske feil), men positivt i forhold til administrative tiltak (f.eks. kontakt med bedrifter, kontroll, m.m.)

Tabell 5-6: Leveringsgrad av yrkesdata, etter størrelse. Arbeidstakerregisteret pr. uke 31-2004

| | Antall foretak | Uvektet | Vektet med antall ansatte |
|-----------------|----------------|---------|------------------------------|
| | | Gj.sn. | Gj.sn. |
| I alt | 126 377 | 90.8 % | 93.8 % |
| 0 - 9 | 100 663 | 89.8 % | 92.1 % |
| 10 - 99 | 23 017 | 94.3 % | 94.2 % |
| 100 - 999 | 2 473 | 94.8 % | 95.0 % |
| 1 000 - 9 999 | 220 | 94.3 % | 94.4 % |
| 10 000 - 99 999 | 4 | 74.5 % | 82.3 % |

Vi ser til slutt på omfanget fordelt etter leveringsgrad. Grupperingen her er valgt utfra administrative vurderinger. Omlag 20 000 foretak har ikke levert det man kan kalle komplette yrkesdata.

Tabell 5-7: Omfanget av yrkesdataberapportering. Arbeidstakerregisteret pr. uke 31-2004

| Leveringsgrad | Foretak | Ansatte | Mangler yrkeskode |
|-----------------|---------|-----------|----------------------|
| I alt | 126 377 | 2 065 772 | 116 698 |
| A Over 90% | 105 875 | 1 777 384 | 27 601 |
| B Mellom 10-90% | 13 756 | 264 185 | 76 416 |
| C Under 10% | 6 746 | 24 203 | 12 681 |

5.2.5 Strømninger

Periodefilen fra Arbeidstakerregisteret for uke 35 ble ferdig i mellomtiden og vi har undersøkt strømninger i yrkeskodestatus fra uke 31 til uke 35, 2004. Det er viktig å etablere et nivå for endringer, og å overvåke innrapporteringen av yrkesdata av ulike typer. De viktigste faktorene er partielt frafall (manglende innrapportering) og stillingskoder fra offentlige register. Tabellen viser antall og prosent av overganger for personer. Det er beregnet ved å koble filer fra de to ukene på fødselsnummer, etter å ha valgt kun hovedarbeidsforhold for hver person (det er de arbeidstakerforhold som telles i registerbasert sysselsetningsstatistikk). De aller fleste av overgangene skyldes endring av jobb (nytt orgnr.), ikke endring av yrke innen et arbeidstakerforhold.

Tabell 5-8: Yrkesdata i Arbeidstakerregisteret fra uke 31-35 2004, etter hovedarbeidsforhold. Antall og prosent.

| | | Uke 35 | | | |
|--------------|---------------|-----------|---------|---------------|-----------|
| Uke 31 | | I alt | Frafall | Stillingskode | Yrkeskode |
| Koblet fil | I alt | 1 962 682 | 139 498 | 605 567 | 1 217 617 |
| | Frafall | 141 148 | 117 985 | 9 495 | 13 668 |
| | Stillingskode | 604 930 | 9 027 | 593 833 | 2 070 |
| | Yrkeskode | 1 216 604 | 12 486 | 2 239 | 1 201 879 |
| I begge uker | I alt | 1 918 551 | 116 155 | 596 787 | 1 205 609 |
| | Frafall | 117 089 | 114 714 | 715 | 1 660 |
| | Stillingskode | 596 415 | 512 | 593 833 | 2 070 |
| | Yrkeskode | 1 205 047 | 929 | 2 239 | 1 201 879 |
| Nye inn | | 22 662 | 1 874 | 8 780 | 12 008 |
| Gamle ut | Frafall | 1 397 | 1 397 | | |
| | Stillingskode | 8 515 | 8 515 | | |
| | Yrkeskode | 11 557 | 11 557 | | |

| | | Uke 35 | | | |
|--------------|---------------|---------|---------|---------------|-----------|
| Uke 31 | | I alt | Frafall | Stillingskode | Yrkeskode |
| Koblet fil | I alt | 100.0 % | 7.1 % | 30.9 % | 62.0 % |
| | Frafall | 100.0 % | 83.6 % | 6.7 % | 9.7 % |
| | Stillingskode | 100.0 % | 1.5 % | 98.2 % | 0.3 % |
| | Yrkeskode | 100.0 % | 1.0 % | 0.2 % | 98.8 % |
| I begge uker | I alt | 100.0 % | 6.1 % | 31.1 % | 62.8 % |
| | Frafall | 100.0 % | 98.0 % | 0.6 % | 1.4 % |
| | Stillingskode | 100.0 % | 0.1 % | 99.6 % | 0.3 % |
| | Yrkeskode | 100.0 % | 0.1 % | 0.2 % | 99.7 % |
| Nye inn | | 100.0 % | 8.3 % | 38.7 % | 53.0 % |

Det ser ut til at frafallet ikke endrer seg stort. Det er positivt at det er få som endrer seg fra 'yrkeskode' til 'manglende yrkeskode'. At mange av de nye har stillingskode er nok en tilfeldighet akkurat denne uka.

5.3 Metodikk for store feil

Først tar vi for oss store avvik i yrkeskodingen, som har størst betydning for statistikken. Her beskrives ulike metoder for å identifisere enheter som skal undersøkes nærmere i forhold til **leverte yrkeskoder**. Når det gjelder enheter hvor yrkeskode av ulike grunner mangler, behandles dette separat bl.a. ved å kontakte store foretak og samarbeid med Trygdeetaten om informasjon til mindre enheter.

5.3.1 Enhet

Avhengig av omfanget og type av tiltak kan man kontakte enheten, kode om automatisk eller sende til manuell kontroll. Valg av enhet vil være noe forskjellig avhengig av om vi skal undersøke manglende innrapportering av yrkesdata, eller selve yrkesfordelingen. Rapportering til Trygdeetaten kan foregå på ulike måter (papirskjema eller diskett) og ulike nivåer:

- Bedrift, f.eks. enbedriftsforetak.
- Foretak, som leverer for alle sine bedrifter.
- Lønnssentral (felles rapportering for mange små bedrifter).

De administrative og juridiske enhetsnivåene har som nevnt betydning for leveringsmetoder og leveringsgrad. Men i arbeidet med kontrollere de leverte yrkeskodene, er det naturlig å velge enheten *bedrift*. Dette er den minste organisasjonsdelen som er enkelt å identifisere og som skal ha en ensartet primæraktivitet. Derfor blir bedriften utgangspunktet for å modellere yrkesfordelinger.

5.3.2 Variabler

Målet er å undersøke yrkesstruktur i *liknende bedrifter*. For å operasjonalisere denne likheten og gruppere bedriftene ser vi først på innholdet i tre viktige kjennemerker.

| | |
|---|--|
| <p>a) Yrke er definert som grupper av arbeidsoppgaver:</p> <ul style="list-style-type: none"> • Kunnskaper, ferdigheter, kompetansenivå. • Verktøy, maskiner, materialer som brukes. • Hvilke varer eller tjenester som produseres. | <p>b) Næring er definert som grupper av aktiviteter:</p> <ul style="list-style-type: none"> • Hvilke råvarer, m.m. brukes. • Hvilke produksjonsprosesser utføres. • Hvilke varer eller tjenester produseres. |
|---|--|

Både på handlings- og resultatsiden er det felles domener, selv om Standard for yrkesklassifisering legger vekt på uavhengighet. En kan tolke denne uavhengigheten som 'ikke direkte kobling'. Særlig uavhengigheten i forhold til utdanning har vært viktig for å understreke betydning av realkompetanse i motsetning til formell utdanning (utdanningskode fra register). Endel yrker vil være generelle i den betydning at arbeidsoppgavene er de samme uavhengig av næring. Men for mange yrker henger arbeidsoppgavene mer spesifikt sammen med prosess og produkt, slik at bedriftens næring i stor grad kan modellere bedriftens yrkesstruktur. Det er altså ikke snakk om å klassifisere yrket til enkeltpersoner, men å vurdere bedriften som helhet.

| |
|---|
| <p>c) Bedriftens størrelse og organisering som f.eks. organisasjonsform og foretaksstruktur, er ikke standardisert på samme måte og er mindre spesifikke med hensyn til yrkesstruktur. Allikevel hevder vi at antall ansatte vil ha betydning:</p> <ul style="list-style-type: none"> • Antall ledere og spesialisering av ledelsesfunksjoner øker med antall ansatte. Eks.: økonomidirektør. • Omfanget og type av hjelpeaktiviteter er en viktig komponent i yrkesstruktur. Eks.: transportavdeling. |
|---|

5.3.3 Størrelsesinndeling

For revidering ønsker en å finne de som er feilkodet og som derfor skiller seg ut. Men selv om en enhet skiller seg ut, behøver den ikke være feil. Dette kan kalles 'falske positive', og for effektiv revisjon må disse være så få som mulig. Det er ikke urimelig å forvente større yrkesvariasjon når bedriftene er svært ulike i størrelse. Variansen i antall ansatte innen en gruppe bør derfor være minimal. Dette må balanseres mot at gruppestørrelsen ikke bør være for liten. Når vi ser på størrelsesfordelingene finner en at det bør være ulike grenser næringer. En felles størrelsesinndeling vil gi endel tomme og neste tomme strata, og noen som blir urimelig store. Vi bruker i hovedsak median og kvartiler for å få noenlunde jevnt fordeling bedrifter i hver gruppe innen hver næring. Det er viktig å ha tilstrekkelig antall bedrifter å sammenlikne med. En ulempe er at noen grupper kan bli lite homogene, altså så store forskjeller i størrelse at yrkesstrukturen blir mindre sammenliknbar.

Det kan nevnes at i sykefraværstatistikken bedriftsundersøkelse deles det inn i ulike størrelsesgrupperinger, optimalisert på sykefraværshraten. Her har vi ikke en tilsvarende egnet tilleggsvariabel, slik at vi ikke kan nyttiggjøre oss disse erfaringer. Vi må derfor forsøke oss litt fram for å finne en egnet inndeling. Tabellene viser noen hovedtall på nærings- og bedriftsstørrelse i utvalgte næringer med mange ansatte og/eller mange bedrifter. For selve stratifiseringen, vises til vedlegg 1. Tellende arbeidstakerforhold er gyldige hoved- og biarbeidsforhold for bosatte personer.

Tabell 5-9: Antall bedrifter og ansatte i utvalgte næringer. Arbeidstakerregisteret uke 31-2004.

| Næringer | Antall bedrifter | Antall ansatte | Gj.sn. ansatte |
|---|------------------|----------------|----------------|
| 01.220 Saue- og geitehold. Oppdrett av hester | 1 732 | 2 729 | 1.6 |
| 11.100 Utvinning av råolje og naturgass | 48 | 16 410 | 341.9 |
| 11.200 Tjenester tilknyttet olje- og gassutvinning | 209 | 11 988 | 57.4 |
| 22.120 Forlegging av aviser | 351 | 13 808 | 39.3 |
| 35.114 Bygging og reparasjon av oljeplattformer og moduler | 70 | 12 303 | 175.8 |
| 45.211 Oppføring av bygninger | 4 018 | 37 953 | 9.4 |
| 45.230 Bygging av veier, flyplasser og idrettsanlegg | 639 | 15 361 | 24.0 |
| 45.310 Elektrisk installasjonsarbeid | 1 874 | 25 498 | 13.6 |
| 45.330 VVS-arbeid | 1 778 | 13 596 | 7.6 |
| 50.102 Detaljhandel med motorvogner, unntatt motorsykler | 1 280 | 11 554 | 9.0 |
| 50.200 Vedlikehold og reparasjon av motorvogner, unntatt motorsykler | 2 332 | 17 961 | 7.7 |
| 50.500 Detaljhandel med drivstoff til motorvogner | 1 456 | 14 607 | 10.0 |
| 52.110 Butikkhandel med bredt vareutvalg med hovedvekt på nærings- og nytelsesmidler | 5 380 | 61 356 | 11.4 |
| 52.420 Butikkhandel med klær | 3 769 | 21 539 | 5.7 |
| 55.101 Drift av hoteller, pensjonater og moteller med restaurant | 972 | 21 112 | 21.7 |
| 55.301 Drift av restauranter og kafeer | 3 885 | 34 192 | 8.8 |
| 60.211 Rutebiltransport | 236 | 13 028 | 55.2 |
| 60.220 Drosjebiltransport | 2 751 | 6 610 | 2.4 |
| 60.240 Godstransport på vei | 3 487 | 19 849 | 5.7 |
| 61.101 Utenriks sjøfart | 316 | 13 100 | 41.5 |
| 64.110 Posttjenester | 644 | 22 181 | 34.4 |
| 65.120 Bankvirksomhet ellers | 1 089 | 25 028 | 23.0 |
| 70.202 Utleie av egen fast eiendom ellers | 2 861 | 8 577 | 3.0 |
| 72.220 Annen konsulentvirksomhet tilknyttet system- og programvare | 2 005 | 18 163 | 9.1 |
| 74.121 Regnskap og bokføring | 2 218 | 10 676 | 4.8 |
| 74.140 Bedriftsrådgiving | 1 793 | 6 543 | 3.6 |
| 74.209 Annen teknisk konsulentvirksomhet | 1 792 | 12 371 | 6.9 |
| 74.502 Utleie av arbeidskraft | 433 | 20 779 | 48.0 |
| 74.700 Rengjøringsvirksomhet | 1 097 | 19 122 | 17.4 |
| 75.110 Generell (overordnet) offentlig administrasjon og økonomiforvaltning | 842 | 32 625 | 38.7 |
| 75.120 Offentlig administrasjon tilknyttet helsestell, sosial virks., undervisning, kirke, kultur | 1 282 | 19 929 | 15.5 |
| 75.130 Offentlig administrasjon tilknyttet næringsvirksomhet og arbeidsmarked | 980 | 16 697 | 17.0 |
| 75.140 Hjelpetjenester for offentlig administrasjon | 518 | 19 309 | 37.3 |
| 75.220 Forsvar | 141 | 23 870 | 169.3 |
| 75.240 Politi- og påtalemyndighet | 465 | 14 411 | 31.0 |
| 80.102 Grunnskoleundervisning | 3 251 | 95 665 | 29.4 |
| 80.210 Undervisning i allmennfag | 224 | 14 277 | 63.7 |
| 80.220 Undervisning i tekniske og andre yrkesrettede fag | 430 | 22 840 | 53.1 |
| 80.301 Undervisning ved universiteter | 35 | 17 397 | 497.1 |
| 80.302 Undervisning ved statlige høgschooler | 126 | 9 850 | 78.2 |
| 85.111 Alminnelige somatiske sykehus | 92 | 81 464 | 885.5 |
| 85.116 Institusjoner for voksenpsykiatri | 148 | 16 949 | 114.5 |
| 85.118 Somatiske sykehjem | 796 | 73 071 | 91.8 |
| 85.121 Allmenn legetjeneste | 1 727 | 9 474 | 5.5 |
| 85.130 Tannhelsestjenester | 2 156 | 7 281 | 3.4 |
| 85.141 Hjemmesykepleie | 136 | 10 140 | 74.6 |
| 85.313 Omsorgsinstitusjoner for eldre og funksjonshemmede | 675 | 22 630 | 33.5 |
| 85.321 Hjemmehjelp | 615 | 40 551 | 65.9 |
| 85.327 Barnehager | 5 320 | 56 291 | 10.6 |
| 91.310 Religiøse organisasjoner | 1 906 | 10 204 | 5.4 |
| 93.020 Frisering og annen skjønnhetspleie | 2 242 | 9 739 | 4.3 |

Det må bemerkes at næringer som omfatter 75 (off.adm), 80 (undervisning) og 85 (helse- og sos.) er svært store. Stillingskoder dominerer yrkesdata fra disse grupper, og vil derfor ikke inngå i grunnlaget for de revisjonsmetoder som er omtalt her. Det betyr at uansett størrelsen på tiltak i denne omgang vil de ikke ha overveldende betydning for den totale yrkesstatistikken. Men behovet for yrke i Arbeidstakerregisteret er først og fremst på et detaljert nivå, og det er viktig å kvalitetssikre små grupper. Makrotallene er jo allerede godt ivaretatt i AKU. Interessen for og betydning av enkeltyrker er ikke bare avhengig av antall ansatte.

