

Dina Rafat

**Analyse av sammenheng
mellom ektefellers syssel-
setting i en familie**

Notater

Innhold

1.	<i>Innledning</i>	3
2.	<i>Datsett</i>	3
3.	<i>Analyse av Sammenheng mellom Ektefellers/Samboeres Sysselsetting</i>	4
3.1.	<i>Oddsforholdet</i>	4
3.2.	<i>Test av betinget uavhengighet</i>	6
3.3.	<i>Test av homogen sammenheng</i>	7
3.4.	<i>Estimering av felles verdi</i>	8
4.	<i>Korrelasjonsanalyse</i>	8
5.	<i>Sammenheng mellom Ektefellers Sysselsetting mht Alder</i>	9
6.	<i>Estimering av Totalt Antall Sysselsatte blant Ektepar i Norge</i>	10
7.	<i>Frafallsjustering ved Imputering</i>	14
8.	<i>Oppsummering</i>	18
9.	<i>Referanser</i>	19
Vedlegg		
A.	<i>Partielle tabeller for 7 grupper (basert på alder og registersysselsetting)</i>	20
B.	<i>Partielle tabeller for 7 grupper (basert på alder)</i>	23
	<i>De sist utgitte publikasjonene i serien Notater</i>	26

1. Innledning

Målet med dette arbeidet er å se på sammenheng mellom kvinnens og mannens sysselsetting i en familie og estimere det totale antall sysselsatte blant ektepar i Norge. I dette notatet ble samboere med felles barn inkludert i familiebegrepet som anbefalt i Solheim og Zhang (2001).

Problemstillingen er knyttet til omlegging av Arbeidskraftundersøkelsen (AKU).

Arbeidet er bygd opp på følgende måte: I kapittel 2 beskrives datagrunnlaget for våre analyser. Kapittel 3 studerer sammenheng mellom ektefellers/samboeres sysselsetting mht registersysselsetting og alder, korrelasjonsanalyse er presentert i kap. 4 sett i forhold til resultatene fra kapittel 3. Kapittel 5 ser på sammenheng mellom ektefellers sysselsetting bare mht alder. Det totale antall sysselsatte blant ektepar i Norge estimeres i kap. 6. Frafallsjustering ved imputering er gitt i kapittel 7 og samme analyser gjennomføres med data supplert med de imputerte verdiene.

2. Datasett

Datagrunnlaget for denne studien er AKU-tall for andre kvartal år 2001. Det ble konstruert en SAS-datafil som inneholdt følgende variable fra AKU:

- Intervjuobjektets (IOs) fødselsnummer (F_NR);
- IOs familienummer (IOS_FAM);
- alder (alder);
- kjønn (kjønn);
- ekteskapeleg status hentet fra AKU og register (spm120e);
- faktisk sysselsetting, samlet for hoved- og bi-arbeid (sstat);
- registersysselsetting (stmnace).

Fra AKU-datasettet valgte vi par som var gift og samboere med felles barn. De har samme familienummer, som gjør det lettere å identifisere dem som par. Det kan være interessant å se på sysselsettingsmønsteret hos samboere par uten felles barn, men i dette notatet nøyer vi oss med de som har felles familienummer.

Av 20 243 personer i AKU-datasettet for andre kvartal var 10 226 personer gift og 2 833 samboere. De var delt i to datasett mht kjønn, sortert på familienummer og slått sammen etter familienummer. Til slutt fikk vi et datasett som inneholdt 4 691 par med identifisert alder. Videre ble data organisert i en tabell, hvor par var kryssklassifisert ifølge tre aldersgrupper (16-24, 25-44, 45-74) og registersysselsetting (sysselsatt ifølge register og ikke sysselsatt ifølge register). Tabell 1 er grunnleggende for vår analyse.

Tabell 1. Ektefellepar kryssklassifisert på alder og registersysselsetting.

		regsyss kv						Total
		Nei (0)			Ja (1)			
		Alder			Alder			
		16-24	25-44	45-74	16-24	25-44	45-74	
regsyss_m	Alder							
Nei (0)	16-24	1	2	0	0	3	6*	12
	25-44	3	37	6	1	83	25*	155
	45-74	8	19	356*	5	38	277*	703
Ja (1)	16-24	3	3	4*	1	10	27*	48
	25-44	12	324	28	15	982	84*	1 445
	45-74	23	92	403	42	313	1 455*	2 328
Total		50	477	797	64	1 429	1 874*	4 691

Skyggelegginger plasserer par i forskjellige grupper (se kapittel 3).

Kommentar: Tabell 1 viser at vårt datasett inneholder relativt mange par (*) hvor mannen er i alderen 16-24 og kvinnen er i alderen 45-74 (37 par). 6 par er gift, mens resten, 31 par, er samboere. Det kan hende at de 31 par inkluderer slike situasjoner hvor mor og voksen sønn fortsatt har samme familienummer, men har samboer hver for seg. I vårt tilfellet, er det vanskelig å skille mellom de som faktisk er samboere og de, som er telt på feil grunnlag. Derfor, ekskluderte vi denne gruppen fra vår analyse.

3. Analyse av Sammenheng mellom Ektefellers/Samboeres Sysselsetting

For en gitt familie/samboerpar la

a_K = aldersgruppe for kvinne (verdier: 1,2,3)

a_M = aldersgruppe for mann (")

r_K = 1 hvis kvinne er registersysselsatt, 0 ellers

r_M = 1 hvis mann er registersysselsatt, 0 ellers.

$$z = (a_K, a_M, r_K, r_M).$$

Vi skal se på sammenheng mellom mannens og kvinnens faktiske sysselsetting i en familie, gitt z-konfigurasjon.

Tabell 1 viser at vi har fått 36 forskjellige konfigurasjoner. Men mange celler er enten tomme eller inneholder utilstrekkelig antall observasjoner. For senere analyser har vi slått sammen de 36 konfigurasjonene i 7 z-grupper for å unngå tomme celler i partielle tabeller. Gruppene er som følger:

Gruppe 1: Kvinne 16 - 44 år, regsyss. 0. Mann 16 - 74 år, regsyss. 0;

Gruppe 2: Kvinne 45 - 74 år, regsyss. 0. Mann 25 - 74 år, regsyss. 0;

Gruppe 3: Kvinne 45 - 74 år, regsyss. 0. Mann 25 - 74 år, regsyss. 1;

Gruppe 4: Kvinne 16 - 44 år, regsyss. 1. Mann 16 - 74 år, regsyss. 0;

Gruppe 5: Kvinne 45 - 74 år, regsyss. 1. Mann 25 - 74 år, regsyss. 0;

Gruppe 6: Kvinne 16 - 44 år, regsyss. 1. Mann 16 - 74 år, regsyss. 1;

Gruppe 7: Kvinne 45 - 74 år, regsyss. 1. Mann 25 - 74 år, regsyss. 1.

En stor gruppe: kvinne 16-44 år, regsyss. 0, mann 16-74 år, regsyss.1 med 457 observasjoner ble ekskludert fra vår analyse pga. en tom celle i partiell tabell. Etter grupperingen det totale antallet par ble redusert til 4 197.

La

Y_K = 1 hvis kvinne er faktisk sysselsatt, 0 ellers

Y_M = 1 hvis mann er faktisk sysselsatt, 0 ellers.

Y_M og Y_K er to kategoriske variabler.

3.1. Oddsforholdet

Gitt z og at $Y_M = 1$: La $\pi_{1,z} = P(Y_K = 1 | Y_M = 1, z)$ - sannsynlighet for at kvinnen faktisk er sysselsatt.

La begivenheten at kvinne er sysselsatt kalles "suksess". Odds for suksess blant familier hvor mannen er faktisk sysselsatt, $Y_M = 1$, gitt z :

$$\text{odds}_1 = \frac{\pi_{1,z}}{1 - \pi_{1,z}} = \frac{P(\text{suksess})}{P(\text{fiasko})}$$

Tilsvarende gitt at $Y_M = 0$ og gitt z : La $\pi_{0,z} = P(Y_K = 1 | Y_M = 0, z)$.

