

Interne notater

STATISTISK SENTRALBYRÅ

80/24

4. september 1980

SAMMENLIKNING AV INNTEKTSTALL OPPGITT I FRUKTBARHETSUNDERSØKELSEN 1977 MED INNTEKTSTALL FRA SKATTEREGISTERET

Av
Erling Siring

INNHold

	Side
1. Innledning	1
2. Oppsummering av resultatene	1
3. Sammenlikning mellom gjennomsnitt innen delgrupper	2
4. En konsistenstabell	4
5. Spredningen til avvikene i forskjellige delgrupper	6
6. Samvariasjonen mellom inntekt og andre variable	10
Vedlegg 1:	
En formell analyse av tabell 4.1	14
Vedlegg 2:	
Utklipp fra instruks og spørreskjema	15
Referanseliste	16

*) Jeg vil takke Rolf Aaberge som har gitt nyttige kommentarer under arbeidet med manuskriptet.

1. INNLEDNING

I Byråets intervjuundersøkelser ønsker en ofte inntektsopplysninger. En står da overfor valget mellom å spørre etter inntekt eller hente den fra skatteregisteret.

Fordeler og ulemper ved å hente inntektstallene fra skatteregisteret framfor å spørre etter dem skal ikke behandles i dette notatet. Derimot skal vi foreta en sammenlikning mellom inntektstall som ble oppgitt i en intervjuundersøkelse, og inntektstall som er registrert i skatteregisteret for de intervjuede personene. Sammenlikningen er mulig gjort ved at vi har koblet utvalget i Fruktbarhetsundersøkelsen 1977 (heretter forkortet til FU77) med skatteregisteret.

I FU77, der bare kvinner ble intervjuet, ble det spurt om hvor stor nettoinntekt intervjuobjektene (IO-ene) hadde i 1976 før skatt ble trukket fra. De gifte og samboende ble også bedt om å oppgi nettoinntekt for sine samlivspartnere. I instruksen til intervjuerne ble nettoinntekt definert som det beløp en kommer fram til på selvangivelsen etter at fradragspostene er trukket fra. Vedlegg 2 viser utdrag fra instruksen til intervjuerne og fra spørreskjemaet.

Ved koblingen med skatteregisteret skaffet vi oss foruten inntektsopplysninger om IO-ene i FU77, også inntektsopplysninger om ektemennene til de av IO-ene som var gift. Som inntektsbegrep ved koblingen brukte vi "nettoinntekt stat" for de som ikke var sjøfolk, og summen av "nettoinntekt stat" og "inntekt ombord" for de som var sjøfolk.

Til FU77 ble det trukket et utvalg på 5 047 personer, og oppnådd intervju med 4 137 av dem. Av de som ble intervjuet, var det 3 458 personer som vi fikk registrert husholdningsinntekten til både ved intervju og ved "å hente" opplysninger fra skatteregisteret. Disse personene, som utgjør tabellmassen i notatet, kaller vi for nettoutvalget. I noen av tabellene i notatet er ikke hele nettoutvalget tatt med på grunn av at vi har manglet relevante opplysninger om noen av IO-ene.

I dette notatet bruker vi ikke ordet "husholdningsinntekt" i dets helt rette betydning. Vi definerer her husholdningsinntekt som summen av egeninntekt og ektemannens inntekt for de gifte kvinnene, og bare egeninntekt for de ugifte og tidligere gifte kvinnene. I alle tabellene i dette notatet blir inntekt fra FU77 betegnet med IFU, og inntekt fra skatteregisteret med ISR.

Dette arbeidet er bare et forarbeid til et mer omfattende målefeilstudium, og det blir derfor ikke presentert noen målefeilsmodell i dette notatet. En målefeilsmodell vil bli presentert i en senere publikasjon sammen med resultatene fra kvalitetsstudier av andre variable.

2. OPPSUMMERING AV RESULTATENE

I gjennomsnitt for alle i nettoutvalget i FU77 ligger husholdningsinntekten som ble oppgitt i undersøkelsen, ca 11 prosent høyere enn husholdningsinntekten fra skatteregisteret.

I kapittel 3 sammenlikner vi gjennomsnitt av inntektstallene fra FU77 med gjennomsnitt av inntekts-tallene fra skatteregisteret innen delgrupper. Differansen mellom husholdningsinntekt oppgitt i FU77 og husholdningsinntekt fra skatteregisteret er i gjennomsnitt mye større for gifte enn for ugifte og tidligere gifte kvinner. Tabell 3.2 viser at dette har sammenheng med at de gifte kvinnene i gjennomsnitt oppga et altfor høyt tall for ektemannens nettoinntekt i FU77.

Resultatene i kapittel 3 tyder på at alder, inntektsnivå og utdanning har liten effekt på størrelsen til differansen mellom gjennomsnitt av husholdningsinntekter fra FU77 og husholdningsinntekter fra skatteregisteret. Derimot viser resultatene at den nevnte differansen øker sterkt med bostørrelsen for de gifte kvinnene. Dette mener vi er en indikasjon på at mange av IO-ene oppga bruttoinntekt i stedet for nettoinntekt i FU77.

I kapittel 4 presenteres resultater som viser at selv om tallene fra FU77 i gjennomsnitt var høyere enn tallene fra skatteregisteret, forekom store avvik mellom inntektstallene relativt sjeldent.

I kapittel 5 gis resultater som viser at størrelsen til differansen mellom husholdningsinntekt fra FU77 og husholdningsinntekt fra skatteregisteret varierer mye mer blant de gifte kvinnene enn blant de ugifte og tidligere gifte kvinnene. Årsakene til dette mener vi er at mange av de gifte kvinnene hadde problemer med å oppgi et nøyaktig tall for ektemannens nettoinntekt, og at differansen mellom brutto- og netto-husholdningsinntekt varierer mer i størrelse blant ektepar enn blant enslige.

Ellers viser tabellene i kapittel 5 at forekomsten av store avvik mellom de to typene tall for husholdningsinntekt øker litt med alderen til 10 både blant gifte og ikke-gifte. Avvikenes fordeling på pluss- og minussiden varierer imidlertid lite med alderen.

I kapittel 6 har vi prøvd å finne ut om det blir noen forskjell om vi bruker inntektstall fra FU77 eller inntektstall fra skatteregisteret når vi skal analysere samvariasjonen mellom inntekt og andre variable. Resultatene i kapitlet tyder på at det har liten betydning om en bruker inntektstall registrert på den ene eller andre måten i en slik analyse.

I vedlegget har vi benyttet statistiske metoder som bekrefter at personer som oppgir inntekt som ikke samsvarer med opplysningene fra skatteregisteret, har en tendens til å oppgi høyere inntekt enn registrert i skatteregisteret.

3. SAMMENLIKNING MELLOM GJENNOMSNIITT INNEN DELGRUPPER

I FU77 oppga 10-ene i nettoutvalget i gjennomsnitt en husholdningsinntekt på 65 300 kroner for 1976, mens de samme personene i gjennomsnitt hadde en husholdningsinntekt på 58 800 kroner i følge skatteregisteret for samme året. Tabell 3.1 viser hvordan differansen mellom gjennomsnitt av inntektstall fra FU77 (IFU) og gjennomsnitt av inntektstall fra skatteregisteret (ISR) varierer med alder og ekteskapeleg status.