Yrke i Arbeidstakerregisteret skal ha 7-siffer kode. All revisjon av yrke tar sikte på å gi en mer adekvat 7-siffer kode. Men i metodene som gjennomgås her, analyseres det på 4-siffer nivå (standardnivå). Konkret revisjon vil bare bli foretatt dersom det er snakk om endring på 4-siffer nivå. Det vil da gis en passende eller typisk 7-siffer kode utfra den reviderte 4-siffer koden.

Næring brukes i hele notatet om 5-siffer kode (næringsundergruppe). For å kunne modellere yrke er man avhengig av så detaljert næringsbeskrivelse som mulig. Eksempelvis vil vi forvente helt forskjellig yrkesfordeling i gruppene

"45.441 Malerarbeid" og "45.442 Glassarbeid". Hvis vi aggregerte til 4 siffer, ville viktig informasjon om yrke gå tapt. Et annet tydelig eksempel er: "74.871 Inkassovirksomhet o.l." og "74.875 Modellbyråvirksomhet".

5.3.4 Gruppering

I den grad oppdeling og næringskoding i register er korrekt, tenker vi oss at jevnstore bedrifter i samme næring vil ha en noenlunde ensartet aktivitet. Ut fra disse betraktninger ønsker vi å gruppere bedriftene etter næring og størrelse. Vi kan definere operasjonelt ut fra tilgjengelige data:

- (5-1) **Arbeidstakerforhold** = records i Arbeidstakerregisteret som gjelder gyldige forhold og bosatte personer.
- (5-2) **Bedrift** = arbeidstakerforhold med samme organisasjonsnummer bedrift.
- (5-3) **Næring** = prioritert næringskode, fem siffer "p_nace".
- (5-4) **Bedriftsstørrelse** = antall arbeidstakerforhold pr. bedrift. Beregnet tall, ikke ansatte-tall fra enhetsregister eller bedriftsregister.
- (5-5) **Yrke** = fire siffer av den leverte yrkeskoden. Fra "yrk_kode", yrkesvariabelen som ligger i Arbeidstakerregisteret, tar altså ikke med omkodede eller imputerede.
- (5-6) **Gruppe** = En mengde bedrifter med samme næringskode og samme størrelsesgruppe.

5.3.5 Populasjon

Uten at vi kjenner til en optimal avgrensning velges i første omgang å se på bedrifter som oppfyller visse tilleggskrav til antall ansatte, leveringsgrad av yrkeskode på sine ansatte og at det må være tilstrekkelig mange bedrifter å sammenlikne med. Bedrifter som har store mangler i innrapporteringen, vil det uansett være aktuelt å kontakte – hvertfall de som er av en viss størrelse. Men det er greit å holde dem utenfor disse beregningene, pga. skjevt og antakelig informativt frafall. I vedleggene vises tabell med stratifisering etter detaljert næring og størrelsesgruppe.

Hvi vi velger bedrifter med minst 50 ansatte i næringer som har minst 10 bedrifter av den størrelsen som har levert minst 90% yrkeskoder (maksimalt 10% partielt frafall av yrke) får vi en delpopulasjon på 2 253 bedrifter med 295 974 ansatte, eller vel 14% av Arbeidstakerregisteret. Grensene vil bli justert etter hvert som vi får mer empiri. Uansett om det i starten kan bli noe beskjedne andel går vi allikevel grundig gjennom disse fordi det er viktig å starte på et revisjonsopplegg, som kan videreutvikles til å omfatte større deler. Det er interessant å etablere "normalverdier" for yrkesfordelingen i mange viktige næringer og kunnskap om yrkesstruktur, organisering osv., er viktig for andre analyser, f.eks. sykefravær.

Det er også på sin plass å nevne uttrykkelig hvem som faller utenfor i denne omgang, nemlig offentlig sektor og småbedrifter generelt. Det vi kan kalle tilfeldige feil (småfeil), behandles metodisk lengre ut i notatet, men vil ikke bli kontrollert i første del av arbeidet.

5.3.6 Avviksberegninger

En gruppe skal nå bestå av bedrifter med samme næring og omtrent samme størrelse. Vi skal forsøke å identifisere bedrifter med feil i yrkeskodingen ved å måle avvik i yrkesfordeling. Vi kan da diskutere noen egenskaper ved slike fordelinger. Siden alle bedriftene i en gruppe bidrar til gruppas samlede yrkesfordeling, vil ethvert yrke i en bedrift finnes i gruppa. Omvendt kan det imidlertid være yrker i gruppas struktur som ikke forekommer i en gitt bedrift. Dette er en logisk asymmetri. Vi kan også gjøre vurderinger som er asymmetriske: Det er av mindre betydning om et mindre yrke ikke forekommer i enkeltbedrift, selv om det finnes hos de fleste i gruppa. Det kan være av større betydning med en unormalt stor andel av et yrke som ikke er vanlig i resten av gruppa.

Vi forsøker en strategi der vi beregner flere ulike måltall for hver bedrift, for å identifisere ulike typer av avvik og vurdere hvor godt egnet modellene er til dette. Etter forsøk med en rekke ulike beregninger presenteres noen typer som kan være egnet for å identifisere de viktigste formene for avvik.

5.3.6.1 Definisjoner

Betegner:

(5-7) X_b antall hovedarbeidstakerforhold tilsammen i bedrift b

(5-8) $Y_{u,b}$ antall hovedarbeidstakerforhold med yrke u i bedrift b

Yrkesfordeling i bedrift b kan formelt skrives:

$$(5-9) \quad \bar{p}_b = \left[\frac{Y_{b,1}}{X_b}, \frac{Y_{b,2}}{X_b}, \dots, \frac{Y_{b,U}}{X_b} \right]$$

Tilsvarende betegnes:

(5-10) X_g antall hovedarbeidstakerforhold tilsammen i gruppe g

(5-11) $Y_{u,g}$ antall hovedarbeidstakerforhold med yrke u i gruppe g

Yrkesfordeling i gruppe g kan formelt skrives

$$(5-12) \quad \bar{p}_g = \left[\frac{Y_{g,1}}{X_g}, \frac{Y_{g,2}}{X_g}, \dots, \frac{Y_{g,U}}{X_g} \right]$$

Skriver andel for et enkelt yrke u i bedrift b

$$(5-13) \quad p_{u,b} = \frac{Y_{u,b}}{X_b}$$

tilsvarende for gruppe g :

$$(5-14) \quad p_{u,g} = \frac{Y_{u,g}}{X_g}$$

Siden bedriftene i gruppa g presumptivt skal være like med hensyn til aktiviteter, kunne vi anta forventet yrkesfordeling i bedriften:

$$(5-15) \quad \hat{p}_{u,b} = p_{u,g}$$

5.3.6.2 Målinger

Avviket mellom det forventede og det observerte kan da måles ved differansen mellom andelene av et yrke u i en bedrift og gruppa g :

$$(5-16) \quad q_{u,b} = p_{u,b} - \hat{p}_{u,b} = p_{u,b} - p_{u,g}$$

Dette kan brukes direkte for å identifisere enkelte yrkeskoder hvor en bedrift har et stort avvik. Vi er også interessert i å finne de bedrifter som generelt er mest avvikende, og må finne et samlet tall pr. bedrift.

Siden dette er avvik av andeler, så vil generelt $\sum_{u=1}^U q_u = 0$

Derimot vil den vanlige variansen til avvikene i en bedrift øke jo mer "sær" yrkesfordelingen i bedriften er:

$$(5-17) \quad v_b = \text{var}(q_1, q_2, \dots, q_U) = \sum_{u=1}^U \frac{(q_{u,b} - \bar{q}_u)^2}{U-1}$$

Siden generelt $\bar{q} = 0$ kunne vi skrevet enklere:

$$(5-18) \quad v_b = \frac{\sum_{u=1}^U (q_{u,b})^2}{U-1}$$

I praktisk programmering bruker vi den vanlige variansfunksjonen. Det vanlige antall frihetsgrader $U-1$ kan bli litt misvisende, siden det i mange grupper er slik at de fleste yrker har $p=0\%$. Et alternativ er å kalle nevneren U^*-1 , der U^* er antall yrker hvor $p>0\%$ i gruppa, altså frihetsgraden til antall aktuelle yrker. Dette målet øker på samme måten med hvor mange og store avvik bedriften har, men forsterkes litt mer i grupper hvor det er få yrker å velge mellom.

For også å få samme enhet velges standardavviket som et generelt (basis) mål:

$$(5-19) \quad s_b^I = \sqrt{v_b}$$

I visse tilfeller kan det være viktigere å legge vekt på de største avvikene og hvor det er mange andre småyrker som ikke spiller så stor rolle for det totale kvaliteten.

For å finne bedrifter som har store avvik:

$$(5-20) \quad s_b^{II} = \sum_{u=1}^U (q_{u,b})^4$$

Vi så at variansen av andelene er omtrent som kvadratsummen (men avhengig av antall yrker). Fjerdepotens av skjevhetene gir en enda brattere skala, og vi kan si at dette øker betydning av de største yrkene og demper virkningen av smårask. Vi antar dette er egnet for grupper hvor ett eller noen få yrker er dominerende, eller hvor man ønsker å prioritere de største avvikene.

I andre bedriftsgrupper er yrkesfordelingen mer jevn og man kan ikke si at enkelte yrker dominerer. Det kan også være tilfeller der vi ønsker å framheve små men viktige yrker. Den neste beregningen er egnet for å finne bedrifter som har mange eller store avvik også i små yrker. Dette gir særlig utslag i grupper der det er mange og varierte yrker. Man kan si at det vurderer alle yrkene i gruppa som like viktige. Måltallet vi forsøker er summen av absoluttverdiene av avvik:

$$(5-21) \quad s_b^{III} = \sum_{u=1}^U |q_{u,b}|$$

En metode som tilsynelatende er helt forskjellig, men som indikerer noe av det samme, er å ta utgangspunkt i beregning av forventet antall ansatte i et yrke i bedrift:

$$(5-22) \quad \hat{Y}_{u,b} = p_{u,g} \cdot X_b$$

Avviket (nå i *antall ansatte*) blir:

$$(5-23) \quad \delta_{u,b} = Y_{u,b} - \hat{Y}_{u,b} = Y_{u,b} - p_{u,b} \cdot X_b$$

Aggregerer pr. bedrift et tall som kan tolkes som "antall feilplasserte ansatte":

$$(5-24) \quad s_b^{IV} = \frac{\sum_{y=1}^u |\delta_{y,b}|}{2}$$

Dette målet vil likne på det siste i forrige avsnitt, men er på en måte vektet med størrelsen på bedriften. Det betyr at man kan prioritere forhold som er mest utslagsgivende på statistikken, ved å ta de største feilene først.

Vi har nå fem ulike modeller og vil gi eksempler på hvordan dette slår ut for endel konkrete bedrifter og næringer. For oversikten skyld når vi skal sammenlikne modellene, ønsker vi å samordne alle måltallene. I utgangspunktet er en

vanlig standardisering $z_b = \frac{s_b - \mu(s_b)}{\sigma(s_b)}$. Dette vil gi z 'ene gjennomsnitt=0 og standardavvik=1 og positive verdier angir

avvik over gjennomsnittet. Men målene som er foreslått er såpass ulike og har så forskjellig skala at utvelging av de "særeste" bedriftene ikke baseres på felles grenser. Vi går derfor utfra at grenser for akseptabelt avvik i yrkesfordeling må fastsettes empirisk istedenfor å transformere selve målene. For kontroll og eksempler her brukes en enkel nonparametrisk metode ved å rangere bedriftene etter hvert av de ulike målene (synkende). Vi sammenlikner så de ulike rangeringene. Formelt kan vi skrive rangen R til bedriftene b i gruppa g sortert etter et måltall s av typen m :

$$(5-25) \quad R(s^m) = \{b_1, b_2, \dots, b_i, b_{i+1}, \dots, b_{n,g} \mid s_{b,i}^m > s_{b,i+1}^m \mid \forall b \in g\}$$

5.3.7 Eksempler

Før vi ser på endel eksempler, oppsummerer de ulike metodene og innfører noen forkortelser til tabellene. Bemerk også at vi oppgir *rangeringen* til bedriften istedenfor selve måltallene, for å gjøre tabellene mer leselige. "Type 1-5" betegner rangeringen av bedriften i henhold til de ulike aggregeringene, rang=1 er mest avvikende.

Tabell 5-10: Oversikt over de ulike aggregeringsmetodene.

| | |
|--------|---|
| Type 1 | (5-19) $s_b^I = \sqrt{v_b}$ |
| Type 2 | (5-21) $s_b^{III} = \sum_{u=1}^U q_{u,b} $ |
| Type 3 | (5-20) $s_b^{II} = \sum_{u=1}^U (q_{u,b})^4$ |
| Type 4 | (5-24) $s_b^{IV} = \frac{\sum_{y=1}^u \delta_{y,b} }{2}$ |

5.3.7.1 Eksempel: Butikk

Tabellene viser andelen av de største (>1%) yrkene innen grupper av de største (5-11) og minste (5-12) bedriftene i næringen "52.110 Butikkhandel med bredt vareutvalg med hovedvekt på nærings- og nytelsesmidler".