Odds for suksess blant familier hvor $Y_M = 0$, gitt z :

$$\text{odds}_0 = \frac{\pi_{0,z}}{1 - \pi_{0,z}}$$

Det vanligste mål for sammenheng mellom to binære variable er oddsforholdet (Agresti, 1996):

$$\theta_z = \frac{\text{odds}_1}{\text{odds}_0}$$

Dataene fra hver celle i Tabell 1 kan bli presentert i en 2x2 kontingenstabell, hvor radene er faktisk sysselsetting for menn og kolonnene er faktisk sysselsetting for kvinner. For familier med konfigurasjon z : n_{00z} = antall familier hvor ingen er sysselsatt, n_{01z} = antall familier hvor mannen er ikke sysselsatt og kvinnen er sysselsatt, etc.

	$Y_K = 1$	$Y_K = 0$	Totaler
$Y_M = 1$	n_{11z}	n_{10z}	n_{1+z}
$Y_M = 0$	n_{01z}	n_{00z}	n_{0+z}
Totaler	n_{+1z}	n_{+0z}	n_z

Estimering av θ_z :

$$\hat{\pi}_{1,z} = \frac{n_{11z}}{n_{1+z}}, \quad \hat{\pi}_{0,z} = \frac{n_{01z}}{n_{0+z}}$$

og oddsforholdet:

$$\hat{\theta}_z = \frac{\hat{\pi}_{1,z} / (1 - \hat{\pi}_{1,z})}{\hat{\pi}_{0,z} / (1 - \hat{\pi}_{0,z})} = \frac{n_{00z} n_{11z}}{n_{01z} n_{10z}}$$

Når $\theta_z = 1$ så er $\pi_{1,z} = \pi_{0,z}$ og omvendt. Det betyr at Y_M og Y_K er uavhengige, gitt z .

Tabell 2 viser oddsforholdene for hver z -gruppe med 95% konfidensintervall. I gruppe 1, for eksempel, for menn som var sysselsatt er estimert odds for at kvinne var sysselsatt lik $16/23=0,70$. Verdi 0,70 betyr at det var 0,70 kvinner i arbeid for hver kvinne som var uten arbeid. For ikke sysselsatte menn var estimert verdi $6/25=0,24$. Oddsforholdet er lik $\hat{\theta}=0,66/0,24=2,90$. Estimert odds for at kvinnen er sysselsatt er 2,9 ganger større hvis mannen er i arbeid, enn hvis mannen er uten arbeid.

Oddsforholdet er størst i grupper 2 og 7, hvor $\hat{\theta}_z = 5,91$ og $\hat{\theta}_z = 7,66$ henholdsvis. De fleste par i disse gruppene er eldre personer (45-74 år) og hvor begge ikke er registersysselsatt (gruppe 2) eller begge er registersysselsatt (gruppe 7). Det tyder på at blant eldre mennesker er det en stor sammenheng mellom mannens og kvinnens sysselsetting i en familie. Vi ser nærmere på denne sammenhengen i kapittel 5.

Tabell 2. Ektefellers/samboeres sysselsetting, estimert oddsforhold med 95% konfidensintervall.

Gruppe	Y_M	Y_K		Andel av kvinnens syssels. gitt mannens sysselsetting	Oddsforhold	95% konfidensintervall (Oddsforhold)	
		Ja 1	Nei 0				
1	Ja 1	16	23	0,410	2,90	0,97	8,67
	Nei 0	6	25	0,194			
2	Ja 1	12	29	0,293	5,91	2,64	13,22
	Nei 0	21	300	0,065			
3	Ja 1	98	300	0,246	3,27	0,98	10,94
	Nei 0	3	30	0,091			
4	Ja 1	90	2	0,978	1,22	0,11	13,83
	Nei 0	37	1	0,974			
5	Ja 1	86	4	0,956	1,76	0,57	5,40
	Nei 0	196	16	0,925			
6	Ja 1	1294	50	0,963	1,44	0,19	10,98
	Nei 0	18	1	0,947			
7	Ja 1	1437	48	0,968	7,66	3,72	15,77
	Nei 0	43	11	0,796			

Sammenhengen mellom Y_M og Y_K er svak i grupper 4 og 6. I disse grupper er oddsforholdene nær 1, som betyr at mannens og kvinnens sysselsetting er nesten uavhengige. Som vi ser fra tabellen er anslagene veldig usikre, med store konfidensintervaller, som overlapper hverandre. For beregning av konfidensintervallene henvises til Agresti (1996).

3.2. Test av betinget uavhengighet

I Tabell 2 har vi tre variabler: mannens sysselsetting Y_M , kvinnens sysselsetting Y_K og kontroll variabel z som fremstiller $Z=7$ forskjellige grupper. Nå kan vi undersøke sammenheng mellom Y_M og Y_K samtidig som vi kontrollerer for z . For $2 \times 2 \times Z$ tabeller, null hypotesen at Y_M og Y_K er betinget uavhengige gitt z , betyr at alle $\theta_z = 1$ i hver partiell tabell.

Under null hypotesen, estimert forventning og betinget varians av n_{11z} , gitt marginaler, er

$$\hat{\mu}_{11z} = \hat{E}(n_{11z}) = \frac{n_{1+z}n_{+1z}}{n_z}$$

$$\text{Var}(n_{11z}) = \frac{n_{1+z}n_{0+z}n_{+1z}n_{+0z}}{n_z^2(n_z - 1)}$$

En Cochran-Mantel- Haenszel (*CMH*) statistikk summerer avviket fra null hypotesen for de 7 partielle tabeller (Agresti, 1996):

$$CMH = \frac{\left[\sum_z (n_{11z} - \hat{\mu}_{11z}) \right]^2}{\sum_z \text{Var}(n_{11z})}$$

Tabell 2 viser oddsforhold for hver partiell tabell. Alle oddsforholdene viser en positiv sammenheng og faller i samme retning, så det er rimelig å bruke *CMH* statistikk. Vi får $CMH = 40,52$ med $DF = 1$. Der er en sterk indikasjon mot betinget uavhengighet ($P < .0001$) med antall observasjoner $n = 4 197$. Dvs., vi kan påstå betinget avhengighet, med stor grad av sikkerhet.

Videre skal vi teste separat nullhypotesene at $\theta_z = 1$ for alle z . For en gitt z , la $\pi_{ij} = P(Y_M=i, Y_K=j)$ - sannsynlighet for at (Y_M, Y_K) faller i cellen i rad i og kolonne j .

I 2x2 kontingenstabeller, har nullhypotesen av statistisk uavhengighet følgende form:

$$H_0: \pi_{ij} = \pi_{i+} \pi_{+j} \quad \text{for alle } i \text{ and } j.$$

For å teste H_0 identifiserer vi $\mu_{ij} = n\pi_{ij} = n\pi_{i+}\pi_{+j}$ som forventet antall. Her μ_{ij} er forventningsverdien av n_{ij} , hvis vi antar uavhengighet.

$$\text{Estimert forventet antall} \quad \hat{\mu}_{ij} = \frac{n_{i+}n_{+j}}{n}, \text{ under } H_0.$$

For å teste uavhengighet i $I \times J$ kontingenstabeller, er Pearson kjikvadrat statistikk brukt:

$$X^2 = \sum \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}}.$$

Stor forskjell mellom observert antall og forventet antall gir stor X^2 verdi og bevis mot nullhypotesen. Tabell 3 viser kjikvadratstatistikk for alle Z grupper.

Tabell 3. Kjikvadratstatistikk.

Gruppe	Antall observ.	Kjikvadratstatistikk		
		DF	Verdi	P-verdi
1	70	1	3,7637	0,0524
2	362	1	22,6643	<,0001
3	431	1	4,0974	0,0429
4	130	1	0,0250	0,8744
5	302	1	0,9835	0,3213
6	1363	1	0,1238	0,7249
7	1539	1	41,5114	<,0001

Resultatene i Tabell 3 stemmer bra overens med konfidensintervallene i Tabell 2.

