Statistics Norway
Statistisk sentralbyrå

Ole Villund

Yrke i sysselsettingsstatistikken

Seksjon for arbeidsmarkedsstatistikk

Innhold

1 Innledning

Vi tar sikte på at data til den registerbaserte sysselsettingsstatistikken som publiseres sommeren 2004 skal inneholde yrke for alle sysselsatte, slik at det kan lages detaljert statistikk for yrke. Ved innføringen av et nytt kjennemerke er det ikke bare viktig at kvaliteten er så god som mulig, men også at kvaliteten kan kvantifiseres og at produksjonsprosessen dokumenteres. Det er videre ønskelig at ved revisjon og justering av data skal effektene av dette kunne måles. Formålet med dette notatet er å bidra til dokumentasjonen av yrke i registerbasert offisiell statistikk.

Hensikten med yrke i sysselsettingsstatistikken er å gi yrkesfordelinger på et detaljert nivå, og for små områder. Det er ikke noe selvstendig mål å publisere aggregerte nivåer som allerede blir ivaretatt av rutinemessig AKU-statistikk. For brukere av vår statistikk er det verdt å merke seg at antall sysselsatte i AKU og den registerbaserte sysselsettingsstatistikken er like for lønnstakere og selvstendig næringsdrivende. Det er også god kvalitet på viktige variabler som yrkesstatus. Men hvis vi publiserer yrkesfordeling for deler av de sysselsatte vil summene ikke stemme overens med tilsvarende summering i publisert yrkesstatistikk i AKU. Dette skyldes både ulike kildedata og ulikheter i yrkesfordelingene på detaljert nivå. Årsakene til dette vil bli forsøkt forklart i de neste kapitler.

Sysselsettingsstatistikken baserer seg nå på en mengde registre, et av de største og viktigste er fortsatt Arbeidstakerregisteret. Det vil derfor ha stor betydning for den totale kvaliteten på yrke i sysselsettings-statistikken, og mange av undersøkelsene tar utgangspunkt i dette registeret. Fra 2001 har det blitt innrapportert yrke til Arbeidstakerregisteret fra arbeidsgivere. Vi har tidligere undersøkt kvaliteten på yrkesvariabelen ved ulike metoder, det henvises til notat 79/2003 (Villund), notat 80/2003 (Villund) og upublisert notat 18.10.2002 av Leiv Solheim. Fram til nå har mye av undersøkelsene dreiet seg om frafall / manglende innrapportering, skjevheter i rapportering og problemer med selve klassifiseringsmetodene. Yrkeskodingen har i oppstartsfasen basert seg i stor grad på automatiske og manuelle metoder for klassifisering utfra tekst. Det var da viktig å bruke metoder som kunne identifisere forskjeller i kodekvaliteter avhengig av kilde til yrkeskoden. Etter hvert som yrke er etablert i Arbeidstakerregisteret, vil de fleste arbeidstakerforhold være innmeldt med yrkeskode fra arbeidsgiver. Også de arbeidstakerforhold som er kodet ved Statistisk sentralbyrå tidligere skal etter årskontrollen være kontrollert av arbeidsgiver. Det er derfor vår vurdering at yrkeskoder i Arbeidstakerregisteret er en mer ensartet datakilde. Metodikken for å måle kvalitet blir derfor noe forskjellig fra tidligere.

1.1 Oversikt over yrkeskoding

Det kan være greit med en kortfattet gjennomgang av data med yrkeskoder som brukes i tidligere og aktuell statistikk, og annet som er nevnt i ulike aktuelle publikasjoner og i forbindelse med kommende statistikk.

Tabell 1-1: Yrke i statistikker generelt

Statistikk	Data	Yrkeskoding	Standard
Folketellingene 1980 og 1990	Fulltelling, skjema/intervju. Egenoppgitt yrke.	Manuell koding	NYK (Nordisk yrkesstandard)
Folke- og boligtellingen 2001	Ikke spurt om yrke Ikke hentet fra register	-	-
Arbeidskraftundersøkelsen	Løpende spørreundersøkelse, familieenhet. Egenoppgitt yrke.	Manuell koding Årsgjennomsnitt 4 siffer.	NYK t.o.m. 1995 STYRK (Standard for yrkesklassifisering 1998) fra 1996
Arbeidstakere	Basert på arbeidsgiveres meldinger til Arbeidstakerregisteret.	Manuell koding, automatisk koding, omkoding, og direkte levering	STYRK
Sykefravær	Hentes fra Arbeidstakerregisteret	ikke publisert yrke ennå	STYRK
Sysselsettingsstatistikken	Basert på AA og andre adm. dataregistre.	*se tabell 2*	STYRK
Lønnsstatistikken	Utvalgsundersøkelse, bedriftsenhet. Arbeidsgiver oppgir yrke.	Manuell koding, direkte levering	STYRK for endel grupper, stillingskoder,m.m.
Ledige stillinger og arbeidsledighet	Registrering av arbeidssøkere og meldinger fra arbeidsgivere.	AETAT	STYRK fra 2002

Tabell 1-2: Muligheter for yrke i Sysselsettingsstatistikken

Kilde til sysselsetting	Kilde til yrke	Andel (2002)
Arbeidstakerforhold fra Arbeidstakerregister/LTO[1] eller kun Arbeidstakerregisteret .	Kan benytte yrkeskoden i Arbeidstakerregisteret direkte, eller bruke samme metodikk som Arbeidstakerregisteret ved frafall (ca. 10%).	81%
Kun LTO, mange småjobber.	Mangler yrkeskode helt.	9%
Selvstendig næringsdrivende, spesielle yrker.	Mangler yrkeskode helt.	7%
"Andre", uensartet gruppe.	Ingen anbefaling ennå.	3%

Tabellen forteller oss at kvaliteten på yrke i Arbeidstakerregisteret har mye å si for publisering av statistikk om yrke for alle sysselsatte. Det neste kapittelet vil gjennomgå endel undersøkelser av kvaliteten på yrke i Arbeidstakerregisteret.

2 Kvalitet på yrke i Arbeidstakerregisteret

De fleste sysselsatte er registrert i Arbeidstakerregisteret. Det er derfor viktig å fastslå kvaliteten på yrke i dette registeret. Vi har tidligere funnet endel avvik mellom yrkesfordelingen i Arbeidstakerregisteret og f.eks. AKU-tall. Notat 04/2004 "Kvaliteten i arbeidsmarkedsdelen i Folke- og boligtellingen 2001" av Aslaug Hurlen Foss beskriver metodikk for konsistensanalyse av data fra sysselsettingsregistre og Arbeidskraftundersøkelsen (AKU) som er koblet på personnivå. Variablene som er undersøkt er sysselsetting og feilkilder i datagrunnlaget, yrkesstatus, aktivitetstype, kjønn, alder, utdanning, næring, m.m. Notatet gir inngående beskrivelse av variabler og definisjoner i de to datakildene, og en god presentasjon av metoder for å måle kvaliteten i disse.

I dette notatet utnytter vi deler av den samme metodikken for å undersøke kvaliteten på sysselsetting og yrkeskode i data fra 2002. Det er interessant å bruke de samme metodene som er brukt på 2001-data, for å kunne gi et sammenliknbart kvalitetsmål på yrkesvariabelen i forhold til de andre variablene som er undersøkt.

[1] Lønns- og trekkoppgaveregisteret (LTO). Statistisk sentralbyrå mottar kopi fra Skattedirektoratet hvert år, og kobler dette til Arbeidstakerregisteret ved hjelp av bl.a. fødselsnummer og organisasjonsnummer.

2.1 Data

Data fra Arbeidskraftundersøkelsen (AKU) inneholder en rekke variabler knyttet til arbeidsmarkedet. AKU er en intervjuundersøkelse med familie som primær trekkeenhet. I de trukne familier trekkes alle personer som er 16-74 år. Omtrent 24.000 personer trekkes ut, og hver person er med i 2 år. Intervjuene skjer fortløpende gjennom hele året, hver person intervjues 1 gang pr. kvartal. I de neste tabellene er det brukt reviderte datasett fra 4.kvartal 2002 til tabellbruk ($AKU/data/tabdata/g02/kv04). Det er 21578 personer i dette nettoutvalget.

Data fra Sysselsettingsstatistikken er en statusfil ($SFP/sysdef/wk23/regsys2002k4) som skal dekke omtrent samme tidspunkt. Til denne er det koblet yrkeskode fra Arbeidstakerregisteret. Yrkesdata er valgt ut fra senere sett, fram til uke 52-2003 fordi det i 2002 var endel manglende innrapporteringer på yrke, som siden har blitt komplettert. Denne tidsforskyvning kan ha en viss betydning for personer som har endret yrke i samme arbeidstakerforhold. Det er ikke foretatt imputering av de øvrige, slik det er planlagt i statistikkproduksjon.

I begge datasett er enheten *person.* Det er mange personer som har flere sysselsettingsforhold, f.eks. flere deltids arbeidstakerforhold, eller selvstendig næringsdrivende som er også ansatt i vanlige arbeidstakerforhold, eller har småjobber i tillegg. For disse personene er det i begge datasett valgt ut ett hovedforhold pr. person.

2.2 Variabler

2.2.1 Sysselsettingstype

Sysselsettingstype er valgt som en nøkkelvariabel, fordi den er viktig i statistikk, og fordi den har et kjent nivå og kvalitet. I register er sysselsettingstype omtrent lik 'yrkesstatus', men er noe forenklet pga. kategorier som ikke er så interessante i forhold til *yrkeskode*, og også for å få et annet navn på variabelen for å unngå forvirring. Kort fortalt så er *yrkeskode* en kategorisering av arbeidsoppgaver, 'yrkesstatus' beskriver i hovedsak om man er ansatt eller selvstneidg næringsdrivende. Tabellen viser konverteringen til sysselsettingstype. 'Sstat' i AKU er sysselsettingsstatus/yrkesstatus (for ikke-sysselsatte: hovedsakelig virksomhet). Gruppen "Andre" omfatter personer i litt ulike grupper. Familiearbeidere plasseres i begge data under "ansatte". Pga. av ulikheter i operasjonaliseringen av definisjonen og i datagrunnlaget, samt at det er små grupper, slår vi sammen "andre"-gruppene.

Tabell 2-1: Teknisk definisjon av sysselsettingstype

Sysselsettingstype	Register, "yrkstat"	AKU, "sstat"
1 Ansatt	1 Lønnstakere	101 Vernepliktige 111, 121 Ansatte 113, 123 Familiearbeidere 119, 129 Uoppgitt yrkesstatus
2 Selvstendig	2 Selvstendige	112, 122 Selvstendige
9 Andre	Øvrige (ledige, på tiltak, og ikke-sysselsatte)	Øvrige (arbeidssøkere og utenfor arbeidsstyrken)

2.2.2 Yrkeskode

Yrke blir kodet etter Standard for yrkesklassifisering (STYRK) fra 1998. STYRK er en tilpasset norsk versjon av ISCO88(COM) som er Eurostats versjon av ILO-standarden. Yrkeskode er en kategorivariabel med 353 yrker på det mest detaljerte nivå. I denne analysen benytter vi det mest aggregerte nivå som inneholder 10 *yrkesfelt.* Andelene i yrkesfordelingen i hvert datasett er beregnet av de som har en yrkeskode i det aktuelle settet.

I AKU blir yrke kodet manuelt utfra tekst der intervjuobjektet selv oppgir yrke og arbeidsoppgaver, samt tilleggsopplysninger som næring, utdanning og andre egenskaper ved person og bedrift.

I Arbeidstakerregisteret blir yrkeskode oppgitt av arbeidsgiver, og for en god del arbeidstakerforhold kodet utfra tekst som arbeidsgiver oppgir, og for en enda større del omkodet fra stillingskoder. Stillingskoder er koder som benyttes i offentlige lønnsregistre, og beskriver i varierende grad personens arbeidsoppgaver. Stillingskoder kan derfor være mer eller mindre egnet som grunnlag for yrkesklassifisering. Det samme kan man også si om rapportering av tekst. Rapportert tekst som *ikke beskriver arbeidsoppgaver* utgjør fortsatt en god del av grunnlaget i Arbeidstakerregisteret.

Hovedgrunner til forskjeller i yrkeskoding i AKU og register:

- Det er ulikt kildegrunnlag for yrkeskoden
  - AKU-yrke baseres på opplysninger fra den sysselsatte selv, eller nærmeste familie (omtrent 15% indirekte intervju)
  - Registeryrke baseres på opplysninger fra arbeidsgiver.
- Innenfor registerdata er det forskjeller i kildegrunnlaget
  - Yrkeskoden kan være levert direkte.
  - Omkodet på mange forskjellige måter, fra tekst, stillingskode, m.m.
- Det er mye større frafall av yrkeskode i register enn AKU, og frafallet er neppe tilfeldig. På grunn av ulik kilde, er også frafallsmodellen helt forskjellig i de to datasett. Som en hovedantagelse kan frafallet i AKU modelleres etter personvariabler, i Arbeidstakerregisteret etter bedriftsvariabler.
- AKU har i utgangspunktet yrke på alle sysselsatte, i registerdata finnes kun yrkesdata fra Arbeidstakerregisteret. Det innebærer at selvstendig næringsdrivende og småjobber ikke kan sammenliknes, da yrkesfordelingen i disse gruppene i register ikke kan estimeres på en enkel måte (se siste del av notatet).

2.3 Kobling av datasett

Person er enhet i både AKU og sysselsettingsfil og data kobles ved fødselsnummer. Det koblede datasettet inneholder 21535 records, noe som er over 99%. Det er noe usikkerhet knyttet til valg av viktigste sysselsettingsforhold i hvert av datasettene. Dette skyldes delvis ulikheter i det som kalles operasjonalisering av en formell definisjon. Man tar utangspunkt i samme definisjon, men utfra forskjellige data må man velge ulik prosedyre for å klassifisere samme personen i hvert av settene. I AKU spør man hver person om det samme på en mest mulig ensartet måte. I sysselsettingsregister må man basere seg på Arbeidstakerregisteret og andre registerdata, som kan ha varierende metoder, kilder og kvaliteter for ulike grupper.

Tabellen under viser fordelingen av datakilder i sysselsettingsdata, på de personer som er koblet til AKU-data. I Notat 04/2004 fant man at 'Koblet AA/LTO' var den mest pålitelige kilde, og vi merker oss at dette utgjør 89% av de som er klassifisert som ansatt i register i de koblede data. 'Kun LTO' er personer som blir klassifisert utfra Lønns- og trekkoppgaveregisteret uten kobling til Arbeidstakerregisteret. Disse omfatter mange småjobber som ikke blir innmeldt til Arbeidstakerregisteret. Det er også mange som er midlertidig utmeldt pga. persimisjoner, mv. Det er antatt at yrkesfordelingen i disse vil avvike betydelig fra Arbeidstakerregisteret. Det er imidlertid ikke helt enkelt å bevise dette utfra det foreliggende datasett

Tabell 2-2: Registerkilde i sysselsettingsdata, av de som er koblet til AKU-data.

	Kilde, antall				Kilde, prosent			
	I alt	Ansatt	Selvst.	Andre	I alt	Ansatt	Selvst.	Andre
I alt	21535	14622	1001	5912	100	100	100	100
Andre	5912	-	-	5912	27	-	-	100
Koblet AA/LTO	12981	12981	-	-	60	89	-	-
Kun AA	199	199	-	-	1	1	-	-
Kun LTO	1395	1395	-	-	6	10	-	-
Selvangivelse	1048	47	1001	-	5	0	100	-

2.4 Metode

Vi estimerer andelene i AKU uten bruk av vekter for å kunne sammenlikne kvalitetsmålene med de som er brukt i del 4 og 5 i notat 2004/4, hvor det er estimert uten vekter. Videre fjernes endel records fra AKU-data, som forklares nærmere i forbindelse med tabeller lenger fram. Hvis vi skulle brukt vekter i disse tabeller, måtte de rekalibreres. Generelt kan man si at ved å ikke bruke de individuelle vektene som ligger på AKU-data, får man ikke nytte av stratifisering og kalibrering som er gjort. Det må merkes at denne kompenserende metodikken er laget for å få mest mulig korrekt sysselsettingsnivå, ikke yrkesfordeling. Det kan godt være at de skjevhetene og feilene som vises pga. frafall i AKU og manglende innrapportering i Arbeidstakerregisteret, er større enn eventuelle effekter av variansreduksjon.

Den totale forskjellen mellom yrkesfordelingene skyldes i hovedsak:

- Utvalgsskjevhet, både at AKU ikke er helt representativ for befolkningen når det gjelder yrke, og at Arbeidstakerregisteret ikke er representativ når det gjelder alle sysselsatte.

- Systematisk forskjellig koding: altså skjevhet som skyldes ulikheter i data- eller kodemetode.
- Tilfeldig forskjellig koding:

Under vises teoretiske beregninger av feiltyper.

Formel 2–1: Indikatorvariabel

$$y_{k_i} = \begin{cases} 0 \, hvis \;\; personen \;\; "i" \;\; har \;\; egenskapen \;\; i \, kilden \;\; "k" \\ 1 \, hvis \;\; personen \;\; "i" \;\; ikke \; har \;\; egenskapen \;\; i \, kilden \;\; "k" \end{cases}$$

Formel 2–2: Andel med egenskap i register

$$\hat{p}_{reg} = \sum_{i=1}^{n} y_{reg\,i} \,/\, n$$

Formel 2–3: Andel med egenskap i AKU

$$\hat{p}_{AKU} = \sum_{i=1}^{n} y_{aku\,i} \,/\, n$$

Formel 2–4: Systematiske målefeil

$$\hat{\beta} = \hat{p}_{reg} - \hat{p}_{AKU}$$

Formel 2–5: Relativ systematisk målefeil

$$R_{syst.} = \frac{100\% \cdot \hat{\beta}}{p_{reg}}$$

Formel 2–6: Tilfeldig målefeil

$$\hat{\tau}^2 = (p_I + p_{II}) - (p_I - p_{II})^2$$

der andelene med samsvar og ikke samsvar er gitt ved:

	Har egenskap i AKU	**Har ikke egenskap i AKU**
Har egenskap i reg.	$P_{reg} * P_{aku}$	P_{II}
Har ikke egenskap i reg.	P_I	

Formel 2–7: Relativ tilfeldig målefeil

$$R_{tilf.} = \frac{100\% \cdot \frac{\hat{\tau}}{\sqrt{n}}}{p_{reg.}}$$

Formel 2–8: Enkelt estimat av utvalgsusikkerhet

$$\hat{SE}(\hat{p}_{AKU}) = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$$

2.5 Resultater

2.5.1 Sysselsettingstype

Vi undersøker først denne variabelen for å kunne sammenlikne med året før, og for å kunne sammenlikne med kvalitetsnivå på yrkesvariabelen. Vi kan først konstatere at resultatene fra 2002 er ganske like de i 2001. Det gir

grunn til å tro at datagrunnlaget er relativt stabilt, og vi går videre med denne metoden for å sammenlikne med kvaliteten på yrkesvariabelen.