Tabell 5-11: Eksempel på yrkesstruktur i bedrifter og avvik fra gruppegjennomsnitt.

| | 1210 Adm. Dir. | 1224 Avd.sjef | 5221 Butikkmedarb. | 9133 Kjøkk.ass. | type_1 | type_2 | type_3 | type_4 |
|---------|----------------|---------------|--------------------|-----------------|--------|--------|--------|--------|
| Gruppe | 2 % | 1 % | 84 % | 1 % | - | - | - | - |
| Bedrift | 2 % | 29 % | 0 % | 0 % | 1 | 1 | 1 | 2 |
| Bedrift | 25 % | 4 % | 0 % | 0 % | 2 | 2 | 2 | 1 |
| Bedrift | 3 % | 9 % | 6 % | 0 % | 3 | 3 | 3 | 4 |
| Bedrift | 3 % | 13 % | 40 % | 1 % | 4 | 4 | 5 | 3 |
| Bedrift | 16 % | 1 % | 42 % | 0 % | 5 | 5 | 4 | 5 |
| Bedrift | 7 % | 2 % | 54 % | 2 % | 6 | 6 | 6 | 10 |
| Bedrift | 0 % | 1 % | 75 % | 1 % | 30 | 47 | 8 | 6 |
| Bedrift | 0 % | 1 % | 75 % | 8 % | 35 | 51 | 9 | 7 |
| Bedrift | 0 % | 0 % | 81 % | 0 % | 68 | 71 | 41 | 8 |
| Bedrift | 1 % | 1 % | 85 % | 2 % | 76 | 75 | 77 | 9 |

Tabell 5-12: Eksempel på yrkesstruktur i bedrifter og avvik fra gruppegjennomsnitt.

| | 1210 Adm. Dir. | 5221 Butikkmedarb. | type 1 | type 2 | type 3 | type 4 |
|---------|----------------|--------------------|--------|--------|--------|--------|
| Gruppe | 4 % | 90 % | 0 | 0 | 0 | 0 |
| Bedrift | 3 % | 3 % | 1 | 1 | 1 | 1 |
| Bedrift | 3 % | 52 % | 2 | 2 | 2 | 2 |
| Bedrift | 3 % | 71 % | 3 | 3 | 4 | 4 |
| Bedrift | 3 % | 67 % | 4 | 4 | 3 | 3 |
| Bedrift | 3 % | 77 % | 5 | 5 | 5 | 5 |
| Bedrift | 3 % | 82 % | 6 | 10 | 11 | 6 |
| Bedrift | 0 % | 100 % | 8.5 | 7.5 | 7.5 | 18.5 |

Vi kan merke oss endel egenskaper ved disse gruppene:

- Det er ganske få yrker og et av dem er meget stort. Et så dominerende yrke har mye å si for de fleste avviksmålene. Som vi kunne forvente er det høyest andel i de små bedriftene, da vi går utfra at store bedrifter har mer av spesialiserte støttefunksjoner.
- Bedriftene med mange direktører og ingen medarbeidere er en typisk konsernledelse. Det er altså ikke yrkesstrukturen som er feil, men næringskoden beskriver ikke bedriftens aktivitet på samme måte som for de andre bedriftene.
- Det er stor ulikhet og skjevhet i andelen rengjøringsmedarbeidere. Dette avhenger nok av om bedriften ansetter eget vaskepersonale eller leier inn et vaskefirma – ikke om feilklassifisering av arbeidsoppgaver.
- De ulike modellene karakteriserer disse bedriftene nokså likt, med type 4 som skiller seg mest ut. Det er mulig dette har med variasjonen i bedriftsstørrelsene å gjøre.

For disse strataene spiller det kanskje ikke så stor rolle hvilken modell vi velger, men vi kan hvertfall bruke en enkel metode for å identifisere bedrifter som bør undersøkes nærmere. Særlig de yrker som ikke inngår i dette eksemplet er da viktige å studere nærmere.

5.3.7.2 Eksempel: Bygg

Tabellen viser andelene av de største yrkene innen endel store bedrifter i næringen "45.211 Oppføring av bygninger".

Tabell 5-13: Eksempel på yrkesstruktur i bedrifter og avvik fra gruppegjennomsnitt.

| | 1223 | 2142 | 3111 | 4113 | 7121 | 7122 | 7123 | 7124 | 7125 | 7126 | 7421 | 8331 | 8332 | 9310 | type_1 | type_2 | type_3 | type_4 |
|---------|-----------|----------|-----------|-------|-------|---------|------------|-------------|--------|----------|---------|-------------|-----------|---------|--------|--------|--------|--------|
| | Avd. sjef | Siv.ing. | Bygn.ing. | Sekr. | Murer | Betong. | Jernbinder | Forskaling. | Tømrer | Anl.arb. | Snekker | Anl.maskin. | Kranfører | Hj.arb. | | | | |
| Gruppe | 2 % | 5 % | 5 % | 2 % | 2 % | 4 % | 2 % | 13 % | 29 % | 10 % | 5 % | 2 % | 2 % | 4 % | 0 | 0 | 0 | 0 |
| Bedrift | 1 % | 0 % | 1 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 4 % | 0 % | 82 % | 0 % | 0 % | 1 | 1 | 1 | 12 |
| Bedrift | 0 % | 0 % | 10 % | 0 % | 0 % | 0 % | 0 % | 4 % | 1 % | 77 % | 0 % | 0 % | 0 % | 3 % | 2 | 2 | 3 | 15 |
| Bedrift | 2 % | 1 % | 22 % | 2 % | 0 % | 63 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 3 | 3 | 2 | 11 |
| Bedrift | 1 % | 1 % | 5 % | 0 % | 0 % | 1 % | 0 % | 71 % | 0 % | 0 % | 0 % | 0 % | 7 % | 4 % | 4 | 4 | 4 | 21 |
| Bedrift | 5 % | 0 % | 0 % | 0 % | 0 % | 1 % | 0 % | 0 % | 13 % | 5 % | 50 % | 5 % | 3 % | 3 % | 5 | 6 | 5 | 18 |
| Bedrift | 0 % | 5 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 77 % | 0 % | 3 % | 0 % | 0 % | 3 % | 6 | 5 | 8 | 30 |
| Bedrift | 0 % | 0 % | 3 % | 3 % | 0 % | 0 % | 0 % | 0 % | 73 % | 0 % | 0 % | 0 % | 0 % | 14 % | 7 | 7 | 7 | 27 |
| Bedrift | 1 % | 0 % | 6 % | 1 % | 0 % | 0 % | 0 % | 1 % | 71 % | 0 % | 1 % | 1 % | 3 % | 3 % | 8 | 8 | 13 | 34 |
| Bedrift | 0 % | 3 % | 2 % | 0 % | 0 % | 0 % | 0 % | 0 % | 71 % | 1 % | 4 % | 2 % | 0 % | 7 % | 9 | 9 | 11 | 20 |
| Bedrift | 0 % | 0 % | 2 % | 0 % | 0 % | 6 % | 3 % | 43 % | 5 % | 0 % | 0 % | 0 % | 3 % | 2 % | 10 | 12 | 6 | 2 |
| Bedrift | 0 % | 2 % | 12 % | 1 % | 0 % | 2 % | 8 % | 26 % | 13 % | 0 % | 25 % | 2 % | 1 % | 1 % | 15 | 17 | 14 | 10 |
| Bedrift | 2 % | 9 % | 7 % | 6 % | 0 % | 6 % | 3 % | 1 % | 14 % | 15 % | 4 % | 1 % | 1 % | 0 % | 26 | 27 | 18 | 1 |
| Bedrift | 2 % | 6 % | 3 % | 1 % | 0 % | 0 % | 0 % | 30 % | 30 % | 14 % | 1 % | 0 % | 1 % | 3 % | 34 | 28 | 50 | 9 |
| Bedrift | 3 % | 11 % | 5 % | 1 % | 13 % | 7 % | 1 % | 9 % | 20 % | 12 % | 0 % | 1 % | 1 % | 2 % | 39 | 39 | 41 | 3 |
| Bedrift | 1 % | 6 % | 7 % | 3 % | 0 % | 8 % | 2 % | 8 % | 18 % | 10 % | 3 % | 6 % | 2 % | 12 % | 41 | 40 | 46 | 4 |
| Bedrift | 8 % | 10 % | 8 % | 3 % | 0 % | 1 % | 2 % | 16 % | 21 % | 6 % | 3 % | 0 % | 1 % | 1 % | 48 | 49 | 42 | 8 |
| Bedrift | 4 % | 7 % | 7 % | 1 % | 0 % | 4 % | 1 % | 15 % | 37 % | 2 % | 0 % | 1 % | 1 % | 1 % | 50 | 48 | 48 | 6 |
| Bedrift | 2 % | 3 % | 3 % | 1 % | 3 % | 1 % | 2 % | 25 % | 30 % | 12 % | 1 % | 1 % | 2 % | 1 % | 51 | 43 | 51 | 7 |
| Bedrift | 2 % | 9 % | 5 % | 1 % | 4 % | 2 % | 1 % | 16 % | 33 % | 13 % | 0 % | 0 % | 2 % | 3 % | 52 | 52 | 52 | 5 |

Om yrkene i gruppa:

- Det er flere yrker og jevnere fordeling enn i butikknæringen, det er i mindre grad et enkelt yrke som dominerer.
- Yrkesstrukturen varierer mye mer i denne næringen. F.eks. flere bedrifter mangler tømrere, mens andre har nærmere 80%. Liknende for forskalingsnekkere, de fleste har få eller ingen, men noen har betydelige andeler.

Det kunne jo også bety at spesielt mange har systematiske feil i yrkeskodingen. Men det er meget mulig at disse bedriftene ikke har så ensartede aktiviteter at det er meningsfullt å sammenlikne alle yrkesandelene direkte. En analyse av sammenheng mellom de ulike yrkene tyder på at noe av årsaken kan være at ulike aktivitetstyper er skilt ut i egne bedrifter. Det er ikke overraskende at yrkesstrukturen blir ulik om man oppfører tre- eller betongbygg, og at flere bedrifter spesialisere seg. Enkelte firma opplyser at de har skilt ut tømrere som egen bedrift. For slike tilfeller er det viktigere å finne mer homogene strata, enn å raffinere målemetodene.

5.3.7.3 Eksempel: Elektro

Tabellen viser andelene av de største yrkene innen endel bedrifter i næringen "45.310 Elektrisk installasjonsarbeid".

Tabell 5-14: Eksempel på yrkesstruktur i bedrifter og avvik fra gruppegjennomsnitt.

| | 1210 | 3111 | 3113 | 3119 | 4114 | 4131 | 7241 | 9132 | 9310 | type_1 | type_2 | type_3 | type_4 |
|---------|----------|-----------|---------------|------------|---------|--------|-------------|-------|---------|--------|--------|--------|--------|
| | Adm.dir. | Bygn.ing. | Elektro.tekn. | Andre ing. | Kontor. | Lager. | Elektriker. | Vask. | Hj.arb. | | | | |
| Gruppe | 2 % | 5 % | 3 % | 1 % | 3 % | 1 % | 71 % | 1 % | 2 % | 0 | 0 | 0 | 0 |
| Bedrift | 0 % | 0 % | 69 % | 6 % | 0 % | 0 % | 0 % | 0 % | 0 % | 1 | 1 | 1 | 1 |
| Bedrift | 0 % | 17 % | 13 % | 0 % | 0 % | 0 % | 17 % | 0 % | 0 % | 2 | 2 | 2 | 2 |
| Bedrift | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 94 % | 0 % | 0 % | 3 | 3 | 5 | 4 |
| Bedrift | 0 % | 0 % | 0 % | 6 % | 12 % | 3 % | 58 % | 3 % | 0 % | 4 | 4 | 3 | 3 |
| Bedrift | 0 % | 3 % | 0 % | 13 % | 0 % | 0 % | 60 % | 3 % | 7 % | 5 | 9 | 4 | 5 |
| Bedrift | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 85 % | 3 % | 0 % | 6 | 8 | 6 | 6 |
| Bedrift | 0 % | 0 % | 0 % | 0 % | 6 % | 0 % | 85 % | 3 % | 0 % | 7 | 5 | 10 | 8 |
| Bedrift | 3 % | 0 % | 0 % | 0 % | 3 % | 3 % | 85 % | 0 % | 3 % | 8 | 6 | 25 | 17 |
| Bedrift | 3 % | 6 % | 0 % | 0 % | 6 % | 0 % | 61 % | 3 % | 6 % | 9 | 12 | 7 | 9 |
| Bedrift | 3 % | 6 % | 0 % | 0 % | 0 % | 0 % | 85 % | 0 % | 0 % | 10 | 7 | 17.5 | 10 |
| Bedrift | 6 % | 0 % | 0 % | 0 % | 9 % | 0 % | 68 % | 3 % | 0 % | 18 | 27 | 8 | 7 |

Som forventet dominerer elektrikerne denne næringsgruppen, og avviket i andelen av dette yrket har mye å si for bedriftens måltall. Det er endel innslag av ingeniører/tekniker som er vanskelig å vurdere som direkte feil. Dette kan nok være en gruppe som er egnet for "type 3" som jo prioriterer de største avvikene.

5.3.7.4 Eksempel: Bank

Tabellen viser andelen av de største yrkene innen mellomstore (gj.sn. 32 ansatte) bedrifter i næringen "65.120 Bankvirksomhet ellers".

