

*Marit Melstrøm*

**Lønnsstatistikk for ansatte i varehandel 2009**

Klargjøring og analyse

© Statistisk sentralbyrå, november 2010 Ved bruk av materiale fra denne publikasjonen skal Statistisk sentralbyrå oppgis som kilde.	<b>Standardtegn i tabeller</b>	<b>Symbol</b>
ISBN 978-82-537-7969-0 Trykt versjon	Tall kan ikke forekomme	.
ISBN 978-82-537-7970-6 Elektronisk versjon	Oppgave mangler	..
ISSN 1891-5906	Oppgave mangler foreløpig	...
Emne: 06.90	Tall kan ikke offentliggjøres	:
Trykk: Statistisk sentralbyrå	Null	-
	Mindre enn 0,5 av den brukte enheten	0
	Mindre enn 0,05 av den brukte enheten	0,0
	Foreløpig tall	*
	Brudd i den loddrette serien	—
	Brudd i den vannrette serien	
	Desimaltegn	,

## Forord

Notatet dokumenterer og beskriver prosessen med å klargjøre, kvalitetssikre, analysere og utarbeide lønnsstatistikken for ansatte i varehandel for 2009. Statistikken gir en oversikt over lønnsnivået og endringer i lønn per år for heltidsansatte. Statistikken, i sin nåværende form ble produsert første gang i 1997, og er sammenlignbar tilbake til dette året. Det har imidlertid blitt publisert lønnsstatistikk for heltidsansatte i varehandel i alle år fra 1957, med unntak for årene 1963 og 1965.

## Sammendrag

I lønnsstatistikk for ansatte i varehandel benyttes et sett kontroll- og revisjonsrutiner som i utgangspunktet er standardiserte for alle lønnsstatistikkene.

Effekten av å kontrollere og revidere rådatamaterialet er i flere tilfeller betydelig. Den er imidlertid vanskelig å måle og å dokumentere fordi revisjon og oppretting foregår både i databasen og deretter på selve publiseringsfilen ved hjelp av SAS-rutiner. Vi har dermed ikke en tilgjengelig kopi av fullstendig ubehandlede rådata.

Tellingstidspunktet er 1. september, med unntak av NHOs medlemmer som telles per 1. oktober og følger oktobernæringenes utsendelse og purredatoer. Næringen består også av foretak og bedrifter som tilhører kvartalsvis lønnsstatistikk.

Når det gjelder datainngangen var den tilfredsstillende for denne næringen i 2009. Vi fikk inn 95,3 prosent av alle bedriftene i næringen. Apotekbransjen hadde svarprosent lik 99,0 mens den for HSH var 96,2 og for SAMFO 98,9. 96,4 prosent av dataene på nedlest fil ble godkjente, men dette er etter at betydelige korreksjoner er utført på Oracle-basen. Godkjente observasjoner betyr at dataene er funnet i orden ut ifra gitte kriterier i våre godkjenningsrutiner og at de dermed blir benyttet i statistikkproduksjonen.

Et viktig trekk som skaper utfordringer i produksjonen av denne statistikken er mangelfull/feilaktig rapportering av avtalt arbeidstid for timelønnte. Næringen har mange timelønnte arbeidstakere som tilkalles ved behov, og som ikke har noen avtalt arbeidstid. Det er ofte store avvik mellom innrapportert avtalt arbeidstid og faktisk arbeidet tid. Mange enheter er kontaktet for å forsøke å innhente riktig informasjon om avtalt arbeidstid. I og med at mange ansatte ikke har noen slik avtale har vi benyttet informasjonen om ”faktisk arbeidet tid” i statistikkproduksjonen.

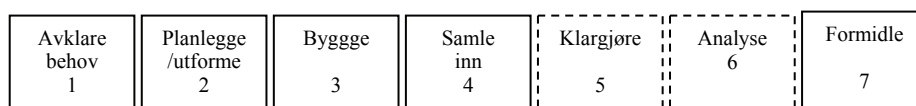
## Innhold

<b>Forord</b> .....	<b>3</b>
<b>Sammendrag</b> .....	<b>4</b>
<b>1. Innledning</b> .....	<b>6</b>
<b>2. Populasjon og utvalg</b> .....	<b>6</b>
<b>3. Datafangst</b> .....	<b>6</b>
3.1. Datainngang .....	6
3.2. Brutto inkommet .....	8
<b>4. Behandling av data fra Oracle-basen til endelig produksjonsfil</b> .....	<b>8</b>
4.1. Tilrettelegging av endelig produksjonsfil .....	9
<b>5. Effekten av kontroll- og revisjonsrutinene</b> .....	<b>17</b>
5.1. Arbeidstid.....	17
5.2. Lønnsvariablene .....	17
5.3. Oppsummering .....	20
<b>Litteraturliste:</b> .....	<b>21</b>
<b>Vedlegg A: Prosesdiagram lønnsstatistikk</b> .....	<b>22</b>
<b>Vedlegg B: Spredningsmål i lønnsstatistikken</b> .....	<b>23</b>
<b>Tabellregister</b> .....	<b>25</b>

## 1. Innledning

Notatet er i utgangspunktet todelt, der første del tar for seg datafangsten, mens del 2 tar for seg kontroll- og revisjonsopplegget i SAS og den effekten dette har. Datafangst og kontroll- og revisjonsrutiner henger nøye sammen. Kvaliteten på datafangsten virker inn på hvordan kontroll- og revisjon skal utføres.

Oppgavegiverne oppfordres til å levere data elektronisk via Altinn eller på CD. Det er også mulig å levere papirskjema for de som ikke kan sende inn elektronisk. Bearbeidingen av dataene kan deles inn i fire trinn: Mottak, registrering, innlesing til Oracle-tabell og etablering av produksjonsfil. Det foregår kontroll og revisjon i alle disse trinnene. Dette notatet beskriver populasjonen, utvalget, viktige hovedtrekk ved selve datainnsendingen og datainngangen. Det beskriver også rutiner og kontroller som gjøres i SAS etter at datagrunnlaget er lastet ned fra Oracle-databasen 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 informasjon om populasjon og utvalg 2009, se notat 2009/20). 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. Det er fulltelling for HSH (Handels- og Servicenæringens Hovedorganisasjon) og SAMFO (Samvirkeforetakenes Forhandlingsorganisasjon) sine medlemmer. SSB samler inn data og produserer forhandlingsstatistikk på oppdrag for de nevnte organisasjonene. Forhandlingstabellene for HSH-området gjelder engroshandel, detaljhandel og apotekbransjen og forhandlingstabellene for SAMFO-området gjelder detaljhandel. Fulltellingsområdene identifiseres ved hjelp av medlemsregistre fra de nevnte organisasjoner.

## 3. Datafangst

### 3.1. Datainngang

Status ved frigiving viser at vel 93 prosent av foretakene var kommet inn.