Tabell 2-3: Sysselsettingstype i register og AKU, 2002k4. Andeler og systematiske og tilfeldige feil

	AKU		Register		Systematiske feil + utvalgskjevhet		Tilfeldige feil	
	antall	andel	antall	andel	feil	relative	feil	relative
Ansatt	14258	0.66208	2115000	0.6540	-0.0081	-1.24 %	0.0020	0.30 %
Selvstendig	1037	0.04815	152000	0.0470	-0.0012	-2.46 %	0.0016	3.34 %
Andre	6240	0.28976	967083	0.2990	0.0093	3.10 %	0.0017	0.57 %

2.5.2 Yrke

Den første tabellen viser analyse av alle personer i de koblede data. Det er betydelige skjevheter når vi bruker data som er kun er koblet på personnivå. Vi vet at det er store forskjeller i yrkesfordelingen mellom ansatte/lønnsmottakere og selvstendige næringsdrivende. Det er derfor interessant å justere dette for å få en mer realistisk sammenlikning av yrkesfordelingen. De to påfølgende tabellene viser data med kun ansatte i hvert sett, og kun de som er definert som ansatte både i AKU og register. Det siste betyr at både de selvstendige og de som er krysskoblet blir holdt utenfor beregningene. Dette utfra en forestilling om at en person ofte har ulike yrker i sitt arbeidstakerforhold og sitt selvstendigforhold. Dette vil imidlertid ikke framgå direkte av de benyttede data, da det her kun er valgt hovedsysselsettingsforholdet i begge sett. Imidlertid vil forskjellen mellom tabelldel *b* og *c* indirekte vise at man ikke alltid klarer å treffe samme jobb.

Tabell 2-4: Yrkesfelt i register og AKU, 2002k4. Andeler og systematiske skjevheter i ulike utvalg.

a: Alle sysselsatte, ukorrigert	AKU		Register		Systematiske feil, m.m.	
	antall	andel	antall	andel	feil	relative
1 Lederyrker	1 163	7.5 %	139 553	7.0 %	-0.005033	-7.15 %
2 Akademiske yrker	1 664	10.8 %	192 192	9.7 %	-0.010971	-11.32 %
3 Høyskoleyrker	3 589	23.3 %	422 604	21.3 %	-0.019591	-9.20 %
4 Kontor- og kundeservice	1 233	8.0 %	186 610	9.4 %	0.014155	15.04 %
5 Salgs- og serviceyrker	3 466	22.5 %	511 875	25.8 %	0.033391	12.94 %
6 Bønder, fiskere o.l.	538	3.5 %	20 656	1.0 %	-0.024462	-234.90 %
7 Håndverkere o.l.	1 737	11.3 %	191 202	9.6 %	-0.016202	-16.81 %
8 Operatører, sjåfører m	1 229	8.0 %	184 729	9.3 %	0.013466	14.46 %
9 Andre yrker	807	5.2 %	134 002	6.8 %	0.015247	22.57 %
I alt	15 426	100 %	1983 423	100 %		

b: Kun ansatte i hvert sett	AKU		Register		Systematiske feil, m.m.	
	antall	andel	antall	andel	feil	relative
1 Lederyrker	1 126	7.9 %	137 124	7.2 %	-0.007441	-10.34 %
2 Akademiske yrker	1 535	10.8 %	186 960	9.8 %	-0.010129	-10.33 %
3 Høyskoleyrker	3 446	24.3 %	413 390	21.7 %	-0.026057	-12.02 %
4 Kontor- og kundeservice	1 199	8.5 %	179 600	9.4 %	0.009695	10.29 %
5 Salgs- og serviceyrker	3 291	23.2 %	491 718	25.8 %	0.025958	10.06 %
6 Bønder, fiskere o.l.	165	1.2 %	18 061	0.9 %	-0.002157	-22.76 %
7 Håndverkere o.l.	1 519	10.7 %	183 672	9.6 %	-0.010726	-11.13 %
8 Operatører, sjåfører m	1 124	7.9 %	173 409	9.1 %	0.011735	12.90 %
9 Andre yrker	781	5.5 %	122 336	6.4 %	0.009121	14.21 %
I alt	14 186	100 %	1906 270	100 %		

c: Kun ansatte i register	AKU		Register		Systematiske feil	
	antall	andel	antall	andel	feil	relative
1 Lederyrker	1 098	8.2 %	137 124	7.2 %	-0.010345	-14.38 %
2 Akademiske yrker	1 478	11.1 %	186 960	9.8 %	-0.012677	-12.93 %
3 Høyskoleyrker	3 316	24.8 %	413 390	21.7 %	-0.031625	-14.58 %
4 Kontor- og kundeservice	1 149	8.6 %	179 600	9.4 %	0.008116	8.61 %
5 Salgs- og serviceyrker	3 015	22.6 %	491 718	25.8 %	0.03202	12.41 %
6 Bønder, fiskere o.l.	111	0.8 %	18 061	0.9 %	0.001157	12.21 %
7 Håndverkere o.l.	1 439	10.8 %	183 672	9.6 %	-0.011479	-11.91 %
8 Operatører, sjåfører m	1 061	8.0 %	173 409	9.1 %	0.011462	12.60 %
9 Andre yrker	678	5.1 %	122 336	6.4 %	0.01337	20.83 %
I alt	13 345	100 %	1906 270	100 %		

Neste tabell viser samsvaret på mikronivå, altså yrkeskoden til en person i register og yrkeskoden i AKU til samme person. Det totale samsvaret i gruppene ligger altså mellom 60-78%, selv på det mest aggregerte nivå av yrkeskoder.

Tabell 2-5: Yrkesfelt i register og AKU, 2002k4. Antall og prosent.

		AKU-yrker (forklaringsnøkkel til venstre)								
Register-yrke	I alt	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***
I alt	12322	1016	1341	3033	1079	2810	95	1328	996	624
1 Ledere	975	601	71	155	43	54	7	17	24	3
2 Akademiske	1203	78	830	252	20	9	1	10	2	1
3 Høyskole	2844	125	359	2062	118	84	3	67	16	10
4 Kontor- og kunde.	1151	53	29	161	751	92	.	9	32	24
5 Salgs- /serv.	3086	97	26	267	76	2412	7	47	33	121
6 Bønder, fiskere .	87	2	.	4	3	6	59	4	5	4
7 Håndverkere .	1196	26	13	64	18	29	7	924	97	18
8 Operatører, sjåfører	1129	26	12	47	23	32	8	188	739	54
9 Andre	651	8	1	21	27	92	3	62	48	389

	I alt	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***
1 Ledere	100	62	7	16	4	6	1	2	2	0
2 Akademiske	100	6	69	21	2	1	0	1	0	0
3 Høyskole	100	4	13	73	4	3	0	2	1	0
4 Kontor- og kunde.	100	5	3	14	65	8	.	1	3	2
5 Salgs- /serv.	100	3	1	9	2	78	0	2	1	4
6 Bønder, fiskere .	100	2	.	5	3	7	68	5	6	5
7 Håndverkere .	100	2	1	5	2	2	1	77	8	2
8 Operatører, sjåfører	100	2	1	4	2	3	1	17	65	5
9 Andre	100	1	0	3	4	14	0	10	7	60

Tabell 2-6: Yrkesfelt i register og AKU, 2002k4. Sammenlikning av systematiske og tilfeldige feil.

	AKU andel	Register andel	Systematiske feil	Tilfeldige feil
1 Lederyrker	8.2 %	7.2 %	-14.38 %	4.54 %
2 Akademiske yrker	11.1 %	9.8 %	-12.93 %	3.31 %
3 Høyskoleyrker	24.8 %	21.7 %	-14.58 %	1.69 %
4 Kontor- og kundeservice	8.6 %	9.4 %	8.61 %	3.43 %
5 Salgs- og serviceyrker	22.6 %	25.8 %	12.41 %	1.46 %
6 Bønder, fiskere o.l.	*0.8 %*	*0.9 %*	*12.21 %*	*22.31 %*
7 Håndverkere o.l.	10.8 %	9.6 %	-11.91 %	3.14 %
8 Operatører, sjåfører m	8.0 %	9.1 %	12.60 %	3.68 %
9 Andre yrker	5.1 %	6.4 %	20.83 %	4.73 %
I alt	100 %	100 %		

Tabell 2-6 viser beregninger som er gjort på data med kun personer som er ansatte i register og AKU, jf. tabell 2-3. Det vises tydelig at de tilfeldige feilene er små sammenliknet med de systematiske skjevhetene. For den minste gruppen, yrkesfelt 6, er nok resultatet usikkert, men for de øvrige kan vi trygt slå fast at det eksisterer en reell systematisk forskjell i kodingen. Utvalgsusikkerheten er til sammenlikning mindre enn de andre feiltypene.

2.6 Konklusjon så langt

Det er fortsatt betydelige forskjeller i yrkeskodingen i AKU og register. Dette har også tidligere undersøkelser vist. Selv når rapporteringsgraden til Arbeidstakerregisteret er blitt bedre, og selv når vi forsøker å isolere liknende ansettelsesforhold i begge kilder, finner vi store forskjeller. Dette skyldes både datagrunnlaget, datakildene, og metodene for klassifisering. I AKU har man konkret spørsmål om arbeidsoppgaver, i tillegg til yrkestittel. Dette gjør at de som koder yrke kan foreta en mer individuell vurdering av yrkeskoden. Videre er det litt ulikt innhold i variabler for næring og utdanning, hvor det foretas manuell koding for endel i AKU, mens det i register kun kobles på informasjon fra andre registre.

Utfra en faglig vurdering kan man betrakte yrkeskode på en person som mest korrekt når den er klassifisert utfra komplette data i AKU. Allikevel må vi konkludere med at de forskjellene i yrkeskodingen vi finner, ikke kan fjernes helt med de data- og metodemessige ulikheter som eksisterer. Behovet for detaljerte yrkesdata og yrkesfordeling i små områder, gjør at vi velger å publisere registerbaserte yrke selv om det for enkelte brukere kan framstå som inkonsistent med publiserte AKU-yrker.

3 Revisjon av yrkeskoder innen bestemte grupper

Som ledd i kvalitetsarbeidet med yrke i Arbeidstakerregisteret er det interessant å sammenlikne med andre uavhengige kilder til yrkesdata. En slik kilde er yrkesdata fra lønnsstatistikken. I denne delen ser vi på en gruppe arbeidstakere hvor vi finner yrkesdata fra Arbeidstakerregisteret, lønnsstatistikken og AKU. Dette har utgangpunkt i utveksling av data med Seksjon for lønnsstatistikk, som mottar yrkestitler og -koder på mange arbeidstakere.

Formålet med sammenlikningen her er i første rekke å identifisere enkelte yrker som er problematiske å klassifisere, for om mulig å finne metoder til omkoding. Det kan også være aktuelt å bruke dette til revisjon av yrkeskoden til enkeltpersoner. Det kan også si noe om kvaliteten på disse delene av Arbeidstakerregisteret. Vi ser her på yrkeskoder på det mest detaljerte nivå (4 siffer), og på små grupper. De resultater man kommer fram til er ikke nødvendigvis representative for yrkeskoding av arbeidstakerforhold generelt.

3.1 Finansnæringen

3.1.1 Sammenlikning av yrke i lønnsdata og Arbeidstakerregisteret

Seksjon 260 har fått tilgang til 42 110 records med yrkeskoder lønnsstatistikken. Vi har koblet disse data til Arbeidstakerregisteret, og ser på samsvar i yrkeskodingen. Kobler på personnummer og organisasjonsnummer bedrift, vi går utfra at samme jobben. Vi kjenner ikke til i hvilken grad man endrer yrke innen samme ansettelsesforhold. Det er 26 946 personer (64%) som kan kobles på denne måten og som har yrkeskode i begge datakilder. Tabellen viser hvor mange som har samme yrke på ulike detaljeringsnivå hvor 0 betyr helt ulik og de andre tallene at fra 1-4 siffer i koden er like.

Tabell 3-1: Samsvar av yrke i finansnæringen, etter næring og antall like siffer i koden.

Antall	*I alt*	*0*	*1*	*2*	*3*	*4*
I alt	26 946	7 002	402	1 196	1 195	17 151
Finansiell tj. u/ forsikr. og pensjonsf.	17 579	4 549	214	740	636	11 440
Forsikring og pensjonsfond u/ off. trygdeordn.	6 255	1 336	100	262	248	4 309
Hjelpevirksomhet for finansiell tjenesteyting	3 112	1 117	88	194	311	1 402
Prosent	***I alt***	***0***	***1***	***2***	***3***	***4***
I alt	100	26	1	4	4	64
Finansiell tj. u/ forsikr. og pensjonsf.	100	26	1	4	4	65
Forsikring og pensjonsfond u/ off. trygdeordn.	100	21	2	4	4	69
Hjelpevirksomhet for finansiell tjenesteyting	100	36	3	6	10	45

Vi ser at samsvaret er nokså likt mellom næringene og at rundt ¾ av alle har lik eller liknende yrkeskode i de to datakildene. For den minste gruppen er det noe dårligere samsvar, uten at vi kjenner noen spesiell grunn til det. Generelt er det vanskelig å klassifisere yrker innen tjenestenæringer med varierte arbeidsoppgaver og bruk av generelle titler. En annen ting som vi ser her, som også har vist seg i andre undersøkelser, er at de aller fleste har enten helt like 4 siffer yrkeskode, eller helt ulike. Det betyr at sammenlikning på 1-3 siffer ikke gir så mye mer informasjon.

Den neste tabellen viser samsvar i yrkeskodene etter likhet i teksten som er oppgitt som yrkestittel. Tekstlikhet er beregnet ved en innebygd SAS-funksjon som beregner staveavstanden. Dette måltallet er slik at 0 betyr helt lik, og jo høyere tall, desto mer ulike er tekstene. I tabellen er målet på tekstlikhet gruppert. Det er kun 11269 (42%) som har oppgitt tekst i begge datasettene, de øvrige er kodet utfra levert stillingskode eller yrkeskode. Det er tydelig at ulik tekst om samme jobb/person, kan være en årsak til ulike yrkeskoder. Allikevel er det mange som har svært lik tekst og helt ulik yrkeskode. I utgangspunktet skal to personer i samme næring med samme yrkestittel få lik yrkeskode, noe som vil bli fulgt opp med revisjonstiltak.

Tabell 3-2: Samsvar i yrkeskode etter likhet i yrkestittel. Antall og prosent.

	I alt	*0*	*1*	*2*	*3*	*4*
I alt	11269	3067	159	623	656	6764
0	4551	928	56	254	114	3199
15	722	185	14	35	25	463
30	574	174	10	27	28	335
45	1107	213	19	27	92	756
60	1058	270	9	62	120	597
75	3257	1297	51	218	277	1414
	I alt	***0***	***1***	***2***	***3***	***4***
I alt	100	27	1	6	6	60
0	100	20	1	6	3	70
15	100	26	2	5	3	64
30	100	30	2	5	5	58
45	100	19	2	2	8	68
60	100	26	1	6	11	56
75	100	40	2	7	9	43

3.1.2 Kvalitet på enkelte yrker

Vi bruker samme metoder som i kapittel 1 på data som er koblet fra Arbeidstakerregisteret og lønnsdata. Det er her mer tvilsom hva som kan kalles riktig og feil kode. Systematiske "feil" vil betegne i hvor stor grad arbeidstakerforhold innen et bestemt yrke i register er kodet til et annet yrke i lønnsdata, enn omvendt. Tilfeldige "feil" vil her betegne i hvor stor grad det er kodet ulikt, men like mye fra det ene til det andre som omvendt.

Tabell 3-3: Utvalgte yrkeskode i de to datakilder. Antall, andeler og relative størrelser.

