

Interne notater

STATISTISK SENTRALBYRÅ

81/25

7. juli 1981

EKSEMPLER PÅ FRAFALLSEFFEKTER I HUSHOLDNINGS- OG PERSONUNDERSØKELSER^{*)}

Av

Erling Siring

INNHold

	Side
1. Innledning	2
2. Bruken av sentrale dataregistre	2
2.1. Frafallseffekter i Valgundersøkelsen 1969	3
2.2. " i Forbruksundersøkelsen 1973	4
2.3. " i Fruktbarhetsundersøkelsen 1977	4
2.4. Nektene sammenlignet med annet frafall	7
3. Indirekte metoder for analyse av frafallseffekter	9
3.1. Sammenligning av kjente fordelinger i frafallet med fordelinger i hele utvalget	9
3.2. Oppfølging med intervjuere i postundersøkelser ...	11
3.3. Analyse av data fra gjenbesøk	15
3.4. Effekter av frafall på multivariabel analyse	17
4. Referanser	24

*) Innlegg på Nordisk statistisk seminar om frafall februar 1981

1. Innledning

I de fleste utvalgsundersøkelser vil en få et frafall som skyldes delvis at uttrukne person nekter å svare, og delvis at intervjuerne av en eller annen grunn ikke får kontakt med de uttrukne personer. Av flere grunner er det viktig å finne ut om slikt frafall fører til skjevheter i resultatene. Dersom frafallet ikke fører til skjevheter, kan en helt se bort fra det, og la være å sette inn tiltak for å redusere det. Dersom det på den andre siden fører til skjevheter i resultatene, er det viktig å få et så godt innblikk som mulig i hvor store slike skjevheter er, og kanskje si litt om hva de skyldes.

Det ligger i sakens natur at det ikke er mulig å gi nøyaktige mål for skjevheten som skyldes frafall for samtlige variable i noen undersøkelse. Likevel er det etter vår oppfatning i mange tilfeller mulig å få et ganske godt bilde av de mest alvorlige skjevheter som frafallet medfører med enkle metoder.

I dette notatet skal vi gi eksempler på hvordan en kan gå fram for å få et inntrykk av effektene av frafallet i utvalgsundersøkelser, både når en har tilleggsinformasjon fra andre kilder enn utvalget, og når en bare har resultater fra utvalgsundersøkelsen. I noen tilfeller får en bare et grovt inntrykk av frafallets effekt, mens i andre tilfeller får en et meget nøyaktig uttrykk for skjevheten.

I avsnitt 2 i notatet presenterer vi noen eksempler på bruken av sentrale dataregistre for å få informasjon om frafallseffekter. I avsnitt 3 gir vi eksempler på hvordan en kan gå fram for å vurdere frafallseffekter når en ikke har opplysninger fra andre kilder enn selve utvalgsundersøkelsen.

En hypotese vi har arbeidet en del med i Statistisk Sentralbyrå går ut på at den skjevhet som innføres av det frafall som skyldes at folk nekter å svare, er vesentlig mindre enn den skjevheten som innføres av annet frafall. I avsnitt 2.4 skal vi presentere resultater som støtter denne hypotesen.

2. Bruken av sentrale dataregistre

I Norge eksisterer det en del dataregistre, som gjør det mulig å skaffe enkelte opplysninger om frafall. Muligheten til å koble med sentrale dataregistre er likevel bare blitt brukt i noen ganske få tilfeller på grunn av kostbar databehandling.

Vi regner med å gjøre mer bruk av denne metoden i framtiden når vi får flere registre og data blir lettere tilgjengelig.

I dette avsnittet skal vi presentere tre eksempler på undersøkelser der en har brukt registerdata for å vurdere frafallseffekter.

2.1. Frafallseffekter i Valgundersøkelsen 1969

I 1969 foretok Byrået en intervjuundersøkelse for å studere forholdet mellom stemmegiving ved Stortingsvalget 1968 og andre variable, som kjønn, alder, inntekt, utdanning og yrke. Utvalget var på 2 999 personer, og frafallet var på 9,9 prosent. Utvalget ble koblet med valgregisteret, som viste om en person hadde stemt ved Stortingsvalget eller ikke. Tabell 1 viser prosentandelen som stemte ved valget, blant frafallet, de som svarte, og hele utvalget i forskjellige aldersgrupper.

Tabell 1. Andel som stemte ved valget blant frafallet, de som svarte og hele utvalget. Prosent

	Alder					
	Alle	20-24 år	25-29 år	30-49 år	50-69 år	70-79 år
Frafall	71	59	56	72	78	74
De som svarte	90	85	88	92	92	85
Hele utvalget	88	81	84	90	91	84

Tabellen viser at valgdeltakelsen var lavere blant frafallet enn blant de som svarte, og at forskjellen mellom frafallet og resten av utvalget var større blant yngre enn blant eldre mennesker. Totalt var valgdeltakelsen to prosent høyere for de som svarte i utvalget enn for hele utvalget. Thomsen (1971).

