

Harald Lunde

**Lønnsstatistikk for ansatte i
oljesektoren 2006**
Klargjøring og analyse

Notater

Innhold

<i>Sammendrag</i>	2
1. Innledning	3
2. Populasjon og utvalg	3
3. Datafangsten	3
3.1. Datainnngang	4
3.2. Brutto inkommet	4
4. Behandling av data fra Oracle- basen til endelig produksjonsfil	5
4.1. Tilrettelegging av endelig produksjonsfil	6
4.1.1. Program 1 nedlesing fra Oracle-basen.....	6
4.1.2. Program 2 arbeidstidsforhold	7
4.1.3. Program 3 beregning av månedslønn	8
4.1.4. Program 4 yrkeskoding.....	9
4.1.5. Program 5 godkjenning av observasjoner.....	10
4.1.6. Program 6 vekting.....	11
4.1.7. Program 7 Vektjustering.....	12
4.1.8. Frafall.....	12
5. Effekten av kontroll- og revisjonsrutinene	13
5.1. Arbeidstid.....	13
5.2. Lønnsvariablene	14
6. Vedlegg	18
Vedlegg 1. Prosessdiagram lønnsstatistikk	18
Vedlegg 2. Spredningsmål i lønnsstatistikken	20
Figurer	
Figur 1 Datainnngang i olje- og gassutvinning og bergverksdrift. Antall innkvitterte og på tabell, 2006.....	4

Sammendrag

I lønnsstatistikk for olje- og gassutvinning og bergverksdrift benyttes et sett kontroll- og revisjonsrutiner som i utgangspunktet er standardiserte for alle lønnsstatistikkene.

Effekten av å kontrollere/revidere rådatamaterialet er i de fleste tilfeller betydelig, spesielt innen enkelte yrkesgrupper. En sammenligning mellom det ferdig reviderte datamaterialet og rådata fra oppgavegiverne viste klare forskjeller både med tanke på gjennomsnittsverdier og standardavvik i materialet.

Av de innkomne observasjonene i 2006 var det om lag 96 prosent som ble godkjent til frigivning. Den tilsvarende prosentandelen i 2005 var om lag 97 prosent. I år har det blitt lagt noe mer vekt på å benytte opplysninger innhentet til kvartalsvis lønnsindeks som utfyllende informasjon. En effekt ved dette er at feil i bonus og uregelmessige tillegg ble avdekket for en del foretak. Dette har ført til at en rekke foretak måtte levere nye tall, eller at tallene har blitt korrigert.

Datafangsten i 2006 var god med tanke på hastighet og samlet datainnngang, men det ble brukt en del ressurser på å kontrollere og justere for feilrapportering av bonus, uregelmessige tillegg og overtid. Disse variablene har i en rekke tilfeller blitt rapportert for en for kort periode. For enkelte grupper i næringa utgjør disse lønnsartene en relativt stor andel av lønna, og det får dermed store konsekvenser ved feilrapportering.

Arbeidstiden i oljenæringa har også over tid gjennomgått endringer. Trenden har gått mot at stadig flere heltidsansatte offshorearbeidere får en avtalt arbeidstid ned mot på 28 timer. Dette er i utgangspunktet under heltidsgrensa i lønnsstatistikken, som normalt sett er satt til 33 timer per uke.

Forhold som bør følges opp særskilt ved utarbeiding av lønnsstatistikk for olje- og gassutvinning og bergverksdrift:

- Hittil-i-år verdier for bonus, uregelmessige tillegg og overtid. Ikke alle rapporterer for hele perioden fra 1.1 til tellingstidspunktet for næringa.
- Næringa er preget av få og store foretak som må følges tett i datafangsten.
- Den store lønnsforskjellen mellom olje- og gassutvinning og bergverk gjør at det samlede lønnsnivået i næringa er følsomt for skjevheter i det estimerte antallet (sum av vekter) mellom næringene.
- Arbeidstiden. Det må sjekkes ut om flere grupper får kortere arbeidstid enn 33 timer per uke.

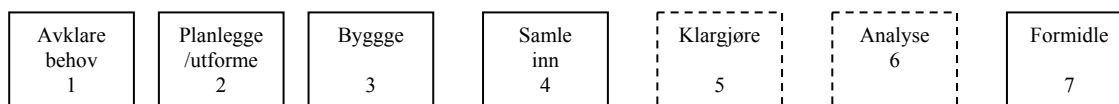
1. Innledning

Dette notatet dokumenterer arbeidsprosessene med å klargjøre, kvalitetssikre og analysere datagrunnlaget til lønnsstatistikken for olje- og gassutvinning og bergverksdrift i 2006.

Notatet er i utgangspunktet todelt, der første del tar for seg datafangsten til lønnsstatistikken, mens del 2 tar for seg kontroll- og revisjonsopplegget og den effekt dette har. Årsaken til denne todelingen er at datafangst og kontroll- og revisjonsrutiner henger tett sammen, og at de til en viss grad er betinget av hverandre. Dvs. kvaliteten på datafangsten virker inn på hvordan kontroll- og revisjonen skal legges opp, og motsatt.

I lønnsstatistikken er det anledning til å svare enten elektronisk (Altinn eller på diskett/CD) eller på papirskjema. Bearbeidningen av data frem til ferdig reviderte tall kan deles inn i fem trinn: Mottak, registrering, innlesning til Oracle-tabell, og etablering av produksjonsfil. Dataene blir kontrollert og revidert på hvert av disse trinnene, men dette notatet konsentrerer seg, i tillegg til datafangsten, om den siste delen kontroll- og revisjonsopplegget. Den skjer etter nedlesing av data fra Oracle-tabellen ved etablering av endelig produksjonsfil (se figur vedlegg 1).

Statistisk sentralbyrås virksomhetsmodell for standardiserte arbeidsprosesser i statistikkproduksjonen er en modell med 7 statistikkprosesser som bygger på hverandre:



Dette notatet beskriver prosess 5 og 6, klargjøring og analysering av datagrunnlaget.

2. Populasjon og utvalg

Det trekkes et utvalg av foretak fra populasjonen i Bedrifts- og foretaksregisteret (for mer om populasjon og utvalg 2006, se notat 2006/63). Foretakene blir stratifisert etter antall sysselsatte. Det er totaltelling blant foretak med sysselsetting over en viss grense, mens trekkeandelene er mindre i strata med færre sysselsatte. Foretak med mindre enn fem sysselsatte trekkes ikke. Personutvalget består av alle sysselsatte i bedriftene i de uttrukne foretakene.