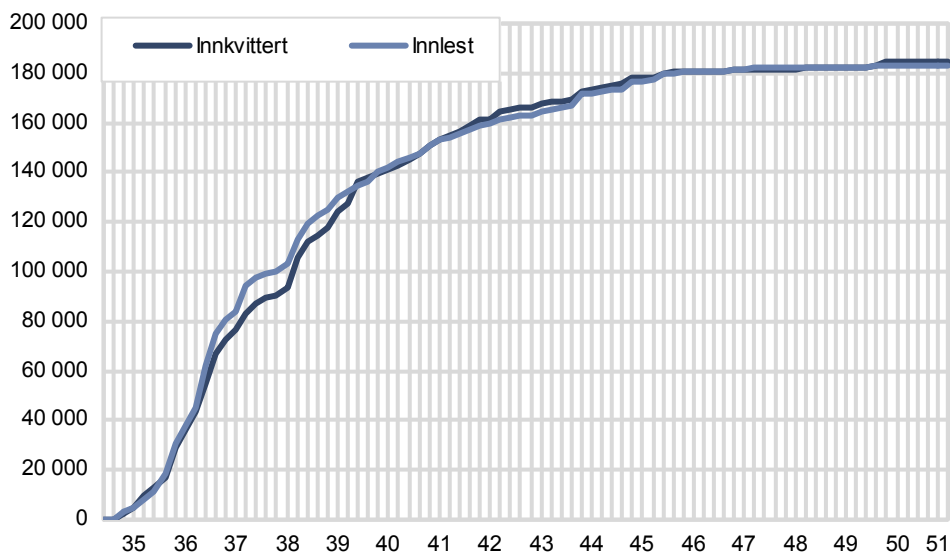
**Datafangsten per 1. september 2009, opptellinger fra vårt mottaksregister**

Dato	Hendelse	Antall foretak (trekkeenheter)
21.08.09	Utsending	4 172
08.09.09	Svarfrist	
22.09.09	Vedtak om tvangsmulkt	1 634
09.10.09	Siste frist for å unngå tvangsmulkt	
25.01.10	Fil til SI	168
Antall fritak: 5		

**Datafangsten per 1. oktober 2009**

18.09.09	Utsending	484
09.10.09	Svarfrist	
30.10.08	Vedtak om tvangsmulkt	195
07.11.08	Siste frist for å unngå tvangsmulkt	
12.03.10	Fil til SI	14
Antall fritak: 1		

Ved svarfrist var det kvittert inn om lag 47 prosent av foretakene, 50 prosent av bedriftene og 57 prosent av de ansatte i næringen. Innlesing til Oracle-basen har holdt et høyt og jevnt tempo gjennom hele høsten. Figur 1 viser innkvitterte og innleste observasjoner fra utsendingen i uke 33 til frigivingen i uke 51. Forholdet mellom innkvitterte data og data som er lest inn på tabell har blitt bedre de siste to årene. Dette har sammenheng med at stadig flere observasjoner kommer inn via Altinn og blir lest automatisk inn i databasen. Dette frigir også tid til å lese inn data som kommer som filvedlegg på e-post eller på CD. I perioder ser vi at det er flere observasjoner innlest enn innkvittert. Dette skyldes at vi får inn data fra bedrifter som vi ikke har med i utvalget.

**Figur 1. Datainnfang i varehandel. Innkvitterte og innleste observasjoner per uke. 2009**

### 3.2. Brutto inkommet

**Tabell 1. Innkvitteingsstatus i varehandel totalt, HSH, HSH Apotek og SAMFO i 2008 og 2009.**

	2009		2008	
	Antall innkvittert inkl. unntakskvittert	Prosent innkvittert	Prosent innkvittert	
Foretak totalt .....	4 295	93,4		95,2
Foretak i HSH .....	2 881	93,9		
Foretak i HSH Apotek .....	51	96,1		
Foretak i SAMFO .....	107	93,6		
Bedrifter totalt .....	13 166	95,9		95,8
Bedrifter i HSH .....	8 845	95,8		95,2
Bedrifter i HSH Apotek .....	410	99,0		98,8
Bedrifter i SAMFO .....	1 162	98,4		98,9
Ansatte (BoF) totalt .....	181 962	96,8		94,4
Ansatte i HSH .....	131 945	96,2		
Ansatte i HSH Apotek .....	4 100	99,3		
Ansatte i SAMFO .....	18 199	99,1		

Tabell 2 er en oversikt over andelen som benytter seg av de ulike innrapporteringsformatene i perioden 2004 til 2009. Hovedvekten ligger fortsatt på innrapportering via kravspesifikasjon (CD eller diskett), men rapportering via Altinn er økende. Bare 1,7 prosent kom inn via utfylling av papirskjema i 2009. Vi må fortsette å fronte Altinn som innleveringsmåte.

**Tabell 2. Rapporteringstype i prosent i perioden 2004-2009. Andel observasjoner**

Næring	Andel				
	A	I	K	M	O
	Andel i prosent				
I alt 2009 .....	3,7	40,7	53,3	0,7	1,7
I alt 2008 .....	3,4	38,1	54,9	1,0	2,6
I alt 2007 .....	8,7	29,2	56,7	1,7	3,8
I alt 2006 .....	7,9	4,1	64,7	2,6	20,8
I alt 2005 .....	9,1	1,8	60,8	3,5	24,8
I alt 2004 .....	10,8	1,0	56,5	4,5	27,3

O- Optisk lest skjema, M- Manuell registrering, K- Elektronisk rapportering, I- Altinn, A – Annet.

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

Å innhente lønnsinformasjon på individnivå fra næringslivet har vist seg å medføre at ikke alle opplysninger som kommer inn 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 til tabell 11), før det kontrolleres mer på makronivå etterpå. 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 næringshovedområde til næringshovedområde. Opplegget består av:

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

Det er viktig å merke seg at alle justeringer som gjøres på grunnlag av mikrokontroller er maskinelle. Dvs. det brukes totalt sett lite tid på mikrokontroller/mikrorevisjon.

I korte trekk er rutinene bygget opp som følger:

- Trinn 1: Nedlesing fra Oracle-basen og tilrettelegging av SAS-fil. 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 og merker/fjerner dubletter i databasen.
- Trinn 5b: Beregner gjennomsnittlig arbeidstid for heltidsansatte i næringshovedområdet og lager heltidsekvivalent variabler.
- Trinn 6: Materialet vektet og sjekkes opp mot populasjonen (Bedrifts- og foretaksregisteret).

Se vedlegg 1 for skjematisk oversikt over disse rutinene.

## 4.1. Tilrettelegging av endelig produksjonsfil

### Program 1 nedlesing fra Oracle-basen

Tabell 3 viser gyldige og ugyldige fødselsnummer i varehandel i 2008 og 2009. D-nummer er fødselsnummer til utenlandske arbeidere, og av totalen utgjør utenlandske arbeidere 1,1 prosent i 2009. Alle observasjoner tas med videre i statistikkgrunnlaget, uavhengig av om de har gyldig fødselsnummer eller ikke.