	Register		Lønnsdata		Systematiske		Tilfeldige	
	antall	andel	antall	andel	feil	relative	feil	relative
1210 ADMINISTRERENDE DIREKTØRER	321	1.2 %	467	1.4 %	-0.002	-21 %	0.003	26 %
1226 PRODUKSJONSDIREKTØRER INNEN FINANS	2464	9.1 %	2107	6.5 %	0.027	29 %	0.003	3 %
1231 FINANS-, ØKONOMI- OG ADMINISTRASJON	678	2.5 %	1321	4.1 %	-0.016	-62 %	0.003	10 %
1234 MARKEDS- OG INFORMASJONSDIREKTØRER	302	1.1 %	245	0.8 %	0.004	33 %	0.003	26 %
2130 SYSTEMUTVIKLERE OG PROGRAMMERERE	548	2.0 %	901	2.8 %	-0.007	-36 %	0.003	14 %
2512 PERSONAL- OG ORGANISASJONSKONSULENT	438	1.6 %	219	0.7 %	0.010	58 %	0.003	18 %
2519 MARKEDSANALYTIKERE OG ANDRE FORRETN	372	1.4 %	258	0.8 %	0.006	42 %	0.004	27 %
2541 SOSIAL- OG SIVILØKONOMER	654	2.4 %	1199	3.7 %	-0.013	-52 %	0.004	16 %
3411 FINANSMEGLERE	881	3.3 %	878	2.7 %	0.006	17 %	0.003	9 %
3412 FORSIKRINGSKONSULENTER	4551	16.9 %	4252	13.1 %	0.038	22 %	0.002	1 %
3415 TEKNISKE OG KOMMERSIELLE SALGSREPR.	272	1.0 %	664	2.0 %	-0.010	-103 %	0.003	25 %
3418 BANKFUNKSJONÆRER	11212	41.6 %	13023	40.1 %	0.015	4 %	0.002	1 %
3419 MARKEDSFØRINGS- OG REKLAMEKONSULENT	256	1.0 %	354	1.1 %	-0.001	-15 %	0.003	33 %
3432 REVISORER (IKKE STATSAUTORISERTE) O.L.	267	1.0 %	290	0.9 %	0.001	10 %	0.004	36 %
4113 SEKRETÆRER	395	1.5 %	381	1.2 %	0.003	20 %	0.003	19 %
4114 KONTORMEDARBEIDERE	397	1.5 %	426	1.3 %	0.002	11 %	0.004	25 %
4121 ØKONOMIMEDARBEIDERE OG REVISJONSASS.	304	1.1 %	326	1.0 %	0.001	11 %	0.003	28 %
4212 POST- OG BANKKASSERERE	301	1.1 %	787	2.4 %	-0.013	-117 %	0.002	20 %

3.1.3 Sammenlikning lønnsdata og Arbeidskraftundersøkelsen (AKU)

Kobler de samme data fra lønnsstatistikken og et AKU-datasett med personer fra hele 2003. Årsaken til at vi tar med hele året, er at finansnæring er såpass liten at et AKU-utvalg fra kun et kvartal blir beheftet med stor utvalgsusikkerhet. Metoden skaper på den annen siden en viss tidsforskjell, som kan ha betydning for de som endrer yrke. Siden det kobles kun på fødselsnummer, blir det en viss usikkerhet knyttet til valg av jobb. For å bøte på dette velges kun de som har relevant næringskode i AKU-data. Det gir et delutvalg på 1233 personer, hvorav 1043 (85%) kan kobles. At det er lavere koblingsandel enn for hele utvalget skyldes nok ulikheter i næringskodene.

Vi benytter samme metodikk som i avsnitt 1 og undersøker samsvaret på detaljert nivå. De tilfeldige feilene er langt større enn eventuelle systematiske forskjeller. Eksempel på dette i tabellen under, som kun viser de største yrkene. En tilsvarende analyse på 1-siffer nivå viser samme tendens. Det er for små grupper til at vi kan trekke noen slutninger om de enkelte yrker i dette materialet.

Tabell 3-4: Sammenlikning av yrkesfordelingene, utvalgte 4-siffer yrkeskoder.

	AKU		Lønnsdata		Systematiske		Tilfeldige	
	antall	andel	antall	andel	feil	relative	feil	relative
1226 PRODUKSJONSDIREKTØRER INNEN FINANS	135	0.12943	73	0.06999	0.059444	46 %	0.06259	48 %
1231 FINANS-, ØKONOMI- OG ADMINISTRASJON	36	0.03452	38	0.03643	-0.001918	-6 %	0.09024	261 %
2130 SYSTEMUTVIKLERE OG PROGRAMMERERE	27	0.02589	20	0.01918	0.006711	26 %	0.10304	398 %
3411 FINANSMEGLERE	45	0.04314	26	0.02493	0.018217	42 %	0.08453	196 %
3412 FORSIKRINGSKONSULENTER	190	0.18217	172	0.16491	0.017258	9 %	0.06386	35 %
3418 BANKFUNKSJONÆRER	390	0.37392	399	0.38255	-0.008629	-2 %	0.07627	20 %

Det totale samsvaret, altså hvor mange som har nøyaktig samme yrkeskode i begge datasett er ca. 70%. Dette er ikke så ulikt det vi finner i andre sammenlikninger av detaljert yrkeskode. Vi kan slå fast at det er betydelige forskjeller i yrkeskodingen i alle tre datasett som er sammenliknet.

3.1.4 Sammenkobling av alle tre datasett

Forsøker å koble sammen Arbeidstakerregisterdata, AKU-data og Lønnsdata. Får da noen færre records, til sammen 829 personer. Sammenlikner yrkeskodene på tvers av alle tre kilder. Tabellen viser at det er kun 45%

som har lik yrkeskode i alle tre datakilder. På den annen side så har nærmere 90% sammenfallende yrkeskoder i minst to kilder.

Tabell 3-5: Samsvar i yrke (4-siffer yrkeskode) i alle tre datasett.

Samsvar	Antall	Prosent
Ingen er like	97	11.7
AKU og register	115	13.9
Lønn og register	144	17.4
Lønn og AKU	98	11.8
Alle tre er like	375	45.2
	829	100

3.1.5 Konklusjon for finansnæringen

Utfra den spesielle gruppen som er undersøkt anbefales at man ser nærmere på enkelte yrker, som f.eks. 3418 (bankfunksjonærer) og 1226 (avdelingsdirektører). Det vi har brukt tidligere er visuell gjennomgang av alle detaljer knyttet til arbeidstakerforhold for utvalgt personer med disse kodene. Slike metoder kan ofte gi ideer om hvordan man kunne tenke seg en eventuell omkoding. For å kunne komme med konkrete anbefalinger om å fastsette nye yrkeskoder på arbeidstakerforhold må følgende kartlegges:

- Identifisere hvilke grupper som skal omkodes, f.eks. ved de metoder som er brukt her.
- Hvilke kriterier som skal ligge til grunn, altså hvilke variabler vi har til rådighet i den aktuelle datakilde. Det vil f.eks. ikke kunne gjøres utfra tekst, siden dette ikke omfatter alle arbeidstakerforhold.
- Hvilken yrkeskode man skal kode *til*, altså hvilket alternativ er mest korrekt.
- Metodikk for å måle effekten av revisjon.
- Kostnader.

Vi kan ikke i øyeblikket komme med konkrete anbefalinger om revisjon av yrkeskoder i denne gruppen.

3.2 Varehandel

Vi foretar endel av de samme analysene av ansatte i varehandelen, etter å ha mottatt data med 150 065 records fra lønnsstatistikken. Innen de aktuelle næringene finner vi 308 999 arbeidstakerforhold i register, hvorav 112 928 (75%) kan kobles ved hjelp av fødselsnummer og bedriftsnummer.

Tabell 3-6: Samsvar av yrke i varehandelen, etter næring og antall like siffer i koden. Prosent og tall.

	I alt	*0*	*1*	*2*	*3*	*4*
I alt	100	22	3	4	4	67
50 Handel med, vedlikehold og reparasjon	100	20	3	4	4	68
51 Agentur- og engroshandel, unntatt motorvogner	100	40	4	6	9	41
52 Detaljhandel, unntatt med motorvogner	100	13	2	2	1	82

	I alt	*0*	*1*	*2*	*3*	*4*
I alt	112 928	24 932	2 909	4 454	4 434	76 199
50 Handel med, vedlikehold og reparasjon	13 954	2 833	410	620	607	9 484
51 Agentur- og engroshandel, unntatt motorvogner	35 675	14 116	1 382	2 289	3 175	14 713
52 Detaljhandel, unntatt med motorvogner	63 299	7 983	1 117	1 545	652	52 002

Innen detaljvarehandelen, som er den største næringen, må man si at samsvaret er meget godt: 87% har lik eller liknende yrkeskode. I agentur- og engroshandel er det over 14 000 personer som har helt forskjellig yrkeskode. Her vet vi fra tidligere at det har vært problemer med å kode salgsarbeid hvor yrkeskoden tilsvarer kompetanse på høyskolenivå. Å skille disse fra arbeidsoppgaver innen salg med lavere kompetansekrav er i mange tilfeller ikke enkelt utfra yrkestittelen alene. Den neste tabellen viser sammenlikning av andelene av de største yrkene på det mest detaljerte nivå. Andelene er målt i hvert datasett for seg, altså ikke koblet. Forskjellene i andeler avspeiler da både systematiske avvik i kodingen og skjevheter i utvalget. Selv om det er betydelige avvik på enkeltyrker, kan vi slå fast at det totale samsvaret er bedre enn f.eks. for finansnæringen.

Tabell 3-7: Sammenlikning av yrkesfordelingene, utvalgte 4-siffer yrkeskoder.

	Register		Lønnsdata		Systematiske feil + utvalgsskjevhet	
	andel	antall	andel	antall	feil	relative
1210 ADMINISTRERENDE DIREKTØRER	2.43 %	7 495	1.52 %	2 280	0.91 %	37.36 %
1224 PRODUKSJONSDIREKTØRER INNEN VAREHAN	0.94 %	2 897	2.91 %	4 365	-1.97 %	-210.30 %
1231 FINANS-, ØKONOMI- OG ADMINISTRASJON	0.66 %	2 041	0.53 %	790	0.13 %	20.30 %
1233 SALGSDIREKTØRER	0.73 %	2 267	0.95 %	1 432	-0.22 %	-30.07 %
1234 MARKEDS- OG INFORMASJONSDIREKTØRER	0.74 %	2 288	0.42 %	625	0.32 %	43.75 %
1239 ANDRE SPESIALDIREKTØRER	0.20 %	625	0.60 %	902	-0.40 %	-197.20 %
1314 LEDERE INNEN VAREHANDEL	2.53 %	7 817	3.25 %	4 877	-0.72 %	-28.47 %
2130 SYSTEMUTVIKLERE OG PROGRAMMERERE	0.60 %	1 843	0.79 %	1 192	-0.20 %	-33.18 %
3228 RESEPTARER	0.29 %	902	0.55 %	822	-0.26 %	-87.65 %
3415 TEKNISKE OG KOMMERSIELLE SALGSREPRE	5.01 %	15 488	5.49 %	8 239	-0.48 %	-9.54 %
3419 MARKEDSFØRINGS- OG REKLAMEKONSULENT	0.66 %	2 032	0.54 %	812	0.12 %	17.72 %
3431 FUNKSJONÆRER INNEN ADMINISTRASJON	0.10 %	319	1.79 %	2 683	-1.69 %	-1632 %
4113 SEKRETÆRER	0.93 %	2 878	0.92 %	1 383	0.01 %	1.05 %
4114 KONTORMEDARBEIDERE	2.97 %	9 176	1.55 %	2 329	1.42 %	47.74 %
4121 ØKONOMIMEDARBEIDERE OG REVISJONSASS	1.28 %	3 947	0.97 %	1 458	0.31 %	23.94 %
4131 LAGERMEDARBEIDERE OG MATERIALFORVAL	5.32 %	16 446	7.71 %	11 569	-2.39 %	-44.85 %
5137 APOTEKTEKNIKERE	1.28 %	3 967	2.39 %	3 587	-1.11 %	-86.19 %
5221 BUTIKKMEDARBEIDERE O.L.	49.27 %	152 234	47.20 %	70 833	2.07 %	4.19 %
5224 SELGERE (ENGROS)	5.72 %	17 684	1.99 %	2 992	3.73 %	65.16 %
7217 BILSKADEREPARATØRER	0.54 %	1 653	0.50 %	752	0.03 %	6.33 %
7231 BILMEKANIKERE	4.44 %	13 732	3.73 %	5 597	0.71 %	16.07 %
8323 LASTEBIL- OG VOGNTOGFØRERE	0.51 %	1 582	0.17 %	255	0.34 %	66.81 %
9132 RENGJØRINGSPERSONALE I BEDRIFTER OL	1.19 %	3 678	0.88 %	1 320	0.31 %	26.10 %
9142 BILKLARGJØRERE O.L.	0.64 %	1 966	0.25 %	371	0.39 %	61.14 %

Hvis man skal revidere yrkeskoder på personnivå, kunne det være aktuelt å prioritere følgende grupper:

5224 SELGERE (ENGROS)

5221 BUTIKKMEDARBEIDERE O.L.

4114 KONTORMEDARBEIDERE

3431 FUNKSJONÆRER INNEN ADMINISTRASJON

1224 PRODUKSJONSDIREKTØRER INNEN VAREHANDEL

4131 LAGERMEDARBEIDERE OG MATERIALFORVALTERE

En sammenlikning av teksten som er oppgitt som yrkestittel i de to datakildene, gir ingen klare indikasjoner på at ulik ordbruk skulle være en hovedårsak til forskjellene i yrkeskoder. Tabellen nedenfor viser at det er et visst synkende samsvar i koder når teksten blir mer ulik. Den viser indirekte også andre viktige forhold:

- Det er relativt få som har tekst i begge kilder som kan sammenliknes.
- Mange har helt ulik tekst, selv om de i utgangspunktet skal være oppgitt av samme arbeidsgiver .

Tabell 3-8: Samsvar i yrkeskode (0 til 4 siffer) og likhet av yrkestittel (0-75). Antall og prosent.

	I alt	*0*	*1*	*2*	*3*	*4*
I alt	36 728	9 135	1 016	1 801	1 534	23 242
0	14 110	3 084	373	782	616	9 255
15	2 310	435	51	88	91	1 645
30	3 242	594	53	102	95	2 398
45	3 430	887	56	167	86	2 234
60	3 574	941	116	159	126	2 232
75	10 062	3 194	367	503	520	5 478

	I alt	*0*	*1*	*2*	*3*	*4*
I alt	100	25	3	5	4	63
0	100	22	3	6	4	66
15	100	19	2	4	4	71
30	100	18	2	3	3	74
45	100	26	2	5	3	65
60	100	26	3	4	4	62
75	100	32	4	5	5	54

3.2.1 Kvalitet på yrkeskodingen i varehandel

En nærmere analyse av feiltypene i koblete data vises i tabellen under. Det er klart det finnes påfallende forskjeller i yrkesfordelingen på dette nivå, da den systematiske skjevheten er større enn den tilfeldige for alle de største yrkene.

Tabell 3-9: Sammenlikning av yrkesfordelingene, utvalgte 4-siffer yrkeskoder. Koblet data.

	Register		Lønnsdata			
	antall	andel	antall	andel	Systematiske feil	Tilfeldige feil
1210 ADMINISTRERENDE DIREKTØRER	3 208	0.02841	1 977	0.01751	38 %	3 %
1224 PRODUKSJONSDIREKTØRER INNEN VAREHAN	1 426	0.01263	3 332	0.02951	-134 %	8 %
1314 LEDERE INNEN VAREHANDEL	1 817	0.01609	4 134	0.03661	-128 %	3 %
3415 TEKNISKE OG KOMMERSIELLE SALGSREPRE	5 214	0.04617	6 722	0.05952	-29 %	2 %
4114 KONTORMEDARBEIDERE	2 713	0.02402	1 853	0.01641	32 %	3 %
4121 ØKONOMIMEDARBEIDERE OG REVISJONSASS	1 447	0.01281	1 209	0.01071	16 %	7 %
4131 LAGERMEDARBEIDERE OG MATERIALFORVAL	8 503	0.0753	9 441	0.0836	-11 %	1 %
5137 APOTEKTEKNIKERE	2 936	0.026	2 973	0.02633	-1 %	1 %
5221 BUTIKKMEDARBEIDERE O.L.	54 176	0.47974	49 320	0.43674	9 %	0 %
5224 SELGERE (ENGROS)	5 857	0.05186	2 165	0.01917	63 %	1 %
7231 BILMEKANIKERE	4 653	0.0412	4 509	0.03993	3 %	1 %

Tabell 3-10: Mikrokonsistens av yrkesfelt (1-siffer yrke). Koblet data.

		Yrkesfelt i lønnstatistikken (se forklaring til venstre)								
Yrkesfelt i AA	I alt	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***
I alt	100	12	2	14	14	49	0	6	1	2
1 Lederyrker	100	75	2	12	4	6	-	1	0	0
2 Akademiske yrker	100	19	46	24	10	1	-	1	0	-
3 Høyskoleyrker	100	9	6	63	10	5	-	5	0	0
4 Kontor- og kundeservic	100	3	0	13	75	4	0	1	1	2
5 Salgs- og serviceyrker	100	6	0	7	3	83	-	0	1	0
6 Bønder, fiskere o.l.	100	14	-	11	-	61	-	7	7	-
7 Håndverkere o.l.	100	2	0	3	3	2	0	87	0	1
8 Operatører, sjåfører m	100	1	0	5	17	7	-	12	55	2
9 Andre yrker	100	1	0	3	5	17	0	9	3	62

For de største gruppene (salg/service og håndverkere) er samsvaret på godt over 80%. Det dårligste samsvaret finner vi i små grupper, som akademikere, sjåfører og hjelpearbeidere.

4 Yrkeskoding i AKU

Kvalitet på yrkeskoding i Arbeidskraftundersøkelsen (AKU) er interessant i seg selv, men også i forhold til å bruke utvalgsdata til å predikere yrke på sysselsatte uten yrkeskode. Hvis vi skal bruke AKU-data som grunnlag for prediksjon av yrke bør vi kunne dokumentere kvalitet på yrkeskodingen i AKU generelt, og usikkerhets-faktorer ved metoden for prediksjon. Det er flere grunner til å undersøke mulighetene for å bruke AKU-data. Sysselsatte totalt, og summene for sysselsettingstypene er like i sysselsettingsstatistikken og AKU. I tillegg er det godt samsvar mellom sysselsettingsstatus og aktivitetsstatus i de to kildene. For næring er kvaliteten noe dårligere, men ikke alarmerende. Det gjør at man i utgangspunktet vil tro at AKU er representativt også når det gjelder yrke. AKU er en svært stor undersøkelse, med lav utvalgsusikkerhet når det gjelder viktige variabler. Den store utvalgsstørrelsen og den høye svarprosenten, gjør at vi kan analysere mindre grupper – noe som er nødvendig i forhold til yrkesvariabelen på et detaljert nivå.