3. Datafangsten

Datainngangen per 1. oktober 2006 var relativt god, og endelig status før frigivning av statistikk viste at om lag 94 prosent av de trukne foretakene var kommet inn.

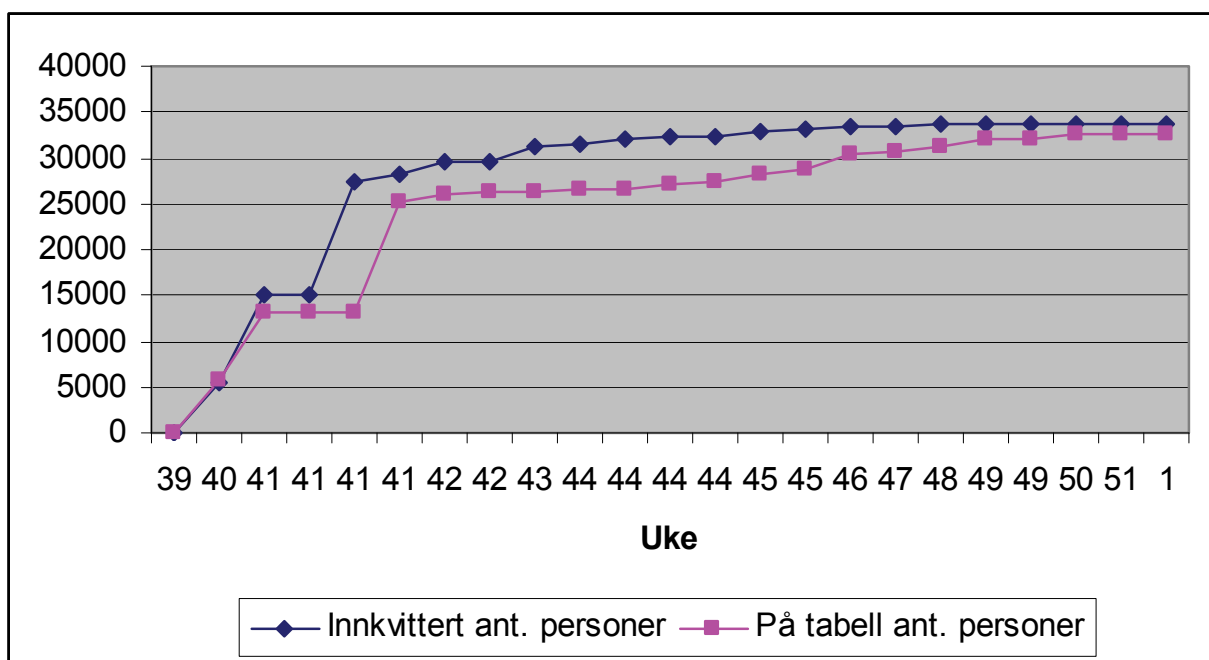
3.1 Datainnngang

Datafangsten per 1. oktober 2006

Dato	Hendelse	Antall foretak (trekkeenheter)
18.09.06 (uke 38)	Utsending	142
16.10.06 (uke 42)	Svarfrist	94 (66 prosent)
27.10.06 (uke 43)	Varsel om tvangsmulkt	26
17.11.06 (uke 46)	Ny svarfrist	
24.11.06 (uke 47)	Vedtak om tvangsmulkt	18
08.12.06 (uke 49)	Siste frist for å unngå tvangsmulkt	
19.01.07 (uke 3)	Frigivning av statistikk	133 (94 prosent)
Antall fritak gitt	Ingen	

Innkvitteringen av foretak innen olje- og gassutvinning og bergverksdrift gikk i hovedsak bra, og det meste av materialet kom inn på et relativt tidlig tidspunkt. Næringa er preget av få og store foretak, og derfor burde det være mulig å ha god oversikt over oppgavegiverne.

Figur 1 Datainnngang i olje- og gassutvinning og bergverksdrift. Antall innkvitterte og lest inn i Oracle-tabellen, 2006.



Ved svarfrist i uke 42 var om lag 30 000 observasjoner innkvitterert, og 25 000 av disse var innlest på tabell.

3.2 Brutto inkommet

I tabell 1 finner vi endelig antall foretak, bedrifter og sysselsatte som er innkvittert i olje- og gassutvinning og bergverksdrift i 2005 og 2006. Andelen foretak og bedrifter innkvittert i 2006 var på omtrent samme nivå som i 2005, mens andelen sysselsatte innkvittert er noe høyere i 2006 enn i 2005. Dette kan tyde på at det har vært fokus på å hente inn de største foretakene.

Tabell 1 Endelig antall foretak, bedrifter og sysselsatte innkvittert i olje- og gassutvinning og bergverksdrift i 2006 og prosentvis andel innkvittert i 2005 og 2006

Datainnngang			
Enhet	2006		2005
	Antall	Prosent	Prosent
	Innkvittert	Innkvittert	Innkvittert
Foretak	133	94	92
Bedrifter	217	81	83
Sysselsatte	33 696	100	87

Tabell 2 er en oversikt over andelen som benytter seg av de ulike innrapporteringsformatene i perioden 2004 til 2006. Den klare hovedvekten ligger på innrapportering via kravspesifikasjon (CD eller diskett), mens rapportering via Altinn er lav, men økende.

Tabell 2 Andel observasjoner etter innrapporteringsformat i perioden 2004 til 2006

KODE	Prosent		
	2006	2005	2004
I alt	100	100	100
Excelfiler	6,3 %	9,0 %	11,2 %
Altinn	0,9 %	0,1 %	0,0 %
Kravspesifikasjon	88,5 %	86,0 %	86,0 %
Manuelt registrert	2,3 %	2,7 %	0,8 %
Optisk lest	2,0 %	2,3 %	2,0 %

4. Behandling av data fra Oracle-basen til endelig produksjonsfil

Å innhente lønnsinformasjon på individnivå fra næringslivet har vist seg å medføre at alle opplysninger som kommer inn ikke nødvendigvis er korrekte. Det kan være ulike typer feil i materialet, fra relativt ubetydelige unøyaktigheter, til mer graverende tilfeller som medfører at observasjonen må forkastes. Det er derfor utarbeidet et system med kontroll- og feilrettingsrutiner for lønnsstatistikken, der hovedmålsettingen er å avdekke eventuell systematikk i typen feil og mangler i det som blir innrapportert. I tillegg er det en målsetting å minimere muligheten for at enkeltobservasjoner med åpenbare feil og mangler skal få uønsket innvirkning i statistikken. Systemet er bygget opp med enkle logiske mikrokontroller først (tabell 3-9), før det kontrolleres mer på makronivå etterpå (tabulering bl.a). Allerede ved innlesing av rådata, enten det måtte være fra Altinn-portalene, fra filer basert på kravspesifikasjonen for elektronisk innrapportering, eller fra optisk leste skjema blir det kontrollert for logiske feil og mangler. Dette skjer litt forenklet sagt ved at observasjoner merkes i Oracle-basen, gradert etter betydningen av feilen/mangelen. Ved betydelige feil og mangler blir oppgavegiveren kontaktet med det samme, og oppfordret om å levere nye data.