X^2 er størst i grupper 2 og 7, hvor oddsforholdene også har en stor verdi. I gruppe 2, $X^2=22,66$ med $P<,0001$ og i gruppe 7 er $X^2=41,51$ med $P<,0001$. Kjikvadrat testen bekrefter at data inneholder klar bevis mot null hypotesen at mannens og kvinnens sysselsetting er uavhengige i disse 2 grupper. P verdi<,0001 betyr at, under H_0 , sjansen for å få verdi av X^2 lik og mer enn beregnet verdi er veldig liten.

X^2 er minst i grupper 4 ($X^2=0,025$) og 6 ($X^2=0,1238$). Det betyr at i disse grupper ser ektefellers sysselsetting ut til å være omtrent uavhengig. Vi har intet bevis for å motsi nullhypotesen. Det kan sees tydeligere hvis vi analyserer prosentdelen i hver rad i tabeller for grupper 4 og 6 (Vedlegg A). Her er prosentfordeling i rad 1 nesten lik prosentfordelingen i rad 2. Menn som ikke er sysselsatt har 2,63% ektefeller som ikke er sysselsatt og 97,37% som er sysselsatt. Menn som er i arbeid har 2,17% ektefeller som ikke er sysselsatt og 97,83% har ektefeller som er sysselsatt. Vi konkluderer at i gruppe 4 (og i gruppe 6) ser det ut til å være liten sammenheng mellom mannens og kvinnens sysselsetting.

3.3. Test av homogen sammenheng

I kapittel 3.1 så vi at oddsforholdene for 7 grupper er ulike. Men er det statistisk bevis på at de er forskjellige? Er dataene så nøyaktig at vi kan påstå at θ_z -ene er ulike?

Vi kan teste nullhypotesen at alle oddsforholdene mellom Y_M og Y_K er like på hvert z-nivå:

$H_0: \theta_{Y_M Y_K(1)} = \theta_{Y_M Y_K(2)} = \dots = \theta_{Y_M Y_K(7)}$. Det er en test av homogen sammenheng for $2 \times 2 \times Z$ tabeller.

La $\{\hat{\mu}_{11z}, \hat{\mu}_{10z}, \hat{\mu}_{01z}, \hat{\mu}_{00z}\}$ være estimert forventet antall i z partiell tabell, som har oddsforhold lik Mantel-Haenszel estimat $\hat{\theta}_{MH}$ av felles oddsforhold (se nedenfor).

En test observator, Breslow-Day observatoren har følgende form:

$$\sum \frac{(n_{ijz} - \hat{\mu}_{ijz})^2}{\hat{\mu}_{ijz}},$$

hvor summen er tatt over alle cellene i $2 \times 2 \times Z$ -tabellen. Jo mindre statistikken er desto mindre er bevis mot H_0 .

I vår tilfelle SAS rapporterer Breslow-Day statistikk lik 9,39, basert på $df=Z-1=7-1=6$, P-verdi = 0,15. Det er en viss grad av bevis at oddsforholdene er forskjellige.

3.4. Estimering av felles verdi

Det er mer informativt å estimere graden av sammenhengen enn bare å teste hypotesen. Når sammenhengen ser ut til å være stabil over partielle tabeller kan vi estimere felles verdi av Z oddsforhold, hvis det antas likhet.

I $2 \times 2 \times Z$ tabell, anta at $\theta_{Y_M Y_K(1)} = \theta_{Y_M Y_K(2)} = \dots = \theta_{Y_M Y_K(7)}$. Mantel-Haenszel estimat av felles verdi er lik

$$\hat{\theta}_{MH} = \frac{\sum_z (n_{11z} n_{00z} / n_z)}{\sum_z (n_{10z} n_{01z} / n_z)} = \frac{\sum_z a_z \hat{\theta}_z}{\sum_z a_z},$$

hvor $a_z = \frac{n_{01z} n_{10z}}{n_z}$, et veiet gjennomsnitt av $\hat{\theta}_1, \dots, \hat{\theta}_7$.

I vår studie blir Mantel-Haenszel oddsforholdsestimatet lik $\hat{\theta}_{MH} = 3,57$ med 95% konfidens intervall (2,31; 5,53). Odds for at kvinne er sysselsatt er 3,5 ganger så stor når mann er sysselsatt enn når mann ikke er sysselsatt. I vår tilfelle varierer oddsforholdene i partielle tabeller ganske mye, derfor gir ikke $\hat{\theta}_{MH}$ en nyttig oppsummering av de betingede sammenhengene.

4. Korrelasjonsanalyse

Selv om korrelasjonskoeffisienten ikke er et velegnet mål for sammenheng for binære variable, så kan man estimere hver korrelasjonskoeffisient ρ_z ved:

$$\hat{\rho}_z = \frac{n_{00z} n_{11z} - n_{10z} n_{01z}}{\sqrt{n_{1+z} n_{0+z} n_{+0z} n_{+1z}}}$$

SAS programmet gir oss følgende verdier av estimatet (Tabell 4):

Tabell 4. Korrelasjonskoeffisient for 7 grupper.

Gruppe	Pearson korrelasjon		Tilnærmet konfidensintervall	
	Verdi	ASE	Verdi - 2ASE	Verdi +2ASE
1	0,2319	0,1118	0,0083	0,4555
2	0,2502	0,0756	0,099	0,4014
3	0,0975	0,0346	0,0283	0,1667
4	0,0139	0,0909	-0,1679	0,1957
5	0,0571	0,0511	-0,0451	0,1593
6	0,0095	0,0318	-0,0541	0,0731
7	0,1642	0,0512	0,0618	0,2666

Vi ser at Pearson korrelasjon ikke gir samme inntrykk av grad av avhengighet for de forskjellige gruppene som oddsforholdene i Tabell 2. Allikevel fordeler den gruppene på en lignende måte som oddsforholdene. Dvs., det observeres en viss grad av korrelasjon i grupper 1, 2 og 7, hvor oddsforholdene er størst, og ingen korrelasjon i grupper 4 og 6, hvor oddsforholdene er minst.

5. Sammenheng mellom Ektefellers Sysselsetting mht Alder

I våre tidligere analyser var det en sterk indikasjon på sammenheng mellom mannens og kvinnens sysselsetting blant eldre mennesker (kapittel 3.1). For å sjekke det nærmere konstruerte vi nye z grupper, hvor vi kontrollerte bare for alder.

$$z = (a_k, a_m)$$

Tabell 5 viser en fordeling av faktisk sysselsetting blant ektefeller og samboer med felles barn mht alderen.

Nå får vi 9 forskjellige konfigurasjoner. Celler hvor mann er 16-24 og kvinner er 16-24 og 25-44 er slått sammen i gruppe 1. Gruppe hvor mann er i alderen 16-24 og kvinne 45-74 er ekskludert fra analyse pga. feil grunnlag for telling (se kapittel 2). Resten av gruppene er som følger:

Gruppe 1: kvinne 16-44, mann 16-24

Gruppe 2: kvinne 16-24, mann 25-44.

Gruppe 3: kvinne 25-44, mann 25-44.

Gruppe 4: kvinne 45-74, mann 25-44.

Gruppe 5: kvinne 16-24, mann 45-74.

Gruppe 6: kvinne 25-44, mann 45-74.

Gruppe 7: kvinne 45-74, mann 45-74.

Se Vedlegg B for partielle tabeller for 7 forskjellige grupper.

Tabell 5. Kryssklassifisering på alder.

Mann alder	Kvinne / alder			Total
	16 - 24	25 - 44	45 - 74	
16 - 24	5	18	37	60
25 - 44	31	1 426	143	1 600
45 - 74	78	462	2 491	3 031
Total	114	1 906	2 671	4 691

Estimert oddsforhold med 95% konfidensintervall og kji kvadratstatistikk er rapportert i Tabell 6.

Tabell 6. Estimert oddsforhold med 95% konfidensintervall og kji kvadratstatistikk.