Tabell 3.1. Husholdningsinntekt fra skatteregisteret og differansen mellom husholdningsinntekt oppgitt i FU77 og husholdningsinntekt fra skatteregisteret i grupper for ekteskapeleg status og alder. Gjennomsnitt i 1 000 kr

Alder	Type tall	Total	Ekteskapeleg status				
			Ugift	Gift	Enke	Separert	Skilt
Alle	ISR	58,8	19,6	76,9	30,6	18,5	32,0
	IFU-ISR	6,5	2,3	8,6	-7,5	1,8	2,5
	Antall personer	3 458	917	2 335	31	57	118
18-24 år	ISR	25,1	13,0	58,5	-	13,5	16,8
	IFU-ISR	4,1	2,4	8,8	-	0,7	1,0
	Antall personer	961	682	255	-	14	10
25-34 år	ISR	67,3	36,5	74,7	23,5	15,1	31,1
	IFU-ISR	7,9	2,5	9,1	-10,6	0,8	2,7
	Antall personer	1 531	191	1 260	9	27	44
35-44 år	ISR	78,7	48,2	85,9	33,6	28,7	35,1
	IFU-ISR	6,9	1,6	7,9	-6,3	4,4	2,4
	Antall personer	966	44	820	22	16	64

Tabellen viser at i gjennomsnitt for alle i nettoutvalget er inntektstallene oppgitt i FU77 6 500 kroner, eller 11,1 prosent, høyere enn inntektstallene funnet i skatteregisteret. I gjennomsnitt er differansen mellom inntektstallet oppgitt i FU77 og inntektstallet fra skatteregisteret mye større for de gifte enn for de ugifte og tidligere gifte kvinnene. Enker skiller seg ut ved at de i gjennomsnitt oppga lavere inntektstall i FU77 enn det vi fant i skatteregisteret. Ellers viser tabell 3.1 at mens inntektsnivået øker med alderen, varierer differansen mellom gjennomsnittene av inntektstallene ved de to registreringsmåtene lite med alderen innenfor de tre største gruppene for ekteskapeleg status.

Tabell 3.1 og andre tabeller (som ikke er med i notatet) har vist oss at differansen mellom gjennomsnitt av de to typene tall for husholdningsinntekt endrer seg lite med inntektsnivået når en justerer for ekteskaphyppighet. Dette er grunnen til at vi har valgt å bruke den absolutte størrelsen til differansen mellom inntektene ved de to registreringsmåtene, i stedet for å bruke størrelsen til differansen relativt til inntektsnivået i våre studier.

Som nevnt over er differansen mellom husholdningsinntektene ved de to registreringsmåtene spesielt stor for de gifte kvinnene. I tabell 3.2 kan vi sammenlikne både gjennomsnitt av tall oppgitt som egeninntekt og gjennomsnitt av tall oppgitt som ektemenns inntekter, med tall fra skatteregisteret.

Tabell 3.2. Inntektstall fra skatteregisteret, og differansen mellom inntektstall fra FU77 og inntektstall fra skatteregisteret i grupper for alder. Gjennomsnittstall i 1 000 kr

Kvinnens alder	Gifte kvinners egeninntekter			Ektemenns inntekter		
	ISR	IFU-ISR	Antall personer	ISR	IFU-ISR	Antall personer
Alle	15,2	2,5	2 727	62,1	6,0	2 274
18-19 år	10,4	1,6	25	45,0	4,6	6
20-24 år	15,4	2,1	359	46,5	6,8	235
25-29 år	14,8	2,1	712	55,4	6,7	619
30-34 år	14,4	2,6	695	65,5	6,6	617
35-39 år	15,8	3,3	511	68,4	6,1	443
40-44 år	16,6	2,4	425	71,0	2,9	354

De gifte kvinnene oppga i gjennomsnitt 6 000 kroner mer i inntekt for ektemennene sine i FU77 enn det vi har fra skatteregisteret, mens de i gjennomsnitt oppga 2 500 kroner mer i egeninntekt i FU77 enn det de hadde oppgitt til likningsmyndighetene. Det siste tallet er omtrent av samme størrelse som det tilsvarende tallet for de ugifte.

Ellers viser tabell 3.2 at differansen mellom gjennomsnitt av inntektstall oppgitt i FU77 for ektemenn og inntektstall fra skatteregisteret er minst i den eldste aldersgruppen. Når det gjelder differansen mellom gjennomsnittene av de to typene tall for de gifte kvinnenes egeninntekt, ser vi at den øker svakt med alderen opp til nest eldste aldersgruppe.

For å se om utdanning har noen effekt på differansen mellom gjennomsnitt av inntektstallene ved de to registreringsmåtene, har vi delt utvalget opp i delgrupper etter IO-enes høyeste fullførte utdanning og alder. Tabell 3.3 viser hvordan differansen mellom de to typene tall for husholdningsinntekt varierer med utdanning.

Tabell 3.3. Husholdningsinntekt fra skatteregisteret og differansen mellom husholdningsinntekt oppgitt i FU77 og husholdningsinntekt fra skatteregisteret i grupper for utdanning og alder. Gjennomsnittstall i 1 000 kr

Alder	Type tall	Høyeste fullførte utdanning			
		Totalt	Folkeskole Framhaldskole Ungdomsskole	Gym I Artium Gym II	Universitet Høyskole
Alle	ISR	58,8	52,0	56,2	78,7
	IFU-ISR	6,5	5,7	6,6	7,6
	Antall personer	3 446	808	2 081	557
18-29 år	ISR	41,7	36,0	41,4	51,6
	IFU-ISR	5,6	5,7	5,3	7,0
	Antall personer	1 750	330	1 163	257
30-44 år	ISR	76,4	63,1	74,9	101,9
	IFU-ISR	7,5	5,7	8,3	8,2
	Antall personer	1 696	478	918	300

Andelen av gifte varierer lite med utdanning innenfor de to aldersgruppene i tabell 3.3, slik at det ikke er noen grunn til å kontrollere for ekteskapelig status i tabellen. Vi ser at differansen mellom husholdningsinntekt fra FU77 og husholdningsinntekt fra skatteregisteret i gjennomsnitt er større i kategorien for "høy" utdanning enn i kategorien for "lav" utdanning i begge aldersgruppene. Etter vår mening skyldes imidlertid ikke dette så mye forskjellen i utdanning i seg selv, som andre bakenforliggende faktorer.

Vi har en hypotese om at mange av IO-ene oppga bruttoinntekt i stedet for nettoinntekt i FU77, og at det først og fremst er størrelsen på fradragspostene i selvangivelsen som har effekt på størrelsen til differansen mellom inntekt fra FU77 og inntekt fra skatteregisteret. Ut fra denne hypotesen er resultatene i tabell 3.3 rimelige siden personer med høy utdanning vanligvis har mer gjeld enn personer med lav utdanning.