Denne dokumentasjonen tar for seg kvalitetssikringen av datamaterialet *etter* at den innledende mikro-kvalitetssikringen er gjennomført i databasen. Det er med andre ord allerede ”ryddet opp” i en del av de mest åpenbare feil og mangler.

Tilrettelegging av data fra Oracle-basen fram til endelig produksjonsfil gjøres ved hjelp av fastlagte SAS-rutiner. Rutinene er lagt opp slik at de foregår i 5 trinn eller operasjoner, som i stor grad er standardisert og tilnærmet like fra næringsområde til næringsområde. Opplegget består av:

- Rutiner for *merking* av observasjoner med mulige feil
- Rutiner for *justering/korriger*ing av observasjoner med opplagte feil
- Rutiner for *å holde utenfor* observasjoner med opplagte feil

I korte trekk er rutinene bygget opp som følger:

Trinn 1: Nedlesing fra Oracle og tilrettelegging av sasfil. Observasjoner uten verdier på relevante variable tas ut.

Trinn 2: Grunnleggende kontroll og justering av bakgrunnsvariablene arbeidstid og avlønningsperiode.

Trinn 3: Etablering av lønnsbegrepene (lønn per måned)

Trinn 4: Godkjenning av yrkeskoder

Trinn 5: Endelig godkjenning av observasjoner. Tar ut observasjoner med ekstremverdier i en eller flere lønnsvariable

Deretter følger trinn 6, der materialet vektet og sjekkes opp mot populasjonen (BoF).

Se vedlegg 1 for skjematisk oversikt over disse rutinene.

4.1 Tilrettelegging av endelig produksjonsfil

4.1.1 Program 1 nedlesing fra Oracle-basen

Tabell 3 viser gyldige og ugyldige fødselsnummer innen olje- og gassutvinning og bergverksdrift i 2005 og 2006. D-nummer er fødselsnummer til utenlandske arbeidere, og av totalen utgjør utenlandske arbeidere 2,9 prosent i 2006. Alle observasjoner tas med videre i statistikkgrunnlaget, uavhengig av om de har gyldig fødselsnummer eller ikke.

Tabell 3 Gyldige og ugyldige fødselsnummer. 2005 og 2006

Fødselsnummer	D- nummer	Antall	
		2005	2006
Ugyldige fødselsnummer og	ikke d-nummer	32	60
Gyldige fødselsnummer og	ikke d-nummer	26 437	28 370
Mangelfullt eller ugyldig fødselsnummer og	ugyldige d-nummer	0	5
Mangelfullt eller ugyldig fødselsnummer og	gyldig d-nummer	536	839
Mangelfullt eller ugyldig fødselsnummer og	ikke d-nummer	10	15

4.1.2 Program 2 arbeidstidsforhold

Program 2 kontrollerer og justerer lønnsperioder og arbeidstid. Kontrollene viser at enkelte observasjoner plasseres i feil lønnsperiode og må omplasseres slik at lønnsopplysningene får riktig håndtering videre i kontrollene.

Alle ansatte har en avlønningsperiode:

- Lønnsperiode 1: Månedslønn 12 ganger i året
- Lønnsperiode 2: 14-dagerslønn
- Lønnsperiode 3: Lønn hver uke
- Lønnsperiode 4: Månedslønn 12,5 ganger i året
- Lønnsperiode 5: Månedslønn 13 ganger i året
- Lønnsperiode 6: Timelønn
- Lønnsperiode 7: Provisjons- eller prosentlønn

Tabell 4 Status etter kontroller og endringer

Status observasjoner	Frekvens	Prosent
I alt	29 289	100,0
Disse må sjekkes. Fastlønte med timelønn i tillegg	240	0,8
Disse må sjekkes. Høy timefortjeneste og lav arbtid	1	0
Disse må sjekkes. Timelønte med fastlønn i tillegg	4	0,0
Månedslønte som mangler arbtid	50	0,2
Obs med lønnsperiode 6 som endres til 1	3	0,0
Observasjoner med annen lønnsperiode som gis lønnsperiode 6	29	0,1
Observasjoner som er uendret	28 372	96,9
Sjekkes. Heltid og fastlønn<10000 (kommafeil?)	425	1,5
Sjekkes. Timelønte med arbtid>45.	86	0,3
Timelønte uten arbtid som gis arbtid=timeant/4.33	79	0,3

Tabell 4 viser at så mye som 97 prosent av materialet slipper igjennom den første kontrollen som gjøres. Største feilpost er observasjoner med veldig lave lønnsutbetalinger i forhold til arbeidstid.

Tabell 5 Antall observasjoner fordelt på lønnsperiode før og etter kontroll og korreksjoner av arbeidstid

Lønnsperiode	Antall	
	Før	Etter
I alt	29 289	29 289
1	24 010	23 640
2	17	11
4	1	0
5	3 872	3 872
6	1 385	1 765
7	1	1
Annet	3	0

4.1.3 Program 3 beregning av månedslønn

Dette programmet foretar først og fremst en beregning av fastlønn, bonus og uregelmessige tillegg. Det foretas også korreksjon og justering av overtidstimer og overtidsgodtgjørelse.