**Tabell 3. Gyldige og ugyldige fødselsnummer, 2008 og 2009**

Fødselsnummer	D-nummer	Antall	
		2008	2009
<b>I alt</b>		<b>177 863</b>	<b>186 702</b>
Gyldige fødselsnummer	og ikke d-nummer	176 159	184 496
Mangelfullt eller ugyldig fødselsnummer	og gyldig d-nummer	1 521	2 098
Ugyldige fødselsnummer	og ugyldig d-nummer	195	108

### Program 2 arbeidstidsforhold

Programmet kontrollerer og justerer lønnsperiode og avtalt 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 viser at over 97,4 prosent av materialet slipper igjennom den første kontrollen som gjøres, mens 2,6 prosent av observasjonene fikk et merke om at de er endret eller må sjekkes.

**Tabell 4. Status etter kontroller og endringer**

Status observasjoner	Frekvens	Prosent
<b>I alt</b> .....	<b>186 702</b>	<b>100,00</b>
Deltid iflg. timeantall/4.33, men arbeidstid mer enn 37,5, ikke rettet .....	93	0,05
Disse må sjekkes. Fastlønte med timelønn i tillegg .....	3 561	1,91
Disse må sjekkes. Høy timefortjeneste og lav arbeidstid .....	2	0,00
Disse må sjekkes. Timelønte med fastlønn i tillegg .....	80	0,04
Har høye bonuser og er provisjonslønte .....	4	0,00
Observasjoner uten lønnsperiode som gis lønnsperiode 7 .....	3	0,00
Mam Observasjoner med annen lønnsperiode r som lønnsperiode 6 .....	3	0,00
Månedslønte som mangler arbeidstid .....	146	0,08
Observasjoner med annen lønnsperiode r som gis lønnsperiode 6 .....	49	0,03
Observasjoner som er uendret. ....	181 931	97,44
Sjekkes. Heltid og fastlønn<10000 (kommafeil?) .....	592	0,32
Sjekkes. Timelønte med arbeidstid>45. ....	238	0,13

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

Lønnsperiode	Frekvens	
	Før	Etter
<b>I alt</b> .....	<b>186 702</b>	<b>186 702</b>
0.....	3	0
1.....	102 744	102 688
2.....	530	530
3.....	14	14
4.....	93	93
5.....	82	82
6.....	82 486	82 538
7.....	750	757

Lønnsperiode 0 er ugyldig variabelverdi. Observasjoner med utfylte lønnsvariable blir kodet om enten til lønnsperiode 1 (hvis fastlønn er større enn timelønn), eller til lønnsperiode 6. Svært få observasjoner endrer lønnsperiode i SAS-rutinen, de fleste korreksjonene utføres direkte i Oracle-basen etter hvert som dataene kommer inn.

*Avtalt arbeidstid* Etter hvert som vi har fått mer og mer data innrapportert på elektronisk form, har vi sett at det er store avvik mellom avtalt arbeidstid og faktisk betalte timer for en gruppe timelønte personer. Noen rapporterer 37,5 timer i feltet for avtalt arbeidstid, mens feltet for betalte timer gjerne viser et langt lavere timeantall. Vi har også noen eksempler på at rapportert avtalt arbeidstid per uke er lavere enn antall betalte timer, og at avtalt arbeidstid ikke er fylt ut i det hele tatt. Utgangspunktet for lønnsstatistikken er å stole på at innrapportert avtalt arbeidstid er riktig, da vi skal lage statistikk for avtalt arbeidstid og ikke for betalt arbeidstid.

Vi har derfor de siste årene hatt stor fokus på at avtalt arbeidstid må være riktig. Vi har kontaktet mange foretak og vi har fått hjelp fra HSH og SAMFO (de bekrefter at alle ansatte skal ha en avtalt arbeidstid i sine kontrakter). Arbeidstid fra AA-registeret (Arbeidstaker-/arbeidsgiverregisteret er lagt inn i database og benyttes til kontrollformål.

Kontakt med mange oppgavegivere viser at en del timelønte deltidsansatte og ekstrahjelper *ikke har en avtalt arbeidstid*, men blir tilkalt ved behov. Her rapporteres 0 eller blank i avtalt arbeidstid, mens noen f.eks. benytter 20 timer da de skal jobbe minst 20 timer osv.

Vi benytter to hovedprinsipper:

1. For alle observasjoner som ikke har rapportert avtalt arbeidstid benytter vi betalte timer siste lønnsutbetaling sammen med informasjon om heltid eller deltid for å beregne avtalt arbeidstid (se tabell 6, 01-04).

2. For alle observasjoner med avtalt arbeidstid  $\geq 33,0$  hvor heltid/deltid sier deltid og betalte timer sier deltid (se tabell 6, 05), erstatter vi avtalt arbeidstid med informasjon fra betalte timer.

**Tabell 6. Observasjoner som får beregnet avtalt arbeidstid ut ifra betalte timer.**

Tidstest	Antall	Prosent
<b>I alt</b> .....	<b>8 152</b>	<b>100,0</b>
01 Deltidsansatte timelønnte, arbeidstid = 0, timelønn > 0 og antall betalte timer < 143 .....	6 211	76,2
02 Deltidsansatte timelønnte, arbeidstid = 0, timelønn > 0 og antall betalte timer > 143 .....	571	7,0
03 Heltidsansatte timelønnte, arbeidstid = 0, timelønn > 0 og antall betalte timer < 143 .....	36	0,4
04 Deltidsansatte timelønnte, arbeidstid = 0, timelønn > 0 og antall betalte timer > 143 .....	41	0,5
05 Deltidsansatte timelønnte, arbeidstid $\geq 33,0$ , betalte timer > 0 og (betalte timer / 4,33) < 33 .....	1 293	15,8

Tabell 7 viser arbeidstidsfordelingen for alle observasjoner uavhengig av om dataene er godkjent til statistikkproduksjon eller ikke. Gjennomsnittlig arbeidstid for heltidsansatte er 37,5 timer i 2009, 37,4 timer både i 2006, 2007 og 2008 og 37,3 i 2005.

**Tabell 7. Antall observasjoner fordelt på arbeidstidsgrupper og gjennomsnittlig arbeidstid for heltidsansatte i 2005, 2006, 2007, 2008 og 2009**

Timer per uke	Antall				
	2009	2008	2007	2006	2005
<b>I alt</b> .....	<b>186 695</b>	<b>177 671</b>	<b>163 042</b>	<b>156 677</b>	<b>145 609</b>
- 4.0 .....	8 650	8 292	6 575	6 868	4 344
4.1-10.0 .....	23 614	23 412	20 762	18 456	12 754
10.1-15.0 .....	14 286	14 573	12 536	10 969	10 474
15.1-20.0 .....	12 405	11 312	10 691	10 688	11 303
20.1-25.0 .....	8 935	8 173	8 304	8 305	9 017
25.1-32.9 .....	16 684	15 428	14 342	13 348	14 259
33.0 - høyere .....	102 121	96 492	89 832	88 043	83 458
Gjennomsnittlig arbeidstid for heltidsansatte <sup>1</sup> .....	37,5	37,4	37,4	37,4	37,3

<sup>1</sup>Tallene er rettet 18.11.2010.