I FU77 ble IO-ene spurt om hvor store boutgifter de hadde pr. måned medregnet renter og avdrag på boliglån. For å prøve å underbygge hypotesen vår viser vi i tabell 3.4 hvordan differansen mellom gjennomsnittene av husholdningsinntektene ved de to registreringsmåtene varierer med boutgiftene for de gifte kvinnene.

Tabell 3.4. Husholdningsinntekt fra skatteregisteret og differansen mellom husholdningsinntekt oppgitt i FU77 og husholdningsinntekt fra skatteregisteret. Gifte kvinner. Gjennomsnittstall i 1 000 kr

Type tall	Boutgifter pr. mnd.							
	Totalt	0	1-399 kr	400-599 kr	600-799 kr	800-1 199 kr	1 200-1 999 kr	2 000 eller mer
ISR	76,9	73,8	61,2	67,8	74,9	77,3	76,7	94,9
IFU-ISR	8,6	-1,0	4,7	7,8	6,6	7,9	10,2	14,2
Antall personer	2 333	12	167	307	348	508	529	325

Tabell 3.4 viser at gjennomsnittet til differansen mellom inntektene ved de to registreringsmåtene øker sterkt med boutgiftene. Siden det opplagt er høy korrelasjon mellom boutgifter og gjeldsrenter, som utgjør det meste av forskjellen mellom brutto- og nettoinntekt, gir resultatene i tabell 3.4 støtte til vår hypotese om at mange av IO-ene oppga bruttoinntekt i stedet for nettoinntekt i FU77.

4. EN KONSISTENSTABELL

I kapittel 3 sammenliknet vi gjennomsnitt av inntektstallene oppgitt i FU77 med gjennomsnitt av inntektstallene registrert av likningsmyndighetene for de samme personene. Mens kapittel 3 ikke ga noen informasjon om samsvaret mellom de to typene inntektstall for enkeltpersoner, gir dette kapitlet slik informasjon ved at det presenteres en såkalt konsistenstabell.

I dette kapitlet har vi laget en inntektsgruppering. Deretter er hvert individ (IO) gruppert etter inntekten oppgitt i FU77 og inntekten registrert i skatteregisteret. En konsistenstabell er da en krysstabell med gruppekoder laget på grunnlag av den ene typen inntektstall i forspalten, og med gruppekoder laget på grunnlag av den andre typen inntektstall i hodet.

I tabell 4.1 er gruppekodene laget på grunnlag av inntektene fra skatteregisteret i forspalten, og gruppekodene laget på grunnlag av inntektene fra FU77 i hodet. Inntektsgrupperingen i tabellen er følgende:

Inntektsgruppe 1 er:	inntekter \leq 20 000
" 2 er:	20 000 < inntekter \leq 40 000
" 3 er:	40 000 < inntekter \leq 60 000
" 4 er:	60 000 < inntekter \leq 80 000
" 5 er:	80 000 < inntekter \leq 100 000
" 6 er:	100 000 < inntekter \leq 125 000
" 7 er:	125 000 < inntekter

Tabell 4.1. Kvinner etter husholdningsinntekt oppgitt i FU77 og husholdningsinntekt fra skatteregisteret. Hyppighetstall og prosenttall

ISR	IFU77								
	I alt	1	2	3	4	5	6	7	
	Hyppighetstall								
I alt	3 458	656	322	636	715	509	353	267	
1	677	<u>606</u>	47	15	4	3	-	2	
2	419	45	<u>232</u>	107	29	3	2	1	
3	726	4	40	<u>425</u>	207	36	13	1	
4	707	1	2	74	<u>411</u>	170	42	7	
5	445	-	-	10	55	<u>238</u>	113	29	
6	310	-	1	3	8	52	<u>156</u>	90	
7	174	-	-	2	1	7	27	<u>137</u>	
A Antall i klassen i FU77	3 458	656	322	636	715	509	353	267	
B Antall i klassen etter ISR	3 458	677	419	726	707	445	310	174	
C Antall likt klassifisert	2 205	606	232	425	411	238	156	137	
	Prosenttall								
I alt	100,0	19,0	9,3	18,4	20,7	14,7	10,2	7,7	
1	19,6	<u>17,5</u>	1,4	0,4	0,1	0,1	-	0,1	
2	12,1	1,3	<u>6,7</u>	3,1	0,8	0,1	0,1	-	
3	21,0	0,1	1,2	<u>12,3</u>	6,0	1,0	0,4	-	
4	20,4	-	0,1	2,1	<u>11,9</u>	4,9	1,2	0,2	
5	12,9	-	-	0,3	1,6	<u>6,9</u>	3,3	0,8	
6	9,0	-	-	0,1	0,2	1,5	<u>4,5</u>	2,6	
7	5,0	-	-	0,1	-	0,2	0,8	<u>4,0</u>	
A/N = Andel i klassen i FU77	100,0	19,0	9,3	18,4	20,7	14,7	10,2	7,7	
B/N = Andel i klassen etter ISR	100,0	19,6	12,1	21,0	20,4	12,9	9,0	5,0	
A/N - B/N = Nettoavvik	-0,6	-2,8	-2,6	0,3	1,8	1,2	2,7	
C/A = Andel likt klassifisert	63,8	92,4	72,0	66,8	57,5	46,8	44,2	51,3	
$\frac{A-B}{A}$ = Relativt nettoavvik	-3,2	-30,1	-14,2	1,1	12,6	12,2	34,8	

Tabell 4.1 viser at marginalfordelingene med hensyn til inntekt fra FU77 og inntekt fra skatteregisteret avviker litt fra hverandre. Det er flere som blir klassifisert til de høyeste inntektsgruppene når en bruker tallene fra FU77, enn når en bruker tallene fra skatteregisteret. Fra konsistenstabellen ser vi at det først og fremst er de personer som er klassifisert til en inntektsgruppe høyere etter tallene fra FU77 enn etter tallene fra skatteregisteret, som forårsaker forskjellen mellom marginalfordelingene.

En desto sterkere konsentrasjon av tabellmassen omkring hoveddiagonalen i tabell 4.1, betyr at det er et desto bedre samsvar mellom de to typene inntektstall. La X_{ij} betegne antall IO-er med registerinntekt i inntektsgruppe i og inntekt oppgitt i FU77 i inntektsgruppe j (celle (i,j) i tabellen), og la N betegne antallet i nettoutvalget. Når en skal vurdere samsvaret mellom de to typene inntektstall, gir følgende to estimerte mål nyttig informasjon

$$S_1 = \frac{1}{N} \sum_{i=1}^7 X_{ii} = 0,638$$

$$S_2 = \frac{1}{N} \sum_{i=1}^7 (X_{ii-1} + X_{ii} + X_{ii+1}) = 0,936 \quad (X_{10} = X_{78} = 0)$$

S_1 forteller at 63,8 prosent av IO-ene i nettoutvalget oppga en inntekt som lå i samme inntektsgruppe som registrert i skatteregisteret. S_1 er identisk med "Andel likt klassifisert" på nest nederste linje i kolonnen "I alt" i tabell 4.1. S_2 forteller at 93,6 prosent av IO-ene i nettoutvalget oppga en inntekt som lå i enten en gruppe lavere, samme gruppe, eller en gruppe høyere enn inntekten registrert i skatteregisteret. S_1 og S_2 antyder at samsvaret mellom de to typene inntektstall er sterkt, og følgelig at korrelasjonen mellom de to typene inntektstall er høy.