Gruppe	Antall observ.	Verdi (Oddsforholdet)	95% konfidensintervallet (Oddsforholdet)		Kji kvadratstatistikk		
					DF	Verdi	Prob.
1	23	4,25	0,22	83,52	1	1,03	0,3106
2	31	2,40	0,34	16,97	1	0,80	0,3714
3	1 426	1,43	0,72	2,85	1	1,04	0,3082
4	143	1,13	0,34	3,76	1	0,04	0,8464
5	78	5,58	0,94	32,97	1	4,32	0,0376
6	462	3,34	1,52	7,31	1	10,00	0,0016
7	2 491	6,13	5,03	7,48	1	356,99	<,0001

Her ser vi en stor sammenheng mellom ektefellers sysselsetting i gruppe 7, hvor både mann og kvinne er i aldersgruppe 45-74. Oddsforhold er lik 6,13 med et lite konfidensintervall. Også kji kvadratstatistikk ($X^2 = 356,98$ med $P < ,0001$) bekrefter at data inneholder klar bevis mot null hypotesen om statistisk uavhengighet.

Det er nesten ingen sammenheng mellom mannens og kvinnens sysselsetting i grupper 3 og 4, hvor begge er i alderen 25-44 (gruppe 3) og kvinne er 45-74 mens mann er 25-44 (gruppe 4). I begge tilfeller er oddsforholdet nær 1, som peker på statistisk uavhengighet.

Konklusjon: Det er en stor sammenheng mellom ektefellers sysselsetting i en familie, hvor begge er i alderen 45-74. Sammenhengen er størst blant godt voksne mennesker og minst blant par som er i "karrierealderen" (25-44 år).

6. Estimering av Totalt Antall Sysselsatte blant Ektepar i Norge

Vi betegner nå z -gruppene (mht alder og registersysselsetting) som etterstrata. For å estimere totalt antall sysselsatte blant ektepar og samboer med felles barn i Norge danner vi etterstratifiserings-estimatoren, \hat{t}_{est} :

$$\hat{t}_{est} = \sum_{z=1}^Z N_z \bar{y}_z$$

hvor \bar{y}_z er andelen faktisk sysselsatte i etterstratum z i utvalget og N_z er populasjonsstørrelsen på etterstratum z , dvs.

$$\bar{y}_z = \frac{(2n_{11z} + n_{10z} + n_{01z})}{2n_z} = \frac{(n_{1+z} + n_{+1z})}{2n_z}.$$

For å finne ut antall familier i hvert etterstratum var Familiefilen koblet opp med Arbeidstakerregister (AT). Det var 830 502 familier i alderen 16-74 år i Norge i 2 kvartal 2001, dvs 1 661 004 personer som var gift eller samboer med minst ett felles barn, fordelt på 8 etterstrata på følgende måte (se Tabell 7). Her inkluderte vi en gruppe (etterstrata 0) hvor kvinne er i alder 16-44, regsyss. 0 og mann 16-74 år, regsyss.1. Den gruppen var ekskludert fra våre tidligere analyser pga. en tom celle i partiell tabell. Tabell 7 inkluderer ikke par, hvor mann er i alderen 16-24 og kvinne er i alderen 45-74. Det er totalt 7 par (dvs. 14 personer) i populasjonen som er utelukket. Estimert antall personer blant ektepar

og samboer med felles barn som var faktisk sysselsatt ($N_z \bar{y}_z$) fordeler seg på hvert etterstratum som vist i Tabell 7 og det etterstratifiserte estimatet blir $\hat{t}_{est} = 1\,289\,153$.

Tabell 7. Estimert antall sysselsatte fordelt på etterstrata (mht registersysselsetting og alder).

Etterstrata	Populasjonsstørrelse N_z	Andel faktisk sysselsatte \bar{y}_z	$N_z \bar{y}_z$
0	183 548	0,7068	129 731,72
1	40 476	0,4357	17 635,39
2	176 494	0,1022	18 037,69
3	138 914	0,5789	80 417,31
4	55 352	0,8423	46 622,99
5	103 702	0,6159	63 870,06
6	514 340	0,9743	501 121,46
7	448 164	0,9633	431 716,38
Total	1 660 990		1 289 153,00

AKU utvalget er et stratifisert utvalg etter fylke. Vi skal betrakte det som et tilnærmet enkelt tilfeldig stratifisert utvalg og skal estimere variansen til etterstratifiseringsestimatorene ved hjelp av Bootstrap-metoden.

Vi betrakter y_{zi} som tilnærmet uavhengig identisk fordelte (uif) innen for hvert strata (fylke). Utvidelse av den klassiske algoritmen for bootstrap for stratifisert utvalg, gitt denne egenskapen, er som følger (Rao and Wu, 1988):

1. Vi trekker et enkelt tilfeldig utvalg $\{y_{gi}^*\}_{i=1}^{n_g}$ med tilbakelegging fra det observerte utvalget $\{y_{gi}\}_{i=1}^{n_g}$ i strata g , uavhengig for hvert stratum. Estimatorene $\hat{\theta}^*$ (i vår tilfelle \hat{t}_{est}^*) er beregnet på grunnlag av bootstraputvalget etter samme formel som $\hat{\theta}$ (\hat{t}_{est}) er beregnet på grunnlag av det opprinnelige utvalget.
2. Punkt 1 gjentas uavhengig et stort antall B ganger som gir tilsvarende estimatorene $\hat{\theta}^{*1}, \dots, \hat{\theta}^{*B}$
3. Bootstrapestimatorene til variansen til $\hat{\theta}$ blir da

$$v_b(a) = \frac{1}{B-1} \sum_1^B (\hat{\theta}^{*b} - \hat{\theta}^*)^2,$$

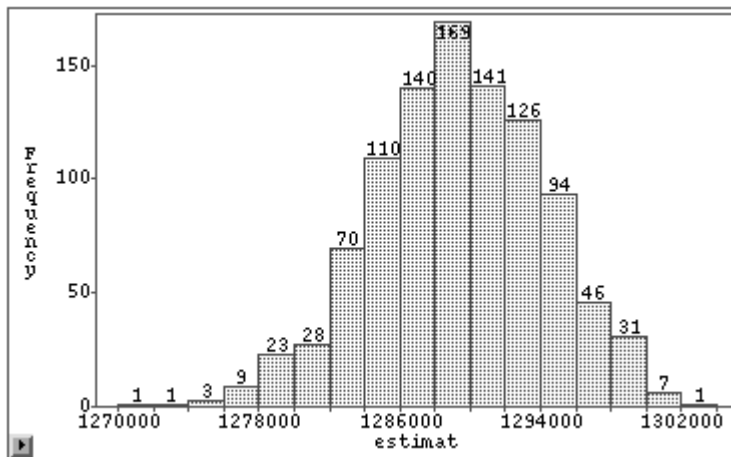
hvor $\hat{\theta}^* = \sum \hat{\theta}^{*b} / B$

1000 bootstrap utvalg ble trukket fra ektepar og samboer med felles barn i det opprinnelige AKU utvalget. Vi begrenser oss til den populasjonen, pga. at nettopp den er av interesse. Da er variansestimateret betinget, gitt den utvalgsstørrelsen vi har fått. For hvert bootstraputvalg beregnes estimatoren \hat{t}_{est} og bootstrap variansestimateret til \hat{t}_{est} blir $4\,903^2$ (Tabell 8).

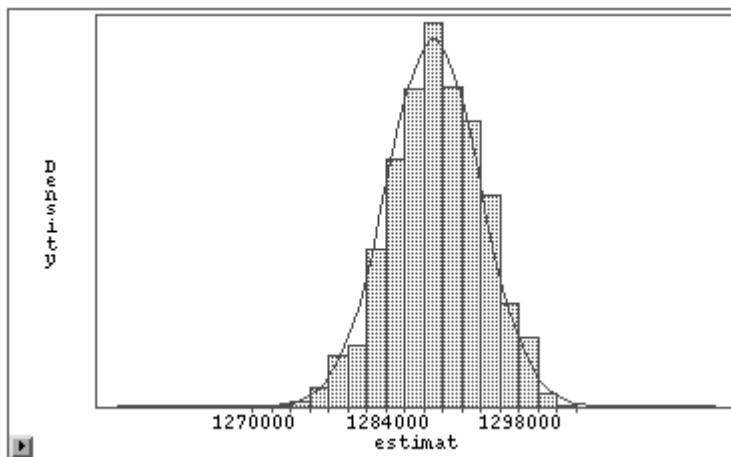
Tabell 8. Oppsummering av bootstrapresultatene.