Det er endel ting som bør nevnes av usikkerhetsmomenter:

- Det er laget et trekke- og estimeringsopplegg med hensikt å få mest mulig korrekte sysselsettingstall, både når det gjelder nivå og endringer. Det kan tenkes at dette opplegget gir noe utslag for andre variabler, som f.eks. yrke.
- AKU bruker familie som trekkeenhet. Det finnes en viss intrafamiliær korrelasjon innen spesielle yrker, f.eks. legefamilier og psykologpar. Dette gir en viss skjevhet i yrker på detaljert nivå, og har gitt uforholdsmessig store svingninger på kort sikt i enkelte yrker.
- AKU-estimatene har utvalgsusikkerhet, skjevhet og andre utvalgsegenskaper.
- Metodiske egenskaper som følge av manuell yrkesklassifisering. Det er store fordeler med en individuell behandling, og en totalvurdering av alle opplysninger. På den annen side vil det være en viss mengde tilfeldige feil.
- Egenskaper som følge av selve prediksjonsmetodikken. Her vil det i første rekke bli en overvurdering av store yrker, og undervurdering av allerede små.

4.1 Kvalitet og utvalgsusikkerhet generelt

Selv om AKU i utgangspunktet er en meget stor undersøkelse, vil det fort bli små tall når vi skal identifisere grupper som er ensartede i forhold til yrkesrelaterte variabler. Vi må derfor sette en grense for akseptabel usikkerhet, uttrykt ved et minimumsantall personer i en gruppe i et utvalg. Det kan være greit først å se på den estimerte usikkerheten som er oppgitt i AKU-dokumentasjon. Tabell 3-1 viser at hvis vi bruker årsgjennomsnitt så vil det kreve et antall tilsvarende 5000 personer i populasjonen for å få et standardavvik på 10%. Med en populasjonsstørrelse på omtrent 3 200 000 og utvalg på omtrent 24 000, betyr det at en ønsket "minimumsgruppe" i utvalget må være på minst 75 personer for å få et "slingringsmonn" på 10% (altså +/- 5%).

Det neste som kan være et problem med utvalgsundersøkelser er stort eller ikke-tilfeldig frafall. Svarandelen i AKU er meget god, så frafallets størrelse i seg selv er ikke et betydelig problem. Tabell 3-2 viser frafallet i de

senere årene angitt i prosent av bruttoutvalget. Frafallet er ganske jevnt når det gjelder kjønn, men ikke f.eks. for alder. Variabler som kan gjenfinnes for personen i register, gjør at man kan justere for eventuelle skjevheter. En kjenner til at når det gjelder sysselsetting og arbeidsledighet er frafallet avhengig av undersøkelsesvariabelen selv. I det nye estimeringsopplegget er dette tatt hensyn til, med justering ved hjelp av sysselsettingsstatus fra register. Det er ikke enkelt å svare på hvilke konsekvenser dette har for yrkesvariabelen.

Yrkeskode på de sysselsatte må sies å være nesten komplett, det partielle frafallet her ligger på rundt 0.2%. Viktige hjelpevariabler til yrkeskoding som næring og utdanning må også sies å være ganske fullstendige, som det framgår av de videre tabeller. Dette uten at vi har undersøkt den interne kvaliteten på disse variablene spesielt.

Tabell 4-1: Størrelsen på standardavviket til estimatet. Fra AKU-dokumentasjon.

	Kvartalstall		Årsgjennomsnitt	
Antall i pop.	Antall	Prosent	Antall	Prosent
5 000	800	16	500	10
7 000	900	12.9	600	8.6
10 000	1 100	11	700	7
20 000	1 600	8	1 100	5.5
30 000	1 900	6.3	1 300	4.3
40 000	2 200	5.5	1 500	3.8
50 000	2 500	5	1 700	3.4
60 000	2 700	4.5	1 800	3
70 000	2 900	4.1	1 900	2.7
100 000	3 500	3.5	2 300	2.3
200 000	4 800	2.4	3 200	1.6
300 000	5 800	1.9	3 900	1.3
400 000	6 600	1.7	4 400	1.1
500 000	7 200	1.4	4 800	1
1 000 000	9 100	0.9	6 100	0.6
1 700 000	9 600	0.6	6 400	0.4
2 000 000	9 100	0.5	6 100	0.3

Tabell 4-2: Frafallet i AKU pr. kvartal, 2000-2003. Prosent.

Kvartal	Frafall	Kvartal	Frafall	Kvartal	Frafall	Kvartal	Frafall
2000-1	10.0	2001-1	12.2	2002-1	10.9	2003-1	11.4
2000-2	12.9	2001-2	16.4	2002-2	13.8	2003-2	13.7
2000-3	11.0	2001-3	14.1	2002-3	10.7	2003-3	11.0
2000-4	12.5	2001-4	12.6	2002-4	10.5	2003-4	10.9

Tabell 4-3: Partielt frafall av yrkeskode hos sysselsatte i AKU pr. kvartal, 2000-2003. Antall i utvalget.

Kvartal	Frafall	Yrkeskode
2000-1	18	15 336
2000-2	16	15 085
2000-3	15	15 356
2000-4	24	14 869
2001-1	25	14 940
2001-2	25	14 339
2001-3	32	14 754
2001-4	31	14 984
2002-1	29	15 264
2002-2	27	14 857
2002-3	26	15 392
2002-4	27	15 274
2003-1	24	14 995
2003-2	25	14 603
2003-3	26	15 130
2003-4	25	15 032

Tabell 4-4: Partielt frafall av utdanningskode i AKU pr. kvartal, 2000-2003.

Kvartal	Antall tegn i utdanningskode			
	1	2	4	5
2000-1	.	4 364	.	10 972
2000-2	.	4 085	.	11 000
2000-3	.	3 984	.	11 372
2000-4	.	3 560	.	11 309
2001-1	3	3 382	58	11 497
2001-2	1	3 131	67	11 140
2001-3	1	3 308	60	11 385
2001-4	1	3 048	47	11 888
2002-1	.	2 249	41	12 974
2002-2	1	2 222	34	12 600
2002-3	.	2 412	24	12 956
2002-4	1	2 221	15	13 037
2003-1	.	2 126	11	12 858
2003-2	.	2 518	5	12 080
2003-3	.	2 594	.	12 536
2003-4	.	2 444	.	12 588

Antall av sysselsatte med yrke i utvalget. Utdanningskode blir koblet fra register over befolkningens høyeste utdanning. Her beholder man 5 av 6 siffer, og dette utgjør for tiden nærmere 85% av de om er sysselsatte og har yrkeskode. De hvor man ikke får koblet på utdanning, blir kodet manuelt – dette er massen med 2 siffer. De øvrige kombinasjonene er rene feil og utgjør ubetydelige mengder.

4.2 Manuell yrkeskoding i AKU

I AKU kodes hver person manuelt utfra hva intervjuobjektet har opplyst om yrke og arbeidsoppgaver. I tillegg brukes opplysninger om bedriften, som næring og størrelse, og i noen grad personens utdanning. For å si noe om kvaliteten på den manuelle yrkeskodingen i AKU må man skille ut faktorer som ikke har med den usikkerheten i estimatene som kommer av at man trekker et mer eller mindre tilfeldig utvalg, og eventuelle effekter av at man justerer estimatene for å optimalisere andre variabler enn yrke. Det er to egenskaper som direkte sier noe om effektene av det manuelle kodearbeidet:

- I hvilken grad man klarer å klassifisere arbeidsoppgavene til hver person i henhold til standarden. Dette kan man kalle gyldighet eller validitet. Dette er ikke mulig å måle uten å ha en uavhengig kilde til

informasjon om personens arbeidsoppgaver. Det har vi ikke, og vi tar det for gitt at AKU-yrkeskoden er den best mulige og bruker den som ideal i sammenlikning med andre yrkesdata, som f.eks. i del 1 og 2.

- Et annet spørsmål er i hvilken grad man utfra de samme opplysninger klarer å gi samme yrkeskode hver gang. Dette sier noe om hvor stødig eller pålitelig metoden er, og kan kalles reliabilitet. Dette bør kunne måles.

4.2.1 Kvalitet

For å kvantifisere kvaliteten av yrkeskoding i AKU kan vi forsøke å måle *reliabiliteten,* hvor pålitelig eller konsekvent kodingen er. Det kan tenkes ulike måter å måle denne:

- Samme data kodes manuelt 2 ganger, uten at den som koder på nytt kjenner den gamle koden. Denne metoden er mest intuitiv og teoretisk solid, men resurskrevende.
- Sammenlikne koder i to kvartaler for samme person. På grunn av den nåværende rutinen vil denne metoden ikke gi mening (med visse unntak). Sysselsatte blir spurt om de har samme stilling som sist – hvis de svarer ja, kopierer man koden fra forrige kvartal for å spare intervjutid og kodetid. Unntaket er hvis det første intervjuet var indirekte.
- Man kan ta datasett som er yrkeskodet allerede og sammenlikne 'ensartede grupper'. Spørsmålet er hva som konstituerer ensartede grupper. Et hovedproblem er "jo mer ensartet – desto mindre".

Vi forsøker å måle reliabilitet / konsistens i kodingen på denne måten: Hvis man kan finne grupper hvor gitte yrkeskarakteriserende variabler er nokså like, vil vi kunne se på spredning i hver gruppe – altså om yrkesfordelingen er konsentrert om ett bestemt yrke. De variablene som kan være aktuelle er yrkestittel (tekst), arbeidsoppgaver (tekst), næring (kode), utdanning (kode), bedriftsstørrelse (over eller under 10 ansatte). Kort sagt er premissene for en slik undersøkelse at hvis to personer har samme tekst (tittel og arbeidsoppgaver) og er ansatt i liknende bedrifter, har de såpass like arbeidsoppgaver at de skal få samme yrkeskode.

4.2.2 Forsøk med tekst

Yrkestittel og arbeidsoppgaver er tekstvariabler, og det kan være greit med en kort orientering om dette. Tekstvariabler skiller seg fra nummeriske og kategoriske variabler som er vanlige i datasett som vanligvis behandles her. Man kan beskrive fritekst som "nesten kontinuerlig" fordi det er nærmest uendelig antall muligheter, og "ikke-ordinal" fordi det ikke er gitt en betydningsmessig sorteringsrekkefølge. Typisk egenskaper ved tekstdata:

- Mange ord brukes svært sjelden, og noen svært få brukes ofte.
- Diverse tegn og ortografisk variasjon gir mange "unødige" kategorier.
- Små forskjeller i skrivemåte kan ha stor semantisk betydning.
- To ord med lik skrivemåte kan bety ulike ting (homonymer).
- To ord som skrives helt forskjellig kan bety det samme (synonymer).

Med tanke på den store variasjonen ønsker vi å øke utvalgsstørrelsen. Som et forsøk tar vi med alle AKU-data med yrkeskode fra de siste 16 kvartaler. Av 240 210 records velges den siste pr. person, som gir til sammen 51 545 personer med yrke. I denne undersøkelsen vil en derfor ikke kunne si noe om kvalitetsendring fra år til år.

Tabellen viser frekvensfordelingen for tekster og tekstgrupper ('soundex'-funksjonen i SAS), og illustrerer noe av vanskene både ved å klassifisere utfra tekst og å måle kvalitet på anvendelsen av tekst. Bruk av tekstgruppering gir nokså lite effekt på spredningen, men brukes f.eks. i automatiske kodeprogrammer.

Tabell 4-5: Frekvensanalyse av tekstbruk. Data fra AKU 2000-2003.

Mål	Tekst	Tekstgruppe
Sum	51 545	51 545
Antall	11 651	8 686
Gjennomsn.	4.42	5.93
Standardav.	33.17	41.12
Max	1 402	1 489
99%	62	92
95%	10	15
90%	4	6
Q3	1	2
Median	1	1

Vi ser altså at blant de rundt 50.000 personene brukes over 10.000 forskjellige yrkestitler. Av alle disse titlene brukes 99% færre enn 63 ganger. Tre fjerdedeler brukes kun 1 gang. Det er klart at dette får betydning for en analyse av reliabilitet/konsistens. Hvis vi bruker tekst i sammenlikningen vil gruppene hvor både tekst og næring skal være like, fort bli små. De fleste gruppene vil være konsistent med hensyn på yrke, rett og slett fordi de består av 1 person. Slike grupper kan meget vel være riktig kodet, men kan ikke brukes for å måle stabiliteten i kodingen.

Vi forsøke å operasjonalisere en type 'yrkeshomogene grupper' på denne måten: de personer som har lik kombinasjon av tekst og næring, og som er minst 75 personer i utvalget. Det delutvalget vi får da inneholder 49 grupper med 8195 personer. Yrkesfordelingen regnes ut ved formelen nedenfor, som gir summen av kvadratene av andelene av hvert yrke i hver gruppe.

Formel 4–1: Ensartethet i yrke innen en gruppe

$$m_g = \sum_{1}^{h} (p_{yrke_{hg}})^2$$

Der h er antall forskjellige yrkeskoder som forekommer i gruppa g. Tallet m blir høyere jo mer ensartet yrkesfordelingen er innen gruppen. Det ideelle er at alle i en gruppa har samme yrke, eller hvertfall konsentrert om noen få yrker. Den høyeste verdien $m_{max} = 1$, er hvis alle i gruppa har samme yrkeskode. I de undersøkte gruppene ligger verdiene mellom 0.5 og 1.0. Gjennomsnittet for alle gruppene i delutvalget er 0.944. Hvis vi vekter med gruppestørrelsene blir det 0.953. Medianen er 0.99, så det er tydelig at noen få grupper er spesielt lave. Denne verdien betyr *ikke* "95% er riktig kodet". Det kan tolkes som sjansen for at en person i gruppa fikk sin opprinnelig yrkeskode, hvis man omfordelte de eksisterende yrkeskodene innen gruppa helt tilfeldig. Jo nærmere 1 desto bedre, og 0.95 i dette forsøket tolkes som: for vanlige yrkestitler i AKU er yrkeskodingen veldig konsekvent.

I grupper med spesielt lave verdier, ville man gå utfra at det kunne skyldes klassifisering utfra flere variabler, ikke nødvendigvis at kvaliteten er spesielt dårlig. Dette gjør seg f.eks. gjeldende for administrative ledere, hvor bedriftsstørrelse/sektor har betydning for yrkeskoden. Det samme gjelder til en viss grad yrker som krever høyere utdanning. Istedenfor å innføre flere variabler, som ville gitt enda mindre grupper, gjør vi den samme analysen etter å ha fjernet alle med leder- og akademikerkoder. En får da 40 grupper med 5 657 personer, og finner da at det vektede gjennomsnittet stiger til 0.958. Det er jo ingen voldsom økning, m.a.o. ville man ikke fått så stor effekt av å trekke inn flere variabler – men man ville fått flere grupper som ble for små. Ved manuell kontroll av enkelte yrker benyttes selvsagt alle kjennemerker man har.

Siden de delutvalg man analyserer her ikke er tilfeldig utvalgt, kan man ikke trekke direkte slutninger om *hele* AKU-yrkeskodingen. Man kan anta at koderne vil være sikrere og mer konsekvent når de velger yrkeskode for en vanlig yrkestittel enn for en sjelden tittel. På den annen side kan mange av de sjeldne tekstene også være enkle å klassifisere, f.eks. fordi de er lengre og inneholder mer informasjon, eller fordi de beskriver spesielle yrker på en entydig måte.

Alt i alt går vi utfra at yrkeskodingen i AKU er det beste sammenlikningsgrunnlaget vi har for yrkeskoding andre steder.

5 Yrke i sysselsettingsstatistikken

Det er et ønskelig at Sysselsettingsstatistikken skal inneholde yrkeskode for alle sysselsatte. Vi har studert yrkeskoding i Arbeidstakerregisteret inngående, fordi dette er den største gruppen av sysselsatte. I tillegg til de i Arbeidstakerregisteret med yrkeskode må vi finne yrke på følgende grupper:

- De som mangler yrkeskode i Arbeidstakerregisteret (ca.10% av dette registeret)
- Sysselsatte som er definert utfra registre som ikke inneholder yrkesdata. (ca. 21% av alle)
  - De som er definert som sysselsatte utfra Lønns- og trekkoppgaveregisteret (LTO) alene – altså ikke koblet mot Arbeidstakerregisteret. Det er antatt at dette er mange småjobber, da det i Arbeidstakerregisteret er det krav om minst 4 timer pr. uke og minst 6 dager.
  - Selvstendig næringsdrivende. Disse defineres utfra Selvangivelsesregisteret, hvor man bruker næringsinntekt som et ledd i selve definisjonen.

Det er grunn til å anta at yrkesfordelingen i disse gruppene vil avvike fra vanlige arbeidstakerforhold.

5.1 Hovedgrupper av sysselsatte

Det kan være interessant å se nærmere på yrkesfordelingen etter sysselsettingstype, nemlig ansatte og selvstendige. Det er ikke noen prinsipiell grunn til å lage et eget opplegg for hver gruppe, hvis man kunne predikere yrke utfra andre, fullstendig determinerende variabler. Det er tydelige forskjeller i yrkesfordelingen etter sysselsettingstype. Det er også slik at datakildene / grunnlagsregistrene er helt forskjellige for de to gruppene i sysselsettingsstatistikken, og det er derfor høyst aktuelt å behandle disse hver for seg. I tabellene er den relative forskjellen valgt å sammenlikne med *ansatte*.

Tabell 5-1: Aggregert yrkesfordeling etter sysselsettingsstatus. Årsgjennomsnitt AKU 2002.