S_1 og S_2 vil variere med måten en grupperer inntektene på, slik at de bare kan brukes som grove mål for samsvar og korrelasjon. Et mål for korrelasjon anvendt på rådata er den vanlige empiriske korrelasjonskoeffisienten. Den er beregnet til å være 0,915 mellom de to typene inntektstall.

Figur 4.1 er et diagram der IO-ene er plottet inn etter de to typene tall for husholdningsinntekt (regnet i 100 kroner). Vertikalaksen i figur 4.1 viser inntektsbeløpene oppgitt i FU77, mens horisontalaksen viser inntektsbeløpene fra skatteregisteret. De få IO-ene med inntektstall høyere enn 200 000 kroner er av praktiske grunner utelatt fra diagrammet. Figuren illustrerer tydelig det som tabell 4.1 antyder, nemlig at de aller fleste av observasjonene ligger på et belte omkring diagonalen med beltets midtlinje noe over diagonalen.

I vedlegget vurderer vi om den tendensen til å oppgi høyere inntekt i FU77 enn til likningsmyndighetene som tabell 4.1 antyder, er signifikant. Resultatet av den formelle analysen bekrefter den ovenfor nevnte tendensen.

5. SPREDNINGEN TIL AVVIKENE I FORSKJELLIGE DELGRUPPER

I dette kapitlet skal vi studere hvordan spredningen til differansen mellom husholdningsinntekt oppgitt i FU77 og husholdningsinntekt fra skatteregisteret varierer med forskjellige variable.

Vi har delt IO-ene i nettoutvalget inn i grupper etter størrelsen på differansen mellom de to typene tall for husholdningsinntekt. La D betegne differansen mellom inntekt oppgitt i FU77 og inntekt fra skatteregisteret. Gruppeinndelingen er følgende:

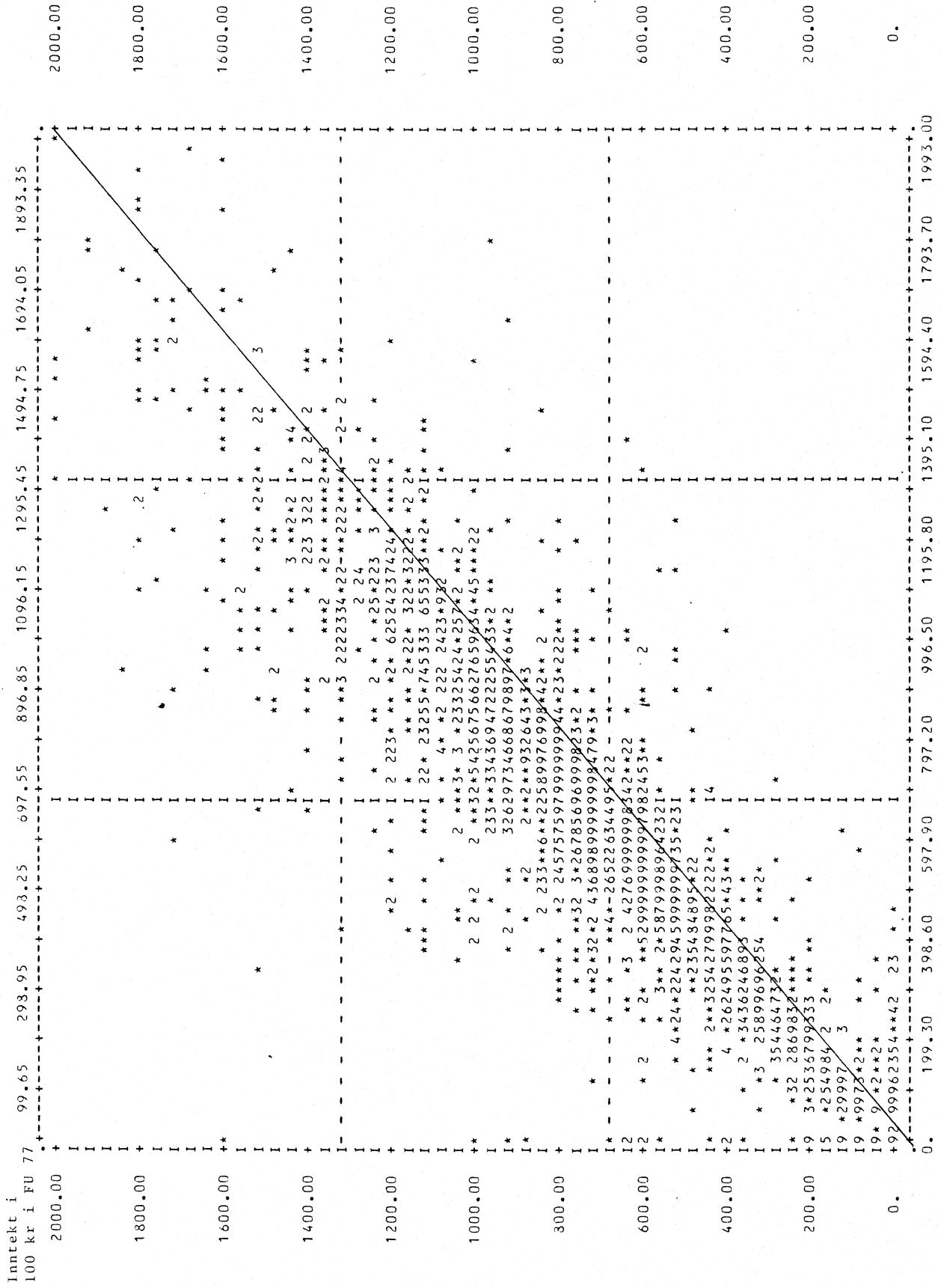
Gruppe 1:	IO-er med	$D \leq$	-20 000
Gruppe 2:	IO-er med	$-20\ 000 < D \leq$	-15 000
Gruppe 3:	IO-er med	$-15\ 000 < D \leq$	-10 000
Gruppe 4:	IO-er med	$-10\ 000 < D \leq$	-5 000
Gruppe 5:	IO-er med	$-5\ 000 < D \leq$	5 000
Gruppe 6:	IO-er med	$5\ 000 < D \leq$	10 000
Gruppe 7:	IO-er med	$10\ 000 < D \leq$	15 000
Gruppe 8:	IO-er med	$15\ 000 < D \leq$	20 000
Gruppe 9:	IO-er med	$20\ 000 < D \leq$	25 000
Gruppe 10:	IO-er med	$25\ 000 < D \leq$	30 000
Gruppe 11:	IO-er med	$D >$	30 000

Tabell 5.1 viser hvordan spredningen til differansen mellom inntektene ved de to registreringsmåtene varierer med ekteskapelig status.