N	Mean	Std Dev	Minimum	Maksimum
1000	1 289 341,79	4 902,84	1 271 727,95	1 303 174,77

Figur 2 viser at histogram over 1000 etterstratifiseringsestimater tilnærmer seg normalkurven, dvs. at etterstratifiseringsestimatorene er tilnærmet normalfordelt.



Figur 1. Histogram over etterstratifiseringsestimater (mht alder og registersyssestetting) basert på 1000 bootstrapvalg.



Figur 2. Histogram over etterstratifiseringsestimater med tilpasset normalkurve.

En forholdsvis liten bootstrapvarians peker på at den beregnede etterstratifiseringsestimatorene \hat{t}_{est} er ganske nøyaktig.

Vi foretar samme analyser med z-grupper som tar hensyn bare til alderen som etterstrata. Familiefilen gir oss 1 667 856 personer (833 928 familier) fordelt på aldersgrupper på følgende måte (Tabell 9). Det er en forskjell på 3 426 familier mellom Familiefilen og Arbeidstakerregisteret.

Fordeling av det estimerte antallet ektefeller og samboer som er faktisk sysselsatt ($N_z \bar{y}_z$) på hvert etterstratum er vist i Tabell 9 og det etterstratifiserte estimatet blir $\hat{t}_{est} = 1\,351\,598$.

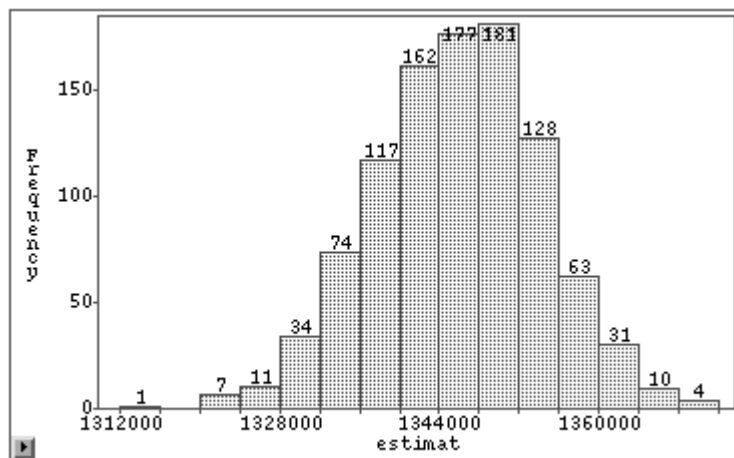
Tabell 9. Estimert antall sysselsatte fordelt på etterstrata (mht alder).

Etterstrata	Populasjonsstørrelse N_z	Andel faktisk sysselsatte \bar{y}_z	$N_z \bar{y}_z$
Mann 16-24, kvinne 16-24	6 110	0,7000	4 277
Mann 16-24, kvinne 25-44	1 960	0,8889	1 742,22
Mann 16-24, kvinne 45-74	14	0,9324	13,05
Mann 25-44, kvinne 16-24	18 690	0,7097	13 264,29
Mann 25-44, kvinne 25-44	623 580	0,8945	557 792,31
Mann 25-44, kvinne 45-74	19 844	0,8287	16 444,72
Mann 45-74, kvinne 16-24	270	0,8141	219,81
Mann 45-74, kvinne 25-44	145 232	0,8885	129 038,63
Mann 45-74, kvinne 45-74	852 156	0,7379	628 805,91
Totall	1 667 856		1 351 597,90

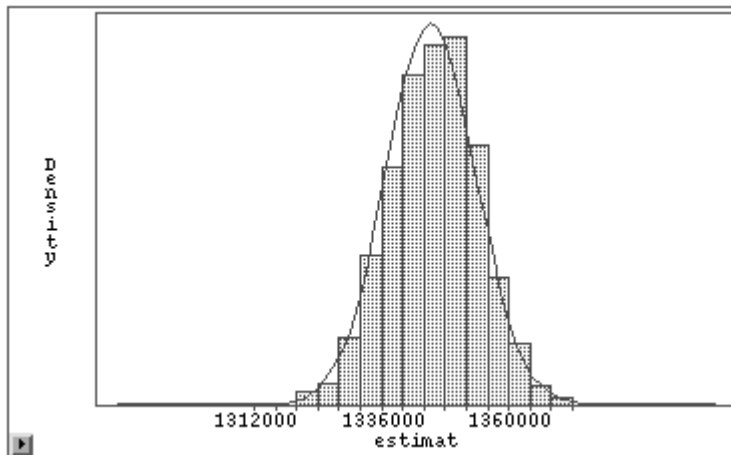
Bootstrapestimering av variansen til etterstratifiseringsestimatorene basert på 1000 utvalg gir et standardavvik på 8 484 (Tabell 10).

Tabell 10. Oppsummering av bootstrapresultatene.

N	Mean	Std Dev	Minimum	Maksimum
1000	1 345 758,94	8 483,54	1 314 652,37	1 371 036,61



Figur 3. Histogram over etterstratifiseringsestimer (mht alder) basert på 1000 bootstraputvalg



Figur 4. Histogram over etterstratifiseringsestimater med tilpasset normalkurve.

Bootstrapestimatoren til standardavvik er nesten dobbelt så stor når vi etterstratifiserer bare på alderen. Som regel får vi større reduksjon av usikkerheten i estimatet når strata er homogene, dvs. lite variasjon innen strata og stor variasjon mellom strata. Hvis vi sammenligner andelen faktisk sysselsatte i Tabell 7 og Tabell 9, viser det seg at det er større variasjon i andelen faktisk sysselsatte når vi etterstratifiserer både mht registersyssestetting og alder. Hvis vi etterstratifiserer bare etter aldersgrupper (Tabell 9) er variasjonen mellom strata veldig liten.

Konklusjon: Vi får en bedre etterstratifiseringsestimator ved å etterstratifisere etter både alder og registersyssestetting. Mer forventningsrett estimator oppnås fordi registersyssestetting er i høy grad korrelert med faktisk sysselsetting. Når registersyssestetting droppes fra etterstratifisering får vi høyere varians og skjevhet i estimatet.

7. Frafallsjustering ved Imputering

I AKU utvalget for andre kvartal år 2001 var det trukket ut 24 237 personer, 3 965 av disse var enhetsfrfall. 1 940 personer var gifte eller samboere med felles barn. Etter sortering og kobling med familie-filen og AT register fikk vi 461 par i frafallsgruppen med kjent alder og registersyssestetting status, som var fordelt på aldersgrupper på følgende måte:

Tabell 11. Enhetsfrfall fordelt på etterstrata.

		regsyss_kv						Total
		Nei (0)			Ja (1)			
		Alder			Alder			
		16-24	25-44	45-74	16-24	25-44	45-74	
regsyss_m	Alder							
Nei (0)	16-24	0	0	0	0	0	0	
	25-44	0	19	1	0	7	28	
	45-74	0	5	62	0	3	95	
Ja (1)	16-24	0	0	0	0	0	0	
	25-44	2	39	0	1	91	138	
	45-74	0	5	38	0	31	200	
Total		2	68	101	1	132	461	

Nå kjenner vi antallet i frafallsgruppen som tilhører etterstratum z , m_z . For hvert etterstratum imputerer vi sysselsettingsfordelingen for de m_z ekteparene som i svargruppen, dvs. fyller inn de estimerte verdier av (Y_k, Y_m) . La

m_{00z}^* = antall imputerte familier hvor ingen er sysselsatt, etc.

	$Y_K = 1$	$Y_K = 0$	Totaler
$Y_M = 1$	m_{11z}^*	m_{10z}^*	m_{1+z}^*
$Y_M = 0$	m_{01z}^*	m_{00z}^*	m_{0+z}^*
Totaler	m_{+1z}^*	m_{+0z}^*	m_z

Imputeringsverdiene er bestemt ved:

$$\frac{m_{ijz}^*}{m_z} = \frac{n_{ijz}}{n_z} \Rightarrow m_{ijz}^* = m_z \frac{n_{ijz}}{n_z}, \quad i, j = 0, 1.$$

Imputeringsverdiene ble avrundet til et helt tall slik at summen i hvert etterstratum stemmer med antallet m_z .