Månedslønnsbegrepet består av en summering av avtalt lønn, uregelmessige tillegg og bonus. For å beregne månedslønn må det derfor først gjøres en beregning av de øvrige lønnslementer som månedslønna bygger på:

Avtalt lønn

Avtalt lønn er beregnet lønn siste måned for alle, både fastlønte, timelønte og provisjonslønte. Dette for å få sammenlignbare størrelser uavhengig av lønnsperiode. Innrapportert lønn siste avlønningsperiode knyttet til variabel for avlønning benyttes i utregningen av avtalt lønn:

Avlønningsperiode 1	= Oppgitt fastlønn siste periode
Avlønningsperiode 2	= Oppgitt fastlønn per 14-dager multiplisert med 2.165 (antall 14-dagers-perioder i en måned).
Avlønningsperiode 3	= Oppgitt fastlønn per uke multiplisert med 4.33 (antall uker i en måned).
Avlønningsperiode 4	= Oppgitt fastlønn per uke multiplisert med (12,5/12) .
Avlønningsperiode 5	= Oppgitt fastlønn per uke multiplisert med (13/12) .
Avlønningsperiode 6	= Hovedregel er timesats (enten beregnet eller direkte oppgitt) multipliser med antall timer i en måned (avtalt arbeidstid per uke multiplisert med antall uker i en måned).
Avlønningsperiode 7	= Hovedregel: (bonus hittil i år + bonus 4. kvartal i fjor)/12 + evt oppgitt fastlønn siste periode. Bonusfeltene tømmes etter beregning av avtalt lønn.

Uregelmessige tillegg

Innhentes som aggregert størrelse fra 1. januar til tellingstidspunktet. Gjennomsnittsberegning per måned for denne perioden.

Bonus

Innhentes både for 4. kvartal i fjor, samt som aggregert størrelse fra 1. januar til tellingstidspunktet. Gjennomsnittsberegning per måned for hele denne 12-månedperioden.

Månedslønn

Månedslønna framkommer ved en summering av de beregnede månedsverdiene for avtalt lønn, uregelmessige tillegg og bonus.

Overtidsgodtgjørelse

Inngår ikke i månedslønnsbegrepet. Innhentes som aggregert størrelse fra 1. januar til tellingstidspunktet. Gjennomsnittsberegning per måned for denne perioden.

For overtid gjøres en spesiell kontroll av forholdet mellom innrapporterte overtidstimer i forhold til overtidsgodtgjørelsen.

Tabell 6 Oversikt over overtidstimer og overtidsbetaling

Status observasjoner	Frekvens	Prosent
I alt	29 289	100,0
Observasjoner uten overtidstimer eller betaling	11 992	40,9
Observasjoner med flere overtidstimer enn betaling	48	0,2
Observasjoner med verdier på betaling	17 220	58,8
og av disse mangler det timer på	29	0,1

På basis av denne kontrollen blir overtidsgodtgjørelsen justert på observasjoner som åpenbart bare har fått innrapportert *overtidstillegg*, og ikke hele godtgjørelsen, for utført overtidsarbeid. Dette gjøres ved at det beregnes en lønn per time innen normal arbeidstid for den aktuelle observasjonen, som så *legges til* overtidsgodtgjørelsen per time.

4.1.4 Program 4 yrkeskoding

I program 4 gjennomføres en logisk kontroll av gyldighet av innrapporterte yrkeskoder. Oppgavegiverne skal rapportere yrkeskoder i henhold til standard for yrkesklassifisering (Styrk). I kvalitetssikringsarbeidet før frigivning brukes relativt mye tid på å kvalitetssikre yrkeskodingen. Blant annet sjekkes forholdet mellom yrkeskode og yrkesbetegnelse, og ved feil og mangler kontaktes oppgavegiverne i så stor grad som mulig.

1. siffer i yrkeskoden betegnes som yrkesfelt:

1. Lederyrker
2. Akademiske yrker
3. Høyskoleyrker
4. Kontor- og kundeserviceyrker
5. Salgs- og serviceyrker
6. Yrker innen jordbruk, skogbruk og fiske (ikke i bruk)
7. Håndverkere o.l.
8. Operatører og sjåførere
9. Yrker uten krav til utdanning

Innrapporteringa av yrkeskoder så slik ut i 2005 og 2006:

Tabell 7 Fordeling av observasjoner på 1-siffer STYRK, 2005 og 2006

St1	2005		2006	
	Frekvens	Prosent	Frekvens	Prosent
I alt	29 208	100	33 420	100
1	3 172	11	3 408	10
2	6 320	22	7 154	21
3	5 018	17	5 044	15
4	1 477	5	1 520	5
5	355	1	484	1
6	1	0	-	0
7	3 156	11	3 747	11
8	8 355	29	10 123	30
9	972	3	1 187	4
X	382	1	753	2

Som en kontroll av innrapporterte yrkeskoder til lønnsstatistikken kobles det på 7-sifret yrkeskode fra AA-registeret (Arbeidstaker-/arbeidsgiverregisteret) med oppdatering nærmest mulig tellingstidspunktet. Det viser seg å være relativt store avvik mellom hva som ligger i AA-registeret og hva som rapporteres til lønnsstatistikken (se tabell 8).

Tabell 8 Fordeling av antall med 1-siffer styrk i lønnsstatistikken sammenlignet med AA-registeret

Lønns- statistikken	AA- registeret										
	I alt	1	2	3	4	5	6	7	8	9	X
I alt	32 725	1 603	1 779	3 569	140	446	44	2 293	5 405	4 222	12 224
1	3 408	1 268	116	206	20	12	0	18	112	162	1 494
2	7 154	102	1 396	953	60	19	1	32	63	821	3 707
3	5 044	97	188	1 941	87	56	3	90	169	592	1 821
4	1 520	14	45	85	837	16	0	9	54	137	323
5	484	4	2	19	3	230	0	4	4	33	185
7	3 747	14	9	77	52	31	12	1 878	357	433	884
8	10 123	101	22	268	78	70	25	249	4 526	1 339	3 445
9	1 187	3	1	20	3	12	3	13	119	653	360
X	58	0	0	0	0	0	0	0	1	52	5

Det er foreløpig et problem at mange ansatte ligger uten yrkeskode i AA-registeret. Mens hele 12 000 observasjoner ligger uten kode i AA-registeret, er det i lønnsstatistikken 58 observasjoner. Andel uten kode i registeret er særlig høy blant de som er kodet på nivå 2 (akademiske yrker) i lønnsstatistikken, med over 50 prosent. I lønnsstatistikken er det ingen med kode på nivå 6 (yrker innen jordbruk, skogbruk og fiske), mens i registeret er det 44 observasjoner med kode på dette nivået.

En direkte sammenligning av den 4-sifrede yrkeskoden fra den ferdige reviderte fila fra lønnsstatistikken og fra AA-registeret, viser at snaut 11 100 observasjoner hadde identisk yrkeskode i disse to kildene, altså ikke mer enn omtrent 1/3 av observasjonene.