Med "Uoppgitt i prosent" i tabellene i dette kapitlet menes prosentandelen av de som ble intervjuet i FU77, som vi ikke har husholdningsinntekten til ved begge registreringsmåtene. Tabellen viser at denne prosentandelen er 16,5 for alle som ble intervjuet. Det kan nevnes at det var 13,6 prosent av alle som ble intervjuet som ikke oppga husholdningsinntekten i undersøkelsen.

Tabell 5.1 viser at 41,4 prosent av IO-ene i nettoutvalget oppga en inntekt som avvek mindre enn 5 000 kroner fra inntekten som vi fant i skatteregisteret. 2,8 prosent av IO-ene oppga en inntekt som var minst 20 000 kroner mindre enn inntekten fra skatteregisteret, mens 6,9 prosent oppga en inntekt som var mer enn 30 000 kroner større enn inntekten fra skatteregisteret.

Figur 4.1. Samvariasjon mellom husholdningsinntekt i skatteregisteret og FU 77



Inntekt i 100 kroner i skatteregisteret

Tabell 5.1. Kvinner i ulike grupper for ekteskkelig status, etter differansen mellom husholdningsinntekt oppgitt i FU77 og husholdningsinntekt fra skatteregisteret. Prosent

IFU-ISR	Alle	Ekteskkelig status		
		Ugift	Gift	Før gift
I alt	100,0	100,0	100,0	100,0
1	2,8	1,0	3,2	6,3
2	1,8	1,0	1,9	3,4
3	3,3	3,2	3,4	2,4
4	5,6	6,1	5,4	5,8
5	41,4	63,7	31,6	52,9
6	14,7	14,9	14,8	11,7
7	10,0	5,1	11,9	10,7
8	6,8	2,2	8,8	3,9
9	4,1	1,2	5,4	1,9
10	2,6	0,4	3,7	-
11	6,9	1,2	9,7	1,0
Antall personer (nettoutvalg)	3 456	917	2 333	206
Uoppgitt	681	63	604	14
Uoppgitt i prosent	16,5	6,4	20,6	6,4

I tabellen ser vi at spredningen til differansen mellom inntekt oppgitt i FU77 og inntekt fra skatteregisteret er størst for de gifte, og minst for de ugifte. Blant de tidligere gifte kvinnene ser vi at det var en større tendens til å oppgi for lite i inntekt i FU77 i forhold til registerinntekten, enn i de to andre gruppene for ekteskkelig status. Det var f.eks. 6,3 prosent av dem som oppga en inntekt i FU77 som var minst 20 000 kroner mindre enn inntekten fra skatteregisteret.

Det neste vi skal se på er hvordan spredningen til differansen mellom de to typene inntektstall varierer med alder. Vi har delt utvalget inn i to grupper ettersom IO-ene var gift eller ikke, for å kunne vurdere hvilken effekt alder har på spredningen til differansene når vi justerer for effekten av ekteskkelig status.

Tabell 5.2. Ugifte og før gifte kvinner i ulike aldersgrupper, etter differansen mellom husholdningsinntekt oppgitt i FU77 og husholdningsinntekt fra skatteregisteret. Prosent

IFU-ISR	Alle	Alder					
		18-19 år	20-24 år	25-29 år	30-34 år	35-39 år	40-44 år
I alt	100	100	100	100	100	100	100
1	2,0	-	1,4	1,2	1,9	8,2	8,2
2	1,4	0,4	0,9	1,8	2,8	2,4	4,9
3	3,0	1,1	3,4	6,1	3,8	1,2	1,6
4	6,1	4,5	6,2	4,8	5,7	11,8	8,2
5	61,7	74,9	63,6	58,8	54,7	41,2	39,3
6	14,3	13,5	14,6	13,3	17,9	11,8	16,4
7	6,1	2,6	5,0	7,9	8,5	12,9	11,5
8	2,5	1,1	2,5	1,8	4,7	4,7	3,3
9	1,3	0,7	1,8	0,6	-	2,4	3,3
10	0,4	-	0,2	1,8	-	-	-
11	1,2	1,1	0,5	1,8	-	3,5	3,3
Ant. personer (nettoutvalget)	1 123	267	439	165	106	85	61
Uoppgitt	77	26	28	9	3	5	6
Uoppgitt i prosent	6,4	8,9	6,0	5,2	2,8	5,6	9,0

Tabell 5.2 viser ugifte og før gifte kvinner i ulike aldersgrupper etter differansen mellom inntekt oppgitt i FU77 og inntekt registrert i skatteregisteret.

Tabell 5.2 viser at spredningen til differansen mellom inntekt oppgitt i FU77 og inntekt fra skatteregisteret øker med alderen. Spesielt er prosentandelen i gruppe 1 (de med minst differans) mye større i de to eldste aldersgruppene enn i de andre aldersgruppene.

Tabell 5.3 svarer til tabell 5.2 med data bare om de gifte kvinnene i utvalget til FU77.

Tabell 5.3. Gifte kvinner i ulike aldersgrupper etter differansen mellom husholdningsinntekt oppgitt i FU77 og husholdningsinntekt fra skatteregisteret. Prosent

IFU-ISR	Alle	Ålder					
		18-19 år	20-24 år	25-29 år	30-34 år	35-39 år	40-44 år
I alt	100	100	100	100	100	100	100
1	3,2	-	2,0	1,7	3,2	3,5	6,3
2	1,9	-	2,4	1,6	1,4	1,1	4,1
3	3,4	-	1,6	2,5	3,8	2,4	6,8
4	5,4	-	6,0	4,4	5,0	5,7	7,1
5	31,6	60,0	31,2	36,8	30,6	29,4	27,3
6	14,8	-	19,6	15,1	15,2	13,2	12,6
7	11,9	20,0	14,4	12,0	9,8	14,3	10,7
8	8,8	20,0	7,6	9,1	10,2	8,8	6,6
9	5,4	-	2,8	5,2	5,9	6,4	5,7
10	3,7	-	4,0	3,3	5,1	3,1	2,5
11	9,7	-	8,4	8,2	9,8	11,9	10,4
Ant. personer (nettutvalg)	2 333	5	250	634	625	453	366
Uoppgitt	604	22	129	123	132	97	101
Uoppgitt i prosent	20,6	81,5	34,0	16,2	17,4	17,6	21,6

Tabell 5.3 viser at det er spesielt mye uoppgitt i de to yngste aldersgruppene. Dette kommer av at det var spesielt mange nygifte i disse to aldersgruppene, og at de hadde problemer med å oppgi ektemannens inntekt. Dette hadde sammenheng med at intervjuetidspunktet var i oktober 1977, mens en spurte om inntektene for 1976.

Hvis en ser bort fra aldersgruppen 18-19 år, der det er svært få observasjoner, ser vi at aldersgruppen 25-29 år er den aldersgruppen der differansene mellom de to typene inntektstall ligger nærmest 0. I de tre eldste aldersgruppene er spredningen til differansene mellom inntektene ved de to registreringsmåtene omtrent like stor.

Det neste vi skal se på er hvordan spredningen til differansen mellom husholdningsinntekt oppgitt i FU77 og husholdningsinntekt fra skatteregisteret varierer med boutgiftene. Tabell 5.4 viser dette for de gifte kvinnene i utvalget.