Basert på de imputerte data får vi et nytt datasett:

	$Y_K = 1$	$Y_K = 0$	Totaler
$Y_M = 1$	$m_{11z}^* + n_{11z}$	$m_{10z}^* + n_{10z}$	$m_{1+z}^* + n_{1+z}$
$Y_M = 0$	$m_{01z}^* + n_{01z}$	$m_{00z}^* + n_{00z}$	$m_{0+z}^* + n_{00z}$
Totaler	$m_{+1z}^* + n_{+1z}$	$m_{+0z}^* + n_{+0z}$	$m_z + n_z$

Vi ser at

$$m_{ijz}^* + n_{ijz} = n_{ijz} \frac{m_z + n_z}{n_z}$$

Vi foretar samme analyser som i kapitlene 3 og 6 med data supplert med de imputerte verdiene. Tabell 12 viser oddsforholdene for alle etterstrata med 95% konfidensintervall. Oddsforholdet varierer lite fra de basert på svarutvalget. Konfidensintervaller har blitt for korte for data supplert med de imputerte verdiene sammenlignet med Tabell 2.

Tabell 12. Estimert oddsforholdet med 95% konfidensintervall (supplerte data).

Etterstrata	Y_M	Y_K		Andel av kvinnens syssels. gitt mannens sysselsetting	Oddsforhold	95% konfidensintervall (Oddsforhold)	
		Ja 1	Nei 0				
0	Ja 1	211	288	0,423			
	Nei 0	0	4	0			
1	Ja 1	21	31	0,404	2,88	1.11	7.43
	Nei 0	8	34	0,190			
2	Ja 1	14	34	0,292	5,80	2.76	12.19
	Nei 0	25	352	0,066			
3	Ja 1	107	326	0,247	3,61	1.09	12.01
	Nei 0	3	33	0,083			
4	Ja 1	97	2	0,979	1,21	0.11	13.75
	Nei 0	40	1	0,976			
5	Ja 1	93	5	0,949	1,49	0.53	4.14
	Nei 0	213	17	0,926			
6	Ja 1	1410	55	0,962	1,28	0,17	9,72
	Nei 0	20	1	0,952			
7	Ja 1	1559	52	0,968	7,66	3,83	15,28
	Nei 0	47	12	0,797			

En Cochran-Mantel-Haenszel (*CMH*) statistikk summerer avviket fra nullhypotesen at mannens og kvinnens sysselsetting er betinget uavhengige gitt Z . *CMH* er større enn i kapittel 3. *CMH*=47,66 med $DF=1$ tyder på betinget avhengighet med stor grad av sikkerhet ($P<.0001$ med antall observasjoner $n=5\ 115$).

Tabell 13 viser kjikvadratstatistikk for alle etterstrata. X^2 er størst i grupper 2 og 7 og større enn i samme grupper i svarutvalget. X^2 er minst i grupper 4 og 6 og mindre enn i svarutvalget.

Tabell 13. *Kjikkvadratstatistikk.*

Etterstrata	Antall observ.	Supplerte data			Svarutvalget		
		Kjikkvadratstatistikk			Antall observ.	Kjikkvadratstatistikk	
		DF	Verdi	Prob.		Verdi	Prob.
0	503	1	2,9136	0,0878			
1	94	1	4,9583	0,0260	70	3,7637	0,0524
2	425	1	25,9447	<,0001	362	22,6643	<,0001
3	469	1	4,9659	0,0259	431	4,0974	0,0429
4	140	1	0,0243	0,8762	130	0,0250	0,8744
5	328	1	0,5755	0,4481	302	0,9835	0,3213
6	1486	1	0,0580	0,8097	1363	0,1238	0,7249
7	1670	1	45,2166	<,0001	1539	41,5114	<,0001

Resultatene fra Cochran-Mantel-Haenszel (*CMH*) statistikk, Breslow-Day statistikk og Mantel-Haenzel estimat av felles verdi er gitt i Tabell 14. Som vi ser gir supplerte data sterkere tester enn svarutvalget.

Tabell 14. *CMH, Breslow-Day and Mantel-Haenzel tester.*

	Svarutvalget			Supplerte data		
	Verdi	P-verdi	n	Verdi	P-verdi	n
CMH test	40,52	<.0001	4197	47,66	<.0001	5115
Breslow-Day test	Verdi	P-verdi	df	Verdi	P-verdi	df
	9,39	0,15	6	12,32	0,09	7
Mantel-Haenzel estimat	Verdi	95% konfidensintervall		Verdi	95% konfidensintervall	
	3,57	2,3	5,5	3,61	2,4	5,4

Konklusjon: Alle tester som er utført med data supplert med de imputerte verdiene forsterker tilsynelatende konklusjonene som er gjort i kapittel 3. Imidlertid, det er kapittel 3 som gir oss de riktige analyser. Med imputering øker vi utvalgsstørrelsen og betrakter de imputerte data som faktiske data, uten å ta hensyn til imputeringsusikkerhet. Derfor blir analysene for presise med imputerte data, dvs. sterkere tester, kortere konfidensintervaller etc. Denne øvelsen er en god illustrasjon på feilbruk av imputering. Man kan ikke konstruere ny data og foreta statistiske analyser uten å ha mål på usikkerheten i selve de imputerte verdier som estimater for de sanne verdiene i frafallsgruppen (Bjørnstad, 2000).

For å estimere det totale antall sysselsatte blant ektepar i Norge basert på supplerte data danner vi etterstratifiseringsestimatorene, \hat{t}_{est}^s . Estimert antall personer blant ektefeller som var faktisk sysselsatt fordeler seg på hvert etterstratum som vist i Tabell 15 og det etterstratifiserte estimatet blir $\hat{t}_{est}^s = 1\ 288\ 696$. Noter at verdiene \hat{t}_{est} og \hat{t}_{est}^s er tilnærmet like og differansen skyldes avrundingen foretatt på de imputerte verdiene.

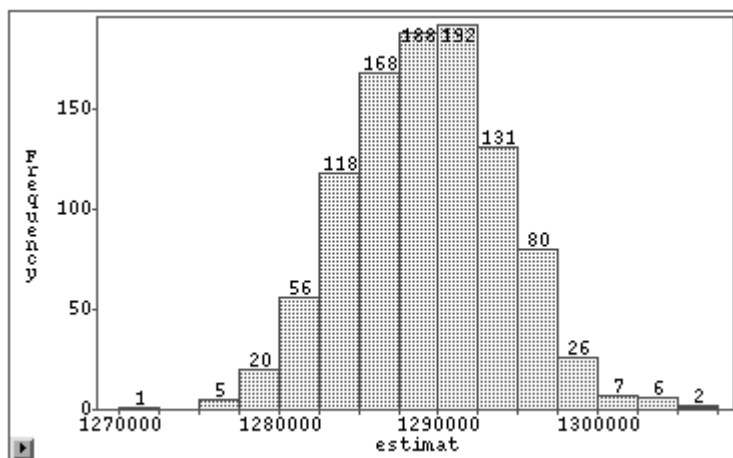
Tabell 15. Estimert antall sysselsatte basert på data supplert med de imputerte verdiene.