4.1.5 Program 5 godkjenning av observasjoner

I program 5 lages en oversikt over godkjente observasjoner og observasjoner som forkastes i statistikken. Lønnsopplysninger har maksimums- og minimumsgrenser. Disse grensene er ikke blitt endret i 2006 i forhold til 2005. Hovedtankegangen bak dette kontrollopplegget er som følger:

- Det skal settes romslige grenser for å sikre at observasjoner med antatt korrekte verdier beholdes.
- Grensene skal likevel være satt slik at observasjoner med åpenbart ekstreme verdier, som med stor grad av sikkerhet er feilaktig, holdes utenfor datagrunnlaget.
- Grenseverdiene settes på grunnlag av en rekke forhold, som blant annet erfaring fra tidligere år, kommunikasjon med forhandlingsverden (f.eks minstelønn innen de ulike tariffområdene) og oppgavegivere, samt statistiske mål på spredning. Grenseverdiene gjennomgås og justeres år for år.
- Underveis i kvalitetssikringsarbeidet sjekkes observasjoner som faller ut med feilkoder, i stå stor grad som mulig.

Tabell 9 Antall observasjoner fordelt etter koder for grenseverdier (ok-koder) i 2006

Ok-kode	Frekvens	Prosent
I alt	33 420	100,0
Ok=0 - Godkjente observasjoner	32 046	95,9
Ok=2 - Avtalt lønn 0 kroner	32	0,1
Ok=3 - Heltidsansatte med månedslønn<15000 kroner	851	1,6
Ok=7 - Ledere (heltid) som ikke har 20 000<månedslønn<200 000	50	0,2
Ok=8 - St1 = 2,3 og månedslønn over 110 000 kroner	3	0,0
Ok=10 - St1 > 5 og månedslønn over 65 000 kroner	193	0,6
Ok=11 - Overtid (ot2) over 50 000 kroner	3	0,0
Ok=12 - Tillegg (tl) over 65 000 kroner	11	0,0
Ok=13 - Bonus (bon) over 100 000 kroner	7	0,0
Ok=14 - Deltidsansatte med arbtid<=1	97	0,3
Ok=15 - Deltidsansatte, st1=1 og timefortjeneste>600	2	0,0
Ok=17 - Deltidsansatte, st1=2,3 og timefortjeneste>500	5	0,0
Ok=18 - Deltidsansatte, st1=2,3 og timefortjeneste<80	11	0,0
Ok=19 - Deltidsansatte, st1=4,5 og timefortjeneste>250	16	0,1
Ok=20 - Deltidsansatte, st1=4,5 og timefortjeneste<60	4	0,0
Ok=21 - Deltidsansatte, st1=7,8 og timefortjeneste>250	20	0,1
Ok=22 - Deltidsansatte, st1=7,8 og timefortjeneste<60	2	0,0
Ok=23 - Deltidsansatte og timefortjeneste>250	11	0,0
Ok=24 - Deltidsansatte og timefortjeneste<50	3	0,0
Ok=25 - Timelønnte med utilstrekkelig info om timer	32	0,1
Ok=99 Dubletter	21	0,1

Den vanligste ”feiltypen” er for lav lønn i forhold til arbeidstiden, feilkode 3. Dette skyldes i mange tilfeller at foretak som leverer elektronisk ikke har ”renset” lønssystemet sitt for personer som ikke er ansatt på tellingstidspunktet, men som har fått en eller annen form for godtgjørelse i løpet av året. Samlet sett holdes 96 prosent av observasjonene innen rammene av øvre og nedre grenseverdier.

4.1.6 Program 6 vektning

Programmet stratifiserer populasjonen etter antall ansatte i foretakene og beregner vekter. Vektemetoden er basert på justert etterstratifisert invers trekkesannsynlighet, beskrevet i notat av Grini (2003).

Tabell 10 Antall foretak i populasjon og utvalg og invers trekkesannsynlighet (Vekt)

Stratum	Antall			Vekt 2006	Vekt 2005
	Foretak i utvalget	Foretak i pop.	Ansatte i pop.		
CA01	22	22	25 356	1	1
CA02	11	20	2 659	1,81818	1,4375
CA03	16	96	1 847	6	6
CB01	34	35	2 369	1,02941	1,03226
CB02	18	38	617	2,11111	1,85185
CB03	20	87	625	4,35	4,04762

4.1.7 Program 7 Vektjustering

Prinsippet for justering av vektene er at sum oppblåst antall i lønnsstatistikken skal reflektere sum antall ansatte fra Bedrifts- og foretaksregisteret for hver næring (Grini 2003). Dette betyr at gjennomsnittlig lønn for et næringshovedområde skal reflektere et veid gjennomsnitt av hver av næringene som området består av. For å oppnå dette må hver av næringenes relative andel av den estimerte sysselsettinga i undersøkelsen, gjenspeile populasjonsandelene fra Bedrifts- og foretaksregisteret. I tillegg må sammenhengen mellom de to siste årganger av lønnsstatistikken kontrolleres. Denne justeringen er spesielt viktig i områder med store lønnsforskjeller mellom næringer, der avvik vil få større betydning for det endelige lønnsnivået.

Tabell 11 Vektjustering 2006. Heltids- og deltidsansatte

Næring	Vektet antall		Vektet andel		BoF andel	Justering
	2005	2006	2005	2006		
I alt	34 400	36 133	1	1	1	
Bergverk	3 800	3 752	0,110465116	0,1038386	0,108044733	1,063728838
Olje	30 600	32 381	0,889534884	0,8961614	0,891955267	0,992615347

I 2006 ble vektjusteringen gjort direkte mot 2005-årgangen av statistikken, da populasjonen i BoF ikke endret seg av betydning. Oppblåst andel av de ansatte i de to delnæringene er svært lik begge år, og samtidig veldig nær populasjonsandelene.

Effekten av vekteprosedyrene på det beregnede lønnsnivået i næringa vises i tabell 12. Vi ser at vektjusteringen som gjøres, der betydningen av oljenæringa relativt til bergverksnæringa dras noe ned, får den følge at det samlede lønnsnivået justeres ned med om lag 100 kroner.

Tabell 12 Effekt av vektning 2006. Gjennomsnittlig månedslønn for heltidsansatte. Kroner

	Månedslønn		
	Uvektet	Ujustert vekt	Justert vekt
I alt	47 047	46 995	46 882
Bergverk	48 521	48 750	48 750
Olje	31 434	30 848	30 848

4.1.8 Frafall

Vi har to hovedtyper frafall, enhetsfracfall og partielt frafall. Enhetsfracfall defineres her som frafall av bedrifter, mens partielt frafall er frafall av personer.