Tabell 5.4 viser at spredningen til differansen mellom de to typene inntektstall er større i gruppene for boutgifter høyere enn 800 kr, enn i gruppene for boutgifter lavere enn 800 kr. Videre ser vi at tendensen til å oppgi for mye i inntekt i FU77 i forhold til "registerinntekten" øker med økende boutgifter. Blant de av IO-ene i FU77 som oppga at de hadde 2 000 kr eller mer i boutgifter pr. måned, var det hele 20,4 prosent som oppga en inntekt i FU77 som var mer enn 30 000 kr høyere enn "registerinntekten".

Vi har prøvd å finne ut om variablene utdanning og inntekt hadde noen effekt på spredningen til differansen mellom inntektene ved de to registreringsmåtene. Etter å ha foretatt en grov justering for alder og ekteskapslyppighet fant vi at spredningen endret seg lite med både økende utdanning og inntekt. Når det gjaldt inntekt, forekom imidlertid store avvik mellom de to typene inntektstall noe hyppigere blant IO-er med høy "registerinntekt" enn blant IO-er med lav "registerinntekt".

Tabell 5.4. Gifte kvinner i ulike grupper for boutgifter, etter differansen mellom husholdningsinntekt oppgitt i FU77 og husholdningsinntekt fra skatteregisteret. Prosent

IFU-ISR	Totalt	Boutgifter pr. mnd. i kr						
		0	1-399	400-599	600-799	800-1 199	1 200-1 999	2 000 eller mer
Totalt	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
1	3,0	16,7	5,4	1,3	2,9	3,0	2,5	3,7
2	1,9	-	1,8	2,0	1,4	2,0	1,9	2,2
3	3,3	8,3	1,2	2,9	5,5	3,5	2,5	3,4
4	5,3	16,7	7,8	4,6	4,0	4,7	5,3	6,5
5	31,8	16,7	35,9	35,8	39,9	32,5	28,5	21,7
6	14,9	16,7	19,8	18,6	14,4	14,8	13,2	12,1
7	12,1	8,3	13,2	14,3	11,2	13,6	11,9	8,4
8	9,0	8,3	6,0	8,5	7,2	10,2	11,0	7,7
9	5,6	-	1,8	3,3	4,6	5,3	7,2	8,7
10	3,7	-	1,8	2,9	1,7	2,6	6,2	5,3
11	9,6	8,3	5,4	5,9	7,2	7,9	9,8	20,4
Antall personer (nettoutvalg)	2 194	12	167	307	348	508	529	323
Uoppgitt	464	8	52	81	73	102	94	54
Uoppgitt i prosent	17,5	40,0	23,7	20,9	17,3	16,7	15,1	14,3

6. SAMVARIASJON MELLOM INNTEKT OG ANDRE VARIABLE

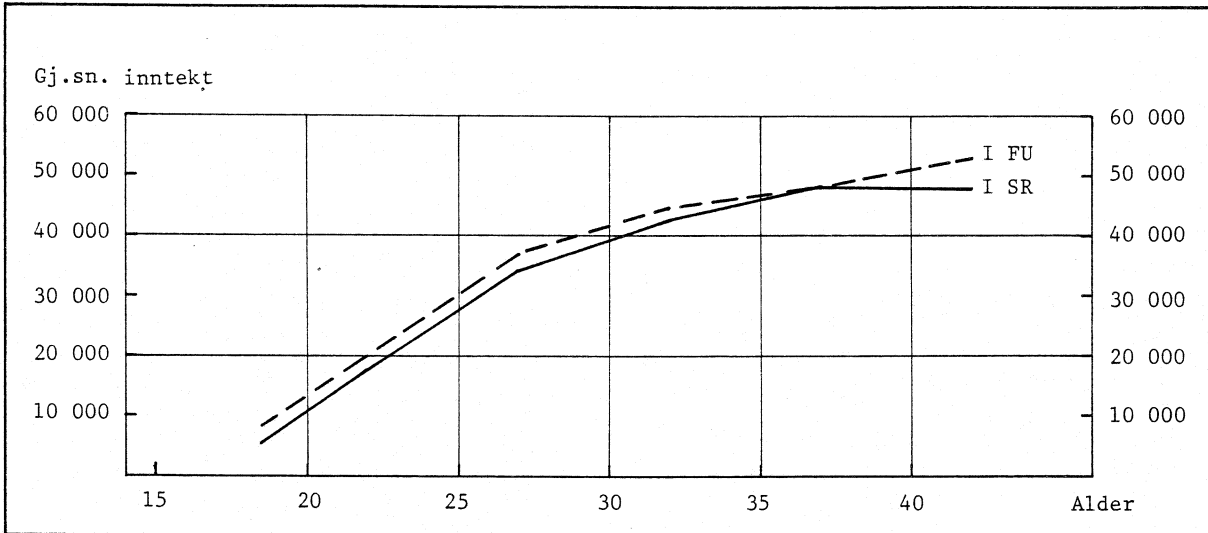
I dette avsnittet skal vi se på om det blir noen forskjell om vi bruker inntektstallene oppgitt i FU77 eller inntektstallene fra skatteregisteret når vi skal studere samvariasjonen mellom inntekt og andre variable.

Vi vil først se på samvariasjonen mellom gjennomsnittsinntekt og alder. Siden det er husholdningsinntekt vi har konsentrert oss om, har vi valgt å se på ugifte og gifte hver for seg. Før gifte er ikke tatt med. Vi har delt IO-ene inn i 6 aldersgrupper og beregnet gjennomsnittsinntektene i hver av aldersgruppene. Disse har vi plottet inn på figur 6.1 og figur 6.2 henholdsvis, og tegnet kontinuerlige kurver på grunnlag av 6 punkter. De prikkede linjene viser hvordan gjennomsnittsinntekten varierer med alder når vi bruker tall fra FU77, mens de helt optrukne linjene viser det samme når vi bruker tallene fra skatteregisteret.