### Program 3 beregning av månedslønn

Dette programmet foretar først og fremst en beregning av avtalt lønn, bonus og uregelmessige tillegg. Det foretas også korreksjoner 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 fastlønnte, timelønnte og provisjonslønnte. Dette for å få sammenlignbare størrelser uavhengig av lønnsperiode. Innrapportert lønn siste avlønningsperiode knyttet til variabel for avlønning benyttes i utregningen av avtalt lønn:

Avlønnsperiode 1	=	Oppgitt fastlønn siste periode (månedslønn)
Avlønnsperiode 2	=	Oppgitt fastlønn per 14-dager multiplisert med 2.165 (antall 14-dagers-perioder i en måned).
Avlønnsperiode 3	=	Oppgitt fastlønn per uke multiplisert med 4.33 (antall uker i en måned).
Avlønnsperiode 4	=	Oppgitt fastlønn per uke multiplisert med (12,5/12) .
Avlønnsperiode 5	=	Oppgitt fastlønn per uke multiplisert med (13/12) .

Avlønnsperiode 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).
Avlønnsperiode 7 =	Hovedregel: (bonus hittil i år + bonus 4. kvartal i fjor)/12 + evt. oppgitt fastlønn siste periode. Bonusfeltene tømmes etter beregning av avtalt lønn.
<i>Uregelmessige tillegg</i>	Innhentes som aggregert størrelse fra 1. januar til tellingstidspunktet. Gjennomsnittsberegning per måned for denne perioden.
<i>Bonus</i>	Innhentes både for 4. kvartal i fjor, samt som aggregert størrelse fra 1. januar til tellingstidspunktet. Gjennomsnittsberegning per måned for hele denne 12-månedperioden.
<i>Månedslønn</i>	Månedslønna framkommer ved en summering av de beregnede månedsverdiene for avtalt lønn, uregelmessige tillegg og bonus.
<i>Overtidsgodtgjørelse</i>	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 8. Oversikt over overtidstimer og overtidbetaling**

	Frekvens	Prosent
<b>I alt</b> .....	<b>186 702</b>	<b>100,0</b>
Observasjoner uten overtidstimer eller betaling .....	106 774	57,2
Observasjoner med flere overtidstimer enn betaling .....	190	0,1
Observasjoner med verdier på betaling .....	77 636	41,6
av disse mangler det timer på .....	2 102	1,1

På basis av denne kontrollen blir overtidsgodtgjørelsen justert maksimelt 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.

#### Program 4 yrkeskoder

I program 4 gjennomføres en logisk kontroll av gyldighet av innrapporterte yrkeskoder. Fram til tellingen i 2008 ble det innhentet ”stillingskoder” knyttet til forhandlingssystemet i denne næringen, men fra 2008 ble det rapportert yrkeskoder i henhold til Standard for yrkesklassifisering, C521 (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.

**Tabell 9. Fordeling av observasjoner på 1. siffer STYRK<sup>1</sup> i 2008 og 2009**

	2008		2009	
	Frekvens	Prosent	Frekvens	Prosent
<b>I alt</b> .....	<b>177 671</b>	<b>100,0</b>	<b>186 695</b>	<b>100,0</b>
1 .....	18 666	10,5	20 502	11,0
2 .....	4 952	2,8	5 060	2,7
3 .....	15 658	8,8	16 765	9,0
4 .....	21 054	11,9	20 950	11,2
5 .....	104 820	59,0	109 425	58,6
6 .....	120	0,0	128	0,1
7 .....	8 089	4,6	9 671	5,2
8 .....	1 966	1,1	2 158	1,2
9 .....	2 216	1,3	1 980	1,1
X .....	130	0,0	56	0,0

<sup>1</sup> 1-siffer yrkeskoder i følge C521:

1: Administrative ledere og politikere

2: Akademiske yrker

3 Yrke med kortere høgskole- og universitetsutdanning og teknikere

4: Kontor- og kundeserviceyrker

5: Salgs-, service-, og omsorgsyrker

6: Yrke innen jordbruk, skogbruk og fiske

7: Håndverkere og lignende

8: Prosess- og maskinoperatører, transportarbeidere mv.

9: Yrker uten krav til utdanning

Verdien X er samle verdien for ugyldige innrapporterte yrkeskoder

Som en kontroll av innrapporterte yrkeskoder til lønnsstatistikken kobles det på 7-sifret yrkeskode fra AA-registeret med oppdatering nærmest mulig tellingstidspunktet.

**Tabell 10. Fordeling av antall med 1-siffer STYRK i lønnsstatistikken sammenlignet med AA-registeret**

I alt, STYRK i Lønnsstatistikken	I alt, AA-STYRK	I alt, AA-STYRK										
		0	1	2	3	4	5	6	7	8	9	
<b>I alt, STYRK i Lønnsstatistikken</b> ....	<b>186 702</b>	<b>595</b>	<b>18 198</b>	<b>4 963</b>	<b>16 110</b>	<b>19 386</b>	<b>91 974</b>	<b>1 738</b>	<b>9 068</b>	<b>2 329</b>	<b>23 906</b>	
1.....	20 503	52	15 996	166	662	339	2 298	13	122	34	821	
2.....	5 060	38	134	4 072	306	100	80	1	22	9	298	
3.....	16 766	53	6 269	417	12 525	591	1 551	9	173	46	775	
4.....	20 950	43	331	118	655	16 498	1 205	5	212	230	1 653	
5.....	109 430	365	1 025	153	1 616	1 483	86 390	63	411	304	17 620	
6.....	128	-	2	-	-	-	22	71	-	-	33	
7.....	9 671	25	71	9	268	167	163	7	8 039	96	826	
8.....	2 158	4	5	3	54	172	99	3	45	1 578	195	
9.....	1 980	14	6	2	22	35	165	1	39	29	1 667	
X .....	56	1	2	23	2	1	1	-	5	3	18	

Sammenligner vi de ulike nivåer i AA-registeret med lønnsstatistikken ser vi at det på nivå 1 i AA-registeret er en del som havner på nivå 5 i lønnsstatistikken. På nivå 3, nivå 4 og spesielt på nivå 9 i AA-registeret er det også mange observasjoner som havner på nivå 5 i lønnsstatistikken. I AA-registeret er det 595 observasjoner som det ikke er kodet yrke på, mens det for lønnsstatistikken er 56 observasjoner.

En direkte sammenligning av godkjente observasjoner på lønnsfila viser at for 132 745 observasjoner (74 prosenten) er 4-siffer AA-STYRK lik 4-siffer STYRK i lønnsstatistikken. Etter at vi gikk over til STYRK-koder for denne bransjen fra 2007 til 2008 var det samme forholdet 68 prosent i 2008.

### Program 5 godkjenning av observasjoner

I program 5 lages en oversikt over godkjente observasjoner og observasjoner som forkastes i statistikken. Lønnsopplysninger for både heltids- og deltidsansatte har maksimums- og minimumsgrenser. For timelønte og provisjonslønte ble maksimumsgrensen for månedslønna hevet til 65 000 kroner og månedslønna for ansatte med 1-siffer STYRK = 4 eller 5 hevet til 70 000 kroner.

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 (for eksempel 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 så stor grad som mulig.