Enhetsfracfall skyldes at bedrifter ikke har innrapportert til oss, eller at det er feil i datamaterialet vi har fått innrapportert. Under enhetsfracfall av bedrifter kommer også de bedriftene vi har valgt å holde ute manuelt. Totalt sett var enhetsfracfallet på 9 foretak med 49 bedrifter i 2006.

Partielt frafall kan blant annet skyldes mangelfull innrapportering, feil i data eller at personer ikke er blitt innrapportert. Denne type frafall er omtalt i kapittel 4.1.5 hvor det er blitt satt minimums- og maksimumsgrenser for de ulike lønnsvariablene. I olje- og gassutvinning og bergverksdrift er de fleste observasjonene som ikke ble godkjent, heltidsansatte med månedslønn mindre enn 15 000 kroner. Det partielle frafallet var totalt sett på mindre enn 5 prosent av observasjonene i 2006.

5. Effekten av kontroll- og revisjonsrutinene

Måling av effekten av kontrollopplegget gjøres her ved å sammenligne datagrunnlaget før og etter at alle sasrutinene er kjørt. Dette kapitlet tar for seg hvilken virkning rutinene har på de mest sentrale variablene i lønnsstatistikken.

5.1 Arbeidstid

Arbeidstiden er grunnleggende bestemmende for to forhold: a) om en person er hel- eller deltidsansatt, og b) arbeidstiden benyttes for å beregne månedslønn for en representativ måned. Det legges derfor relativt mye ressurser inn i arbeidet med denne variabelen, og det er et absolutt krav i innrapporteringen at feltet må være utfyllt. Mye av kontrollene går av den grunn ut på sjekke sammenheng mellom arbeidstid og lønn.

Tabell 13 Effekt av kontroll og revisjon av avtalt arbeidstid per uke

	Antall observasjoner		Gjennomsnittlig avtalt arbeidstid	
	Før revisjon	Etter revisjon	Før revisjon	Etter revisjon
I alt	33 696	32 046	34,6	34,2
Heltidsansatte	31 080	31 494	34,9	34,8
Deltidsansatte	2 616	552	19,6	27,5

Som det går fram av tabellen er det flere observasjoner på den ureviderte fila (rådata) enn på den reviderte fila (publiseringsfila) grunnet at alle ikke-godkjente observasjoner her er holdt utenfor etter kriteriene beskrevet i kapittel 5.1.5. I tillegg til dette er også observasjoner helt uten innrapportert arbeidstid holdt utenfor den ferdig reviderte fila. Disse observasjonene registreres som deltidsansatte inntil de sjaltes ut, derav nedgangen i antall deltidsansatte fra urevidert til revidert fil. Vi ser også at det er flere heltidsansatte på den reviderte fila, altså blir en del observasjoner flyttet fra del- til heltid i revisjonen. Feil/manglende arbeidstid er hovedårsaken til dette. Av de som i utgangspunktet på rådatafila lå som deltidsansatte, har hele 1 916 observasjoner fått justert arbeidstiden slik at de er flyttet til heltid, mens 58 stykker er flyttet motsatt vei.

5.2 Lønnsvariablene

Responsvariablene i lønnsstatistikken er lønnsvariablene som frigis i den offisielle statistikken, og er derfor også de vi først og fremst ønsker å måle revisjonseffekten av. Dette gjelder først og fremst månedslønn, som består av avtalt lønn, uregelmessige tillegg og bonus. I tillegg kommer tall for overtidsgodtgjøring, men dette inngår ikke i månedslønna.

Revisjonseffekten på disse variablene defineres her som forskjellen mellom revidert og urevidert verdi, dvs. forskjellen på verdi på endelig publiseringsfil og verdi på rådatafil. Problemet i dette tilfellet er at responsvariablene ikke finnes direkte på rådatafila fordi de ikke innrapporteres direkte fra oppgavegiverne. De er avledede variable som beregnes i løpet av kontrollrutinene, på grunnlag av innrapporterte variable.

For å forenkle vurderingen av revisjonsmålingen tar vi bare for oss heltidsansatte i det følgende. Det er relativt sett få deltidsansatte i denne næringa, og i tillegg blir resultatene noe vanskeligere å tolke når lønna for de deltidsansatte skal inkluderes.

Først ser vi næringa som helhet, og tar med alle lønnsvariable, med verdier for gjennomsnitt, median, nedre og øvre kvartil, samt minimum og maksimum.

Tabell 14 Effekt av kontroll og revisjon av lønnsvariable. Heltidsansatte. Kroner

	Revidert	Urevidert	Avvik			
Antall	31 494	31 080	414			
	Gjennomsnitt			Median		
	Revidert	Urevidert	Avvik	Revidert	Urevidert	Avvik
Månedslønn	46 882	45 583	1 299	42 917	42 672	245
Avtalt lønn	40 205	38 941	1 264	36 351	35 525	826
Uregelmessige tillegg	4 824	4 740	84	1 568	1 317	251
Bonus	1 853	2 073	-220	707	722	-15
Overtid	3 075	2 677	398	887	651	236

	1. kvartil			3. kvartil		
	Revidert	Urevidert	Avvik	Revidert	Urevidert	Avvik
Månedslønn	34 978	33 489	1 489	54 987	54 926	61
Avtalt lønn	29 093	28 209	884	47 972	48 100	-128
Uregelmessige tillegg	0	0	0	7 642	7 209	433
Bonus	0	0	0	2 257	2 292	-35
Overtid	0	0	0	4 423	3 844	579

	Min			Max		
	Revidert	Urevidert	Avvik	Revidert	Urevidert	Avvik
Månedslønn	15 000	-	-	488 840	1 298 660	-809 820
Avtalt lønn	5 500	-	-	377 500	377 500	0
Uregelmessige tillegg	0	-	-	63 980	775 578	-711 598
Bonus	0	-	-	213 250	458 463	-245 213
Overtid	0	-	-	49 603	51 385	-1 782

Som det går fram av tabellen gjelder det for alle variablene, med unntak av uregelmessige tillegg, at de største endringene skjer i nedre enden av skalaen. Dvs det endres/fjernes flere observasjoner med åpenbart for lave lønnsverdier enn motsatt. Gjennomsnittlig månedslønn er derfor høyere etter enn før revisjon, og avviket er mer enn 1 300 kroner, eller omtrent 2,9 prosent av urevidert verdi. Et problem med denne tabellen er imidlertid at den tar for seg tall på et aggregert nivå, dvs næringshovedområdet som helhet. Kontroll- og revisjonsopplegget er imidlertid bygget opp og tilpasset til at den endelige lønnstatistikken skal utarbeides på et mer disaggregert nivå. Det er derfor tatt høyde for at verdi på lønnsvariablene vil variere etter gitte kjennemerker, f eks næring og yrke.