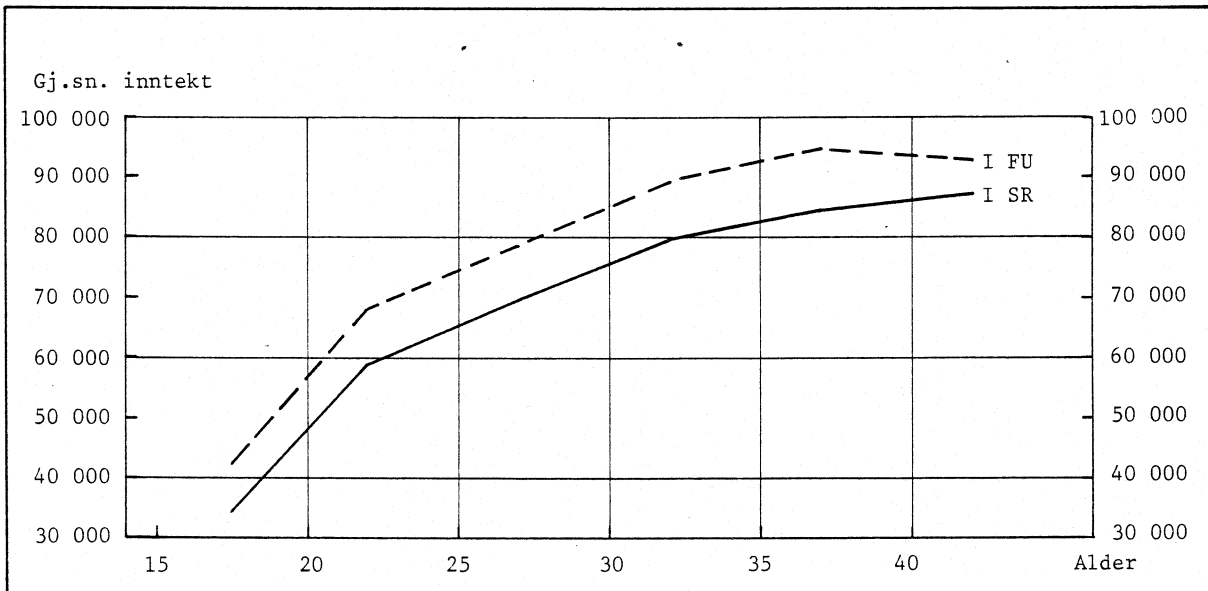
Når det gjelder den øverste delen av figur 6.1, må det sies at kurvene her er plottet på grunnlag av få observasjoner. Ellers kan en si om både figur 6.1 og figur 6.2 at de prikkede linjene er nesten parallelle med de helt optrukne linjene. Det betyr at selv om inntektstallene fra FU77 jevnt over er høyere enn registertallene, antyder figurene at samvariasjonen mellom gjennomsnittsinntekt og alder blir omtrent den samme uansett hvilken av de to typene inntektstall en bruker.

Det neste vi skal se på er samvariasjonen mellom gjennomsnittsinntekt og utdanning. Som i kapittel 3 har vi delt IO-ene inn i 3 utdanningsgrupper og to aldersgrupper. I figur 6.3 har vi plottet inn gjennomsnittsinntektene fra tabell 3.3. Søylen viser tallene fra FU 77, mens de dobbelt-skraverte delene av søylene viser tallene fra skatteregisteret.

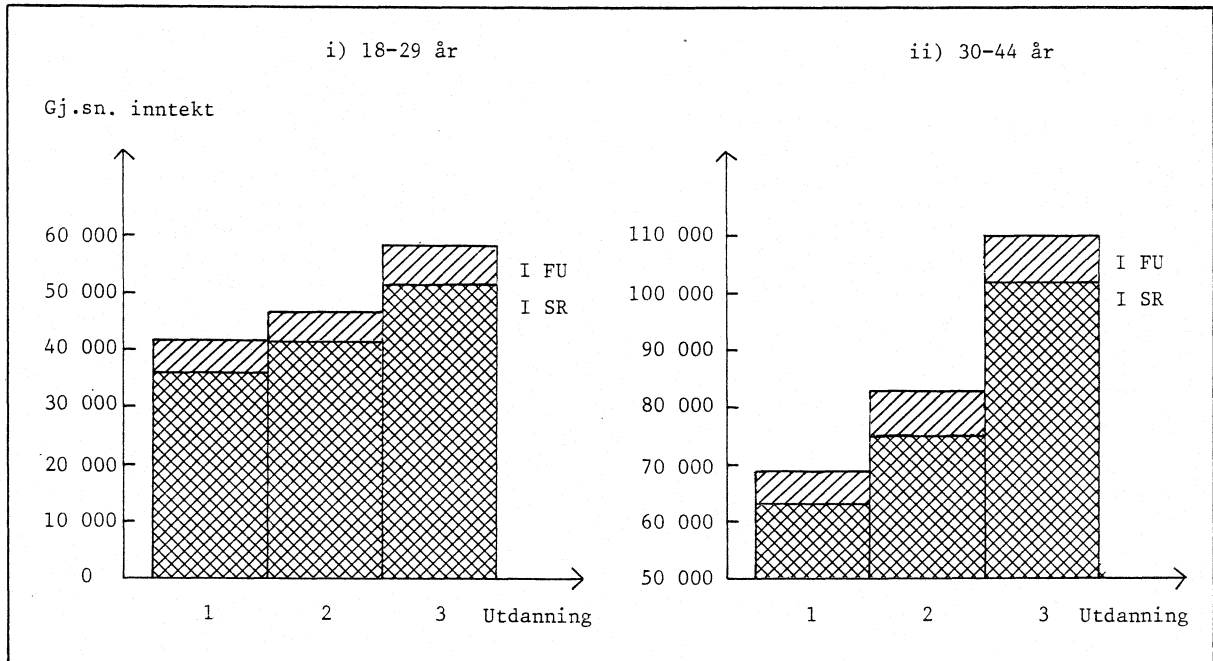
Figur 6.1. Gjennomsnittsinntektenes variasjon med alder når vi bruker inntektstall fra FU77 og skatteregisteret henholdsvis. Ugifte kvinner



Figur 6.2. Gjennomsnittsinntektenes variasjon med alder når vi bruker inntektstall fra FU77 og skatteregisteret henholdsvis. Gifte kvinner



- 6.3. Gjennomsnittsinntektenes variasjon med utdanning når vi bruker tall fra FU77 og skatregisteret henholdsvis.
 i) Aldersgruppen 18-29 år ii) Aldersgruppen 30-44 år



I begge aldersgruppene er differansen mellom gjennomsnitt av inntektstall oppgitt i FU77 og gjennomsnitt av tall fra skatregisteret større i utdanningsgruppe 3 enn i utdanningsgruppe 1. Forskjellen er imidlertid så liten at det ser ut til å ha liten betydning om en bruker den ene eller andre type inntektstall når en skal analysere samvariasjonen mellom inntekt og utdanning.

Ved hjelp av SPSS-programmet har vi fått beregnet noen mål for samvariasjon mellom inntekt og andre variable. Vi har valgt å konsentrere oss om følgende to mål:

i) ρ = den vanlige korrelasjonskoeffisienten

ii) $\gamma = \frac{P-Q}{P+Q}$, der P er antall kordante par og Q er antall diskordante par av IO-er. (Med et kordant par av IO-er menes to IO-er som har den samme relative ordningen med hensyn til begge variablene. Med et diskordant par av IO-er menes et par av IO-er der det ene IO-et har høyest verdi med hensyn til den ene variabelen, og det andre IO-et har høyest verdi med hensyn til den andre variabelen.) γ og andre samvariasjonsmål er drøftet i Haldorsen (1976).

Vi har beregnet de to målene for samvariasjon mellom inntekt og følgende andre variable: alder, utdanning, boutgifter, om IO-ene har bil eller ikke og om IO-ene har farge-TV eller ikke. De to sistnevnte variablene kan bare anta to verdier. I beregningene har vi brukt omkodede gruppeverdier for de forskjellige variablene. Vi har benyttet 11 inntektsgrupper med gruppeverdier fra 1 til 11, de samme 6 aldersgruppene som i figur 6.1 med gruppeverdier lik den laveste alder i hver gruppe, 7 utdanningsgrupper med gruppeverdier fra 1 til 7, og 7 grupper for boutgifter med gruppeverdier fra 1 til 7.