Tabell 5-15: Eksempel på yrkesstruktur i bedrifter og avvik fra gruppegjennomsnitt.

| | 1226 | 1231 | 2512 | 3412 | 3418 | 4114 | 4212 | 5163 | 9132 | type 1 | type 2 | type 3 | type 4 |
|---------|---|------|------|------|------|------|------|------|------|--------|--------|--------|--------|
| | Avd.sjef Øk.dir. Pers.kons. Fors.kons. Bankfunk. Kontor Kasserer. Vaktmester Rengj. | | | | | | | | | | | | |
| Gruppe | 11 % | 4 % | 2 % | 1 % | 62 % | 3 % | 1 % | 2 % | 3 % | 0 | 0 | 0 | 0 |
| Bedrift | 31 % | 19 % | 0 % | 0 % | 25 % | 3 % | 6 % | 0 % | 0 % | 1 | 1 | 1 | 1 |
| Bedrift | 16 % | 6 % | 0 % | 26 % | 29 % | 0 % | 6 % | 0 % | 6 % | 2 | 2 | 3 | 3 |
| Bedrift | 18 % | 3 % | 0 % | 0 % | 27 % | 21 % | 9 % | 6 % | 6 % | 3 | 3 | 2 | 2 |
| Bedrift | 15 % | 6 % | 0 % | 0 % | 36 % | 27 % | 3 % | 3 % | 0 % | 4 | 4 | 6 | 4 |
| Bedrift | 6 % | 0 % | 0 % | 0 % | 94 % | 0 % | 0 % | 0 % | 0 % | 5 | 5 | 4 | 7 |
| Bedrift | 6 % | 0 % | 0 % | 0 % | 94 % | 0 % | 0 % | 0 % | 0 % | 6 | 6 | 5 | 9 |
| Bedrift | 15 % | 6 % | 3 % | 0 % | 33 % | 6 % | 0 % | 3 % | 9 % | 7 | 7 | 7 | 5 |
| Bedrift | 6 % | 0 % | 0 % | 0 % | 90 % | 0 % | 0 % | 0 % | 0 % | 8 | 8 | 8 | 10 |

Det framgår at modellene er samstemte om hvilke bedrifter som skiller seg mest ut. Det er meget store forskjeller i andelen bankfunksjonærer, noe vi ikke har funnet en god forklaring på.

Neste tabell viser noen av de største bedriftene (gj.sn. 240 ansatte) innen samme næringsundergruppe. En finner for det første en annen yrkesstruktur, med flere yrker og større variasjon. Videre gir det endel andre utslag i de ulike måle metodene.

Tabell 5-16: Eksempel på yrkesstruktur i bedrifter og avvik fra gruppegjennomsnitt.

| | 1226 | 1231 | 2130 | 2411 | 2512 | 2541 | 3411 | 3418 | 4113 | 9132 | type 1 | type 2 | type 3 | type 4 |
|---------|--|------|------|------|------|------|------|------|------|------|--------|--------|--------|--------|
| | Avd.sjef Øk.dir. IT-kons. Øk.saksb. Pers.kons. Økon. Fin.megler Bankfunk. Sekr. Rengj. | | | | | | | | | | | | | |
| Gruppe | 13 % | 3 % | 2 % | 1 % | 2 % | 9 % | 2 % | 54 % | 1 % | 1 % | 0 | 0 | 0 | 0 |
| Bedrift | 18 % | 3 % | 0 % | 0 % | 1 % | 39 % | 3 % | 29 % | 0 % | 0 % | 1 | 1 | 1 | 1 |
| Bedrift | 14 % | 4 % | 9 % | 0 % | 0 % | 6 % | 17 % | 25 % | 2 % | 0 % | 2 | 2 | 4 | 11 |
| Bedrift | 1 % | 1 % | 0 % | 3 % | 3 % | 1 % | 0 % | 85 % | 0 % | 1 % | 3 | 3 | 2 | 6 |
| Bedrift | 2 % | 0 % | 0 % | 0 % | 1 % | 0 % | 0 % | 84 % | 0 % | 2 % | 4 | 4 | 3 | 7 |
| Bedrift | 11 % | 0 % | 0 % | 0 % | 1 % | 35 % | 0 % | 46 % | 1 % | 0 % | 5 | 5 | 5 | 15 |
| Bedrift | 10 % | 5 % | 1 % | 0 % | 7 % | 3 % | 0 % | 35 % | 2 % | 0 % | 6 | 7 | 7 | 9 |
| Bedrift | 5 % | 3 % | 4 % | 0 % | 0 % | 2 % | 21 % | 43 % | 2 % | 0 % | 7 | 8 | 8 | 12 |
| Bedrift | 7 % | 1 % | 0 % | 0 % | 0 % | 33 % | 0 % | 57 % | 0 % | 0 % | 8 | 6 | 6 | 24 |
| Bedrift | 2 % | 1 % | 1 % | 0 % | 10 % | 5 % | 0 % | 69 % | 0 % | 0 % | 9 | 9 | 12 | 18 |
| Bedrift | 10 % | 1 % | 0 % | 0 % | 1 % | 0 % | 0 % | 72 % | 5 % | 2 % | 11 | 10 | 9 | 10 |
| Bedrift | 8 % | 2 % | 2 % | 0 % | 0 % | 4 % | 4 % | 40 % | 2 % | 0 % | 13 | 14 | 13 | 4 |
| Bedrift | 11 % | 0 % | 0 % | 0 % | 3 % | 1 % | 1 % | 71 % | 1 % | 0 % | 14 | 13 | 11 | 8 |
| Bedrift | 10 % | 2 % | 0 % | 0 % | 0 % | 1 % | 0 % | 68 % | 2 % | 0 % | 16 | 15 | 15 | 5 |
| Bedrift | 23 % | 0 % | 2 % | 0 % | 2 % | 0 % | 0 % | 64 % | 1 % | 0 % | 18 | 17 | 20 | 3 |
| Bedrift | 13 % | 5 % | 2 % | 3 % | 1 % | 4 % | 1 % | 58 % | 2 % | 3 % | 26 | 26 | 26 | 2 |

5.3.7.5 Eksempel: Restaurant

Tabellen viser yrkesfordelingen i mellomstore bedrifter i næringsundergruppe "55.301 Drift av restauranter og kafeer".

Tabell 5-17: Eksempel på yrkesstruktur i bedrifter og avvik fra gruppegjennomsnitt.

| | 1210 | Adm.dir. | 1224 | Avd.sj. | 4222 | Resepsjon. | 5122 | Kokk. | 5123 | Servitør. | 8321 | Sjåfør. | 9133 | Kj.ass. | 9152 | Vakt. | type_1 | type_2 | type_3 | type_4 |
|---------|------|----------|------|---------|------|------------|------|-------|------|-----------|------|---------|------|---------|------|-------|--------|--------|--------|--------|
| Gruppe | 2 % | 2 % | 1 % | 14 % | 49 % | 3 % | 17 % | 6 % | 0 | 0 | 0 | 0 | | | | | | | | |
| Bedrift | 0 % | 3 % | 0 % | 0 % | 0 % | 0 % | 97 % | 0 % | 1 | 1 | 1 | 1 | | | | | | | | |
| Bedrift | 0 % | 0 % | 0 % | 19 % | 0 % | 3 % | 71 % | 0 % | 2 | 2 | 3 | 3 | | | | | | | | |
| Bedrift | 3 % | 3 % | 0 % | 10 % | 0 % | 0 % | 60 % | 7 % | 3 | 3 | 4 | 5 | | | | | | | | |
| Bedrift | 0 % | 3 % | 25 % | 9 % | 3 % | 0 % | 53 % | 0 % | 4 | 4 | 2 | 2 | | | | | | | | |
| Bedrift | 0 % | 0 % | 0 % | 0 % | 97 % | 0 % | 0 % | 0 % | 5 | 5 | 8 | 9 | | | | | | | | |
| Bedrift | 3 % | 0 % | 0 % | 0 % | 97 % | 0 % | 0 % | 0 % | 6 | 6 | 10 | 11 | | | | | | | | |
| Bedrift | 0 % | 0 % | 0 % | 0 % | 94 % | 0 % | 0 % | 0 % | 7 | 7 | 11 | 10 | | | | | | | | |
| Bedrift | 0 % | 0 % | 0 % | 0 % | 16 % | 0 % | 44 % | 25 % | 8 | 10 | 5 | 4 | | | | | | | | |
| Bedrift | 0 % | 6 % | 0 % | 0 % | 90 % | 0 % | 0 % | 0 % | 9 | 8 | 9 | 8 | | | | | | | | |
| Bedrift | 3 % | 0 % | 0 % | 6 % | 90 % | 0 % | 0 % | 0 % | 10 | 9 | 15 | 14 | | | | | | | | |

Det er flere yrker som viser stor spredning, og er illustrerende for problemstillingen om reelle systematiske feil i forhold til en næringsundergruppe med heterogene aktiviteter. Det er også grunn til å spørre seg om alle bedriftene har riktig næringen i forhold til aktivitet. Det er f.eks. kjent at bedrifter som ikke driver restaurantvirksomhet kaller seg restaurant. De viktigste yrkene kunne kanskje brukes til å splitte opp gruppa etter aktivitet. Et eksempel er hamburgersteder hvor det hverken er spesialisering innen matlaging eller servering, altså hvor det ikke er kokker eller servitører. En annen

type kan være hjemkjøring av matretter, hvor noen kun jobber med telefonbestilling. Det kan samtidig tenkes at det innen denne næringen er betydelige systematiske feil eller mangler i yrkeskodene.

5.3.7.6 Eksempel: olje

Vi ser nå på en næring med jevnere fordeling på flere på ulike yrker, "11.200 Tjenester tilknyttet olje- og gassutvinning". Det blir mer tydelig forskjeller på de ulike målemetodene, og vi kan bruke f.eks. type 4 for å skille ut bedrifter som avviker mer generelt, ikke bare i noen få dominerende yrker.

Tabell 5-18: Eksempel på yrkesstruktur i bedrifter og avvik fra gruppegjennomsnitt.

| Gruppe | 1222 | | 2114 | | 2147 | | 2512 | | 3116 | | 3120 | | 4113 | | 4121 | | 4131 | | 5163 | | 7212 | | 7237 | | 7241 | | 8113 | | 8114 | | 9310 | | type 1 | type 2 | type 3 | type 4 | |
|---------|------|------|-------|------|------|-------|-------|------|-------|------|------|---------|--------|--------|---------|------|------|---------|------|--------|------|------|------|-----|------|------|------|-----|------|-----|------|-----|--------|--------|--------|--------|----|
| | Avd. | Sjef | Geol. | Siv. | ing. | Pers. | kons. | Ing. | Data. | Skr. | Øk. | medarb. | Lager. | Vaktm. | Sveiser | Ind. | mek. | Elektr. | Opr. | Brønn. | Hj. | arb. | | | | | | | | | | | | | | | |
| Bedrift | 0 % | 47 % | 0 % | 0 % | 0 % | 0 % | 53 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 1 | 1 | 4 | 16 |
| Bedrift | 0 % | 0 % | 0 % | 0 % | 0 % | 8 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 14 % | 74 % | 0 % | 0 % | 0 % | 0 % | 2 | 2 | 1 | 6 | | | |
| Bedrift | 71 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 24 % | 0 % | 0 % | 0 % | 0 % | 0 % | 3 | 3 | 2 | 18 | | | |
| Bedrift | 0 % | 0 % | 0 % | 0 % | 0 % | 1 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 98 % | 0 % | 0 % | 0 % | 0 % | 0 % | 4 | 4 | 3 | 21 | | | |
| Bedrift | 0 % | 0 % | 0 % | 0 % | 1 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 26 % | 0 % | 0 % | 2 % | 0 % | 0 % | 0 % | 0 % | 0 % | 5 | 5 | 11 | 15 | | | |
| Bedrift | 3 % | 0 % | 2 % | 6 % | 21 % | 1 % | 7 % | 5 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 1 % | 0 % | 0 % | 22 | 24 | 21 | 5 | | | | | |
| Bedrift | 2 % | 6 % | 4 % | 2 % | 40 % | 11 % | 2 % | 2 % | 5 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 13 % | 0 % | 0 % | 0 % | 0 % | 24 | 22 | 25 | 3 | | | | |
| Bedrift | 0 % | 0 % | 2 % | 0 % | 4 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 3 % | 1 % | 66 % | 13 % | 0 % | 0 % | 28 | 27 | 24 | 4 | | | | |
| Bedrift | 5 % | 0 % | 0 % | 0 % | 1 % | 0 % | 0 % | 0 % | 2 % | 3 % | 1 % | 9 % | 4 % | 63 % | 1 % | 2 % | 32 | 31 | 28 | 1 | | | | | | | | | | | | | | | | | |
| Bedrift | 3 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 0 % | 3 % | 4 % | 1 % | 9 % | 8 % | 41 % | 11 % | 13 % | 39 | 39 | 39 | 2 | | | | | | | | | | | | | | | | |

5.3.8 Prioritering av strata for kontroll

Uansett metode for aggregering av avvik, er det store forskjeller mellom strata. Et stratum der en eller noen få bedrifter skiller seg ut, vil være enklere å kontrollere enn et med mange avvikende. Hvis det ser ut som "alle skiller seg ut" kan vi ikke etablere en statistisk mal for yrkesfordelingen. Dette kan skyldes to forhold, eller en kombinasjon av disse:

- Det er en gruppe bedrifter med store feil og usikkerhet i yrkeskodingen.
- Det er reell forskjeller i arbeidsoppgaver og grunn til ulikhet i yrkeskoding.

Dette kan bare avgjøres ved direkte kontroll, og det kan bli ganske omfattende for visse bransjer. I første fase av å etablere kontrollrutiner, forsøker vi å prioritere de strata som er enklest å håndtere. For å identifisere disse brukes følgende metode:

- Beregner gjennomsnitt av de aggregerte avvikene i gruppa
- Beregner *trimmet* gjennomsnitt, der f.eks. de 10% høyeste verdiene holdes utenfor, av de samme.
- Forholdet på de to gjennomsnittene skal si i hvilken grad gruppa har få/store avvik (ekte utliggere). Måltallet er enhetsløst og uavhengig av størrelsen på bedrifter eller grupper.

Vi har "type 1" pr. bedrift:

$$(5-26) \quad s_b = \text{std}(q_{y,b})$$

Skriver forholdet mellom gjennomsnittene kort:

$$(5-27) \quad k_g = \frac{\bar{s}_b}{\bar{s}_b^*}$$

k_g vil ha verdier fra 1 og oppover. Nedenfor vises noen eksempler. Vi må gjennomføre flere kontroller for å kunne si hvor godt metodene egner seg i forhold til en realitetsvurdering av yrkeskodingen.