Yrkesfelt	Ansatte	Selvstendige	Forskjell	Rel. Forskjell
0 Andre yrker	0.5 %	0.3 %	0.2 %	38.1 %
1 Lederyrker	8.0 %	3.2 %	4.8 %	60.3 %
2 Akademiske yrker	11.2 %	12.4 %	-1.2 %	-11.1 %
3 Høyskoleyrker	23.9 %	12.6 %	11.4 %	47.5 %
4 Kontor- og kundeservice	8.6 %	0.5 %	8.1 %	94.2 %
5 Salgs- og serviceyrker	22.7 %	12.2 %	10.6 %	46.5 %
6 Bønder, fiskere o.l.	1.2 %	30.3 %	-29.1 %	-2429.0 %
7 Håndverkere o.l.	10.5 %	19.3 %	-8.8 %	-84.3 %
8 Operatører, sjåfører m	7.7 %	8.9 %	-1.2 %	-16.0 %
9 Andre yrker	5.7 %	0.3 %	5.4 %	94.7 %

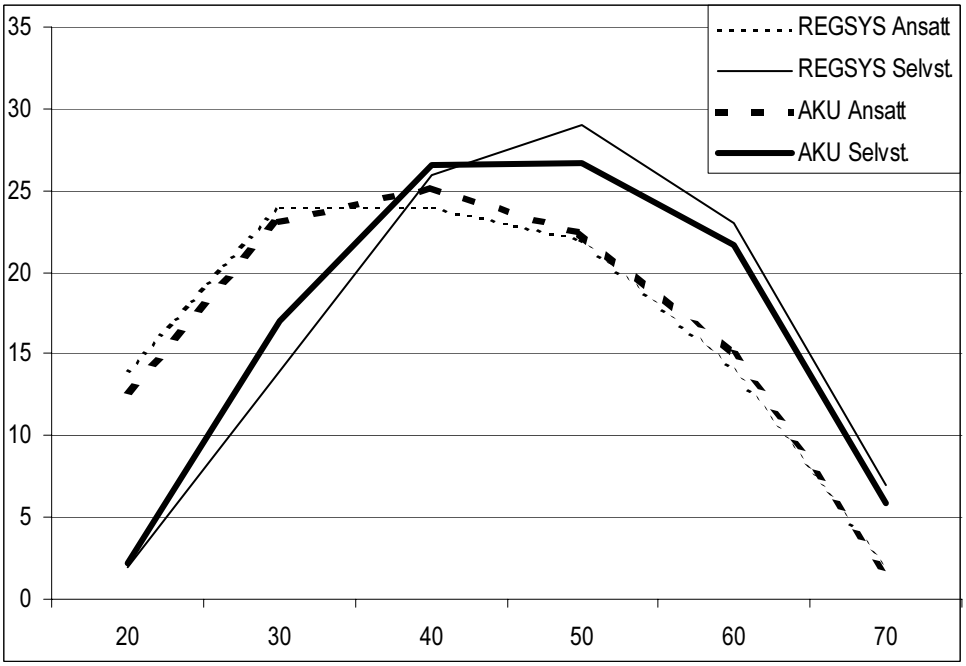
Tabell 5-2: Detaljert yrkesfordeling etter sysselsettingsstatus. Utdrag. Årsgjennomsnitt AKU 2002.

Yrke	Ansatte	Selvstendige	Forskjell	Rel. Forskjell
1210 ADMINISTRERENDE DIREKTØRER	1.04 %	0.31 %	0.73 %	69.78 %
2130 SYSTEMUTVIKLERE OG PROGRAMMERERE	1.67 %	1.00 %	0.67 %	40.30 %
3310 GRUNNSKOLELÆRERE	3.34 %	0.04 %	3.30 %	98.75 %
3320 FØRSKOLELÆRERE	0.97 %	0.02 %	0.95 %	97.61 %
3415 TEKNISKE OG KOMMERSIELLE SALGSREPR.	1.93 %	1.72 %	0.21 %	10.71 %
3432 REVISORER OG REGNSKAPSFØRERE	1.06 %	0.73 %	0.33 %	31.25 %
4113 SEKRETÆRER	2.05 %	0.03 %	2.02 %	98.46 %
4114 KONTORMEDARBEIDERE	1.54 %	0.10 %	1.45 %	93.77 %
4121 ØKONOMIMEDARB. OG REVISJONSASS.	1.05 %	0.16 %	0.89 %	84.59 %
5122 KOKKER	1.04 %	0.72 %	0.32 %	30.78 %
5123 HOVMESTERE, SERVITØRER, O.L.	0.98 %	0.15 %	0.83 %	84.61 %
5131 BARNE- OG UNGDOMSARBEIDERE O.L.	3.00 %	0.41 %	2.59 %	86.28 %
5132 OMSORGSARBEIDERE OG HJELPEPLEIERE	3.35 %	0.02 %	3.33 %	99.34 %
5139 ANNET PLEIE- OG OMSORGSPERSONALE	2.49 %	0.06 %	2.44 %	97.80 %
5163 VAKTMESTERE O.L.	1.14 %	0.18 %	0.97 %	84.59 %
5221 BUTIKKMEDARBEIDERE O.L.	7.17 %	4.60 %	2.57 %	35.82 %
7241 ELEKTRIKERE, ELEKTRONIKERE OL.	1.24 %	0.52 %	0.72 %	58.32 %
9132 RENGJØRINGSPERSONALE I BEDRIFTER OL.	2.84 %	0.16 %	2.69 %	94.44 %
9133 KJØKKEN- OG ANRETNINGSASSISTENTER	1.10 %	0.04 %	1.06 %	96.46 %

5.1.1 Selvstendig næringsdrivende

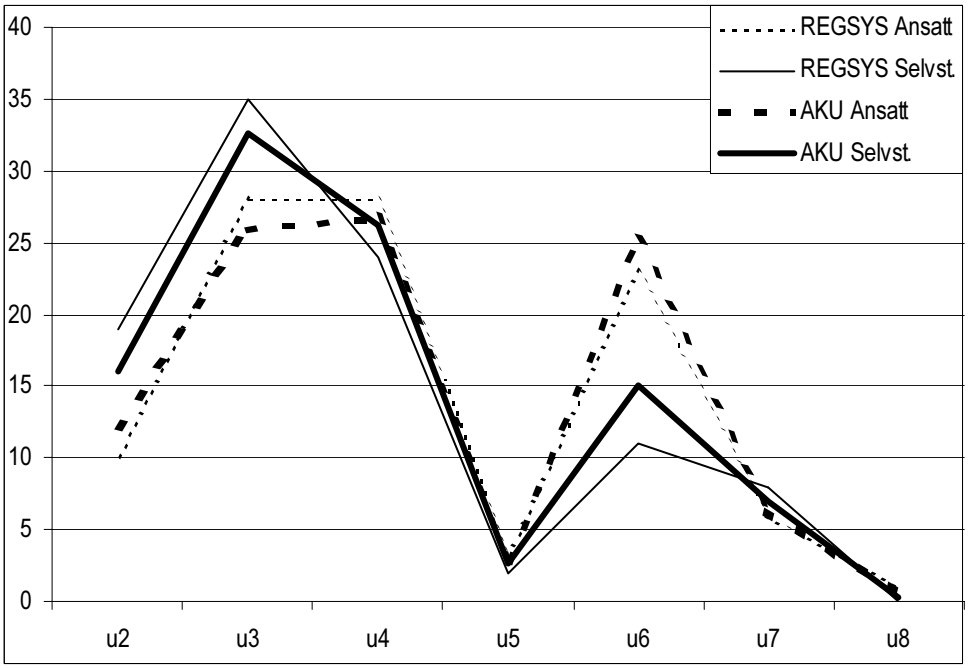
Det er mange som har common-sense oppfatninger av hva som karakteriserer selvstendig næringsdrivende. Det kan derfor være greit med en gjennomgang av noen viktige egenskaper. Vi har tidligere vist at yrkesfordelingen skiller seg i vesentlig grad fra lønnstakere/ansatte. Figuren nedenfor viser aldersprofilen på selvstendige sammenliknet med lønnstakere/ansatte. Det er tatt med data både fra AKU og registersysselsatte. Begge kilder viser at de selvstendige er stort sett noe eldre, og det er særlig få i de yngste aldersgruppene. Man kan ikke si at unge personer er spesielt representative for selvstendig næringsdrivende. Det er også store kjønnsforskjeller, kun en av fire selvstendige er kvinner, blant de ansatte er omtrent halvparten kvinner.

Figur 5-1: Aldersprofil etter sysselsettingsstatus, og -kilde. Andeler innen 10-års gruppering.



Den neste figuren viser hvordan utdanningsnivåene fordeler seg etter sysselsettingsstatus. Nivå 3-4 er videregående skole, 6 er høyskole og 7-8 er lange universitetsutdanninger. Det ser ikke ut til at den typiske selvstendige er akademiker. Vi hører ofte om leger og advokater, men yrkesfordelingen som er vist tidligere at de fleste er bønder og håndverkere. Utdanningsprofilen samsvarer med dette, der det er flere med kortere utdanning og færre med høyere utdanning, særlig på høyskolenivå..

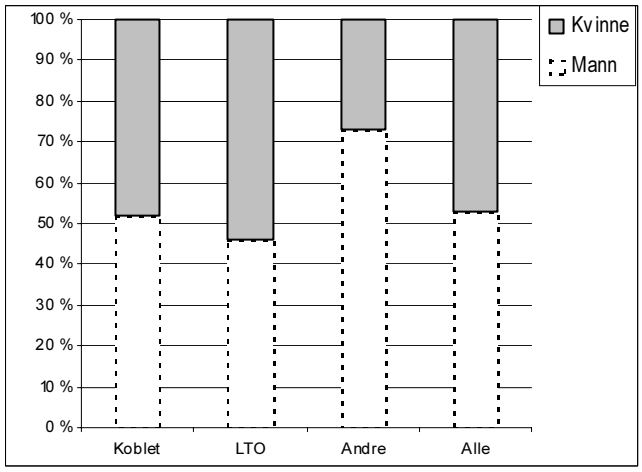
Figur 5-2: Utdanningsprofil etter sysselsettingsstatus, og -kilde. Se forklaring i teksten over.



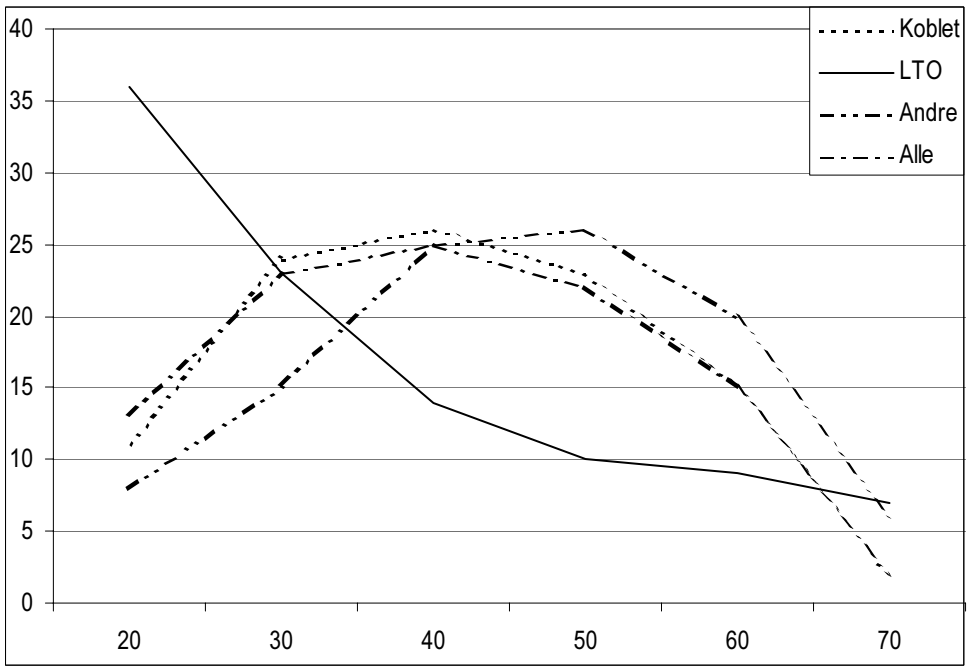
5.1.2 Lønnstakere uten kobling til Arbeidstakerregisteret

Data i dette avsnittet er fra sysselsettingsstatistikken 2002. For enkelthets skyld kaller vi gruppen lønnstakere uten kobling til Arbeidstakerregisteret for "LTO", selv om de fleste andre i LTO kan kobles til Arbeidstakerregisteret. Noen av de som ikke kan kobles til Arbeidstakerregisteret kan kobles til andre registre. I figurene under betyr "koblet" at lønnstakerne gjenfinnes i Arbeidstakerregisteret, uavhengig kobling til andre

registre. Vi finner at det er noe overvekt av kvinner i LTO-gruppen sammenliknet med andre grupper, og alle sysselsatte. Aldersprofilen er helt klart forskjellig, og en av de mest markante forskjellene mellom LTO og alle andre.

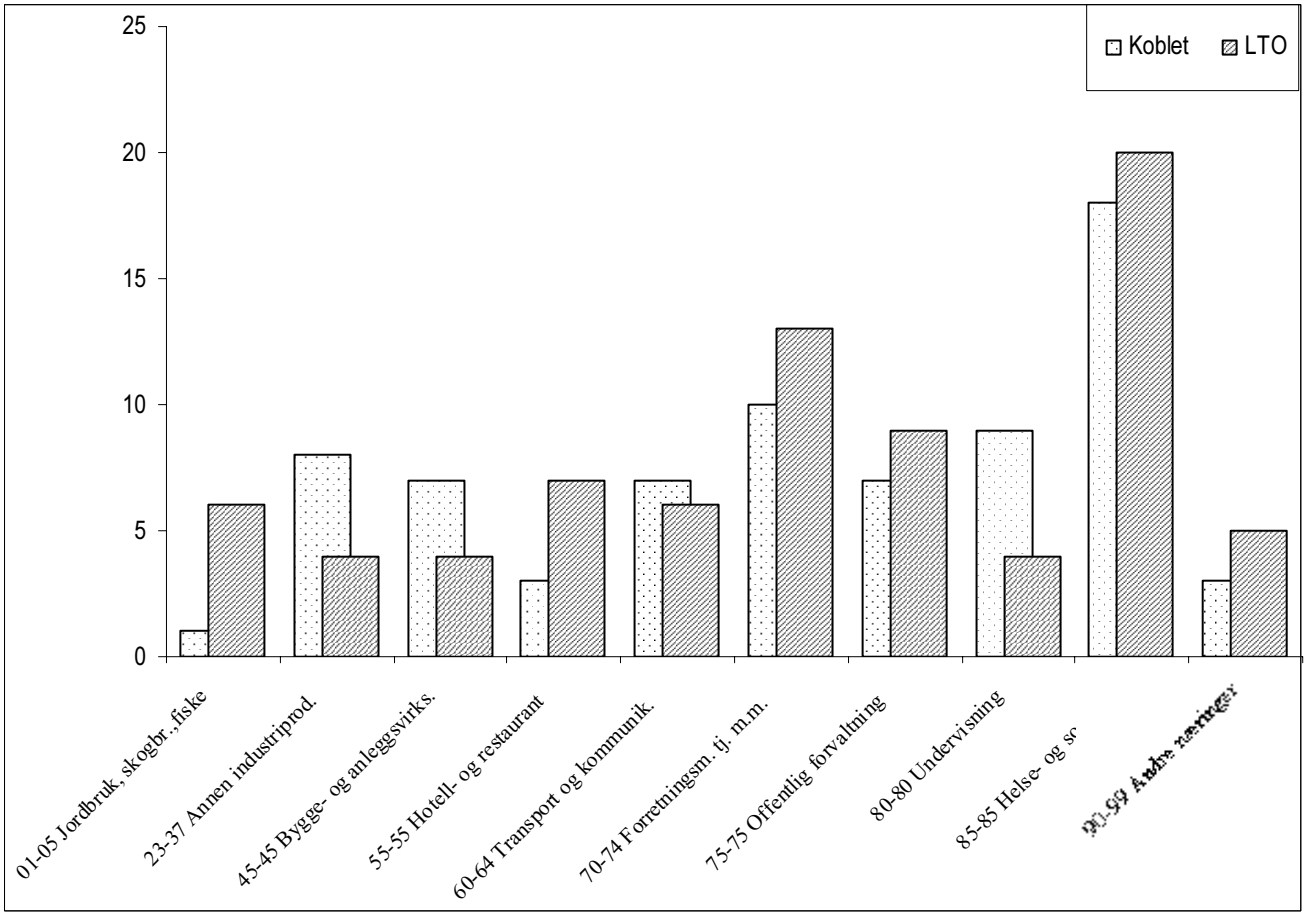


Figur 5-3: Aldersprofil etter kilde til sysselsettingsstatus. Andeler innen 10-års gruppering.



Det er ikke noen utpreget forskjell i utdanningsnivået for LTO-gruppen og det meste kan kanskje forklares ved aldersforskjellene. I visse næringer derimot er det endel forskjeller, som vil få konsekvenser for yrkeskoding. På figuren er det valgt ut visse næringer som illustrerer forskjellene. Med tanke på at LTO-gruppen totalt sett er liten, og derfor vil utgjøre svært få i de korresponderende AKU-data, er det kanskje grunn til å ikke behandle denne gruppe på en spesiell måte ved prediksjon av yrke. Hvis man for alle ansatte legger næring + utdanning/alder til grunn, vil man oppnå omtrent samme yrkesfordeling, men med mindre varians på estimatet, altså mindre usikkerhet.

Figur 5-4: Andel i utvalgte næringer etter kilde til sysselsettingsstatus.



5.2 Valg av variabler

Når vi skal predikere yrke på sysselsatte står vi ovenfor en rekke valg med hensyn til metode og data. Et viktig element i denne sammenheng er å finne de uavhengige variabler som har mest å si for yrkeskoden. Merk at yrkeskode i denne sammenheng er en kategoriell variabel med 352 verdier. Det er lite aktuelt å bruke regresjonsmetoder. Vi tar sikte på å bruke en metode som likner på den som er planlagt for Arbeidstakerregisteret (deterministisk imputering utfra ikke-informative SHG-grupper). Hver sysselsatt blir da tildelt en sannsynlighet for enhver yrkeskode. Sannsynligheten (som desimaltall) er lik gjennomsnittet (andelen av yrket) i den gruppen personen tilhører. Mye av arbeidet vil bestå i å identifisere grupper som har mest mulig ensartede arbeidsoppgaver, definert utfra:

Formel 5–1: Ensartethet i yrke innen en gruppe

$$m_g = \sum_{1}^{h} (p_{yrke_{hg}})^2$$

Der h er antall forskjellige yrkeskoder som forekommer i gruppa g. Tallet m blir høyere jo mer ensartet yrkesfordelingen er innen gruppen. Det vi ønsker er at yrkene skal være så "klumpete" som mulig, altså konsentrert om noen få yrker. Dette målet likner på varians, men har større følsomhet for store andeler. Det ideelle er at èn yrkeskode er dominerende innenfor en gruppe, og jo færre koder desto bedre. En ulempe med varians er at den raskt nærmer seg null når andelene er omtrent like. For vårt formål er det bedre med 4 yrker à 25%, enn 10 yrker à 10%. I begge tilfeller blir variansen lik null. Den minste verdien er hvis alle yrkene i gruppa er like store, da er $m_{min} = 1/h$. I eksempelet vil den ene gruppa ha 1/4 og den andre 1/10. Den høyeste verdien $m_{max} = 1$, er hvis alle i gruppa har samme yrkeskode – altså en ekte svarhomogen gruppe.