I tabell 6.1 kan vi sammenlikne resultatene av beregningene når vi i det ene tilfellet benytter inntektstallene fra FU77 (IFU), og når vi i det andre tilfellet benytter inntektstallene fra skatregisteret (ISR).

Tabell 6.1. Samvariasjonen mellom inntekt og diverse andre variable målt ved ρ og γ når vi benytter inntektstallene henholdsvis fra FU77 og skatteregisteret

Gruppe	Variabel 1	Variabel 2	ρ	γ
Ugifte og før gifte	IFU	Alder	0,456	0,561
	ISR	Alder	0,470	0,598
	IFU	Utdanning	0,185	0,131
	ISR	Utdanning	0,190	0,122
Gifte	IFU	Alder	0,227	0,217
	ISR	Alder	0,250	0,237
	IFU	Utdanning	0,260	0,272
	ISR	Utdanning	0,257	0,254
	IFU	Ikke bil - bil	0,133	0,291
	ISR	Ikke bil - bil	0,121	0,284
	IFU	Ikke F.TV - F.TV	0,206	0,264
	ISR	Ikke F.TV - F.TV	0,213	0,275
	IFU	Boutgifter	0,266	0,241
	ISR	Boutgifter	0,212	0,188

I tabell 6.1 ser vi at det blir en ubetydelig forskjell på ρ - og γ -verdiene om vi bruker inntektstallene fra FU77 eller om vi bruker inntektstallene fra skatteregisteret. Et unntak har vi når det gjelder samvariasjonen mellom boutgifter og inntekt. Vi ser at korrelasjonen mellom inntekt og boutgifter blir større når vi bruker inntektstallene fra FU77 enn når vi bruker inntektstallene fra skatteregisteret. Det er et resultat som er rimelig ut fra vår antagelse om at mange av IO-ene oppga bruttoinntekt i stedet for nettoinntekt i FU77.

Heller ikke når det gjelder korrelasjonen mellom inntekt og boutgifter målt ved ρ og γ , kan en si at forskjellen blir særlig stor om en bruker inntektstallene registrert på den ene eller andre måten. Det ser derfor ut til å ha liten betydning om en bruker inntektstall fra skatteregisteret eller inntektstall oppgitt til intervjuere, når en skal analysere samvariasjonen mellom inntekt og andre variable.

EN FORMELL ANALYSE AV TABELL 4.1

Vi skal nedenfor gi en formell vurdering av materialet i tabell 4.1. Når en analyserer en tabell som tabell 4.1, gjør en det for å kunne trekke konklusjoner om hele populasjonen. Vi lar "FU-inntekten" betegne det beløpet som en vilkårlig person i populasjonen (kvinner 18-44 år) ville ha oppgitt som inntekt, dersom hun hadde blitt intervjuet i FU77.

Tabell 4.1 antyder at de personene som har "FU-inntekten" i en annen inntektsgruppe enn "registerinntekten", er i mindretall. Vi vil her bruke statistiske metoder for å kunne trekke en av følgende tre konklusjoner om disse personene:

- i) Det er en tendens til å oppgi høyere "FU-inntekt" enn "registerinntekt".
- ii) Det er tendens til å oppgi lavere "FU-inntekt" enn "registerinntekt".
- iii) Ikke si noe.

Dette blir kalt et "tre-desisjonsproblem". For å etablere en desisjonsregel for problemet må en bestemme en modell for tabellen.

Tabell 4.1 kan betraktes som resultatet av en multinomisk forsøksrekke med $N = 3\,458$ uavhengige forsøk, der det for hvert forsøk er like mange mulige utfall som celler i tabellen. La p_{ij} betegne sannsynligheten i hvert forsøk for å bli klassifisert til celle (i,j) i tabellen, og la X_{ij} være antall personer som er blitt klassifisert til celle (i,j) . (Celle (i,j) er for IO-er med registrerinntekt i inntektsgruppe i og inntekt oppgitt i FU77 i inntektsgruppe j .) Vi sier at modellen for tabellen er multinomisk.

Vi skal nå vise hvordan tre-desisjonsproblemet ovenfor kan bli uttrykt ved parametrene i den multinomiske modellen for tabellen. Ved å benytte modell (3.9), Aaberge (1980), med $a_{ij} = 1$ for alle $i < j$, dvs.:

$$\frac{p_{ij}}{p_{ji}} = \theta, \quad i < j$$

følger opplagt at $\theta > 1$ medfører konklusjon (i) og at $\theta < 1$ medfører konklusjon (ii).

Gjennom denne modellen har vi redusert problemet til å bestemme om $\theta > 1$, $\theta < 1$ eller eventuelt ikke si noe. Vi kan derfor benytte tre-desisjonsutgaven av metoden som er presentert i kapittel 3.2, Aaberge (1980). Vi velger nivået for metoden lik $\epsilon = 0,005$. Siden antall observasjoner er stort, benytter vi den asymptotiske utgaven av metoden, som gir følgende desisjonsregel for θ :

En påstår at $\theta < 1$ dersom

$$T = \frac{Z - \mu}{\sigma} \leq k_{\epsilon},$$

og at $\theta > 1$ dersom

$$T = \frac{Z - \mu}{\sigma} \geq k_{1-\epsilon},$$

der k_{ϵ} er ϵ -fraktilen i den standardiserte normalfordelingen, $Z = \sum_{i < j} X_{ij}$, $\mu = \frac{1}{2} \sum_{i < j} (X_{ij} + X_{ji})$ og $\sigma^2 = \frac{1}{2} \mu$.

Innsetting av tallene fra tabell 4.1 gir:

$$T = 16,6 > k_{0,995} = 2,58.$$

En kan altså påstå at $\theta > 1$ som medfører at en kan trekke konklusjon i).

Utdrag fra instruksen til intervjuerne:

S p m. 136- Nertoinntekten er det beløp som en kommer fram til på
138 selvangivelsen etter at fradragspostene er trukket fra.

Vi nøyer oss med anslått kronebeløp. Det er ikke nødvendig at IO leter etter, eller slår opp i selvangivelsen for å få beløpet nøyaktig.

Er IO svært usikker, f.eks. på mannens inntekt (skjematype 2), kan De eventuelt notere "mellom 40 000 og 50 000" e.l. Men vi vil helst ha det omtrentlige beløp i kroner.

Utdrag fra spørreskjemaet fra FU77:

136. Omtrent hvor stor var Deres netto inntekt i 1976 før skatt var trukket fra?

728-733
Ca. Kr → 137

734
1 Liknet felles → 138
2 Vet ikke → 137
3 Ønsker ikke å oppgi → 137

Referanser

Aaberge, R. (1980): "Eksakte metoder for analyse av to-vegstabellar". Stensil. (Blir gitt ut i serien Rapportar fra Statistisk Sentralbyrå).

Haldorsen, T. (1976): "Forelesninger om avhengighetsmål i kontingenstabellar av universitetslektor Harald Goldstein". Arbeidsnotat IO 76/27.