**Tabell 11. Antall observasjoner fordelt etter koder for grenseverdier (ok-koder) i 2009**

Ok-kode	Antall observasjoner	Prosentvis andel
<b>I alt</b>	<b>186 702</b>	<b>100,00</b>
Ok= 0 - Godkjente observasjoner	179 930	96,37
Ok= 1 - Lave verdier på lønnsvariablene	27	0,01
Ok= 3 - Timelønnte med utilstrekkelig info om timer	129	0,07
Ok= 4 - Alder=missing	10	0,01
Ok=10 - Heltidsansatte med ialt < 10 500 kroner	838	0,45
Ok=11 - Heltidsansatte, fast månedslønn (fmlonn) over 200 000 kroner	6	0,00
Ok=12 - Heltidsansatte, timelønnte med fmlonn over 65 000 kroner	23	0,01
Ok=13 - Heltidsansatte, provisjonslønnte med fmlonn over 65 000 kroner	8	0,00
Ok=14 - Heltidsansatte, st1=1 og i alt> 250 000 kroner	20	0,01
Ok=15 - Heltidsansatte, st1=1 og i alt< 17 000 kroner	127	0,07
Ok=16 - Heltidsansatte, st1 = 2,3 og i alt over 100 000 kroner	78	0,04
Ok=17 - Heltidsansatte, st1 = 4,5 og i alt over 70 000 kroner	125	0,07
Ok=18 - Heltidsansatte, st1 > 5 og i alt over 80 000 kroner	12	0,01
Ok=19 - Heltidsansatte, Overtid (ot2) over 50 000 kroner	10	0,01
Ok=20 - Heltidsansatte, tillegg (tl) over 65 000 kroner	1	0,00
Ok=21 - Heltidsansatte, bonus (bon) over 85 000 kroner	21	0,01
Ok=30 - Deltidsansatte med arbtid<=1	1195	0,64
Ok=31 - Deltidsansatte, st1=1 og timefortjeneste> 750	10	0,01
Ok=32 - Deltidsansatte, st1=1 og timefortjeneste< 115	19	0,01
Ok=33 - Deltidsansatte, st1=1 og ialt/(arbtid/37.5) >= 17000	6	0,00
Ok=34 - Deltidsansatte, st1=1 og ialt/(arbtid/37.5) <= 150000	22	0,01
Ok=35 - Deltidsansatte, st1=2,3 og timefortjeneste> 500	31	0,02
Ok=36 - Deltidsansatte, st1=2,3 og timefortjeneste< 80	23	0,01
Ok=37 - Deltidsansatte, st1=4,5 og timefortjeneste> 300	655	0,35
Ok=38 - Deltidsansatte, st1=4,5 og timefortjeneste< 60	292	0,16
Ok=39 - Deltidsansatte, st1=7,8 og timefortjeneste> 250	16	0,01
Ok=40 - Deltidsansatte, st1=7,8 og timefortjeneste< 60	10	0,01
Ok=41 - Deltidsansatte, timefortjeneste> 250	30	0,02
Ok=42 - Deltidsansatte, timefortjeneste< 50	11	0,01
Ok=43 - Deltidsansatte, ot>50 000 kroner	1	0,00
Ok=45 - Deltidsansatte, bon> 85000 kroner	1	0,00
Ok=90 - Heltidsdubletter	1194	0,64
Ok=91 - Deltidsdubletter	1646	0,88
Ok=92 - Heltid-deltidsdubletter	175	0,09

Den vanligste ”feiltypen” er observasjoner som er holdt utenfor, enten på grunn av dårlig innrapporterte data eller fordi de har ekstremverdier i noen lønnsarter, for eksempel ok=10 med 838 observasjoner og ok=30 med 1 195 observasjoner. Ser vi på disse observasjonene er det veldig sannsynlig at det er avtalt arbeidstid som er rapportert feil. En sjekk av observasjonene som faller ut på mange av de andre kodene gir også en pekepinn på at det er avtalt arbeidstid som er det store problemet i rapporteringen. Ingen hele foretak er holdt utenfor i denne næringen.

Sjekk av observasjoner med ulike ok-koder kan bidra til at vi finner systematiske feil i dataene som for eksempel skyldes programfeil hos leverandører. Vi kan gjøre rettelser i dataene eller eventuelt gjøre maskinelle opprettinger i programmene for å korrigere feilene.

### Program 5b heltidsekvivalenter

Inkludering av heltidsekvivalenter i de næringsspesifikke frigivingene ble første gang gjort i 2008. Program 5b regner ut ekvivalentvekter og ekvivalentlønnsarter for disse.

I denne næringen er det mange deltidsansatte (se tabell 7). Oppgavegivere som leverer data elektronisk ved å sende inn fil ifølge kravspesifikasjonen for innrapportering av lønnsstatistikk eller via Altinn rapporterer en kode for heltid eller deltid. Vi ser denne koden i sammenheng med rapportert avtalt arbeidstid og lønnsperiode, se pkt. 4.1.2. Vi kontrollerer også alle lønnsarter for både heltids- og deltidsansatte i forhold til maksimums- og minimumsgrenser, se pkt 4.1.5.

Grenseverdiene sees også i sammenheng med hvilket yrke den ansatte har. Observasjoner som faller ut blir kontrollert for om mulig å kunne rettes opp eller bli bekreftet riktig slik at de kan godkjennes og benyttes i statistikken.

Følgende oversikt viser resultatet etter kontrollene vi gjør, altså hvor mange heltids- og deltidsansatte som ikke blir godkjent.

	Heltidsansatte	Deltidsansatte
<b>Totale obs i næringen .....</b>	<b>102 121</b>	<b>84 574</b>
Godkjente obs .....	99 607	80 316
Ikke godkjente obs .....	2 514	4 258
% ikke godkjente obs .....	2,5	5,0

Vi får en unik forståelse av heltidsekvivalenten for næringen i og med at vi beregner gjennomsnittlig antall avtalte timer per uke for heltidsansatte. Dette gjennomsnittet blir utgangspunktet for videre behandling og omregning av deltidsansattes lønn.

Det er kun de deltidsansattes månedslønn vi regner om. Utgangspunktet for omregning er å multiplisere lønna til den enkelte deltidsansatte til hva den ville vært dersom han eller hun jobbet heltid.

Beregningen vises med et eksempel. En butikkekspeditor jobber deltid med en avtalt arbeidstid per uke på 20 timer og har 10 000 kroner i månedslønn. Gjennomsnittlig arbeidstid for heltidsansatte i varehandel er 37 timer per uke. 10 000 kroner deles på 20 timer og multipliseres med 37 timer (gjennomsnittlig antall timer for heltidsansatte i næringen). Vedkommende får 18 500 kroner i lønn dersom han hadde jobbet heltid.

### Program 6 vekting

Programmet stratifiserer populasjonen etter antall ansatte i foretakene og beregner vekter.

Vektemetoden er basert på justert etterstratifisert invers trekkesannsynlighet, beskrevet i notat av Grini (2003/74). I varehandel trekkes det et utvalg med 3 ulike strata for hver av undernæringene; 45 motorkjøretøytjenester, 46 engroshandel og 47 detaljhandel. I tillegg har vi fulltelling for HSHs og SAMFOs medlemmer og lager ett strata for hver av disse (0701 og 0301).

I SSBs statistikk vekter vi for frafall for HSH og SAMFOs medlemmer. I forhandlings-statistikken vi produserer på oppdrag for disse organisasjonene benytter vi imidlertid ikke vekting.