Ved oppdrag og levering av mikrodata kan det bli aktuelt å gi yrkeskode på personnivå. Den må settes ved stokastisk imputering. Det betyr at man deler ut tilfeldige yrkeskoder, hvor sannsynligheten for å få en kode er lik gruppegjennomsnittet. Forventningen til at en person skal få riktig yrkeskode er da gitt ved en funksjon tilsvarende formelen over. Man kan si enkelt at jo mer klumpete yrkene fordeler seg i gruppen, desto større sjanse er det for at hver person får riktig yrkeskode. For enkelte grupper vil det være spredning på flere ulike yrker, slik at mange personer i gruppa sannsynligvis ikke har riktig yrkeskode. Dette vil kunne gi utslag når man gir yrkesfordeling på detaljert nivå eller for små områder. Det betyr også at yrkestall i slike tabeller er *estimater*, altså at det har en viss usikkerhet selv om vi tenker på registerstatistikk som en form for fulltelling uten utvalgsusikkerhet.

Det vil i første rekke være utdanningskode og næringskode som er bestemmende for yrke i en definitorisk forstand. Dette er altså kjennemerker som vi har en faglig oppfatning at arbeidsoppgavene henger sammen med. I tillegg kommer prediktive variabler som kjønn, arbeidstid, alder og lønn. Dette er variabler som ikke blir brukt til klassifisering av yrke, men som kan brukes til å finne sannsynligheten for noen bestemte yrker.

Prinsipielt er det uheldig å bruke de samme uavhengige variabler i imputering som ønsker å lage krysstabeller for senere. Allikevel vil dette bli gjort i noen grad for visse deler av de sysselsatte, noe som må tas med i planleggingen av nye tabeller.

5.2.1 Næring

Næringskoden til en bedrift forteller i stor grad hva som produseres av varer og tjenester. For de sysselsatte som deltar direkte i produksjonen av bedriftens hovedprodukter, vil næringskoden kunne karakterisere arbeidsoppgavene i stor grad. For andre sysselsatte, som f.eks. utfører støttefunksjoner vil næringskoden ikke gi holdepunkter om arbeidsoppgavene og derfor ikke være determinerende for yrkeskode.

En næringskode som beskriver en bestemt produksjonsprosess, vil som regel ha en karakteristisk yrkesfordeling. Næringskoden til bedriften der den sysselsatte jobber, eventuelt med få tilleggsopplysinger vil da kunne legges til grunn for en rimelig yrkeskode. I enkelte andre næringer er arbeidsoppgavene så varierte at de sysselsatte har en veldig spredt yrkesfordeling. Det er dessverre slik at det for mange av disse yrkene heller ikke er andre variabler som er spesielt gode for å bestemme yrkeskode.

Tabellene viser eksempler på næringer med et mål på hvor ensartede arbeidsoppgavene er. Dette er beregnet utfra yrkesfordelingen i hver næring, og er ikke avhengig av størrelsen på selve næringen. Antallet yrker gir også et innblikk i variasjonen av arbeidsoppgaver.

Tabell 5-3: Utvalgte næringer med varierte yrkeskoder. AKU 2003 årsgjennomsnitt.

Antall sysselsatte	Antall yrker	Ensartet	Næringsundergruppe
15 532	48	0.064	11.100 Utvinning av råolje og naturgass
6 110	30	0.091	15.510 Produksjon av meierivarer
6 963	31	0.068	35.111 Bygging og reparasjon av skip og skrog over 100 bruttotonn
13 656	49	0.047	35.114 Bygging og reparasjon av oljeplattformer og moduler
11 180	39	0.077	61.101 Utenriks sjøfart
9 243	47	0.056	73.100 Forskning og utviklingsarbeid innen naturvitenskap og teknikk
13 384	51	0.048	74.209 Annen teknisk konsulentvirksomhet
10 656	49	0.045	74.502 Utleie av arbeidskraft
7 745	36	0.054	74.879 Annen forretningsmessig tjenesteyting ikke nevnt annet sted
32 696	47	0.063	75.110 Generell (overordnet) offentlig administrasjon og økonomiforvaltning
16 274	55	0.066	75.120 Offentlig administrasjon helsestell, sosial virksomhet, undervisning, kirke, kult
15 586	39	0.074	75.130 Offentlig administrasjon tilknyttet næringsvirksomhet og arbeidsmarked
8 064	27	0.08	75.140 Hjelpetjenester for offentlig administrasjon
24 594	72	0.058	75.220 Forsvar
5 766	22	0.096	85.112 Somatiske spesialsykehus

Tabell 5-4: Utvalgte næringer med ensartede yrkeskoder. AKU 2003 årsgjennomsnitt.

Antall sysselsatte	Antall yrker	Ensartet	Næringsundergruppe
4 413	5	0.807	01.121 Dyrking av hagebruksvekster på friland
17 104	5	0.941	01.220 Saue- og geitehold. Oppdrett av hester
9 532	8	0.706	05.011 Hav- og kystfiske
6 368	9	0.716	45.441 Malerarbeid
7 043	6	0.736	52.120 Butikkhandel med bredt vareutvalg ellers
2 632	4	0.728	52.241 Butikkhandel med bakervarer og konditorvarer
3 519	3	0.829	52.431 Butikkhandel med skotøy
1 466	3	0.738	52.452 Butikkhandel med plater, musikk- og videokassetter, CD- og DVD-plater
1 493	4	0.823	52.482 Butikkhandel med gull- og sølvvarer
5 086	10	0.754	52.483 Butikkhandel med fritidsutstyr, spill og leker
14 582	12	0.784	60.211 Rutebiltransport
9 985	3	0.944	60.220 Drosjebiltransport
2 501	3	0.79	80.410 Trafikkskoleundervisning
4 259	2	0.796	85.331 Skolefritidsordninger
14 529	3	0.862	93.020 Frisering og annen skjønnhetspleie

5.2.2 Utdanning

Yrkeskoden skal avspeile faktiske arbeidsoppgaver, og det er ikke alltid klart hvilken sammenheng det er mellom utdanning og faktiske arbeidsoppgaver. En antagelse er at sammenhengen er svært forskjellig for ulike utdanninger, og kanskje særlig lav for såkalte allmennutdanninger. En persons utdanning kan bidra til å bestemme yrkeskode ut fra faglig spesialisering, og i andre tilfeller ut fra utdanningsnivå. Det er jo i yrkesstandarden lagt opp til en viss sammenheng mellom yrkesfelt (1.siffer) og utdanningsnivå.

Det er også slik at sysselsatte som ikke har den formelle utdanningen som normalt kreves i yrket kan utføre de samme arbeidsoppgavene. Dette kan være på bakgrunn av lang erfaring, intern-opplæring, eller mangel på arbeidskraft med formell utdanning. Andre forhold som påvirker sammenhengen mellom utdanning og yrke er rene arbeidsmarkedseffekter, etterspørsel, omskolering, m.m.

For lærlinger er igangværende utdanning viktig. Et hovedproblem er da at det er betydelig produksjonstid på utdanningsregister. Det er kan da bli aktuelt å rett og slett bruke alder istedenfor utdanningskoden, i tillegg til andre opplysninger.

Tabellene viser eksempler på utdanninger som er egnede og uegnede i forhold til yrkeskoding. Utdanningskode 6 siffer er koblet fra register, partielt frafall (manglende kobling) er 1%.

Tabell 5-5: Utvalgte utdanninger som er yrkesrettet. AKU 2003 årsgjennomsnitt.

Antall sysselsatte	Antall yrker	Ensartet	Utdanning
15 557	17	0.713	361202 Hjelpepleierutdanning
6 461	13	0.754	461201 Hjelpepleier, VK II
3 595	5	0.743	469907 Renholdsoperatørfaget, VK II
2 250	2	0.86	661101 Helsesøsterutdanning
1 533	2	0.97	661102 Jordmorutdanning
5 272	7	0.791	665201 Fysioterapeututdanning, treårig
1 651	1	1	669906 Høgskoleingeniørutdanning, bioingeniør, treårig
2 010	4	0.71	669907 Ingeniørutdanning, bioingeniør, toårig
1 535	2	0.963	682302 Politihøgskole, treårig grunnutdanning
2 950	6	0.85	757201 Sivilarkitektutdanning
10 182	5	0.921	763101 Cand.med.-utdanning
2 177	1	1	763102 Utenlandsleger, tilleggskurs
2 836	1	1	764101 Cand.odont.-utdanning
1 346	5	0.719	766101 Cand.pharm.-utdanning
1 087	3	0.701	767101 Cand.med.vet.-utdanning

Tabell 5-6: Utvalgte utdanninger som er generelle. AKU 2003 årsgjennomsnitt.

Antall sysselsatte	Antall yrker	Ensartet	Utdanning
64 155	146	0.031	201101 Framhaldsskoleutdanning
35 176	127	0.028	201102 Folkeskoleutdanning
114 366	195	0.036	201103 Grunnskoleutdanning på ungdomsskoletrinnet
15 816	70	0.049	301104 Folkehøgskoleutdanning, ettårig
50 699	138	0.027	301110 Realskoleutdanning
14 608	72	0.064	341102 Økonomiske og administrative fag, VK I
73 939	146	0.034	343299 Kontorfag, uspesifiserte, videregående, grunnutdanning
23 968	96	0.022	355211 Mekaniske fag, grunnkurs
13 317	83	0.023	359999 Naturvit., håndverk, tekniske fag, andre, uspes., videreg., grunnutd.
30 489	82	0.077	369902 Helse- og sosialfag, grunnkurs
139 078	220	0.032	401101 Allmenne fag, VK II
22 214	91	0.021	401103 Gymnasutdanning
41 110	118	0.038	441106 Økonomiske og administrative fag, VK II
21 409	100	0.018	619902 Forberedende prøver
22 229	104	0.024	999999 Uoppgitt

En måte å redusere usikkerheten på er å aggregere. Når vi vurderer navnet på utdanningene og navnet på yrkene, vil det i mange tilfeller være klart at man må benytte det mest detaljerte nivå på utdanning for å kunne avspeile yrket. Det er med andre ord en avveining mellom faglig sett ensartete grupper, og usikkerheten ved at disse gruppene blir små. En liten analyse av yrkes-ensartetheten i hver utdanningsgruppe forhold til aggregeringsnivå viser noe av dette forholdet. Tallene er vektet med gruppestørrelsen, og tabellen viser frekvensfordelingen i forhold til aggregering av utdanningsvariabelen, 1-6 siffer. Den viser kort sagt:

- De aller fleste *utdanningskodene* er for generelle til alene å kunne predikere yrke.
- For de utdanninger som kan brukes, må det benyttes detaljert utdanningskode.

Hvis vi så ser nærmere på 6-siffer utdanningskodene, finner vi at en ganske liten andel sysselsatte har det vi kan kalle yrkesspesifikke utdanninger. Men for de få det gjelder, kan det være gode grunner til nettopp å bruke utdanning. Videre vil utdanning være mye mer nyttig i sammenheng med andre variabler.

Tabell 5-7: Utdanning, fordeling av yrkesspesifisitet etter aggregeringsnivåer. Data fra AKU 2003.

Aggregering	max	p95	p90	median
Utd.1	0.1148	0.0304	0.0304	0.0271
Utd.2	1.0000	0.2073	0.1676	0.0357
Utd.3	1.0000	0.4775	0.3956	0.0463
Utd.4	1.0000	0.5052	0.4174	0.0543
Utd.5	1.0000	0.5944	0.4491	0.0693
Utd.6	1.0000	0.6729	0.5718	0.0851

Tabell 5-8: Utdanningskode 6 siffer, antall og andeler i yrkesspesifisitetgrupper. Data fra AKU 2003.

Ensartethet	Sum	Andel	Antall koder
I alt	2 268 663	100 %	1 152
0.0	782 340	34 %	36
0.1	560 359	25 %	156
0.2	318 846	14 %	150
0.3	191 071	8 %	142
0.4	100 733	4 %	108
0.5	79 065	3 %	127
0.6	116 595	5 %	79
0.7	40 931	2 %	26
0.8	16 145	1 %	12
0.9	17 230	1 %	6
1.0	45 347	2 %	310

5.2.3 Alder

Det er visse forskjeller i yrkesfordelingen for ulike aldersgrupper. Mye av dette fanges opp av utdanningsnivå men ikke alt. Alder blir for her en indirekte variabel for uformell kompetanse, erfaring, ikke-registrert utdanning, o.l. som ble diskutert i forrige avsnitt. Vi benytter ganske grove inndelinger for ikke å overstratifisere (for små grupper).

5.2.4 Kjønn

I mange yrker er det svært store kjønnsforskjeller. Vi kan ikke her gå inn på årsakene til dette, som kan ligge i tradisjoner, interesser, utdanning, økonomiske forhold m.m.. Men man kan benytte kjønn som tilleggsvariabel til å regne ut sannsynligheten for yrker, der man ikke har nok detaljerte data i faglige variabler som næring, utdanning, o.l.

5.3 Metode

5.3.1 Foreslått justering av partielt frafall i Arbeidstakerregisteret

Vi kan imputere manglende yrke i Arbeidstakerregisteret ved en imputeringsmetode i flere trinn, som er foreslått i notat 79/2003:

- Deterministisk imputering utfra en ikke-informativ SHG-modell.
- Sannsynligheten for et yrke estimeres ved gruppegjennomsnitt (urealistisk verdi) innen personens SHG.
- Imputerer i flere trinn, etter kvalitetsvurdering av hvert trinn og gruppe.

Deterministisk vil si at det er faste tall, ikke tilfeldig valgte koder. Ikke-informativ betyr at man bruker andre variabler enn yrkeskoden selv. SHG betyr 'svarhomogen gruppe' og er egentlig et uttrykk fra utvalgsundersøkelser. Det er en gruppe som kan defineres utfra kjente variabler, og som har en ensartet respons. I vanlig justering av partielt frafall er den teoretisk beste grupperingen den med størst mulig varians mellom gruppene, og minst mulig varians innen hver gruppe.
Noen egenskaper ved den foreslåtte metoden:

- Det er forholdsvis enkelt å lage makrotall utfra den komplette yrkesvariabelen.
- Ingen ekstra estimeringsusikkerhet som ved stokastiske metoder.
- Metoden er mulig å forklare på en intuitiv måte, altså at den virker rimelig for brukerne.
- Verdiene må konverteres for å kunne brukes på mikronivå.
- Programmeringen blir mer omfattende og dokumentasjonen blir mindre sammenhengende.
- Man underslår endel variasjon i data, og metoden kan være mer sårbar for skjevheter.
- Variansen av estimatet må justeres for effekten av imputering for å gi et realistisk mål.

Vi er vant til å tenke på yrke i register som en variabel med et bestemt kode for hver person. Yrkeskode for sysselsatte uten innleverte yrkesdata vil med den forslåtte metoden framstå som en sannsynlighetsfordeling, altså en vektor av 352 verdier. Praktisk sett blir det altså 352 variabler med desimaltall mellom 0 og 1, ikke èn variabel med 352 kategorier. Det innebærer at verdiene ikke gir mening på mikronivå, men på aggregert nivå gi liknende tall som andre metoder.

For å imputere/predikere yrkeskode skulle man ideelt sett identifisert det vi kan kalle 'ikke-informative yrkeshomogene grupper', dvs. en gruppe der har alle samme yrke, og gruppetilhørighet er definert av andre variabler enn yrke. I en slik gruppe ville hver person få en realistisk verdi (1 eller 0). De fleste grupper vil i realiteten ha en viss spredning av yrker. Det gjør at man i enkelte krysstabeller vil kunne observere fiktive sammenhenger eller svingninger. Hvis man skal levere mikrodata må man foreta en stokastisk yrkesklassifisering, som gir en viss tilleggsusikkerhet. Fordelen med den forslåtte metoden er at man slipper denne ekstra usikkerheten i de vanlige tabellene.

5.3.2 Tilpasning til sysselsettingsstatistikken

I formlene beholdes uttrykket S_r (egentlig *responsgruppen*) om de som har yrkeskode, og $i \notin S_r$ om sysselsatte som ikke har yrkeskode, uansett årsak til mangel (partielt frafall eller manglende kilde). Man kan si at selve mekanikken for å tildele yrkeskode blir som ved imputeringen i Arbeidstakerregisteret, men at YHG-gruppene vi velger fra blir annerledes for dem uten yrkesdata i primærkilden.

Formel 5–2: Estimert andel av yrke *Y* innen gruppen *g* når data er hentet fra Arbeidstakerregisteret

$$\hat{Y}_g = \frac{\sum_{i \in s_r, g} y_{i,g}}{n_g}$$

Formel 5–3: Estimert andel av yrke *Y* innen gruppen *g* når data er hentet fra AKU

$$\hat{Y}_g = \frac{\sum_{i \in g} w_i y_{i,g}}{\sum_{i \in g} w_i}$$

Formel 5–4: Imputert yrke for sysselsatt person *i*

$$Y_{i \notin S_r, i \in g} = \hat{Y}_g$$

Formel 5–5: Sum imputerte i yrke *Y*

$$Y^* = \sum_{g \in G} \sum_{i \notin S_r} Y_{i,g}$$

Formel 5–6: Estimert populasjonsandel

$$\tilde{Y} = \frac{Y_{S_r} + Y^*}{N}$$

For å kvantifisere effekten av imputering i Arbeidstakerregisteret beregnes en justering av varians av estimatet. Usikkerhet pga. imputeringen kommer i tillegg til den usikkerheten som allerede finnes pga. frafall. Usikkerheten pga. rent tilfeldig frafall er svært lav i Arbeidstakerregister ettersom rapporteringsgraden har tatt seg opp, og et høyt absolutt antall.