5.3.8.1 Eksempler

Tabell 5-19: Eksempel på stratum med få/tydelige avvik. $k_g=1.43$

| A 45.310 Elektrisk installasjonsarbeid, med rundt 30 ansatte | | |
|--|-----|---|
| ----- | | |
| Bedrift 1 | | |
| 1% | 6% | 3113 ELEKTRONIKK- OG TELEKOMMUNIKASJONSINGENIØRER OG -TEKNIKERE |
| 73% | 21% | 7241 ELEKTRIKERE, ELEKTRONIKERE OL. |
| 4% | 55% | 7242 SERVICE- OG TELEMONTØRER |
| ----- | | |
| Bedrift 2 | | |
| 73% | 7% | 7241 ELEKTRIKERE, ELEKTRONIKERE OL. |
| 4% | 90% | 7242 SERVICE- OG TELEMONTØRER |
| ----- | | |
| Bedrift 3 | | |
| 1% | 33% | 7134 RØRLEGGERE OG VVS-MONTØRER |
| 0% | 17% | 7212 SVEISERE |
| 0% | 13% | 7214 PLATEARBEIDERE |
| 0% | 10% | 7237 INDUSTRIMEKANIKERE |
| 73% | 3% | 7241 ELEKTRIKERE, ELEKTRONIKERE OL. |
| 0% | 7% | 8281 MONTØRER INNEN PRODUKSJON AV TRANSPORTMIDLER |

Tabell 5-20: Eksempel på stratum med mindre klar yrkesstruktur. $k_g=1.08$

| B 72.220 Annen konsulentvirksomhet tilknyttet system- og programvare, med rundt 40 ansatte | | |
|--|-----|---|
| ----- | | |
| Bedrift 1 | | |
| 3% | 5% | 1226 PRODUKSJONSDIREKTØRER INNEN FINANSIELL TJENESTEYTING, EIENDOMSDRIFT OSV. |
| 33% | 3% | 2130 SYSTEMUTVIKLERE OG PROGRAMMERERE |
| 36% | 87% | 3120 DATAINGENIØRER OG -TEKNIKERE |
| ----- | | |
| Bedrift 2 | | |
| 33% | 0% | 2130 SYSTEMUTVIKLERE OG PROGRAMMERERE |
| 36% | 90% | 3120 DATAINGENIØRER OG -TEKNIKERE |
| ----- | | |
| Bedrift 3 | | |
| 33% | 87% | 2130 SYSTEMUTVIKLERE OG PROGRAMMERERE |
| 36% | 0% | 3120 DATAINGENIØRER OG -TEKNIKERE |
| ----- | | |
| Bedrift 4 | | |
| 33% | 3% | 2130 SYSTEMUTVIKLERE OG PROGRAMMERERE |
| 36% | 89% | 3120 DATAINGENIØRER OG -TEKNIKERE |
| 5% | 9% | 3415 TEKNISKE OG KOMMERSIELLE SALGSREPRESENTANTER |
| ----- | | |
| Bedrift 5 | | |
| 33% | 88% | 2130 SYSTEMUTVIKLERE OG PROGRAMMERERE |
| 36% | 0% | 3120 DATAINGENIØRER OG -TEKNIKERE |
| ----- | | |
| Bedrift 6 | | |
| 3% | 8% | 1226 PRODUKSJONSDIREKTØRER INNEN FINANSIELL TJENESTEYTING, EIENDOMSDRIFT OSV. |
| 33% | 0% | 2130 SYSTEMUTVIKLERE OG PROGRAMMERERE |
| 36% | 87% | 3120 DATAINGENIØRER OG -TEKNIKERE |

5.4 Metodikk for små feil

Stikkprøver og andre undersøkelser har vist at det forekommer yrkeskoder som åpenbart er feil. Mange av disse er fordelt på en slik måte at de ikke vil identifiseres med metodene som er skissert i forrige del. Men både stikkprøver og annet har vist at det er mange "slengere", som vi kan kalle statistisk støy, grums eller tilfeldige feil. For aggregert statistikk vil dette ha mindre betydning. Allikevel er det viktig å undersøke og revidere av to grunner:

- Forvaltningsmessig skal alle ha riktig yrkeskode i Arbeidstakerregisteret, og arbeidsgiver skal kontrollere dette på hvert enkelt arbeidstakerforhold. Forekomst av mange tilfeldige feil kan tyde på at slik kontroll ikke blir utført, eller at arbeidsgiver har problemer med rapportere egnet yrkesdata. Begge tilfeller er viktig å oppdage.
- Det vil kunne gi utslag i statistikk for små grupper, som er mer etterspurt. Det samme gjelder levering av mikrodata til forskning og planlegging.

Det er med andre ord flere gode grunner for å utvikle metoder for å måle omfanget av tilfeldige feil, og identifisere enheter til kontroll.

5.4.1 Grunnlag

Flere analyser fra tidligere med helt ulike formål viser at det mange steder forekommer overraskende mange forskjellige yrkeskoder. I noen næringer kan man nok argumentere for at årsaken er at aktivitetene er svært varierte, for andre er det ikke noen påviselig grunn. Siden vi her baserer oss på Arbeidstakerregisteret alene, kan vi ikke definere tilfeldige feil slik som i samsvarsanalysen i notat 2004/46, side 7. Vi må derfor utvikle andre metoder.

Tabellen under viser at i flere næringer forekommer det mange yrkeskoder, hvorav et mindretall utgjør over 1% hver (definert som "store yrker").

Tabell 5-21: Næringer hvor det forekommer mange yrkeskoder med svært små andeler.

| nace | | Forekommende Yrker | Store Yrker | Små Yrker |
|--------|--|-----------------------|----------------|--------------|
| 11.200 | Tjenester tilknyttet olje- og gassutvinning | 141 | 17 | 124 |
| 40.110 | Produksjon av elektrisitet | 141 | 21 | 120 |
| 45.211 | Oppføring av bygninger | 164 | 16 | 148 |
| 45.230 | Bygging av veier, flyplasser og idrettsanlegg | 130 | 14 | 116 |
| 55.301 | Drift av restauranter og kafeer | 124 | 10 | 114 |
| 60.240 | Godstransport på vei | 147 | 11 | 136 |
| 70.202 | Utleie av egen fast eiendom ellers | 192 | 19 | 173 |
| 72.220 | Annen konsulentvirksomhet tilknyttet system- og programvare | 152 | 13 | 139 |
| 73.100 | Forskning og utviklingsarbeid innen naturvitenskap og teknikk | 162 | 18 | 144 |
| 74.140 | Bedriftsrådgiving | 162 | 16 | 146 |
| 74.202 | Byggeteknisk konsulentvirksomhet | 136 | 19 | 117 |
| 74.209 | Annen teknisk konsulentvirksomhet | 173 | 24 | 149 |
| 74.300 | Teknisk prøving analyse | 137 | 19 | 118 |
| 74.502 | Utleie av arbeidskraft | 195 | 19 | 176 |
| 74.879 | Annen forretningsmessig tjenesteyting ikke nevnt annet sted | 189 | 22 | 167 |
| 75.110 | Generell (overordnet) offentlig administrasjon og økonomiforvaltning | 127 | 10 | 117 |
| 80.429 | Annen undervisning | 155 | 16 | 139 |
| 85.334 | Arbeidstrening for ordinært arbeidsmarked | 164 | 21 | 143 |
| 91.330 | Interesseorganisasjoner ellers | 174 | 23 | 151 |

Vi vil ta som utgangspunkt for nærmere kontroll at innen en presumptivt ensartet gruppe bedrifter vil et høyt antall ulike yrkeskoder med svært små andeler indikerer tilfeldige feil. Vi kan nevne noen forutsetninger og vurderinger som må avveies:

- Enkelte yrker *skal være sjeldne*, f.eks. kun en toppleder (adm.dir eller daglig leder) pr. bedrift.
- Noen yrker har lav andel og finnes i de fleste næringer, f.eks. mange bedrifter har nøyaktig 1 vaktmester.
- Alle yrker som forekommer i minst 1 bedrift i et stratum, forekommer i stratumsgjennomsnittet. Man kan derfor ikke sette en nedre grense for andelen direkte.
- Et eller flere yrker som er svært små i gruppa, kan utgjøre en betydelig del av en bedrift. Det er disse vi helst skal ha tak i. Siden vi har stratifisert etter størrelse, er det allikevel grenser for hvor mye en enkeltbedrift kan skille seg ut
- 'Person' eller 'arbeidstakerforhold' er en udelelig størrelse så den minste andelen er alltid 1 person
 $\min(p_b) = \frac{1}{X_b}$ og enhver andel p blir et rasjonalt tall. Det vil derfor være et forventet basisavvik på grunn av avrundingsfeil. Dette gir særlig utslag på mindre bedrifter.

5.4.2 Beregninger

Definerer noen størrelser:

$$(5-28) \quad X_b \quad \text{antall ansatte i bedrift } b$$

$$(5-29) \quad Y_{u,b} \quad \text{antall ansatte med yrke } u \text{ i bedrift } b$$

$$(5-30) \quad p_{u,b} = \frac{Y_{u,b}}{X_b} \quad \text{andel av yrke } u \text{ i bedrift } b$$

$$(5-31) \quad X_g = \sum_{b \in g} X_b \quad \text{antall ansatte i gruppe } g$$

$$(5-32) \quad Y_{u,g} = \sum_{b \in g} Y_{u,b} \quad \text{antall ansatte med yrke } u \text{ i gruppe } g$$

$$(5-33) \quad p_{u,g} = \frac{Y_{u,g}}{X_g} \quad \text{andel av yrke } u \text{ i gruppe } g$$

(5-34) n_g antall bedrifter i gruppe g

På bakgrunn av de forutsetninger nevnt ovenfor kan vi ikke si at et yrke skal forekomme færre enn én gang i en gitt bedrift. Vi kan derfor ikke sette en nedre grense vilkårlig lav. Det er også klart at noen yrker må få lov til å forekomme 1 gang pr. bedrift. For å nærme oss en egnet grenseverdi, forsøker vi å definere en minimumsandel for ethvert yrke i gruppa som en andel som tilsvarer gjennomsnittlig en person pr. bedrift:

$$(5-35) \quad \lambda_g = \frac{1}{\bar{X}_{g,b}} = \frac{n_g}{X_g} \quad \text{siden } \bar{X}_{g,b} = \frac{X_g}{n_g}$$

Vi definerer nå en egen mengde yrker pr. gruppe som "mistenkelige":

$$(5-36) \quad M_g = \left\{ \forall u \in U \mid 0 < p_{u,g} < \lambda_g \right\}$$

I en gitt bedrift er da mengden "grums":

$$(5-37) \quad \gamma_b = \sum_{m \in M_g} Y_{m,b}$$

Vi kan bruke det absolute tallet direkte, eller se på bedrifter med høye andeler. Det er to hovedkilder til feil av typen falske positive:

- Yrker som er og skal være små, som nevnt i eksemplene over.
- Tilfeldige feil i små bedrifter overestimeres. Strata med høy varians (stor variasjon av bedriftsstørrelsene) blir overrepresentert.

I programmeringen operasjonaliserer vi ved å lage en indikatorvariabel:

$$(5-38) \quad I_{u,g} = \begin{cases} 1 & \mid 0 < p_{u,g} < \lambda_g \\ 0 & \mid p_{u,g} \geq \lambda_g \end{cases}$$

I en gitt bedrift er da mengden "grums":

$$(5-39) \quad \gamma_b = \sum_{u=1}^U Y_{u,b} \cdot I_{u,g}$$

5.4.3 Gruppering og avgrensning

Vi velger i første omgang å bruke samme stratifisering som i undersøkelsene av systematiske avvik. Imidlertid er det klart at i noen næringer er det vanskeligere å påvise tilfeldige feil. Vi definerer derfor et mål på *yrkeshomogenitet*:

$$(5-40) \quad s_g = \sum_{u=1}^U (p_{u,g})^2$$

Dette sier ikke noe om hvilke yrker som hører hjemme i gruppa, men måler generelt i hvilken grad det er et eller noen få enkeltyrker som dominerer. Et lavt tall forteller at det vil være vanskelig å skille ut tilfeldige feil. Hvis derimot næringen har høy yrkeshomogenitet, og samtidig svært mange ulike yrkeskoder, kan det være en symptom på forekomst av tilfeldige feil. Eksemplene nedenfor viser utvalgte næringer i forhold til dette.

Tabell 5-22: Eksempel på næringer etter yrkeshomogenitet. Fra Arbeidstakerregisteret uke 35.

| Næringer med spredte/varierte yrkesfordelinger | Ant. ans. | Ant. yrker | Spes. |
|--|------------------|-------------------|--------------|
| 22.150 Forlagsvirksomhet ellers | 618 | 62 | 0.042 |
| 22.320 Reproduksjon av videoopptak | 67 | 27 | 0.051 |
| 24.510 Produksjon av såpe og vaskemidler, rense- og polérmidler | 799 | 71 | 0.045 |
| 29.560 Produksjon av spesialmaskiner ellers | 1 388 | 89 | 0.040 |
| 33.300 Produksjon av industrielle prosessstyringsanlegg | 912 | 72 | 0.052 |
| 35.114 Bygging og reparasjon av oljeplattformer og moduler | 12 025 | 127 | 0.043 |
| 35.115 Innrednings- og installasjonsarbeid utført på borerigger og moduler | 5 652 | 110 | 0.047 |
| 36.500 Produksjon av spill og leker | 42 | 25 | 0.052 |
| 36.630 Industriproduksjon ikke nevnt annet sted | 463 | 71 | 0.053 |
| 50.101 Agentur- og engroshandel med motorvogner, unntatt motorsykler | 1 875 | 82 | 0.046 |
| 51.330 Engroshandel med meierivarer, egg, matolje og -fett | 621 | 63 | 0.046 |
| 51.872 Engroshandel med skipsutstyr og fiskeredskap | 1 487 | 90 | 0.051 |
| 70.112 Utvikling og salg av egen fast eiendom ellers | 831 | 66 | 0.053 |
| 71.320 Utleie av bygge- og anleggsmaskiner | 1 275 | 78 | 0.050 |
| 71.340 Utleie av maskiner og utstyr ellers | 923 | 106 | 0.037 |
| 74.209 Annen teknisk konsulentvirksomhet | 12 099 | 174 | 0.052 |
| 74.879 Annen forretningsmessig tjenesteyting ikke nevnt annet sted | 8 451 | 187 | 0.045 |
| 85.149 Andre helsetjenester | 1 417 | 74 | 0.050 |
| 85.336 Sosiale velferdsorganisasjoner | 3 032 | 111 | 0.049 |
| 91.330 Interesseorganisasjoner ellers | 4 253 | 175 | 0.050 |
| Næringer dominert av få yrker | Ant. ans. | Ant. yrker | Spes. |
| 52.110 Butikkhandel med bredt vareutvalg med hovedvekt på nærings- og nytelsesmidler | 61 154 | 115 | 0.734 |
| 52.120 Butikkhandel med bredt vareutvalg ellers | 8 522 | 66 | 0.685 |
| 52.210 Butikkhandel med frukt og grønnsaker | 108 | 7 | 0.778 |
| 52.242 Butikkhandel med sjokolade og sukkervarer | 467 | 14 | 0.744 |
| 52.410 Butikkhandel med tekstiler og utstyrsvarer | 3 216 | 48 | 0.670 |
| 52.420 Butikkhandel med klær | 21 939 | 74 | 0.675 |
| 52.431 Butikkhandel med skotøy | 3 829 | 29 | 0.720 |
| 52.432 Butikkhandel med reiseeffekter av lær og lærimitasjoner og varer av lær | 667 | 10 | 0.756 |
| 52.442 Butikkhandel med kjøkkenutstyr, glass og steintøy | 1 785 | 28 | 0.676 |
| 52.452 Butikkhandel med plater, musikk- og videokassetter, CD- og DVD-plater | 1 098 | 25 | 0.672 |
| 52.502 Butikkhandel med brukte klær | 12 | 3 | 0.708 |
| 60.220 Drosjebiltransport | 8 110 | 36 | 0.826 |
| 63.304 Guider og reiseledere | 110 | 12 | 0.748 |
| 74.501 Formidling av arbeidskraft | 4 258 | 76 | 0.700 |
| 75.210 Utenrikssaker | 1 424 | 13 | 0.731 |
| 75.300 Trygdeordninger underlagt offentlig forvaltning | 7 404 | 13 | 0.837 |
| 80.301 Undervisning ved universiteter | 18 443 | 47 | 0.883 |
| 80.302 Undervisning ved statlige høyskoler | 10 667 | 22 | 0.786 |
| 80.303 Undervisning ved militære høyskoler | 1 243 | 5 | 0.680 |
| 93.020 Frisering og annen skjønnhetspleie | 9 325 | 46 | 0.743 |

Vi ser at f.eks. at innen '74.879 Annen forretningsmessig tjenesteyting ikke nevnt annet sted' vil det være vanskelig å si hva som er riktig yrkesfordeling, med 187 ulike koder spredt tynt utover. Navnet gir heller ingen indikasjon på hva som ville være en intuitiv fordeling.