**Tabell 12. Antall foretak og personer i populasjon og antall foretak i utvalg og justert etterstratifisert invers trekkesannsynlighet (vekt). Per stata.**

Strata	Antall 2009			Vekt 2009	Vekt 2008
	Foretak i utvalget	Foretak i pop.	Ansatte i pop.		
0301 .....	104	110	17 797	1,0577	1,1900
0701 .....	2 730	2 983	113 464	1,0927	1,0809
4501 .....	78	82	11 325	1,0513	1,0455
4502 .....	187	374	10 979	2,0000	1,7083
4503 .....	137	1 394	12 064	10,1752	7,4971
4601 .....	18	19	8 027	1,0556	0,9500
4602 .....	54	86	10 667	1,5926	1,6667
4603 .....	362	3 501	46 660	9,6713	9,3133
4701 .....	9	10	11 771	1,1111	1,0000
4702 .....	21	47	7 034	2,2381	1,9286
4703 .....	882	5 940	64 580	6,7347	10,1591

**Program 7 justering av vekter**

Prinsippet for justering av vektene er at sum oppblåst antall i lønnsstatistikken skal reflektere sum antall ansatte fra Bedrifts- og foretaksregisteret (BoF) for hver næring (Grini, notater 2003/74). 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 BoF. 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 13. Vektjustering 2009. Heltids- og deltidsansatte**

Næring	Vektet antall		Vektet andel		BoF andel	Justeringsfaktor mot BoF
	2008	2009	2008	2009		
<b>Totalt</b> .....	<b>328 607</b>	<b>317 109</b>	<b>1</b>	<b>1</b>	<b>1</b>	
Motorkjøretøytenester....	54 742	33 633	0,1666	0,1062	0,1133	1,0677
Engroshandel.....	102 126	98 841	0,3108	0,3117	0,3024	0,9703
Detailhandel.....	171 738	184 605	0,5226	0,5821	0,5842	1,0035

I 2009 ble vektjusteringen gjort direkte mot populasjonen i BoF.

Effekten av vekteprosedyrene på det beregnede lønnsnivået vises i tabell 14. Vi ser at vektjusteringen som gjøres, medfører at det samlede lønnsnivået justeres opp med i overkant av 500 kroner.

**Tabell 14. Effekten av vektning 2009. Gjennomsnittlig månedslønn for heltidsansatte. Kroner**

	Månedslønn 2009. Heltidsansatte. Kroner		
	Uvektet	Ujustert vekt	Justert vekt
<b>Totalt</b> .....	<b>33 401</b>	<b>34 017</b>	<b>33 931</b>
Motorkjøretøytenester .....	33 130	32 426	32 426
Engroshandel .....	37 419	38 317	38 317
Detailhandel .....	29 535	28 888	28 888

*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. 303 foretak og 603 bedrifter hadde ikke levert oppgave til lønnsstatistikken ved siste nedlesing før publisering. Innlevering av data umiddelbart etter produksjon av publiseringsfil og ”vasking”



mot BoF medførte at det ved oversendelse av fil til SI var 182 foretak som ble ilagt tvangsmulkt for manglende levering av lønnsstatistikk, eller for å ha levert mangelfulle data.

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. Om lag 3,6 prosent av alle innkomne observasjoner blir forkastet. 1 195 observasjoner faller bl.a. ut pga. avtalt arbeidstid  $\leq 1$ . Ellers er det også noe frafall for deltidsansatte i ulike yrkesgrupper pga. høy eller lav timefortjeneste. Når vi sjekker dataene som ikke blir godkjent (har ok-koder fra 1-92), får vi en pekepinn på at det er rapportering av avtalt arbeidstid per uke som er hovedproblemet i denne næringen.

## 5. Effekten av kontroll- og revisjonsrutinene

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

### 5.1. Arbeidstid

Avtalt arbeidstid er grunnleggende bestemmende for to forhold: a) om en person er heltids- 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 å sjekke sammenheng mellom arbeidstid og lønn. I denne næringen beregnes i en del tilfeller avtalt arbeidstid ut ifra arbeidet tid, se tabell 6.

**Tabell 15. Effekt av kontroll og revisjon av avtalt arbeidstid per uke**

	Antall observasjoner		Gjennomsnittlig avtalt arbeidstid	
	Før revisjon	Etter revisjon	Før revisjon	Etter revisjon
<b>I alt .....</b>	<b>186 695</b>	<b>179 923</b>	<b>28,01</b>	<b>28,10</b>
Heltidsansatte .....	102 121	99 607	37,52	37,45
Deltidsansatte .....	84 574	80 316	14,94	14,96

Antall observasjoner før revisjon er større enn antall observasjoner etter revisjon. Dette skyldes at ikke-godkjente observasjoner i følge kriterier som vist i kapittel 4.1.5, er holdt utenfor i tallene etter revisjon. Vi ser også at fordelingen mellom heltid og deltid forskyves underveis i revisjonen.

### 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. Resultatene blir noe vanskeligere å tolke når lønna for de deltids-

ansatte 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 16. Effekt av kontroll og revisjon av lønnsvariable. Heltidsansatte. Kroner**

	Revidert	Urevidert	Avvik			
Antall .....	99 607	102 776	-3 169			
	Gjennomsnitt			Median		
	Revidert	Urevidert	Avvik	Revidert	Urevidert	Avvik
Månedslønn .....	33 931	33 487	444	30 050	30 000	50
Avtalt lønn .....	32 057	31 514	543	29 000	28750	250
Uregelmessige tillegg .....	395	408	-13	0	0	0
Bonus .....	1 478	1 569	-91	0	0	0
Overtid .....	615	579	36	0	0	0
	1. kvartil			3. kvartil		
	Revidert	Urevidert	Avvik	Revidert	Urevidert	Avvik
Månedslønn .....	25 470	25 034	436	37 632	37 500	132
Avtalt lønn .....	24 684	24 339	345	35 811	35 639	172
Uregelmessige tillegg .....	0	0	0	3	20	-17
Bonus .....	0	0	0	704	667	37
Overtid .....	0	0	0	321	273	48
	Min			Max		
	Revidert	Urevidert	Avvik	Revidert	Urevidert	Avvik
Månedslønn .....	10 515	0	10 515	627 748	1 160830	-533 082
Avtalt lønn .....	863	0	863	266 667	670 800	-403 333
Uregelmessige tillegg .....	0	0	0	58 905	666 225	-607 320
Bonus .....	0	0	0	546 748	546 748	0
Overtid .....	0	0	0	47 588	80 527	-32 939

Tabellen viser tall for næringshovedområdet i helhet og vi ser at revidert månedslønn er høyere enn urevidert månedslønn. For å få et bedre bilde av effekten av kontroll- og revisjonsopplegget vil vi se på forskjellen mellom revidert og urevidert verdi for lønnsvariablene fordelt på næring og yrke. For enkelthets skyld tar vi i tabell 17 bare for oss månedslønn etter næring og yrke.