Ved å se på forskjellen mellom revidert og urevidert verdi for lønnsvariablene fordelt på næring og yrke, vil man få et bedre bilde av effekten av kontroll- og revisjonsopplegget. For å forenkle tar vi i tabell 15 bare for oss månedslønn.

Tabell 15 Effekt av kontroll og revisjon av månedslønn etter næring og yrke. Antall og kroner

	Antall			Gjennomsnitt			Antall med endret verdi
	Urevidert	Revidert	Avvik	Revidert	Urevidert	Avvik	
I alt	31 494	31 080	414	46 882	45 583	1 299	479
			0				
Bergverk i alt	2 717	2 875	-158	30 848	27 540	3 308	221
Administrative ledere	218	215	3	47 729	45 803	1 926	2
Akademiske yrker	60	66	-6	45 632	45 334	298	1
Høgskoleyrker	103	125	-22	39 841	41 004	-1 163	1
Kontoryrker	139	153	-14	29 373	26 227	3 146	13
Salgs- og serviceyrker	27	27	0	29 079	28 283	796	1
Håndverkere	693	723	-30	29 862	26 308	3 554	75
Operatører og sjåførere	1 352	1 363	-11	28 128	24 944	3 184	109
Renholdere og hjelpearbeidere	117	165	-48	24 547	17 456	7 091	19
Andre	8	38	-30	25 497	20 293	5 204	0
			0				
Olje og gassutv. i alt	28 777	28 205	572	48 750	47 862	888	258
Administrative ledere	3 056	3 026	30	75 039	75 524	-485	10
Akademiske yrker	6 846	6 114	732	53 179	53 925	-746	75
Høgskoleyrker	4 742	5 662	-920	48 100	45 832	2 268	87
Kontoryrker	1 195	1 160	35	34 705	33 744	961	12
Salgs- og serviceyrker	425	412	13	47 222	46 000	1 222	1
Håndverkere	2 853	2 949	-96	39 471	37 809	1 662	10
Operatører og sjåførere	8 093	7 879	214	42 000	41 356	644	59
Renholdere og hjelpearbeidere	919	962	-43	35 455	34 596	859	4
Andre	648	41	607	49 481	23 846	25 635	0

Antall observasjoner som får endret månedslønna i løpet av revisjonsprosessen er i underkant av 500 stykker. Dette utgjør om lag 1,5 prosent av det totale antallet heltidsansatte. Relativt sett gjøres det klart flest endringer på lønnsvariablene i bergverksnæringa, der mer enn 221 observasjoner, eller om lag 8 prosent får en justering av lønna. Tilsvarende tall for oljenæringa var under 1 prosent av de heltidsansatte.

Det er også forskjeller mellom de ulike yrkesgruppene, både hvor mange observasjoner lønna blir endret for, og hvor store endringene er. Lønnsforskjellene mellom ureviderte og reviderte tall skyldes her to forhold:

- a) yrkeskoden er endret
- b) lønna er endret

Som det går fram av tabellen er det klart større omfang ev endring i yrkeskoder enn endring i lønnsvariable, og nær 2 100 observasjoner har fått endret yrkesgruppe (1.siffer i yrkeskoden) i revisjonen.

For eksempel er månedslønna justert opp med 2 200 kroner for høgskoleyrker i oljenæringa, som en følge av at mer enn 900 observasjoner er tatt ut av denne gruppa i revisjonsprosessen, i tillegg til at lønna er endret for 87 observasjoner.

Hvis man skal separere effekten av bare lønnsjusteringer ved å se på kun de observasjonene som er med i det endelige reviderte datasettet, viser det seg at månedslønna justeres opp med vel 400 kroner samlet sett.

Som en følge av revisjonsrutinene minskes også den samlede spredningen. I tabell 16 er det beregnet ulike spredningsmål for heltidsansatte i olje- og gassutvinning og i bergverksdrift på ureviderte og reviderte tall for 2006.

Tabell 16 Effekt av kontroll og revisjon av månedslønn etter næring. Spredningsmål og yrke. Antall og kroner

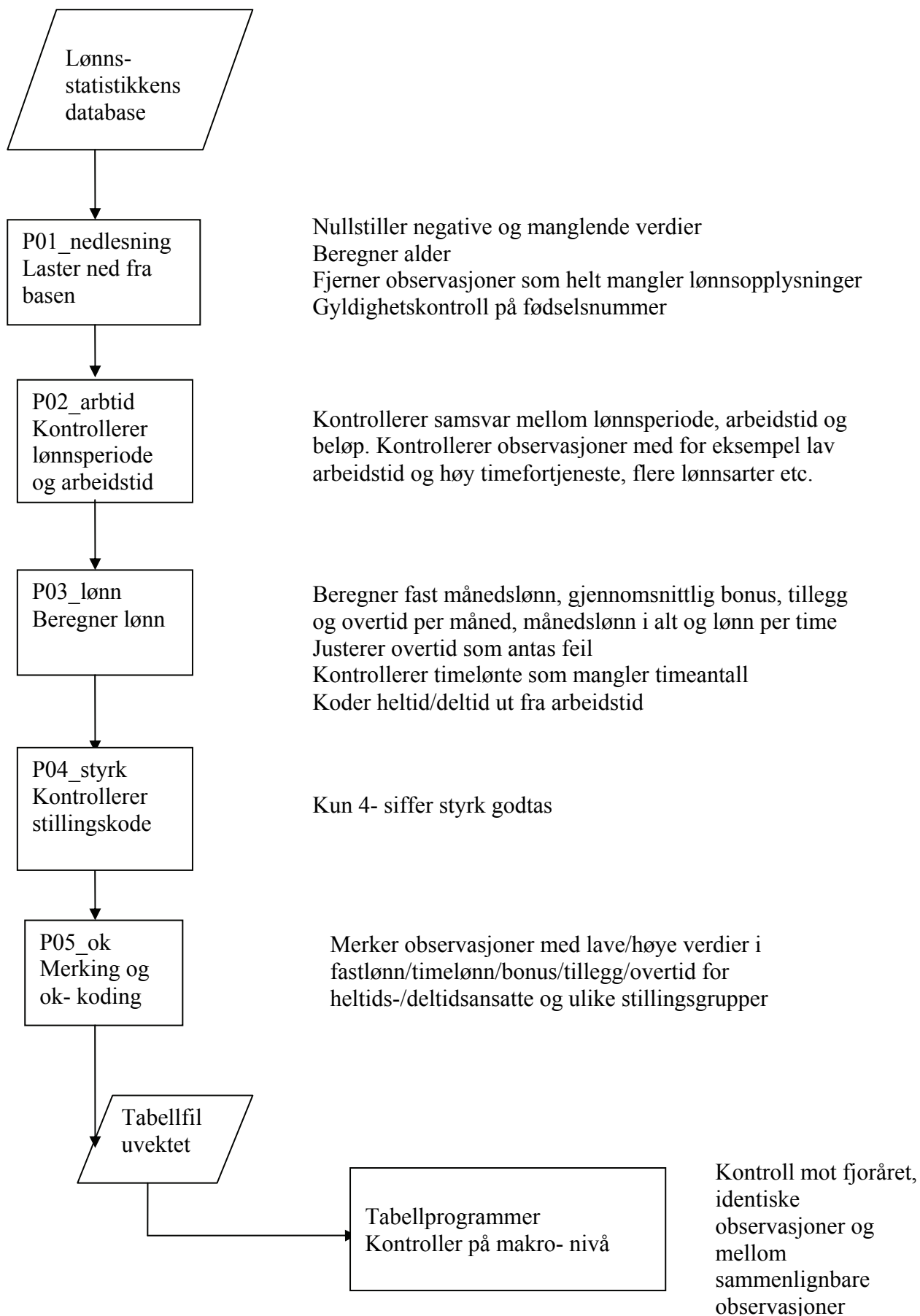
Reviderte tall	Månedslønn				
	Oppblåst antall	Gjennomsnitt	Standardfeil	Relativ standardfeil	Konfidensintervall
I alt	35 419	46 882	144	0,3 %	+/-282
Bergverk	3 696	30 848	219	0,7 %	+/-430
Olje	31 723	48 750	159	0,3 %	+/-312