Det virker påfallende at det på såpefabrikker skal være så stor spredning i yrkeskoder. Videre må vi si at yrkesspredningen større enn forventet innen gravemaskinutleie, uten at vi har detaljkunnskap om de konkrete arbeidsoppgavene her. Liknende betraktninger kan kanskje danne grunnlag for å kontrollere spesielle næringer i senere undersøkelser.

Av de hele 36 yrkeskodene som forekommer i drosjenæringen, regner vi med at de fleste er tilfeldige feil. Med 91% som har koden for taxisjåfør, har vi et betydelig potensial for tilfeldige feil, selv om vi regner med litt administrasjon og støtteapparat i de store drosjesentralene. At vikarbyråer opererer med mange yrkeskoder er ikke overraskende, men det virker påfallende f.eks. at klesbutikker har 74 ulike yrker.

5.4.4 Eksempler

Tabell 5-23: Eksempel på undersøkelse av mulige tilfeldige feil utfra yrkesfordeling i bedrifter.

Stor bedrift inn næringen "65.120 Bankvirksomhet ellers"

| Yrke | Enhet | Gruppe |
|--|--------|--------|
| 1223 PRODUKSJONSDIREKTØRER INNEN BYGGE- OG ANLEGGSVIRKSOMHET | 0.10 % | 0.00 % |
| 1225 PRODUKSJONSDIREKTØRER INNEN TRANSPORT OG KOMMUNIKASJON | 0.10 % | 0.00 % |
| 1239 ANDRE SPESIALDIREKTØRER | 0.10 % | 0.10 % |
| 2221 LEGER | 0.10 % | 0.00 % |
| 2352 STUDIEINSPEKTØRER O.L. | 0.10 % | 0.00 % |
| 2552 SKULPTØRER, KUNSTMALERE OG ANDRE BILLEDKUNSTNERE | 0.30 % | 0.10 % |
| 3141 SKIPSMASKINISTER | 0.10 % | 0.00 % |
| 3152 SIKKERHETSINSPEKTØRER | 0.10 % | 0.00 % |
| 3226 FYSIOTERAPEUTER, ERGOTERAPEUTER O.L. | 0.10 % | 0.00 % |
| 3231 SYKEPLEIERE | 0.10 % | 0.00 % |
| 3431 FUNKSJONÆRER INNEN ADMINISTRASJON | 0.10 % | 0.20 % |
| 3433 HUSØKONOMER | 0.10 % | 0.00 % |
| 3471 DEKORATØRER, DESIGNERE OG REKLAMETEGNERE | 0.10 % | 0.00 % |
| 3491 INFORMASJONSMEDARBEIDERE OG JOURNALISTER | 0.10 % | 0.10 % |
| 4121 ØKONOMIMEDARBEIDERE OG REVISJONSASSISTENTER | 0.20 % | 0.20 % |
| 4212 POST- OG BANKKASSERERE | 0.40 % | 0.20 % |
| 5122 KOKKER | 0.20 % | 0.10 % |
| 5135 LEGESEKRETÆRER | 0.10 % | 0.00 % |
| 5164 VEKTERE | 0.10 % | 0.00 % |
| 8321 BIL-, DRØSJE- OG VAREBILFØRERE | 0.40 % | 0.10 % |
| 9133 KJØKKEN- OG ANRETINGSASSISTENTER | 0.70 % | 0.40 % |
| 9310 HJELPEARBEIDERE I BYGG, ANLEGG, VEDLIKEHOLD OL. | 0.10 % | 0.00 % |

Tabell 5-24: Eksempel på undersøkelse av mulige tilfeldige feil utfra yrkesfordeling i bedrifter.

Mellomstor bedrift innen "36.110 Produksjon av sittemøbler".

| Yrke | Enhet | Gruppe |
|---|---------|--------|
| 1210 ADMINISTRERENDE DIREKTØRER | 1.80 % | 0.70 % |
| 1231 FINANS-, ØKONOMI- OG ADMINISTRASJONSDIREKTØRER | 3.50 % | 0.60 % |
| 1234 MARKEDS- OG INFORMASJONSDIREKTØRER | 1.80 % | 0.60 % |
| 1236 IT-DIREKTØRER | 1.80 % | 0.30 % |
| 1239 ANDRE SPESIALDIREKTØRER | 1.80 % | 0.20 % |
| 2130 SYSTEMUTVIKLERE OG PROGRAMMERERE | 14.00 % | 0.90 % |
| 2141 SIVILARKITEKTER, BY- OG TRAFIKKPLANLEGGERE | 28.10 % | 0.70 % |
| 2352 STUDIEINSPEKTØRER O.L. | 1.80 % | 0.10 % |
| 2512 PERSONAL- OG ORGANISASJONSKONSULENTER | 1.80 % | 0.20 % |
| 3111 BYGNINGSINGENIØRER OG -TEKNIKERE | 1.80 % | 0.10 % |
| 3419 MARKEDSFØRINGS- OG REKLAMEKONSULENTER | 10.50 % | 0.90 % |
| 3431 FUNKSJONÆRER INNEN ADMINISTRASJON | 3.50 % | 0.10 % |
| 3471 DEKORATØRER, DESIGNERE OG REKLAMETEGNERE | 3.50 % | 0.20 % |
| 4113 SEKRETÆRER | 1.80 % | 0.30 % |
| 4121 ØKONOMIMEDARBEIDERE OG REVISJONSASSISTENTER | 5.30 % | 0.90 % |
| 5221 BUTIKKMEDARBEIDERE O.L. | 12.30 % | 0.30 % |

Det er mange av yrkene som man vel neppe kan si er feil utfra disse data, men det er også flere som helt klart ikke hører hjemme i gruppa. Selv om det kan bli litt lange lister, så er allikevel metoden egnet til å finne tilfeldige feil.

5.4.5 Andre vurderinger

For å effektivisere arbeidet ville det vært ønskelig å filtrere ut yrker hvor det er vanskelig å påvise tilfeldige feil. Vi forsøke dette ved å definere *næringsspesifikke* yrker på tilsvarende måte som yrkesspesifikke næringer. Tabellen under viser eksempel på yrker som er henholdsvis generelle og spesifikke i forhold til næring. De førstnevnte vil være mindre interessante i metoden som er utviklet for tilfeldige feil her. Yrker som kan kobles til bestemte næringer (ved statistiske eller evt. yrkesfaglige prinsipper) kan tillegges mer vekt i aggregeringen av mulige tilfeldige feil pr. bedrift.

Tabell 5-25: Eksempel på yrker, etter næringsspesifisitet. Fra Arbeidstakerregisteret uke 35.

| Eksempler på generelle yrke | Antall Ansatte | Antall Næringer | Spes. |
|--|-------------------|--------------------|-------|
| 1210 ADMINISTRERENDE DIREKTØRER | 23 674 | 581 | 0.016 |
| 1222 PRODUKSJONSDIREKTØRER INNEN INDUSTRI, MM. | 9 620 | 358 | 0.017 |
| 1231 FINANS-, ØKONOMI- OG ADMINISTRASJONSDIREKTØRER | 9 449 | 542 | 0.008 |
| 1232 PERSONALDIREKTØRER | 1 971 | 348 | 0.009 |
| 1233 SALGSDIREKTØRER | 5 703 | 395 | 0.012 |
| 1234 MARKEDS- OG INFORMASJONSDIREKTØRER | 6 601 | 449 | 0.011 |
| 1235 INNKJØPS- OG DISTRIBUSJONSDIREKTØRER | 2 954 | 355 | 0.009 |
| 1236 IT-DIREKTØRER | 1 828 | 315 | 0.017 |
| 1239 ANDRE SPESIALDIREKTØRER | 1 820 | 272 | 0.013 |
| 1310 LEDERE INNEN IKKE-SPEISIFISERT NÆRINGSOMRÅDE | 2 356 | 373 | 0.016 |
| 1312 LEDERE INNEN BERGVERKSDRIFT, INDUSTRI, KRAFT, MM. | 2 293 | 288 | 0.016 |
| 2541 SOSIAL- OG SIVILØKONOMER | 7 801 | 418 | 0.020 |
| 3415 TEKNISKE OG KOMMERSIELLE SALGSREPRESENTANTER | 27 563 | 467 | 0.016 |
| 3416 INNKJØPERE | 3 479 | 318 | 0.011 |
| 3431 FUNKSJONÆRER INNEN ADMINISTRASJON | 3 304 | 376 | 0.012 |
| 4113 SEKRETÆRER | 22 830 | 549 | 0.013 |
| 4114 KONTORMEDARBEIDERE | 38 669 | 602 | 0.019 |
| 4132 LOGISTIKERE | 1 999 | 256 | 0.019 |
| 5224 SELGERE (ENGROS) | 18 807 | 401 | 0.019 |
| 7237 INDUSTRIMEKANIKERE | 9 213 | 301 | 0.017 |
| Yrker som er konsentrert i få næringer | Antall Ansatte | Antall Næringer | Spes. |
| 2222 TANNLEGER | 646 | 15 | 0.883 |
| 2523 JURISTER INNEN POLITI OG PÅTALEMYNDIGHET | 34 | 1 | 1.000 |
| 2560 GEISTLIGE YRKER | 1 084 | 18 | 0.900 |
| 3224 OPTIKERE | 866 | 16 | 0.854 |
| 3225 TANNPLEIERE | 249 | 7 | 0.877 |
| 3228 RESEPTARER | 945 | 9 | 0.862 |
| 3310 GRUNNSKOLELÆRERE | 13 802 | 62 | 0.884 |
| 3320 FØRSKOLELÆRERE | 6 579 | 63 | 0.837 |
| 3441 TOLLFUNKSJONÆRER | 442 | 6 | 0.964 |
| 3442 LAVERE SAKSBEHANDLER VED LIGNINGSKONTOR | 44 | 2 | 0.956 |
| 3480 RELIGIØSE YRKER | 1 838 | 31 | 0.887 |
| 4111 STENOGRAFER, REFERENTER O.L. | 32 | 2 | 0.939 |
| 4212 POST- OG BANKKASSERERE | 220 | 12 | 0.869 |
| 5134 TANNLEGESEKRETÆRER | 2 925 | 14 | 0.958 |
| 5141 FRISØRER, KOSMETOLOGER O.L. | 8 512 | 71 | 0.886 |
| 5142 BEGRAVELSESBYRÅ- OG KREMATORIEARBEIDERE | 480 | 4 | 0.967 |
| 5225 INTERVJUERE | 972 | 8 | 0.963 |
| 8143 OPERATØRER INNEN SPON- OG FIBERPLATEPRODUKSJON | 489 | 12 | 0.837 |
| 8311 LOKOMOTIVFØRERE OG LOKOMOTIVKONTROLLØRER | 1 104 | 4 | 0.991 |
| 9141 VINDUSPUSSERE | 49 | 3 | 0.921 |

Det forekommer noen yrker som hører hjemme i statlig og kommunal sektor her, de skulle i prinsippet hatt stillingskode og ikke vært med i delpopulasjonen. Det skyldes delvis feilkoding og endel offentlige bedrifter som leverer yrkeskode allikevel. Det ser ut som noen yrker er mer generelle enn man skulle tro utfra tittelen, som kunne tyde på at yrkeskoden er brukt feil. Vi må imidlertid påpeke at beregningene er utfra *næringsundergruppe* (5-sifret NACE), mens yrkesstandarden grupperer etter 2 siffer og enda mer aggregerte nivåer av næring. Det er mulig at spesifisitetsanalyser i seg selv kan gi grunnlag for å kontrollere enkelte yrker ved egne metoder.

5.5 Konklusjon

Vi har forsøkt flere ulike metoder for å identifisere bedrifter der yrkeskodingen bør kontrolleres. Dette er gjort ved å sammenlikne yrkesfordelingen i bedrifter. Det er varierende i hvilken grad det er mulig å sammenlikne bedriftene ved hjelp av de kjennemerker vi har til rådighet.

Utfra en situasjon med ukjente forhold og begrensede resurser, er det er ikke mulig å fastslå hvilke enheter som skal kontrolleres. Prinsippene som legges til grunn for det videre arbeid er:

- Bruke de data som foreligger i Arbeidstakerregisterfilene.
- Forsøke flere ulike metoder avhengig av delpopulasjonen som studeres.
- Sortere enhetene etter synkende prioritet, utfra parametere som størrelse, frafall og avvik.
- Kontrollere så mange som mulig innenfor de gitte rammer.
- Revisjonstiltak blir i størst mulig grad at arbeidsgiver leverer korreksjoner, og i noen grad automatisk omkoding.
- Det satses på at flest mulig enheter går over til maskinell rapportering.

På bakgrunn av forsøkene forslås å:

- Benytte "Type 1" og "Type 4". De andre typene ser ut til å gi mye av de samme resultatene.
- Benytte metode for tilfeldige feil for de næringer og yrker der dette er egnet. De metoder som er antydnet kan prøves for å avgrense delpopulasjonen av egnede grupper.