**Tabell 17. Effekt av kontroll og revisjon av månedslønn etter næring og yrke. Heltidsansatte. Antall personer og kroner.**

	Antall uvektede observasjoner			Gjennomsnittlig månedslønn		
	Urevidert	Revidert	Avvik	Urevidert	Revidert	Avvik
<b>Varehandel totalt</b>						
<b>Totalt</b> .....	<b>102 776</b>	<b>99 607</b>	<b>3 169</b>	<b>33 487</b>	<b>33 931</b>	<b>444</b>
Lederyrker .....	19 243	18 904	339	46 438	45 927	-511
Akademiske yrker .....	4 507	4 490	17	48 705	46 990	-1 715
Høgskoleyrker .....	15 199	14 935	264	39 789	39 150	-639
Kontoryrker .....	17 782	17 619	163	29 631	29 654	23
Salgs-/serviceyrker .....	33 392	31 910	1 482	25 192	26 734	1 542
Håndverkere .....	9 678	8 939	739	28 718	29 377	659
Operatører og sjåførere .....	1 973	1 943	30	28 217	28 269	52
Yrker uten krav til utdanning .....	870	774	96	23 817	25 172	1 355
Andre yrker .....	132	93	39	34 493	28 387	-6 106
<b>Motorkjøretøytjenester</b>						
<b>Totalt</b> .....	<b>17 050</b>	<b>15 974</b>	<b>1 076</b>	<b>31 741</b>	<b>32 426</b>	<b>685</b>
Lederyrker .....	1 782	1 740	42	47 489	47 592	103
Akademiske yrker .....	187	185	2	41 712	41 644	-68
Høgskoleyrker .....	1 227	1 174	53	36 046	36 404	358
Kontoryrker .....	2 675	2 587	88	30 514	30 783	269
Salgs-/serviceyrker .....	2 456	2 336	120	30 588	31 594	1 006
Håndverkere .....	8 027	7 324	703	28 614	29 336	722
Operatører og sjåførere .....	158	155	3	26 291	26 581	290
Yrker uten krav til utdanning .....	521	466	55	25 223	26 202	979
Andre yrker .....	17	7	10	26 034	23 818	-2 216
<b>Engroshandel</b>						
<b>Totalt</b> .....	<b>42 486</b>	<b>41 556</b>	<b>930</b>	<b>38 843</b>	<b>38 317</b>	<b>-526</b>
Lederyrker .....	6 497	6 352	145	56 201	55 254	-947
Akademiske yrker .....	3 027	2 999	28	49 085	46 849	-2 236
Høgskoleyrker .....	11 649	11 332	317	40 756	40 013	-743
Kontoryrker .....	12 484	12 355	129	29 802	29 743	-59
Salgs-/serviceyrker .....	5 560	5 339	221	31 834	31 834	0
Håndverkere .....	1 371	1 346	25	29 371	29 876	505
Operatører og sjåførere .....	1 646	1 618	28	28 527	28 540	13
Yrker uten krav til utdanning .....	207	189	18	24 755	26 047	1 292
Andre yrker .....	45	26	19	40 709	30 158	-10 551
<b>Detaljhandel</b>						
<b>Totalt</b> .....	<b>43 240</b>	<b>42 077</b>	<b>1 163</b>	<b>27 572</b>	<b>28 888</b>	<b>1 316</b>
Lederyrker .....	10 964	10 812	152	37 113	36 762	-351
Akademiske yrker .....	1 293	1 306	-13	48 619	48 318	-301
Høgskoleyrker .....	2 323	2 429	-106	34 603	34 261	-342
Kontoryrker .....	2 623	2 677	-54	27 436	27 666	230
Salgs-/serviceyrker .....	25 376	24 235	1 141	22 754	24 723	1 969
Håndverkere .....	280	269	11	27 288	27 280	-8
Operatører og sjåførere .....	169	170	-1	26 875	27 136	261
Yrker uten krav til utdanning .....	142	119	23	19 475	21 626	2 151
Andre yrker .....	70	60	10	22 829	26 601	3 772

88 457 heltidsobservasjoner får ikke endret månedslønn i revisjonsprosessen. Det vil si at 11,2 % av heltidsobservasjonene på produksjonsfila er justert.

**Justeringene fordeler seg per næring på følgende måte:**

	Antall ikke endrede obs	Antall obs på produksjonsfila	Prosent endrede obs
<b>Totalt</b> .....	<b>88 457</b>	<b>99 607</b>	<b>11,2</b>
Motorkjøretøytjenester .....	14 787	15 974	7,4
Engroshandel .....	39 646	41 556	4,6
Detaljhandel .....	34 024	42 077	19,1

**Justeringene fordeler seg per yrke på følgende måte:**

	Antall ikke endrede obs	Antall obs på produksjonsfila	Prosent endrede obs
<b>Totalt</b> .....	<b>88 457</b>	<b>99 607</b>	<b>11,2</b>
Lederyrker .....	18 147	18 904	4,0
Akademiske yrker .....	4 312	4 490	4,0
Høgskoleyrker .....	14 123	14 935	5,04
Kontoryrker .....	16 826	17 619	4,05
Salgs-/serviceyrker .....	24 321	31 910	23,8
Håndverkere .....	8 132	8 939	9,0
Operatører og sjåførere .....	1 888	1 943	2,8
Yrker uten krav til utdanning .....	650	774	16,0
Andre yrker .....	58	93	37,6

I tabell 18 er det beregnet ulike spredningsmål for heltidsansatte i varehandel etter næring på ureviderte og reviderte tall for 2009. Som en følge av revisjonsrutinene minskes den samlede spredningen.

**Tabell 18. Effekt av kontroll og revisjon av månedslønn. Spredningsmål per næringsområde. Antall og kroner.**

Reviderte tall .....	Ant obs	Månedslønn			
	Uvektet antall	Gjennomsnitt	Standardfeil	Relativ standardfeil	Konfidensintervall
<b>Totalt</b> .....	<b>99 607</b>	<b>33 931</b>	<b>74</b>	<b>0,22</b>	<b>+/- 145</b>
Motokjøretøytjenester .....	15 974	32 426	119	0,37	+/- 233
Engroshandel .....	41 556	38 317	130	0,34	+/- 255
Detaljhandel .....	42 077	28 888	73	0,25	+/- 143

Ureviderte tall	Ant obs	Månedslønn			
	Oppblåst antall	Gjennomsnitt	Standardfeil	Relativ standardfeil	Konfidensintervall
<b>Totalt</b> .....	<b>102 776</b>	<b>33 487</b>	<b>93</b>	<b>0,28</b>	<b>+/- 182</b>
Motokjøretøytjenester .....	17 050	31 741	153	0,48	+/- 299
Engroshandel .....	42 486	38 843	165	0,42	+/- 323
Detaljhandel .....	43 240	27 572	97	0,35	+/- 190