Ureviderte tall	Månedslønn				
	Oppblåst antall	Gjennomsnitt	Standardfeil	Relativ standardfeil	Konfidensintervall
I alt	34 518	45 583	182	0,4 %	+/-359
Bergverk	3 871	27 540	278	1,0 %	+/-546
Olje	30 648	47 862	203	0,4 %	+/-398

Generelt i lønnsstatistikken ligger den relative standardfeilen eller variasjonskoeffisienten, som den også kalles, mellom 0,1 og 0,5 prosent. I tabellen ser vi at olje- og gassutvinning og bergverksdrift har en relativ standardfeil på rundt 0,3. Dette gir konfidensintervaller på vel 300 kroner i oljenæringa og drøyt 400 kroner i bergverk. Forskjellen i konfidensintervaller mellom revidert og urevidert materiale ligger på drøyt 100 kroner i bergverk og snaut 90 kroner i olje. Altså blir materialet mer komprimert, og vi kan med større sikkerhet anslå en gjennomsnittsverdi etter revisjon enn før.

Vedlegg

Vedlegg 1. Prosessdiagram lønnsstatistikk



Vedlegg 2. Spredningsmål i lønnsstatistikken

I tabell 16 er det blitt beregnet konfidensintervall, standardfeil og relativ standardfeil (CV) for heltidsansatte i olje- og gassutvinning og bergverksdrift i 2006.

Standardavviket benyttes generelt om stokastiske variable mens standardfeilen benyttes når vi snakker om estimater. Standardfeilen for utvalgsgjennomsnittet beregnes ved å ta rota av vekta i foretak a multiplisert med det kvadrerte avviket fra gjennomsnittet i foretak a dividert på rota av n antallobservasjoner minus en. Tar da her hensyn til at dette er et stratifisert tilfeldig klyngeutvalg. I vårt tilfelle er populasjonsgjennomsnittet et rent utvalgsgjennomsnitt. Standardavviket og standardfeilen er absolutte mål for spredning av verdier i et datasett eller i en stokastisk variabel. Den relative standardfeilen beregnes ved å dividere standardfeilen på gjennomsnittlig månedslønn. Vi kan derfor si at den relative standardfeilen er et uttrykk for hvor stor andel standardfeilen for utvalgsgjennomsnittet utgjør av gjennomsnittlig månedslønn.

Variasjonskoeffisienten kalles ofte for den relative standardfeilen og er definert som forholdet mellom standardfeilen til estimatoren og forventningsverdien (gjennomsnittet). I vårt tilfelle ønsker vi å beregne variasjonskoeffisienten til den estimerte gjennomsnittlige månedslønna.

Basert på den generelle definisjonen av gjennomsnittet har vi følgende formel:

$$\bar{X}_i = \frac{\sum_{i=1}^N X_i}{N}$$

der

\bar{X} - gjennomsnittlig månedslønn

i - individ (analyseenhet)

N - totalt antall sysselsatte i populasjonen

X_i - månedslønn for individ i , i populasjonen

Vi kan nå videre definere det estimerte gjennomsnittet for mål populasjonen ved å benytte foretakenes vekter på følgende måte:

$$\hat{\bar{X}} = \frac{\sum_{a=1}^k w_a x_a}{\sum_{a=1}^k w_a}$$

der

\hat{X} - er estimert gjennomsnittlig månedslønn for utvalget

a - foretak (utvalgsenhet)

x_a - er observert gjennomsnittlig månedslønn i foretak a i utvalget

w_a - er den beregnete vekta for foretak a

Standardfeilen (s) til det estimerte gjennomsnittet defineres på følgende måte:

$$s = \sqrt{\frac{\sum_{a=1}^k w_a (x_a - \bar{x})^2}{n-1}}$$

der

n - er antall klynger (som i vårt tilfelle er individer) i utvalget

$(x_a - \bar{x})^2$ - er det kvadrerte avviket fra gjennomsnittet i foretak a i utvalget

Variasjonskoeffisienten (CV) eller den relative standardfeilen som den også kalles kan da beregnes på følgende måte:

$$CV = \frac{s}{\hat{X}}$$

CV uttrykkes ofte som prosentandel

$$CV = \frac{s}{\hat{X}} 100\%$$

95 konfidensintervallet er i statistikken en måte å angi feilmarginen av en måling eller en beregning på. Således inneholder et 95 -konfidensintervall den sanne verdien med en sannsynlighet på 0,95. Ved normalfordelte variabler kan 95 -konfidensintervallene nokså nøyaktig regnes om fra variabelens gjennomsnitt (\hat{X}) og standardfeilen (s) som:

$$CI = [\hat{X} - 1,96 \cdot s; \hat{X} + 1,96 \cdot s]$$