Etterstrata	Populasjonsstørrelse N_z	Andel faktisk sysselsatte \bar{y}_z	$N_z \bar{y}_z$
0	183 548	0,7058	129 548,17
1	40 476	0,4309	17 441,11
2	176 494	0,1024	18 072,99
3	138 914	0,5789	80 417,31
4	55 352	0,8429	46 656,20
5	103 702	0,6159	63 870,06
6	514 340	0,9741	501 018,59
7	448 164	0,9632	431 671,56
Total	1 660 990		1 288 695,9

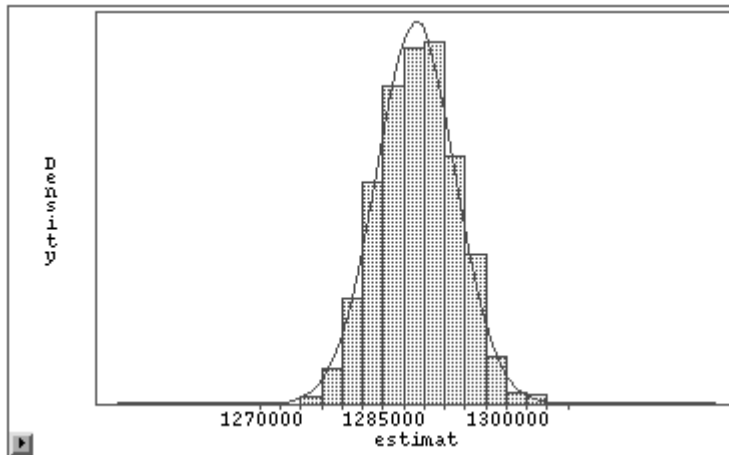
Variansen til etterstratifiseringsestimatorene beregnes som før basert på Bootstrap. Denne gangen foretas imputeringen separat for hvert av 1000 bootstraputvalg. På den måten unngår vi avrunding og skjevhet som kan oppstå pga. den. Bootstrap variansestimatorene til \hat{t}_{est}^s blir 4 919² (Tabell 16).

Tabell 16. Oppsummering av bootstrapresultatene (supplerte data).

N	Mean	Std Dev	Minimum	Maximum
1000	1 289 233,09	4918,97	1 270 943,15	1 307 457,31



Figur 5. Histogram over etterstratifiseringsestimatorer basert på 1000 bootstraputvalg (supplerte data).



Figur 6. Histogram over etterstratifiseringsestimater med tilpasset normalkurve (supplerte data).

Det er en liten forskjell mellom bootstrap variansestimatorene basert på svarutvalget og på supplerte data. Grunnen til det er at det var trukket ut forskjellige bootstraputvalg for hver analyse. Hvis vi trekker ut 1000 utvalg og regner ut etterstratifiseringsestimator både for svarutvalget og for data supplert med de imputerte verdiene for hvert utvalg, får vi $\hat{t}_{est}^S \equiv \hat{t}_{est}$, derfor samme varians. Resultater fra den øvelsen er rapportert i Tabell 17.

Tabell 17. Bootstrapestimatoren til varians (svarutvalget og supplerte data).

Variabel	N	Mean	Std Dev	Minimum	Maksimum
Svarutvalget	1000	1 289 211,78	5 126,99	1271379,61	1305977,40
Supplerte data	1000	1 289 211,78	5 126,99	1271379,61	1305977,40

8. Oppsummering

I dette notatet har vi studert sammenheng mellom ektefellers/samboeres sysselsetting i en familie basert på AKU tall. Analysene ble foretatt både med hensyn til registersyssselsetting og alder. Oddsforholdet, Pearson kji-kvadratstatistikk og korrelasjonsanalyse indikerer en sterk sammenheng mellom ektefellers sysselsetting blant godt voksne mennesker og svak sammenheng blant par som er i "karrierealderen" (25-44 år).

Også det totale antall sysselsatte blant ektepar og samboer med felles barn i Norge ble estimert basert på etterstratifiseringsestimatoren. Variansen til etterstratifiseringsestimatoren var beregnet ved hjelp av Bootstrapmetoden. Det viste seg at en bedre etterstratifiseringsestimatoren oppnås ved å etterstratifisere etter både alder og registersyssselsetting.

Frafalljustering ved imputering ble foretatt og samme analyser var gjennomført med data supplert med de imputerte verdiene. Supplerte data ga oss for presise analyser, som illustrerer at man ikke kan betrakte de imputerte verdiene som faktiske data, uten å ta hensyn til imputeringsusikkerhet.

9. Referanser

Agresti, Alan (1996). An Introduction to Categorical Data Analysis. John Wiley & Sons, INC.

Bjørnstad, Jan F. (2000). En Innføring I Utvalgsundersøkelser. Notater, 2000/50.

Rao, J.N.K. og Wu, C.F.J. (1988). Resampling Inference with Complex Survey Data. Journal of the American Statistical Association. March 1988, Vol.83, s. 231-241.

Solheim, Leiv og Zhang, Li Chun (2001). AKU: Utvalgsplanen, utvalget og husholdninger. Upublisert notat.

Vedlegg A

Partielle tabeller for 7 grupper (basert på alder og registersyssetting)

Frequency - faktisk antall;
Expected - forventet antall gitt uavhengighet;
Percent - prosentdelen av totalen;
Row Pct - prosentdelen i rad;
Col Pct - prosentdelen i kolonne

Gruppe 1: Kvinne 16-44, regsyss. 0. Mann 16-74, regsyss. 0.

	faktsysss_m	faktsysss_kv	
Frequency			
Expected			
Percent			
Row Pct			
Col Pct			
	0	1	Total
0	25	6	31
21.257	9.7429		
35.71	8.57		44.29
80.65	19.35		
52.08	27.27		
1	23	16	39
26.743	12.257		
32.86	22.86		55.71
58.97	41.03		
47.92	72.73		
Total	48	22	70
68.57	31.43		100.00

Gruppe 2: Kvinne 45-74, regsyss. 0. Mann 25-74, regsyss. 0.

	faktsysss_m	faktsysss_kv	
Frequency			
Expected			
Percent			
Row Pct			
Col Pct			
	0	1	Total
0	300	21	321
291.74	29.262		
82.87	5.80		88.67
93.46	6.54		
91.19	63.64		
1	29	12	41
37.262	3.7376		
8.01	3.31		11.33
70.73	29.27		
8.81	36.36		
Total	329	33	362
90.88	9.12		100.00

Gruppe 3: Kvinne 45-74, regsyss. 0. Mann 25-74, regsyss. 0.

		faktsysss_kv		
		faktsysss_m		
Frequency	Expected			
Percent	Percent			
Row Pct	Row Pct			
Col Pct	Col Pct	0	1	Total
0		30	3	33
	25.267	7.7332		
	6.96	0.70		7.66
	90.91	9.09		
	9.09	2.97		
1		300	98	398
	304.73	93.267		
	69.61	22.74		92.34
	75.38	24.62		
	90.91	97.03		
Total		330	101	431
	76.57	23.43		100.00

Gruppe 4: Kvinne 16-44, regsyss. 1. Mann 16-74, regsyss. 0.

		faktsysss_kv		
		faktsysss_m		
Frequency	Expected			
Percent	Percent			
Row Pct	Row Pct			
Col Pct	Col Pct	0	1	Total
0		1	37	38
	0.8769	37.123		
	0.77	28.46		29.23
	2.63	97.37		
	33.33	29.13		
1		2	90	92
	2.1231	89.877		
	1.54	69.23		70.77
	2.17	97.83		
	66.67	70.87		
Total		3	127	130
	2.31	97.69		100.00

Gruppe 5: Kvinne 45-74, regsyss. 1. Mann 25-74, regsyss. 0.

		faktsysss_kv		
		faktsysss_m		
Frequency	Expected			
Percent	Percent			
Row Pct	Row Pct			
Col Pct	Col Pct	0	1	Total
0		16	196	212
	14.04	197.96		
	5.30	64.90		70.20
	7.55	92.45		
	80.00	69.50		
1		4	86	90
	5.9603	84.04		
	1.32	28.48		29.80
	4.44	95.56		
	20.00	30.50		
Total		20	282	302
	6.62	93.38		100.00

Gruppe 6: Kvinne 16-44, regsys. 1. Mann 16-74, regsys. 1.

		faktsyss_kv		
		faktsyss_m		
Frequency	Expected			
Percent	Percent			
Row Pct	Row Pct			
Col Pct	Col Pct	0	1	Total
0	1	18	19	
0.7109	18.289			
0.07	1.32		1.39	
5.26	94.74			
1.96	1.37			
1	50	1294	1344	
50.289	1293.7			
3.67	94.94		98.61	
3.72	96.28			
98.04	98.63			
Total	51	1312	1363	
	3.74	96.26	100.00	

Gruppe 7: Kvinne 45-74, regsys. 1. Mann 25-74, regsys. 1.