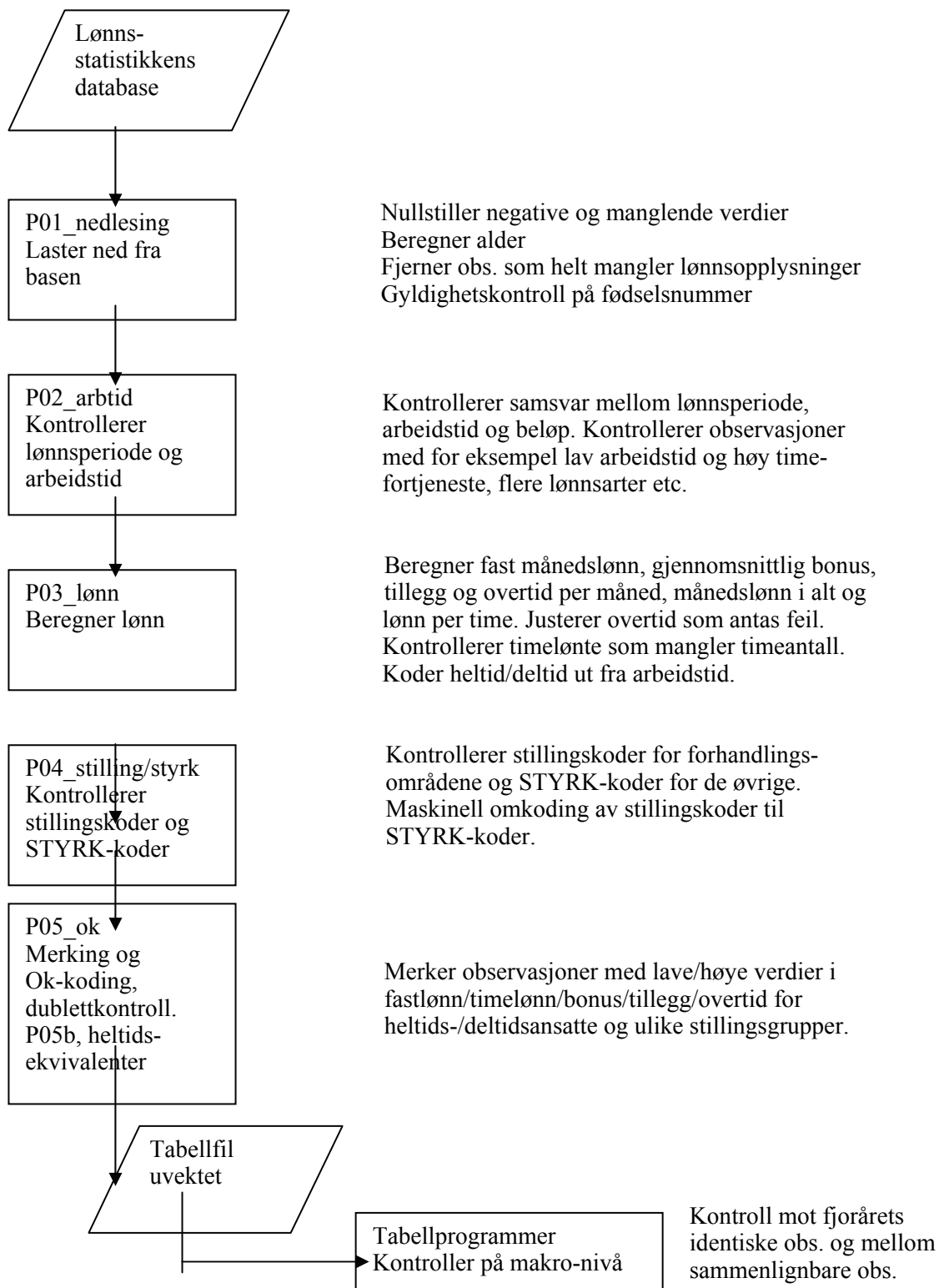
### 5.3. Oppsummering

Som nevnt i begynnelsen av dette notatet er det vanskelig å måle effekten av kontroll og revisjon av dataene. Det som er dokumentert i dette notatet er effekten av maskinelle kontroller vi gjør i SAS på et datagrunnlag hvor det allerede er utført manuell revisjon, også på personnivå. Det ideelle hadde vært om vi kunne sammenlikne data før og etter revisjon. Hva vi skal betrakte som originaldata kan imidlertid være vanskelig å definere fordi det i en del tilfeller er påkrevet å gjøre noen grep allerede i innlesingsrutinen for å få dataene inn på basen. Ser vi imidlertid på data som kommer via Altinn (stadig flere leverer via Altinn) vil det være mulig å lagre data som ikke har vært gjenstand for noen påvirkning fra SSB. Dette bør tilstrebes ved en omlegging av dagens rutiner og vil gi en god dokumentasjon av hvordan SSB påvirker datagrunnlaget gjennom kontroll og revisjon.

**Litteraturliste:**

- |         |   |
|---------|---|
| 2003/74 | Lønnsstatistikk privat sektor 1997-2002. Dokumentasjon av utvalg og beregning av vektorer |
| 2009/20 | Lønnsstatistikk 2009. Etablering av populasjon og utvalg                                  |

## Vedlegg A: Prosesdiagram lønnsstatistikk



## Vedlegg B: Spredningsmål i lønnsstatistikken

I tabell 18 er det blitt beregnet konfidensintervall, standardfeil og relativ standardfeil (CV) for heltidsansatte i varehandel i 2009.

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 antall observasjoner 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:

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

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:

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



## Tabellregister

1.	Innkvitteringsstatus i varehandel totalt, HSH, HSH Apotek og SAMFO i 2008 og 2009.....	8
2.	Rapporteringstype i prosent i perioden 2004-2009. Andel observasjoner.....	8
3.	Gyldige og ugyldige fødselsnummer, 2008 og 2009.....	9
4.	Status etter kontroller og endringer.....	10
5.	Antall observasjoner fordelt på lønnsperiode før og etter kontroll og korreksjoner av arbeidstid.....	10
6.	Observasjoner som får beregnet avtalt arbeidstid ut ifra betalte timer.....	11
7.	Antall observasjoner fordelt på arbeidstidsgrupper og gjennomsnittlig arbeidstid for heltidsansatte i 2005, 2006, 2007, 2008 og 2009.....	11
8.	Oversikt over overtidstimer og overtidsbetaling.....	12
9.	Fordeling av observasjoner på 1. siffer STYRK i 2008 og 2009.....	13
10.	Fordeling av antall med 1-siffer STYRK i lønnsstatistikken sammenlignet med AA-registeret.....	13
11.	Antall observasjoner fordelt etter koder for grenseverdier (ok-koder) i 2009.....	14
12.	Antall foretak og personer i populasjon og antall foretak i utvalg og justert etterstratifisert invers trekkesannsynlighet (vekt). Per stata.....	16
13.	Vektjustering 2009. Heltids- og deltidsansatte.....	16
14.	Effekten av vektning 2009. Gjennomsnittlig månedslønn for heltidsansatte. Kroner.....	16
15.	Effekt av kontroll og revisjon av avtalt arbeidstid per uke.....	17
16.	Effekt av kontroll og revisjon av lønnsvariable. Heltidsansatte. Kroner.....	18
17.	Effekt av kontroll og revisjon av månedslønn etter næring og yrke. Heltidsansatte. Antall personer og kroner.....	19
18.	Effekt av kontroll og revisjon av månedslønn. Spredningsmål per næringsområde. Antall og kroner.....	20