La v^* være variansestimat for den imputerte estimator og svarandelen $\bar{r} = \frac{n}{n+m}$

Da er et forenklet frafallsjustert variansestimat gitt ved:

$v_1 = \frac{v^*}{\bar{r}}$ for deterministisk imputering

Når det gjelder effekten av prediksjon for de øvrige sysselsatte, vil det komme andre usikkerhetsmomenter i tillegg pga. kildene for YHG-estimatene. Variansestimatet v_1 vil nok derfor fortsatt være for lavt.

5.4 Grupperingsstrategi

Vi kan oppsummere bakgrunnen for valg av grupperinger slik:

- Ønske om mest mulig detaljert yrkeskode. Dette er hovedformålet.
- Best mulig kvalitet / minst mulig usikkerhet. Her kommer valg av kilde, størrelse på grupper, osv.
- Modellen bør også vurderes utfra ikke-statistiske kriterier. Her kommer slike ting som yrkesfaglige valg, skjønnsmessige vurderinger, det at modellen skal være intuitivt forståelig for våre brukere.

Før vi tester dette på reelle data, gjør vi noen valg basert på klassifikasjonene for de variablene som brukes. Dette for å begrense valgmulighetene litt, utfra en faglig og IT-resursmessig vurdering. Det betyr altså at vi ikke kjører en ren matematisk optimalisering, men gjør endel nominelle valg. Deretter brukes en trinnvis prosedyre, for å få best mulig resultater innenfor de rammene som er lagt.

Tabell 5-9: Oversikt over variabler.

Uavhengige variabler	**Nivå**	**Kommentar**
Sysselsettingstype/kilde	3 typer	Arbeidstakerregisteret, ansatte/LTO, selvstendige. Vi skal kun gi yrke for sysselsatte, ikke ledige. Skiller heller ikke ut familiearbeidere som egen gruppe.
Utdanning	6,4 og 2 siffer NUS2000-kode	2:, dette er det mest aggregerte som synes forsvarlig. 1. siffer=utdanningsnivå, 2. siffer=fagfelt 4: Utdanningsgruppe 6: Enkeltutdanning, dette er nødvendig for endel yrker. Generelt kan vi si at NUS2000 er ganske ubalansert, dette er vel pga. tilpasning til internasjonal standard.
Næring	5 og 2 siffer SN94-kode	5: Næringsundergruppe 2: Næring. Alene gir dette mange heterogene grupper i forhold til arbeidsoppgaver. Nivåene imellom gir ikke så mye forbedring.
Alder	2 kategorier for AKU-data 4 kategorier for AA-data	Omlag halvdeler ved medianen: 39 år (ansatte) 45 år (selvstendige) Omlag kvartiler: -29, 30-39, 40-49, 50-
Kjønn	2 kategorier	Brukes ikke ved de mest detaljerte nivåer.
Avhengig variabel		
Yrke	4,3,2 og 1 siffer og andre grupperinger	Ønske om detaljering gjør at vi her forsøker ved trinnvis bruk av kombinasjoner av variablene for å gi 4 siffer kode på flest mulig. Ved publisering kan både standardiserte og ikke-standardiserte aggregeringer bli aktuelle.
Datakilde		
AKU årsgjennomsnitt	Estimert størrelse > 10.000	Brukes på grupper uten yrkesdata i kilde. For selvstendige brukes siste 3 år.
Arbeidstakerregisteret	Absolutt størrelse > 100	Brukes på partielt frafall i Arbeidstakerregisteret.

Tabell 5-10: Uavhengige variabler. Aggregeringsnivå.

Tallene angir antall siffer i koden (antall kategorier for kjønn og alder)

	LTO og selvstendige			LTO	selvstendige	i tillegg for Arbeidstakerregisteret	
Type	Næring	Utdanning	Kjønn	Alder	Alder	Kjønn	Alder
$\vec{x}_8$	5	6					
$\vec{x}_7$	5	4				*2*	*4*
$\vec{x}_6$	5		*2*	*2*		*2*	*4*
$\vec{x}_5$	2	6					
$\vec{x}_4$	2	4				*2*	*4*
$\vec{x}_3$	2		*2*	*2*		*2*	*4*
$\vec{x}_2$		2	*2*	*2*		*2*	*4*
$\vec{x}_1$			*2*	*2*	*2*	*2*	*4*

Hver av disse lages for hver kilde/ sysselsettingstype. Det gir til sammen 24 typer av uavhengige variabler. Hver type inneholder et stort antall grupper, avhengig av de kombinasjoner som finnes i kildedata. Hver gruppe vil inneholde et stort eller lite antall yrker. Typen X_1 er kun for sysselsatte som er helt uten yrkesspesifikk informasjon, som f.eks. enkelte som mangler kode eller har ugyldige koder.

Vi lager de ulike typene av grupper utfra variabelkombinasjonene i tabellen. Deretter regnes ut yrkesfordelingen og et kvalitetsmål i hver gruppe. Hver person får den yrkesfordelingen som gruppen vedkommende tilhører. Kvaliteten på hver gruppe innen en type varierer en god del. Den totale kvaliteten for statistikken vil avhenge det innbyrdes forholdet av sysselsettingstypene og fordelingen på de ulike gruppene i reelle data.

Man kan regne ut gjennomsnittskvaliteten på en type, og det er et hovedmønster at kvaliteten er avhengig av detaljeringsgraden. Men det er ikke en entydig rekkefølge på kvalitet fra type 1 til 8. Konkret betyr det f.eks. at enkelte grupper av type 4 kan gi bedre kvalitet enn enkelte av type 6. Derfor blir valg av yrkeskode *ikke* bestemt direkte av typen, men av den enkelte gruppe. Praktisk vil hver person først kobles til flere yrkesfordelinger, så velges den med best kvalitet (altså ikke høyest typenummer).

6 Utprøving på 2002-data

Det er ikke aktuelt å lage yrkesstatistikk for 2002, pga. frafall i Arbeidstakerregisteret og usikker kvalitet, men vi tester ut metodikken på 2002-data inntil 2003-fila er klar. I de neste 3 avsnittene beskrives imputering og predikering på grupper av sysselsatte etter sysselsettingstype.

Kvalitetsmålene ved yrkeskodingen som oppgis her gjelder altså data for 2002. Vi går utfra at kvaliteten blir noe bedre i 2003, da det er mindre frafall og bedre kontroll av yrkeskoder fra arbeidsgivere. Blant annet er det flere bedrifter som har begynt å levere yrkeskoder maskinelt. Til slutt sammenliknes yrkesfordelingen med beregninger fra AKU-data.

6.1 Sysselsatte fra Arbeidstakerregisteret

Justerer for frafall av yrkeskode i Arbeidstakerregisteret hovedsakelig ved hjelp av samme metoder som beskrevet i notat 2003/80. Metoden blir noe modifisert i forhold til det opprinnelige notatet, for å være mer konsistent med metodene for øvrige sysselsatte. I hovedsak grupperes etter de samme variablene som for andre sysselsatte, se forrige avsnitt for detaljer. Bedriftsstørrelse er en av variablene som er foreslått i notatet, men som ikke blir anvendt her. Dette har antakelig bare betydning for spesielle lederyrker, som man har valgt å ikke prioritere i denne omgang.

Innen hver gruppe blir det flere kategorier enn de tilsvarende i AKU-data, ettersom man har en større masse med større variasjon og mindre utvalgsusikkerhet, og man kan lage litt mer homogene grupper. Allikevel må vi sette en grense for liten en gruppe kan være. Det er kjent at det forekommer tilfeldige i registerdata og det er også en ren utvalgsusikkerhet også i register, pga. frafallet.

Tabell 6-1: Nivå på partielt frafall av yrke i Arbeidstakerregistersysselsatte.

Respons	1776833	93.6 %
Frafall	121345	6.4 %
I alt	1898178	100 %

Ansatte definert utfra Arbeidstakerregisteret som mangler yrkeskode er tilsynelatende er lavere enn det som var rapportert for dette året tidligere. Dette skyldes er at man har koblet på yrke i ettertid, når dette senere er levert av arbeidsgiver. Dette kan man si er en form for imputering, og en feilkilde i dette er arbeidstakere som endrer yrkeskode innen samme arbeidstakerforhold.

Tabell 6-2: Kvalitet på yrke i forsøk med Arbeidstakerregisterdata. Etter kjønn, alder og gruppering.

	Gjennomsnittlig kvalitet			Antall personer		
	I alt	Menn	Kvinner	I alt	Menn	Kvinner
I alt	0.984	0.981	0.986	1 898 178	981 937	916 241
16-19 år	0.993	0.993	0.993	66 612	32 368	34 244
20-24 år	0.990	0.991	0.989	146 655	74 913	71 742
25-39 år	0.985	0.984	0.987	706 130	371 779	334 351
40-54 år	0.981	0.977	0.985	674 460	344 722	329 738
55-66 år	0.980	0.976	0.984	290 883	151 147	139 736
67-74 år	0.979	0.977	0.980	13 438	7 008	6 430

	Gjennomsnittlig kvalitet			Antall personer		
	I alt	Menn	Kvinner	I alt	Menn	Kvinner
I alt	0.984	0.981	0.986	1 898 178	981 937	916 241
I	0.041	0.016	0.049	216	51	165
II	0.367	0.450	0.297	9 320	4 215	5 105
III	0.461	0.410	0.562	8 929	5 898	3 031
IV	0.765	0.793	0.731	9 111	4 885	4 226
V	0.829	0.860	0.786	12 474	7 150	5 324
VI	0.692	0.660	0.756	41 230	27 295	13 935
VII	0.912	0.959	0.871	15 912	7 435	8 477
VIII	0.924	0.914	0.933	24 153	11 609	12 544
Har yrke	1.000	1.000	1.000	1 776 833	913 399	863 434

Type I – VIII er grupper spesifisert utfra *x*-variablene nevnt i forrige avsnitt. Vi ser at det er svært få som havner i de mest generelle gruppene, og at det totalt sett ser ut til at de fleste får en velbegrunnet yrkesfordeling. Kvalitetsmålet i tabellen må ikke tolkes som et uttrykk for validiteten av yrkeskodingen, men som et mål på hvor egnet variablene er for å sannsynliggjøre yrkene. En kan godt si at totalkvaliteten på sysselsatte fra Arbeidstakerregister i hovedsak skyldes det lave frafallet. De gruppene som har dårlig kvalitet her, er mest utsatt for tilfeldige feil. Dette kan gi utslag når en analyserer små enheter/regioner. Eksempel på detaljert nivå som antakelig vil bli for usikkert: endringer i antall sykepleiere innen en kommune.

6.2 Yrke for selvstendig næringsdrivende

Selvstendig næringsdrivende er som vi har sett i analysene foran, en variert og spesiell gruppe sysselsatte. Vi mangler yrkesdata i de registre som selvstendig næringsdrivende blir definert utfra. Det har vært foreslått å koble på opplysninger om person fra andre registre, f.eks. arbeidstakerforhold eller stillingskoder. Det er hovdesaklig to usikkerhetsfaktorer knyttet til dette. Det er ikke så store deler vi finner kobling til, som har gode yrkesopplysninger i andre register. Videre vil det være meget usikkert om personen har samme arbeidsoppgaver i sitt sysselsettingsforhold som selvstendig næringsdrivende, som de vedkommende har i andre sysselsettingsforhold. Særlig innen den store gruppen bønder og fiskere vil vi anta at sysselsatte har andre arbeidsoppgaver i ansettelsesforhold. Vi velger derfor å predikere yrke utfra de variablene som er nevnt og som er tilgjengelige og kompatible i både sysselsettingsstatistikksdata og AKU-data. Det betyr at vi velger å se bort fra andre opplysninger som har vært forslått, som lønn og stillingskoder.

Siden selvstendige er en liten gruppe, vil det være få personer i AKU-utvalget – noe som gir mer usikker kvalitet på yrkesfordelingen som gis. For å øke kvaliteten beregnes yrkesfordelingen utfra gjennomsnittet de siste 3 år, altså 12 kvartaler istedenfor 1. Dette berører altså kun den relative fordelingen av yrker innen hver gruppe, ikke nivåtallet for det aktuelle tidspunkt. Endringer i absolutt nivå behøver da ikke bli dårligere. Dette forutsetter den rimelige antagelse at det ikke skjer brå strukturelle endringer *innen gruppene*. Vi kan illustrere dette med et eksempel. Blant selvstendig næringsdrivende med utdanningen "457129 Tømrerfag VKII" som jobber i næringen "45211 Oppføring av bygninger", har 93% yrkeskoden " 7125 Tømrere". Antallet i dette yrket kan ha økt fra 8000 til 9000 fra år 2001 til 2002. Men vi går utfra at det i begge årene er 93% tømrere blant de med nevnte utdanning og næring. På mye lengre sikt vil det selvsagt skje strukturelle endringer, altså forskyvninger av de relative størrelsene av yrkene innen hver gruppe. Vi omberegner da dette hvert år ved hjelp de 3 foregående år.

Tabell 6-3: Kvalitet på yrke i forsøk med selvstendig næringsdrivende. Etter kjønn, alder og gruppering.

	Gjennomsnittlig kvalitet			Antall personer		
	I alt	Menn	Kvinner	I alt	Menn	Kvinner
I alt	0.661	0.672	0.630	152 000	112 653	39 347
16-19 år	0.874	0.879	0.723	332	322	10
20-24 år	0.686	0.706	0.582	2 336	1 960	376
25-39 år	0.640	0.650	0.613	39 556	28 606	10 950
40-54 år	0.662	0.674	0.629	64 661	47 403	17 258
55-66 år	0.679	0.687	0.655	38 066	28 674	9 392
67-74 år	0.668	0.679	0.623	7 049	5 688	1 361

	Gjennomsnittlig kvalitet			Antall personer		
	I alt	Menn	Kvinner	I alt	Menn	Kvinner
I alt	0.661	0.672	0.630	152 000	112 653	39 347
I	0.063	0.057	0.077	4 373	2 910	1 463
II	0.143	0.127	0.192	16 016	11 967	4 049
III	0.449	0.494	0.383	29 249	17 387	11 862
IV	0.544	0.528	0.604	11 136	8 785	2 351
V	0.769	0.751	0.813	3 355	2 376	979
VI	0.812	0.795	0.884	56 983	45 935	11 048
VII	0.952	0.956	0.940	15 894	12 125	3 769
VIII	0.986	0.981	0.999	14 994	11 168	3 826

Vi ser at det for denne gruppen må vi regne med betydelig tilfeldige feil, og at tall for små grupper ikke kan gis.

6.3 Andre lønnstakere

I 2002-data er det 208 938 som er definert som sysselsatte lønnstakere utfra LTO og eventuelt andre registre uten kobling til Arbeidstakerregisteret. Det er ikke mulig å lage en direkte sammenlikning av LTO-gruppen og selvstendige i register med tilsvarende gruppe i AKU. Vi kan få et bilde på kvaliteten ved se på spredningen av yrkene innen de gruppene som er brukt til yrkeskodingen for denne delen. Tabellene viser kvalitetsmålet fordelt etter aggregeringstyper, der VIII er mest detaljert nivå. Vi ser at som forventet fører økende detaljeringsgrad til økt kvalitet og lavere antall. Den totale kvaliteten er ikke spesielt god for denne gruppen. Det skyldes antagelig stor variasjon i arbeidsoppgaver og lite konkrete yrkesrelevante opplysninger om disse sysselsatte.

Tabell 6-4: Kvalitet på yrke i forsøk med LTO-sysselsatte. Etter kjønn, alder og gruppering.

	Gjennomsnittlig kvalitet			Antall personer		
	I alt	Menn	Kvinner	I alt	Menn	Kvinner
I alt	0.380	0.356	0.402	216 822	105 466	111 356
16-19 år	0.414	0.377	0.456	31 937	16 754	15 183
20-24 år	0.382	0.350	0.413	53 160	26 197	26 963
25-39 år	0.378	0.360	0.393	65 531	29 508	36 023
40-54 år	0.372	0.355	0.385	33 467	14 340	19 127
55-66 år	0.353	0.341	0.367	21 806	11 743	10 063
67-74 år	0.352	0.344	0.366	10 921	6 924	3 997

	Gjennomsnittlig kvalitet			Antall personer		
	I alt	Menn	Kvinner	I alt	Menn	Kvinner
I alt	0.380	0.356	0.402	216 822	105 466	111 356
I	0.032	0.013	0.033	260	7	253
II	0.178	0.192	0.175	9 864	2 055	7 809
III	0.239	0.183	0.296	17 141	8 726	8 415
IV	0.237	0.235	0.239	26 410	14 616	11 794
V	0.372	0.355	0.384	34 809	15 044	19 765
VI	0.411	0.381	0.446	72 727	39 356	33 371
VII	0.424	0.400	0.448	37 123	18 358	18 765
VIII	0.628	0.611	0.638	18 488	7 304	11 184

6.3.1 Tilfeldige feil i yrke for alle sysselsatte

De fleste sysselsatte vil ha yrkeskode fra Arbeidstakerregisteret. Denne yrkeskoden skal være kontrollert av arbeidsgiver, men det er en liten mengde arbeidstakerforhold er kodet feil, ikke kontrollert, osv. Størrelsen på

dette kan ikke beregnes her, så vi setter konsistensen i tabellen lik 1. Sysselsatte uten direkte yrkesdata vil for de fleste ha en sannsynlig yrkesfordeling, basert på opplysninger som vi mener er svært relevante for arbeidsoppgavene. Allikevel vil det være en større eller mindre usikkerhet knyttet til yrket til hver enkelt person. Totalt sett kan vi forvente gjennomsnittlig 10% tilfeldige feil når vi ser på yrkeskoden på detaljert nivå pr. person. Dette vil gi utslag i at vi ikke kan gi statistikk for små grupper, etter variabler som ikke inngår i gruppedefinisjonen.