		faktsyss_kv		
		faktsyss_m		
Frequency	Expected			
Percent	Percent			
Row Pct	Row Pct			
Col Pct	Col Pct	0	1	Total
0	11	43	54	
2.0702	51.93			
0.71	2.79		3.51	
20.37	79.63			
18.64	2.91			
1	48	1437	1485	
56.93	1428.1			
3.12	93.37		96.49	
3.23	96.77			
81.36	97.09			
Total	59	1480	1539	
	3.83	96.17	100.00	

Vedlegg B

Partielle tabeller for 7 grupper (basert på alder)

Gruppe 1: Kvinne 16-44, mann 16-24

		faktsyssh_kv		
		faktsyssh_m		
Frequency	Expected	0	1	Total
Percent	Percent			
Row Pct	Row Pct			
Col Pct	Col Pct			
0	1	1	1	2
0.4348	1.5652			
4.35	4.35			8.70
50.00	50.00			
20.00	5.56			
1	4	17	21	
4.5652	16.435			
17.39	73.91			91.30
19.05	80.95			
80.00	94.44			
Total	5	18	23	
	21.74	78.26	100.00	

Gruppe 2: Kvinne 16-24, mann 25-44

		faktsyssh_kv		
		faktsyssh_m		
Frequency	Expected	0	1	Total
Percent	Percent			
Row Pct	Row Pct			
Col Pct	Col Pct			
0	3	2	5	
2.0968	2.9032			
9.68	6.45			16.13
60.00	40.00			
23.08	11.11			
1	10	16	26	
10.903	15.097			
32.26	51.61			83.87
38.46	61.54			
76.92	88.89			
Total	13	18	31	
	41.94	58.06	100.00	

Gruppe 3: Kvinne 25-44, mann 25-44.

		faktsyssh_kv		
		faktsyssh_m		
Frequency	Expected	0	1	Total
Percent	Percent			
Row Pct	Row Pct			
Col Pct	Col Pct			
0	11	36	47	
8.3717	38.628			
0.77	2.52			3.30
23.40	76.60			
4.33	3.07			
1	243	1136	1379	
245.63	1133.4			
17.04	79.66			96.70
17.62	82.38			
95.67	96.93			
Total	254	1172	1426	
	17.81	82.19	100.00	

Gruppe 4: Kvinne 45-74, mann 25-44.

```
faktsyss_m      faktsyss_kv
```

Frequency			
Expected			
Percent			
Row Pct			
Col Pct	0	1	Total
0	4	12	16
	3.6923	12.308	
	2.80	8.39	11.19
	25.00	75.00	
	12.12	10.91	
1	29	98	127
	29.308	97.692	
	20.28	68.53	88.81
	22.83	77.17	
	87.88	89.09	
Total	33	110	143
	23.08	76.92	100.00

Gruppe 5: Kvinne 16-24, mann 45-74.

```
faktsyss_m      faktsyss_kv
```

Frequency			
Expected			
Percent			
Row Pct			
Col Pct	0	1	Total
0	4	2	6
	1.7692	4.2308	
	5.13	2.56	7.69
	66.67	33.33	
	17.39	3.64	
1	19	53	72
	21.231	50.769	
	24.36	67.95	92.31
	26.39	73.61	
	82.61	96.36	
Total	23	55	78
	29.49	70.51	100.00

Gruppe 6: Kvinne 25-44, mann 45-74.

```
faktsyss_m      faktsyss_kv
```

Frequency			
Expected			
Percent			
Row Pct			
Col Pct	0	1	Total
0	11	20	31
	4.8312	26.169	
	2.38	4.33	6.71
	35.48	64.52	
	15.28	5.13	
1	61	370	431
	67.169	363.83	
	13.20	80.09	93.29
	14.15	85.85	
	84.72	94.87	
Total	72	390	462
	15.58	84.42	100.00

Gruppe 7: Kvinne 45-74, mann 45-74.

faktsyssh_m faktsyssh_kv

Frequency			
Expected			
Percent			
Row Pct			
Col Pct	0	1	Total
0	353	251	604
	170.94	433.06	
	14.17	10.08	24.25
	58.44	41.56	
	50.07	14.05	
1	352	1535	1887
	534.06	1352.9	
	14.13	61.62	75.75
	18.65	81.35	
	49.93	85.95	
Total	705	1786	2491
	28.30	71.70	100.00

De sist utgitte publikasjonene i serien Notater

- 2002/9 T.M. Normann: Rekruttering til erfaringskonferanse og undersøkelse om røykevaner blant kvinner i alderen 25-45 år. Dokumentasjonsrapport. 16s.
- 2002/10 J. Holmøy: GERIX 1995-1999. Dokumentasjon, system, data, program. 47s.
- 2002/11 T.M. Normann: Underøking om det lokale sjølvstyret. Dokumentasjonsrapport. 81S.
- 2002/12 L.S. Stambøl: Regionale framskrivinger av sysselsetting og bruttoprodukt ved hjelp av SSBs modellsystem REGARD. Regionale framskrivinger basert på nasjonale anslag med modellene MODAG (1997-2005) og MSG (1995-2020). 35s.
- 2002/13 H. Madsen og L. S. Stambøl: Kontrafaktiske regionale beregninger ved hjelp av SSBs modellsystem REGARD. Regionale beregninger basert på historiske tall på nasjonalt nivå kjørt bakover i tid på grunnlag av modellens basisår (her 1995). 55s.
- 2002/14 V. Hansen og H. Madsen: Månedlig og kvartalsvis elektrisitetsstatistikk. Dokumentasjon av produksjonsrutiner og systembeskrivelse. 41s.
- 2002/15 A. Rolland: Søkelys på det gode liv. 37s.
- 2002/17 D.Rønningen og D. Fredriksen: Beskatningen av pensjonister. 41s.
- 2002/18 D. Rønningen: Overganger fra arbeidsmarkedet til trygd. En litteraturoversikt. 34s.
- 2002/19 F. Gundersen og L. Solheim: Regionalisering av FoU-statistikken. 43.
- 2002/20 L. Vågane: Omnibusundersøkelsen november/desember 2001. Dokumentasjonsrapport. 56s.
- 2002/21 G. Claus, O. Haugen P. M. Holt og E. Knutsen: Regnskapsstatistikk. Næringsoppgaver for ikke-finansielle aksjeselskaper, 1999. Dokumentasjon. 34s.
- 2002/22 M. Takle: Befolkningsstatistikk på rute-nett. Dokumentasjon. 35s
- 2002/23 D. Roll-Hansen, S. Ferstad, M. Stålnacke, P. Tuhus og R. Nøtnæs: En spørreskjemametodisk gjennomgang av datainnsamling gjennom Grunnskolens informasjonssystem (GSI). 109s.
- 2002/24 T. P. Bøe og I. Håland: Dokumentasjon av arbeidskraftundersøkelsen (AKU). 85s.
- 2002/25 A. Akselsen og T. Sandnes: FD - Trygd: Dokumentasjonsrapport. Stønader til enslig forsørger. 1992-2000. 46s.
- 2002/26 E. Rønning: Statistisk sentralbyrå tidsbruksundersøkelse 2000/01. Dokumentasjon og resultater fra intervjuet. 125s.
- 2002/27 S. Myro og C. Torp: Stedsfesting av bedrifter i Bedrifts- og foretaksregisteret. Hovedprosjekt. 37s.
- 2002/28 C. Nordseth og T. Sandnes: FD - Trygd: Dokumentasjonsrapport. Foreløpig uførestønad. 1992-2000. 37s.
- 2002/29 S. Derakhshanfar og T. Sandnes: FD - Trygd: Dokumentasjonsrapport. Økonomisk sosialhjelp. 1992-2000. 36s.
- 2002/30 I. Johansen: Undersøkning om foreldrebetaling i barnehagar, januar 2002. 42s.
- 2002/31 T.M. Køber, H. Moafi, E. Rønning og Ø. Sivertstøl: Bruk av forløpsdatabaser i Statistisk sentralbyrå. 60s.
- 2002/32 T.M. Normann: Omnibusundersøkelsen februar/mars 2002. Dokumentasjonsrapport. 37s.
- 2002/33 S. Reid: Bosettingskriteriene i inntektssystemet til kommunene. Erfaringer med overgang til ny beregningsmåte og nye bosettingskriterier, 2002.