Det er planlagt å analysere effektene av revisjon i forbindelse med publisering av yrkesfordeling i den registerbaserte sysselsettingsstatistikken i 2005.

Kontrollarbeidet vil skje i nært samarbeid mellom Seksjon for Arbeidsmarked og Seksjon for Databearbeiding. Det er utviklet kontroll-lister og -skjema og planlegging av arbeidsrutiner er godt i gang. Det prioriteres å kontrollere store næringer, der det finnes bedrifter med klare avvik. Den aller første delen som undersøkes er mellomstore og større bedrifter innen handel (næringene 50-51-52). Planer, rutiner og resultater av dette vil bli dokumentert av seksjonene.

5.6 Videre utvikling

Stratifisering av bedrifter etter næring og størrelse gir i endel tilfeller strata som ikke er homogene i den grad en kunne ønske for kontroll av yrkeskoder. Hvis en kunne skille ut enda mer ensartede aktiviteter, ville det være et bedre sammenlikningsgrunnlag for yrke. Det er mulig at man kan gjøres ved å bruke mer eller mindre avanserte metoder for å gruppere etter selve yrkesfordelingen. I noen tilfeller, f.eks. bygnæringen kan det være at man kan skilles ut aktivitetsgrupper på bakgrunn av yrkesfordelingen. En feilkilde i dette ligger selvsagt i de systematiske feilene i yrkesfordelingen selv. Det må derfor først følges opp med skjønnsmessige kontroller som bedriftens navn, beskrivelse av virksomhet, osv. Videre er det mulig at en kunne nyttiggjøre andre operasjonaliserbare kjennetegn som institusjonell sektor, organisasjonsform og foretaksstruktur; selv om innføring av flere variabler lett kan gi overstratifisering.

Den siste metoddelen gir også muligheter for å analyser på tvers av enheter, f.eks. å gjennomgå yrkesfordelingen i visse næringer. Dette må kombineres med skjønn for f.eks. å finne næringer som har en påfallende fordeling av yrke, eller omvendt. Med slike metoder blir det også aktuelt å bruke utdanningsdata, samt demografiske kjennetegn. Selv om en tar høyde for betydelige endringer i forhold til tradisjonelle sammensetninger, skulle det være mulig å avdekke feil ved slike analyser. På grunn av slike forhold som f.eks. uformell kompetanse vs registerutdanning, er da også viktig å benytte statistiske metoder, og ikke basere seg på logiske kontroller alene.

Det ligger store utfordringer i å utvikle metoder også for små bedrifter og grupper. For disse områdene er det også flere administrative forhold som gjør at kontroll- og revisjonsarbeidet må legges opp annerledes. Det skyldes ulike datatyper og leveringsmåter, andre rutiner med tilbakemelding til oppgavegiver om endringer, kontroll av egne yrkeskoder, m.m. Videre kan en ikke her benytte metodene som sammenlikner andeler. Vi tar sikte på å forsøke ikke-parametriske metoder som eksistenskontroll, rangering, osv.

6 Vedlegg

6.1 Yrkesstatistikk

Oversikt over publisert statistikk med yrke fra Statistisk sentralbyrå i nyere tid (1996-2003)

Kilde Arbeidskraftundersøkelsen
Populasjon sysselsatte 16-74 år
Referansetid årgjennomsnitt
URL <http://www.ssb.no/yrkeaku/>

| Tabell | Yrkeskode | Kjennemerke | Presisjon | Aggregering |
|--------|-------------|---------------------|-----------------|---|
| 1 | Yrkesfelt | Kjønn | Hele tusen | Yrkesfelt: '0' og '9' slås sammen til '9' |
| 2 | Yrkesfelt | Kjønn | Tidels prosent | Yrkesfelt: '0' og '9' slås sammen til '9' |
| 3 | Yrke | Kjønn | Hele tusen | Publisert: 105 yrker Ikke publisert: 247 yrker |
| 4 | Yrkesfelt | Kjønn, arbeidstid | Hele tusen | Heltid / deltid |
| | Yrkesfelt | Kjønn, arbeidstid | Tidels prosent | Yrkesfelt: '0' og '9' slås sammen til '9' |
| | Yrkesfelt | Kjønn, arbeidstid | Tidels time/uke | "Avtalt/vanlig arbeidstid" |
| 5 | Yrkesområde | Kjønn, alder | Hele tusen | Alle 31 yrkesområder. Alder 4 grupper |
| 6 | Yrkesfelt | Næring | Hele tusen | 2-siffer næring: 14 grupper fra 60 |
| 7 | Yrkesfelt | Kjønn, bostedsfylke | Hele tusen | 19 fylker |
| 8 | Yrkesfelt | Kjønn, sektor | Hele tusen | Sektor: 2 grupper fra 55 sektorer |

Kilde Registerbasert sysselsettingsstatistikk
Populasjon sysselsatte 16-74 år unntatt fylke/kommunal sektor og helseforetak
Referansetid 4. kvartal
URL <http://www.ssb.no/regsyst/>

| Tabell | Yrkeskode | Kjennemerke | Presisjon | Aggregering |
|--------|-----------|---------------------------------------|-----------|--|
| 16 | Yrkesfelt | Arbeidsstedskommune | Hele tall | Yrkesfelt: '0' og '9' slås sammen til '9' |
| 17 | Yrke | Kjønn | Hele tall | 289 yrkesgrupperinger fra 352 yrker |
| 18 | Yrke | Alder | Hele tall | Alder 4 grupper 289 yrkesgrupperinger fra 352 yrker |
| SB | Yrkesfelt | Bostedskommune Arbeidsstedskommune | Hele tall | Yrkesfelt: '0' og '9' slås sammen til '9' |
| SB | Yrke | Kjønn | Hele tall | 289 yrkesgrupperinger fra 352 yrker |

6.2 Datadokumentasjon estimeringsopplegg

Denne dokumentasjonen omfatter kun plassering av filer som inngår i estimeringsopplegget til yrkesfordeling i registerbasert sysselsettingsstatistikk. Navneeksemplene er fra 4. kvartal 2003.

6.2.1 Dokumenter

\\Tor\200\260\Felles\Yrkeskoding i Register\Regsys2003

6.2.2 Programmer

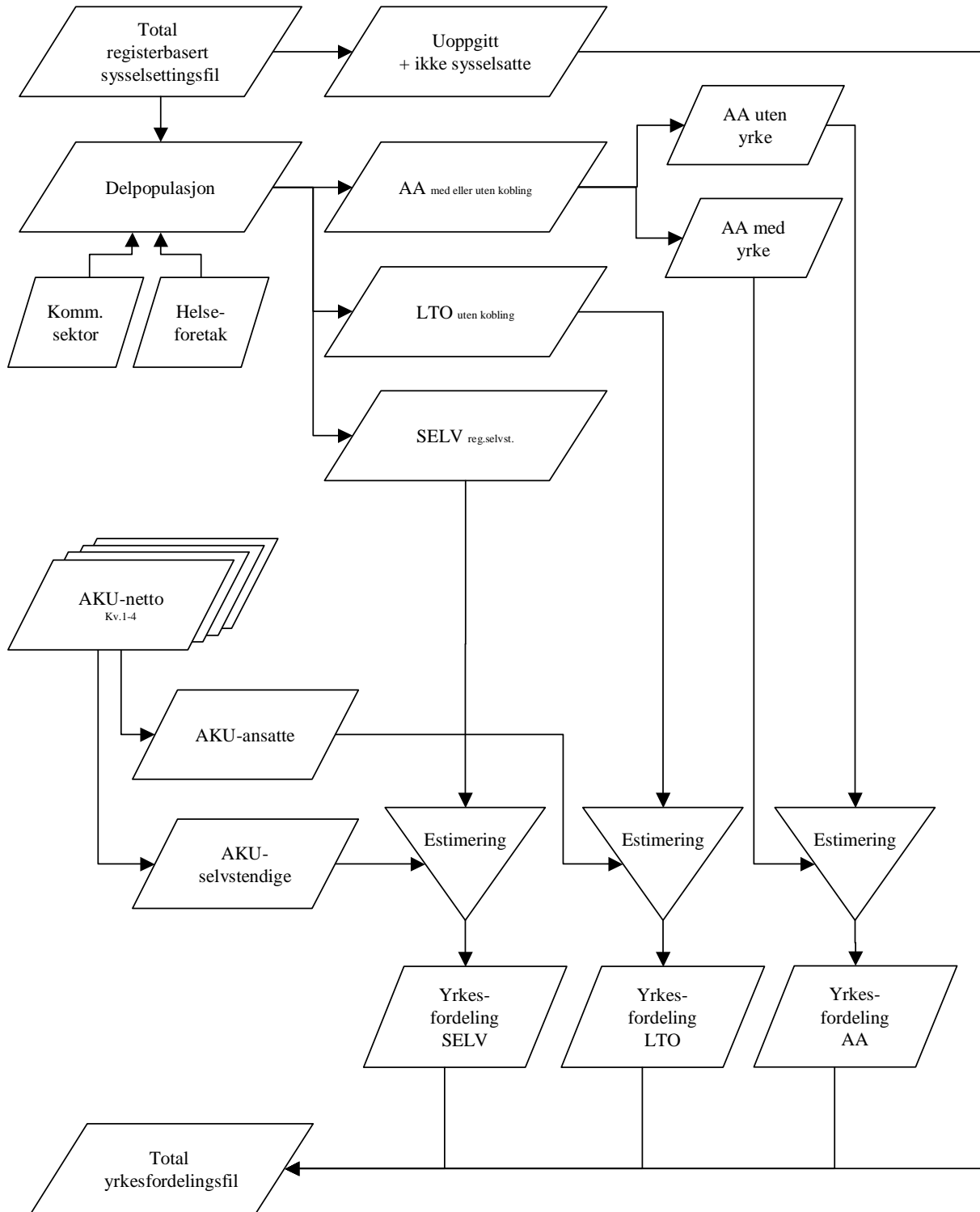
Rutineprogrammer for yrkeskoding \$YRKEREG/prog
Estimeringsopplegg \$SFP/sysdef/prog/yrke

6.2.3 Inndata

| Mappe | Fil | Innhold | Enhet |
|-------------------------|---------------|---|---------------------|
| \$SFP/sysdef/wk23/ | regsyst2003k4 | Sysselsettingsdata, hovedarbeidsforhold | person |
| \$SFP/sysdef/wk23/ | yrke2003 | Sysselsatte med koble på yrke fra Arbeidstakerregisteret | person |
| \$YRKEREG/prog/datasett | g2003 | Yrkesregister del fra Arbeidstakerregister | arbeidstakerforhold |
| \$AKU/data/tabdata/g03/ | kv04 | Utvalgsdata, AKU kvartal | person |
| \$SFP/sysdef/wk23/ | tabell_ | Datasett med aggregerte tall, til tabellbruk | variabelverdi |

6.2.4 Dataflyt

Dataflyt estimering av yrkesfordeling i register



6.3 Revisjonsopplegget

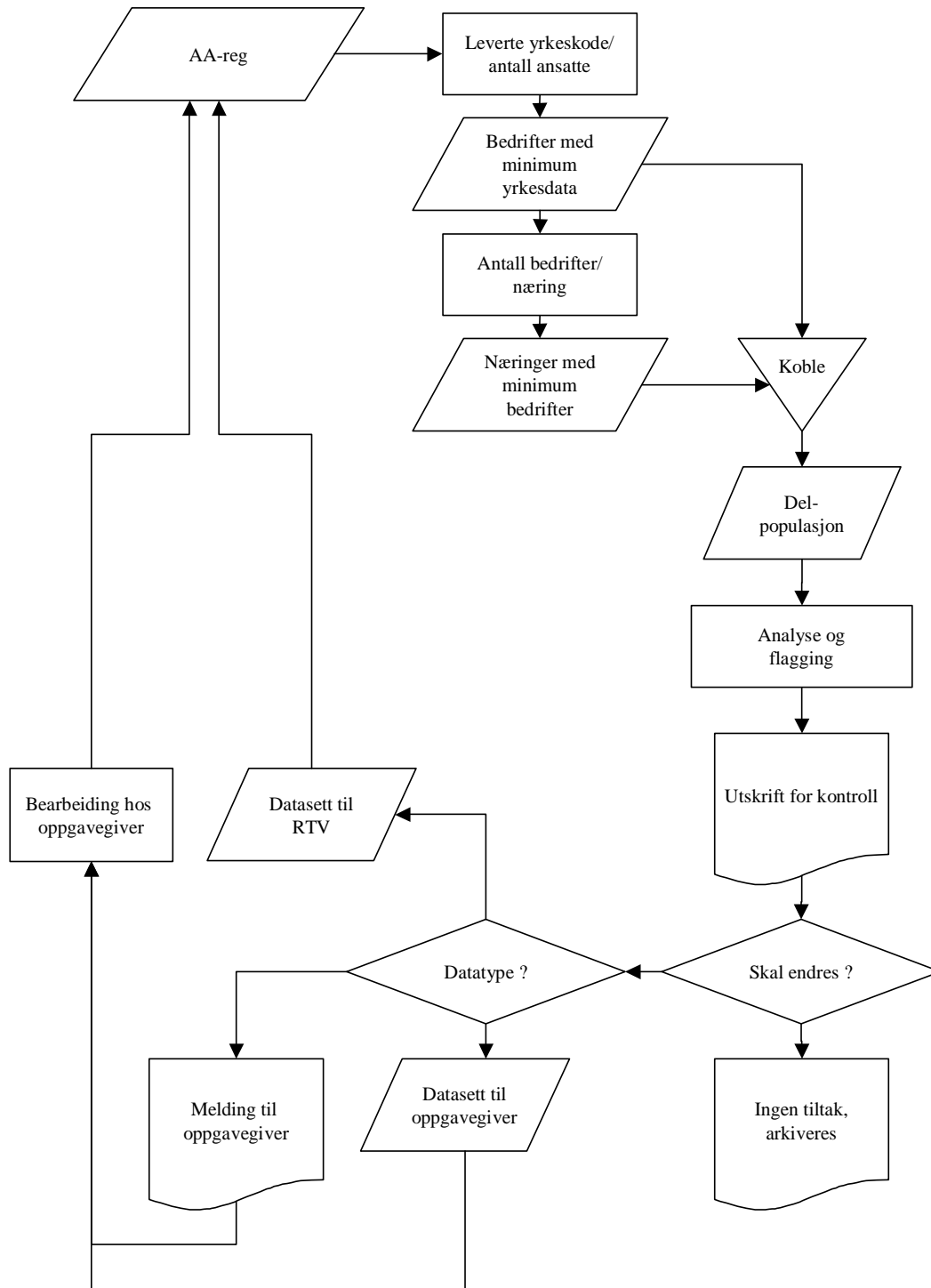
6.3.1 Dataflyt

Databehandling for analysedelen foregår på UNIX server "Ursus"

Programmer: **\$YRKEREG/prog/revisjon**

Inn: **\$ARBTAKER/periode/wk13/per04u40**

Dataflyt revisjonsopplegg



6.3.2 Stratifisering av delpopulasjon av bedrifter

Inndata i forsøkene er fil fra Arbeidstakerregisteret uke 40, en såkalt periodefil som lages hver 4. uke. Referansedato for denne fila er 5. mai 2005. Med de gitte forsøksrammer tar vi ut 4 330 bedrifter med tilsammen 349 693 arbeidstakerforhold (av totalt 2 074 327), inndelt i 142 strata, av 87 næringsundergrupper og 1, 2 eller 4 størrelsesgrupper. Størrelsesgruppene framkommer slik: Næringer med flere enn 80 bedrifter deles etter kvartilene, de med mellom 20 og 80 bare etter median (Q2), mindre næringer deles ikke.