Tabell 6-5: Kvalitet på yrke i hele forsøket. Etter kjønn, alder og gruppering.

	Gjennomsnittlig kvalitet			Antall personer		
	I alt	Menn	Kvinner	I alt	Menn	Kvinner
I alt	0.904	0.897	0.912	2 267 000	1 200 056	1 066 944
16-19 år	0.806	0.783	0.828	98 881	49 444	49 437
20-24 år	0.827	0.823	0.831	202 151	103 070	99 081
25-39 år	0.919	0.919	0.920	811 217	429 893	381 324
40-54 år	0.928	0.920	0.937	772 588	406 465	366 123
55-66 år	0.908	0.894	0.926	350 755	191 564	159 191
67-74 år	0.691	0.667	0.731	31 408	19 620	11 788

	Gjennomsnittlig kvalitet			Antall personer		
	I alt	Menn	Kvinner	I alt	Menn	Kvinner
I alt	0.904	0.897	0.912	2 267 000	1 200 056	1 066 944
I	0.061	0.056	0.068	4 849	2 968	1 881
II	0.212	0.209	0.216	35 200	18 237	16 963
III	0.386	0.394	0.375	55 319	32 011	23 308
IV	0.413	0.423	0.399	46 657	28 286	18 371
V	0.510	0.541	0.482	50 638	24 570	26 068
VI	0.613	0.617	0.603	170 940	112 586	58 354
VII	0.659	0.687	0.623	68 929	37 918	31 011
VIII	0.845	0.865	0.823	57 635	30 081	27 554
Har yrke	1.000	1.000	1.000	1 776 833	913 399	863 434

6.4 Sammenlikning med AKU

Sammenlikner vi yrkesfordelingen for alle sysselsatte i register med yrkesfordelingen i AKU, finner vi de systematiske forskjeller. I den grad vi betrakter yrkesfordelingen i AKU som et korrekt bilde arbeidsoppgavene i populasjonen, vil et avvik fra AKU kunne betraktes som et mål på systematiske feil. Vi vil allikevel påstå at mye av forskjellene skyldes ulikheter i datagrunnlaget. To forskjellige yrkesklassifiseringer av samme sysselsettingsforhold kan begge betraktes som korrekt utfra de opplysninger som er tilgjengelig på det aktuelle tidspunkt og i hver datakilde.

Det er kjent fra tidligere analyser at det er store forskjeller mellom yrke i Arbeidstakerregisteret og AKU. Siden Arbeidstakerregisteret utgjør den største kilden til yrkesdata i sysselsettingsstatistikken, er det ikke overraskende at man finner betydelige systematiske forskjeller også her.

Estimat for sysselsettingsstatistikken er for 4.kvartal 2002, AKU-tallene er årsgjennomsnitt 2002, som er kalibrert med totalsummen. Det avviker derfor noe fra publisert estimat i AKU.

Tabell 6-1: Estimeringsforsøk, de største yrkene. Registersysselsatte og AKU 2002.

Yrke	Estimat register	Estimat AKU	Forskjell
5221 BUTIKKMEDARBEIDERE O.L.	179 415	158 104	13 %
5132 OMSORGSARBEIDERE OG HJELPEPLEIERE	78 707	70 453	12 %
3310 GRUNNSKOLELÆRERE	69 023	70 040	-1 %
9132 RENGJØRINGSPERSONALE I BEDRIFTER OL	67 021	61 145	10 %
5131 BARNE- OG UNGDOMSARBEIDERE O.L.	59 533	63 983	-7 %
4114 KONTORMEDARBEIDERE	58 908	33 704	75 %
5139 ANNET PLEIE- OG OMSORGSPERSONALE	57 933	52 230	11 %
3231 SYKEPLEIERE	46 295	50 041	-7 %
6121 MELKE- OG HUSDYRPRODUSENTER	35 470	28 676	24 %
3415 TEKNISKE OG KOMMERSIELLE SALGSREPRE	32 804	42 828	-23 %
7125 TØMRERE	32 627	40 180	-19 %
4131 LAGERMEDARBEIDERE OG MATERIALFORVAL	30 489	24 536	24 %
4113 SEKRETÆRER	29 779	43 190	-31 %
7241 ELEKTRIKERE, ELEKTRONIKERE OL.	28 920	26 654	9 %
8323 LASTEBIL- OG VOGNTOGFØRERE	28 360	28 310	0 %
1210 ADMINISTRERENDE DIREKTØRER	26 897	22 116	22 %
5163 VAKTMESTERE O.L.	24 476	24 398	0 %

Tabellen viser estimater for de største yrkene, og den relative forskjellen mellom de to kildene. For enkelte yrker vil store forskjeller på detaljert nivå utjevnes mye ved aggregering, f.eks. "Kontormedarbeidere" og "Sekretærer" har forskjeller på henholdsvis 75% og –31%. Hvis vi slår sammen de to yrkene blir forskjellen 15%. En vil anta at det er mange sammenfallende arbeidsoppgaver som gjør at disse er vanskelig å klassifisere likt i de to ulike datakildene. Det samme gjelder yrker innen *salg.* Her er det store systematiske forskjeller på detaljert nivå, som heller ikke kan "maskeres" ved aggregert publisering, f.eks. yrkene "5224 engrosselger" og "3415 salgsrepresentant". Selv om disse fordeles til helt ulike yrkesfelt 3 og 5, antar vi at mange arbeidsoppgaver ikke skiller seg så vesentlig eller at det hvertfall ikke mulig å utlede eventuelle forskjeller ved tilgjengelige registerdata.

Noe av samme problemstillingen kan vi gå utfra ved pleie/omsorgsyrker som begynner med "513". Videre vil bruken av restgrupper utgjøre en kilde til ulikheter, som f.eks. "5139 Annet pleie/omsorgspers." Dette kanskje særlig innen offentlig virksomhet der yrkeskoden konverteres fra stillingskoder, hvor mange havner i "2419 Andre yrker innen off. adm."

Som en hovedregel vil store yrker blir overrepresentert og små yrker underrepresentert. Dette skyldes selve metodikken, der de aller minste gruppene ikke brukes i estimeringen. Små grupper i AKU vil ha betydelig utvalgsusikkerhet og i AKU publiseres ikke yrker med antall under 5000 personer. Det er hele 244 yrker som er for små til å kunne sammenliknes med AKU-tall. Av disse er det 15 som ikke forekommer i det hele tatt i 2002, f.eks. 3473 klovner, mm.; 4214 pantelånere; 7423 kurvmakere, mm..; 7433 buntmakere; 7434 gradører; 7435 parykkmakere; 9141 vinduspussere. Dette betyr ikke nødvendigvis at registertallene for disse yrkene vil være spesielt usikre, men at man ikke har en uavhengig kilde å sammenlikne med. Som nevnt vil små yrker underestimeres pga. av selve metodene som brukes. Resultatet av dette er at vi ved etterspørsel for tall på spesielle yrker må lage egne beregninger for dette. Eksempler på slike yrker er:

1120 toppledere i offentlig forvaltning
1141 ledere i partipolitiske organisasjo
1221 produksjonsdirektører innen jordbr
1311 ledere innen jordbruk, skogbruk
1318 ledere innen renovasjon, personlig
2111 fysikere og astronomer
2112 meteorologer
2121 matematikere og tilsvarende yrker
2122 statistikere
2224 farmasøyter
2225 ernæringsfysiologer
2351 spesialister i utdanningsmetodikk
2531 arkivarer og konservatorer
2532 universitetsbibliotekarer
2542 sosiologer, sosialantropologer, mm
2543 historikere, arkeologer og filosof
2554 koreografer og dansere
3132 operatører av kringkastings- og tel
3144 flygeledere
3151 branninspektører
3212 agroteknikere
3213 skogingeniører, skogkonsulenter o.l
3222 helse- og miljøinspektører
3223 kostholdskonsulenter
3225 tannpleiere
3227 dyrepleiere
3228 reseptarer
3421 handels- og skipsmeglere
3472 sangere og musikere i underholdning
4111 stenografer, referenter o.l.
4112 dataregistrerere (puncheoperatører)
4129 andre tallbehandlere
5113 reiseledere og guider
5121 internatledere o.l.
5142 begravelsesbyrå- og krematoriearbei
5143 slankeverter, solstudioverter o.l.
5149 andre yrker innen personlig tjenest
5169 annet sikkerhetspersonale
5210 mannekenger o.l.
6122 egg- og fjærfeprodusenter
6129 andre dyreoppdrettere og røktere
6412 fangstfolk
7123 jernbindere
7127 tunnel- og fjellarbeidere, sprengn.
7131 taktekkere
7143 sandblåsere o.l.
7144 feiere
7211 støpere
7215 riggere og spleisere
7216 dykkere
7221 smeder
7222 børsemakere, låsesmeder o.l.
7312 musikkinstr.makere og -stemmere
7321 pottemakere og keramikere
7322 glasshåndverkere
7331 kunsthåndverkere i tre ol.
7413 prøvesmakere/kvalitetsbedømmere
7422 trebåtbyggere
7431 vevere, strikkere ol. innen husflid
7435 parykkmakere
7442 skomakere
8112 prosessoperatører (oppredning)
8131 keramiske formere og dekoratører
8132 operatører innen glassproduksjon
8139 andre operatører innen glass,mm.
8143 operatører innen spon- og fiberplat
8222 operatører innen ammunisjons- o.l.
8223 operatører innen gummiproduksjon
8225 operatører innen maling- og lakkpro
8229 operatører innen annen kjemisk-tekn
8252 bokbindere
8254 fotolaboranter
8261 spinne- og nøstemaskinoperatører
8262 veve og hekle-/strikkemaskinoperatø
8264 tekstiloperatører innen fiskeredska
8265 tilskjærere
8267 operatører innen produksjon av sko,
8312 skiftekonduktører
9120 forefallende arbeid for privatperso
9131 rengjøring/husarbeid i privathush
9153 måleravlesere ol.
9210 hjelpearbeidere innen jordbruk, sko

Tabell 6-2: Estimeringsforsøk, utvalgte yrker. Registersysselsatte og AKU 2002.

De mest overestimerte

Yrke	Estimat register	Estimat AKU	Forskjell
4114 KONTORMEDARBEIDERE	58 908	33 704	75 %
3418 BANKFUNKSJONÆRER	22 671	17 509	29 %
4142 POSTBUD OG -SORTERERE	21 028	16 851	25 %
2512 PERSONAL- OG ORGANISASJONSKONSULENT	16 957	13 602	25 %
6121 MELKE- OG HUSDYRPRODUSENTER	35 470	28 676	24 %
4131 LAGERMEDARBEIDERE OG MATERIALFORVAL	30 489	24 536	24 %
5141 FRISØRER, KOSMETOLOGER O.L.	18 463	14 997	23 %
1210 ADMINISTRERENDE DIREKTØRER	26 897	22 116	22 %
2419 ANDRE YRKER INNEN OFFENTLIG ADMINIS	16 772	13 998	20 %
8321 BIL-, DROSJE- OG VAREBILFØRERE	19 070	15 991	19 %

De mest underestimerte

Yrke	Estimat register	Estimat AKU	Forskjell
0111 MENIGE	4 955	10 951	-55 %
3471 DEKORATØRER, DESIGNERE OG REKLAMETE	5 470	10 757	-49 %
1224 PRODUKSJONSDIREKTØRER INNEN VAREHAN	6 540	12 451	-47 %
2320 LEKTORER OG ADJUNKTER I VIDEREGÅEND	12 207	19 755	-38 %
2130 SYSTEMUTVIKLERE OG PROGRAMMERERE	23 208	36 496	-36 %
4113 SEKRETÆRER	29 779	43 190	-31 %
3419 MARKEDSFØRINGS- OG REKLAMEKONSULENT	9 249	12 227	-24 %
3415 TEKNISKE OG KOMMERSIELLE SALGSREPRE	32 804	42 828	-23 %
3119 ANDRE INGENIØRER OG TEKNIKERE	9 788	12 552	-22 %
1227 PRODUKSJONSDIREKTØRER INNEN OFFENTL	9 967	12 685	-21 %

De med minst forskjeller i estimatene

Yrke	Estimat register	Estimat AKU	Forskjell
7237 INDUSTRIMEKANIKERE	10 995	11 853	-7 %
2221 LEGER	14 795	15 669	-6 %
1222 PRODUKSJONSDIREKTØRER INNEN OLJE- O	11 347	12 012	-6 %
9133 KJØKKEN- OG ANRETNINGSASSISTENTER	22 761	23 666	-4 %
1228 PRODUKSJONSDIREKTØRER INNEN UNDERVI	15 729	16 157	-3 %
8331 ANLEGGSMASKINFØRERE	14 304	14 575	-2 %
3310 GRUNNSKOLELÆRERE	69 023	70 040	-1 %
8323 LASTEBIL- OG VOGNTOGFØRERE	28 360	28 310	0 %
5163 VAKTMESTRE O.L.	24 476	24 398	0 %
8322 BUSS- OG SPORVOGNFØRERE	12 855	12 225	5 %

Det er viktig å merke seg at tabellen gjelder forsøk med 2002-data, før man har omgruppert yrker som beskrevet i avsnittet foran. Det er også gjennomført tiltak som vil gi andre resultater for 2003:

- Arbeidsgiver skal ha kontrollert og rettet yrkeskoder for alle sine ansatte.
- Statistisk sentralbyrå har gjennomført revisjoner av yrkeskoder på mange arbeidstakerforhold.

Andre forhold som gjelder sammenliknbarheten med yrkesfordelingen i AKU, vil blir beskrevet i forbindelse med publisering av yrkesfordelingen for 2003.

7 Lister

7.1 Figurer

7.2 Formler

7.3 Tabeller

7.4 Referanser

Notat 04/2004 "Kvaliteten i arbeidsmarkedsdelen i Folke- og boligtellingen 2001" av Aslaug Hurlen Foss 3
notat 79/2003 (Villund) 2
notat 80/2003 (Villund) 2
upublisert notat 18.10.2002 av Leiv Solheim 2

De sist utgitte publikasjonene i serien Notater

2004/21 A. Holmøy og E. Wedde: Undersøkelse om arbeid, livsstil og helse 2003. Dokumentasjonsrapport. 38s.

2004/22 H.C. Hougen og M.A. Gløboden: Samordnet levekårsundersøkelse 2002-tverrsnittundersøkelsen. Dokumentasjonsrapport. 110s.

2004/23 H. Utne: Håndbok for Folke- og boligtelling 2001. 63s.

2004/24 A. Holmøy: Undersøkelse om livsløp, aldring og generasjon (LAG). Dokumentasjonsrapport. Oppdatert versjon av Notat. 2003/88. 129s.

2004/25 A. Vedø: Vekter i undersøkelsen om samvær og bidrag 2002. 13s.

2004/26 A.H. Sætre: Undersøkelsen om samvær og bidrag 2002. Dokumentasjon- og tabellraport. 109s.

2004/27 A. Holmøy: Undersøkelse om Livsløp, aldring og generasjoner (LAG) blant personer fra 80 år og oppover. Dokumentasjonsrapport. 182s.

2004/28 A. Holmøy: Omnibusundersøkelsen januar/februar 2004. Dokumentasjonsrapport. 37s.

2004/29 D.Q. Pham: Sesongjustering for boligprisindeksen. 28s.

2004/30 D.Q. Pham: Sesongjustering prisindeks for kontor- og forretningseiendommer. 14s.

2004/31 M.T. Dzamarija og T. Kalve: Barn og unge med innvandrerbakgrunn. 98s.

2004/32 T. Gulbrandsen og B.O. Lagerstrøm Undersøkelse om dommeravhør og observasjoner av barn i seksuallovbruddsaker. 85s.

2004/33 I. Johansen: Undersøking om foreldrebetaling i barnehagar, januar. 45s.

2004/34 P. Drevland: Offentlig forvaltning i historisk nasjonalregnskap, beregninger for 1949-1969. 17s.

2004/35 E. S. Bjørkli, K. L. Hansen, G. M. Pilskog, T. K. Schjerven og T. Smith: Fristilling og konkurranseutsetting i KOSTRA– bedring av sammenlignbarheten i nøkkeltallene. 104s.

2004/36 A. H. Foss og L. Taule: Museums-statistikken. En gjennomgang av definisjoner, kvalitet og populasjon. 26s.

2004/37 T. E. Haug og T. A. Johnsen: Datagrunnlag for en regional nordisk kraftmarkedsmodell. Produksjonsanlegg, overføringsnett, kraftetterspørsel og -priser. 15s.

2004/38 A. Bruvoll og Ø. Skullerud: Framskrivninger av organisk avfall for 2001-2002. 14s.

2004/39 S.K.Boateng og S. Ferstad: Dokumenta-sjonsnotat for FylkesKOSTRA videregående opplæring. Publisering av 2002-tallene. 197s.

2004/40 A. Finstad, K. Flugsrud, L. Høgset og G. Haakonsen Energiforbruk utenom elektrisitet i norske kommuner - en gjennomgang av datakvalitet. 31s.

2004/41 K. Løyland og T.O. Thoresen: En undersøkelse av den registrerte dagmammavirksomheten. 130s.

2004/42 T. Nygård: Kvalitetsarbeid knyttet til kvartalsvis nasjonalregnskap (KNR) Rapport fra prosjektgruppen . 130s.

2004/43 E.Engelien, G. Haakonsen og M. Steinnes: Støyplage i Norge. Resultater fra førstegenerasjonsmodell for beregning av antall støyutsatte og SPI. 109s.

2004/44 E. Wedde: Mediebruksundersøkelsen 2003. Dokumentasjonsrapport. 32s.

2004/45 A.S. Abrahamsen og D. Rafat: Analyser av populasjonen i UT- prosjektet - ikke-finansielle foretak. 80s.