Tabell 6-1: Stratifisering etter næringsundergrupper og størrelsesgruppene.

| Næringsundergruppe | Størrelsesgruppe | Gjennomsnitt antall ansatte pr. bedrift | | | |
|--|------------------|---|-------|------|-------|
| | | A | B | C | D |
| 01.420 Tjenester tilknyttet husdyrhold | | 42.4 | 146.4 | . | . |
| 11.200 Tjenester tilknyttet olje- og gassutvinning | | 175.9 | . | . | . |
| 15.110 Slakting, produksjon og konservering av kjøtt | | 134.9 | . | . | . |
| 15.130 Produksjon av kjøtt- og fjørfevarer | | 121.7 | . | . | . |
| 15.202 Frysing av fisk, fiskefileter, skalldyr mv. | | 41.2 | 116.2 | . | . |
| 15.209 Bearbeiding og konservering av fisk, mv. | | 55.5 | . | . | . |
| 15.510 Produksjon av meierivarer | | 48.9 | 151.2 | . | . |
| 15.810 Produksjon av brød og ferske konditorvarer | | 37.4 | 81 | . | . |
| 20.101 Saging og høvling av tre | | 57.1 | . | . | . |
| 20.302 Produksjon av bygningsartikler | | 41.6 | 115.6 | . | . |
| 22.120 Forlegging av aviser | | 52.2 | 303.9 | . | . |
| 22.220 Trykking ellers | | 67.1 | . | . | . |
| 26.610 Produksjon av betongvarer for bygge- og anl. | | 61.2 | . | . | . |
| 28.110 Produksjon av metallkonstruksjoner og deler | | 33.3 | 67.6 | . | . |
| 28.750 Produksjon av metallvarer ellers | | 66.3 | . | . | . |
| 32.100 Produksjon av elektronrør mv. | | 83.9 | . | . | . |
| 33.200 Produksjon av måle- og kontrollinstrumenter | | 87.4 | . | . | . |
| 35.111 Bygging og reparasjon av skip og skrog | | 84.5 | . | . | . |
| 35.114 Bygging og reparasjon av oljeplattformer | | 307.2 | . | . | . |
| 35.115 Innrednings- og installasjonsarbeid | | 176.7 | . | . | . |
| 36.110 Produksjon av sittemøbler | | 99.3 | . | . | . |
| 40.110 Produksjon av elektrisitet | | 36.3 | 88.5 | . | . |
| 40.120 Overføring av elektrisitet | | 95.3 | . | . | . |
| 45.110 Riving av bygninger og flytting av masse | | 51.6 | . | . | . |
| 45.211 Oppføring av bygninger | | 33.1 | 42.4 | 59.5 | 156.8 |
| 45.212 Oppføring av andre konstruksjoner | | 115 | . | . | . |
| 45.230 Bygging av veier, flyplasser og idrettsanl | | 75.6 | . | . | . |
| 45.250 Annen spesialisert bygge- og anleggsvirks. | | 69.6 | . | . | . |
| 45.310 Elektrisk installasjonsarbeid | | 32 | 38.5 | 54.2 | 110.8 |
| 45.330 VVS-arbeid | | 37.2 | 63.3 | . | . |
| 50.102 Detaljhandel med motorvogner, unntatt motorsykl. | | 34 | 60.9 | . | . |
| 50.200 Vedlikehold og reparasjon av motorvogner mv. | | 33.1 | 54.1 | . | . |
| 51.390 Engroshandel med bredt utvalg av nærings- | | 56.8 | 148.1 | . | . |
| 51.460 Engroshandel med sykepleie- og apotekvarer | | 63.4 | . | . | . |
| 51.532 Engroshandel med trelast | | 51.4 | . | . | . |
| 51.540 Engroshandel med jernvarer, rørleggerartikler | | 60.9 | . | . | . |
| 51.840 Engroshandel med datamaskiner, tilleggsuts. | | 94.8 | . | . | . |
| 51.850 Engroshandel med maskiner og utstyr til kontor | | 63.8 | . | . | . |
| 51.873 Engroshandel med maskiner og utstyr til olje, mv. | | 46.9 | . | . | . |
| 51.874 Engroshandel med maskiner og utstyr til handel. | | 52 | . | . | . |
| 52.110 Butikkhandel med bredt vareutvalg nær./nyt. | | 31.4 | 36.4 | 47 | 92.7 |

| Næringsundergruppe | Størrelsesgruppe | Gjennomsnitt antall ansatte pr. bedrift | | | |
|--|------------------|---|-------|------|-------|
| | | A | B | C | D |
| 52.120 Butikkhandel med bredt vareutvalg ellers | | 56.8 | . | . | . |
| 52.420 Butikkhandel med klær | | 33.3 | 55.2 | . | . |
| 52.443 Butikkhandel med møbler | | 77.6 | . | . | . |
| 52.451 Butikkhandel med elektriske husholdningsapp. | | 46 | . | . | . |
| 52.461 Butikkhandel med bredt utvalg av jernvarer | | 54.1 | . | . | . |
| 55.101 Drift av hoteller, pensjonater og moteller | | 32.3 | 42.8 | 63.2 | 130.3 |
| 55.301 Drift av restauranter og kafeer | | 31 | 35.1 | 42.5 | 71.4 |
| 55.510 Kantiner drevet som selvstendig virksomhet | | 66.4 | . | . | . |
| 60.100 Jernbanetransport | | 120 | . | . | . |
| 60.211 Rutebiltransport | | 35.2 | 59.8 | 92.4 | 281.5 |
| 60.240 Godstransport på vei | | 34.7 | 73.4 | . | . |
| 62.100 Ruteflyging | | 277.8 | . | . | . |
| 63.301 Reisebyråvirksomhet | | 56.2 | . | . | . |
| 63.401 Spedisjon | | 37.7 | 109.2 | . | . |
| 64.120 Budtjenester ellers | | 147.3 | . | . | . |
| 64.210 Operatørvirksomhet på fasttelefoni | | 191.4 | . | . | . |
| 64.240 Telekommunikasjonsvirksomhet ellers | | 114.6 | . | . | . |
| 65.120 Bankvirksomhet ellers | | 31.9 | 41.1 | 67.2 | 240.3 |
| 66.030 Skadeforsikring | | 38.5 | 181.5 | . | . |
| 70.111 Boligbyggelag | | 64.7 | . | . | . |
| 72.210 Utvikling av standard programvare | | 78.9 | . | . | . |
| 72.220 Annen konsulentvirksomhet tilknyttet system mv. | | 33 | 40.8 | 57 | 231.1 |
| 73.100 Forskning og utviklingsarbeid innen naturvit. | | 93.3 | . | . | . |
| 74.110 Juridisk tjenesteyting | | 74 | . | . | . |
| 74.122 Revisjon | | 89.5 | . | . | . |
| 74.140 Bedriftsrådgiving | | 61.6 | . | . | . |
| 74.202 Byggeteknisk konsulentvirksomhet | | 92.6 | . | . | . |
| 74.209 Annen teknisk konsulentvirksomhet | | 37.7 | 129.5 | . | . |
| 74.400 Annonse- og reklamevirksomhet | | 52.7 | . | . | . |
| 74.502 Utleie av arbeidskraft | | 35.5 | 50.4 | 80.1 | 223.5 |
| 74.600 Etterforskning og vaktjenester | | 48 | 195.9 | . | . |
| 74.700 Rengjøringsvirksomhet | | 33.6 | 46.3 | 73.6 | 216.9 |
| 74.860 Callsentervirksomhet | | 74.9 | . | . | . |
| 74.879 Annen forretningsmessig tjenesteyting | | 72 | . | . | . |
| 80.102 Grunnskoleundervisning | | 31.9 | 38.4 | 45.1 | 59.2 |
| 80.210 Undervisning i allmennfag | | 59.3 | . | . | . |
| 85.114 Rehabiliterings- og opptreningsinstitusjon | | 60.4 | . | . | . |
| 85.118 Somatiske sykehjem | | 36.6 | 55.4 | 83.3 | 160.2 |
| 85.313 Omsorgsinstitusjoner for eldre og funksjonsh. | | 68.2 | . | . | . |
| 85.321 Hjemmehjelp | | 63.3 | . | . | . |
| 85.334 Arbeidstrening for ordinært arbeidsmarked | | 43.8 | 101.3 | . | . |
| 85.335 Varig vernet arbeid | | 51.6 | . | . | . |
| 90.020 Innsamling og håndtering av annet avfall | | 50.8 | . | . | . |
| 91.310 Religiøse organisasjoner | | 57.7 | . | . | . |
| 92.200 Radio og fjernsyn | | 182.4 | . | . | . |
| 93.040 Helsestudio-, massasje- og solstudiovirks. | | 40.3 | . | . | . |

7 Henvisninger

Tor Petter Bø og Inger Håland: 2002/24: *Dokumentasjon av arbeidskraftundersøkelsen (AKU)*

Aslaug Hurlen Foss: 2004/04. *Kvaliteten i arbeidsmarkedsdelen i Folke- og bolig tellingen 2001*

Villund: *notater 2004/46, 2003/80, 2003/79*

Tilleggsskriv om yrke i registerbasert sysselsettingsstatistikk:

http://www.ssb.no/vis/regsys/om_yrke.html

Portal for yrkeskoding:

<http://www.ssb.no/emner/06/yrke/>

De sist utgitte publikasjonene i serien Notater

- | | | | |
|---------|---|---------|--|
| 2004/75 | S. Blom: Holdninger til innvandrere og innvandring 2004. 54s. | 2005/1 | S. Hansen og T. Skoglund: Syssletting og lønn i historisk nasjonalregnskap. Beregninger for 1949-1969. 36s. |
| 2004/76 | A. Rolland: En inspeksjon av Elevinspektørene. 51s. | 2004/2 | FoU og innovasjonstatistikk 2001 og 2002-dokumentasjon. 82s. |
| 2004/77 | A. Rolland: KOSTRA og kvaliteten på de kommunale tjenester. 32. | 2005/3 | M. Steinnes, J. Monsrud, E. Engelién og V.V. Holst Bloch: Samferdsel og miljø. Utvikling av et norsk indikatorsett tilpasset et felles europeisk sammenligningsgrunnlag. 80s. |
| 2004/78 | J.A. Osnes: Beregningsutvalget. Dokumentasjon av SAS-systemet. 98s. | 2005/4 | E. Falnes-Dalheim og A. Falnes-Dalheim: Dokumentasjon av FoB2001. Spesifikasjoner, bearbeiding, flytdiagram for spørreskjemadelen av tellingen. Del I. 117s. |
| 2004/79 | T. Eika og T. Skjerpen. Hvitevarer 2005. Modell og prognose. 18s. | 2005/5 | E. Falnes-Dalheim, A. Falnes-Dalheim: J. Sjørbotten og B. Østvedt: Dokumentasjon av FoB2001. Spesifikasjoner, bearbeiding, flytdiagram for spørreskjemadelen av tellingen. Del II Vedlegg. 146s. |
| 2004/80 | A.K. Johnsen og T. Nøtnes: Biblioteket i fokus? Rapport fra fokusgrupper for bibliotek og informasjonssenteret i Statistisk sentralbyrå. 26s. | 2005/6 | E. Falnes-Dalheim: Bearbeiding av prøvetellingen i Stange 2000. Folke- og bolig tellingen 2001. 126s. |
| 2004/81 | H. Tønseth: Årsrapport 2003. Kontaktutvalget for helse- sosialstatistikk. 12s. | 2005/7 | S. Kwesi Baateng og S. Ferstad: Dokumentasjonsnotat for FylkesKOSTRA vidregående opplæring. Publisering av 2003-tallene. 221s. |
| 2004/82 | I. Håland og G. Næringsrud: Kontantstøtte og Arbeidskraftundersøkelsen (AKU). 28s. | 2005/8 | Ø. Linnestad og O.K. Lien: SM08 Prisindekser. Fraktindeks på utenriks sjøfart. 56s. |
| 2004/83 | L. Vågane: Omnibusundersøkelsen juli /august 2004. Dokumentasjonsrapport. 45s. | 2005/9 | E. Cometa Rauan og R. Johannessen: Forventningsindikator - konsumprisene. November 2004 - mai 2005. 18s. |
| 2004/84 | D. Spilde: Statistikk over energibruk i industrien. Dokumentasjon og brukerveiledning. 53s. | 2005/10 | A.S. Abrahamsen: Analyse av revisjon - Feilkoder og endringer i utenrikshandelstatistikken. 71s. |
| 2004/85 | L. Haakonsen: KVARTS i paksis III. Systemer og rutiner i den daglige driften. 72s. | 2005/11 | A-K. Mevik: Usikkerhet i ordrestatistikken. 22s. |
| 2004/86 | L-C. Zhang og A. Vedø: Omlegging av utvalgsplan for (AKU). 15s. | 2005/12 | A. Akselsen, S. Lien, Ø. Sivertstøl: FD - Trygd. Variabelliste. 56. |
| 2004/87 | F. Strøm: Personer uten registrert inntekt eller formue. En gjennomgang av SSBs datagrunnlag for registerbasert inntekts- og formustatistikk 30s. | | |
| 2004/88 | G. Daugstad og B. Lie: Kvalitativ forstudie til levekårsundersøkelse blant ikkevestlige innvandrere. 138s. | | |
| 2004/89 | S. Lien og Ø. Sivertstøl: Langtidsmottakere av sosialhjelp 1997-1999. 64s. ISSN 0806-3745 | | |