I et forsøk på å studere hvorfor valgdeltakelsen var lavere blant frafallet enn i resten av utvalget, delte en frafallet opp i forskjellige kategorier for frafallsårsak og beregnet valgdeltakelsen i hver av kategoriene. Resultatene presenteres i avsnitt 2.4. Det eneste vi har funnet som delvis forklarer den lave valgdeltakelsen i frafallet, er at det var en del som ikke svarte i undersøkelsen på grunn av at de var psykisk eller fysisk syke, og at valgdeltakelsen naturlig nok var lav blant disse.

2.2. Frafallseffekter i Forbruksundersøkelsen 1973

Forbruksundersøkelsene er husholdningsundersøkelser som gjentas hvert år. De uttrukne husholdningene blir bedt om å føre dagbok over alle sine utgifter i en to ukers periode.

I Forbruksundersøkelsen 1973 var utvalget på 4 705 personer og frafallet på 28,5 prosent. Utvalget ble koblet med skatteregisteret, og husholdningsinntekten til alle husholdningene i utvalget ble funnet. Herigstad (1978). Tabell 2 viser gjennomsnittsinntekten for frafallet, de som svarte og hele utvalget i forskjellige typer husholdninger.

Tabell 2. Husholdningsinntekt for frafallet, de som svarte og hele utvalget i forskjellige grupper for husholdningstype. Gj.snitt i 1 000 kr.

Husholdnings- type	Frafall	De som svarte	Hele utvalget
Alle	36,3	44,4	42,1
Enslige	17,1	18,6	17,9
Ektepar u/barn ..	36,7	38,6	38,0
Ektepar m/barn ..	51,2	47,8	48,3
Antall personer .	1 342	3 363	4 705

I tabellen ser vi at gjennomsnittsinntekten er mye lavere for frafallet enn for resten av utvalget når en ser på alle husholdningstyper samlet. Hvis en derimot ser på de enkelte husholdningstypene hver for seg, er det ikke så stor forskjell mellom frafallet og de som deltok i undersøkelsen. I gruppen ektepar m/barn er gjennomsnittsinntekten faktisk høyere for frafallet enn for resten av utvalget. Dette viser at årsaken til at gjennomsnittsinntekten er så mye lavere for frafallet enn for resten av utvalget, er at frafallet er større blant enslige enn blant andre.

2.3. Frafallseffekter i Fruktbarhetsundersøkelsen 1977

I 1977 ble det utført en fruktbarhetsundersøkelse i Norge. Et utvalg på 5 047 kvinner i alderen 18-44 år ble trukket. For å redusere frafallet til et minimum ble det foretatt inntil 8 forsøk på å oppnå intervju. Det endelige frafallet ble til slutt på 18 prosent.

For å studere frafallseffektene hentet en inn opplysninger fra folkeregistrene og skatteregisteret om henholdsvis antall levende fødsler og inntekt for frafallet. Tabell 3 viser gjennomsnittlig antall levende fødsler for frafallet, de som svarte og hele utvalget i ulike aldersgrupper.

Tabell 3. Gj.snittlig antall levende fødsler blant frafallet, de som svarte og hele utvalget i forskjellige aldersgrupper

Alder	Frafall	Personer som svarte	Hele utvalget
I a l t	1,19	1,57	1,50
18-24 år	0,26	0,40	0,37
25-34 år	1,21	1,68	1,61
35-44 år	2,10	2,55	2,46

Tabellen viser at gjennomsnittlig antall levende fødsler er høyere blant de som svarte, enn blant frafallet i alle aldersgruppene. For å finne ut noe om grunnen til dette har vi delt frafallet opp i forskjellige kategorier for frafallsårsak, og sett på hvordan andelen personer i hver kategori relativt til hele utvalget endrer seg med antall levende fødsler. Resultatene er presentert i tabell 4.

Tabell 4. Utvalgets fordeling på svarkategorien og kategorier for frafallsårsak i forskjellige grupper for antall levende fødsler. Prosent

Antall levende fødsler	Totalt	Frafallsårsak				De som svarte	Antall personer
		Ikke truffet *)	Nekting	Fysisk eller psykisk syk	Annet		
I a l t ...	100,0	6,04	10,20	0,79	0,99	81,97	5 047
0	100,0	10,15	11,51	1,86	1,30	75,19	1 616
1	100,0	6,07	9,86	0,33	1,08	82,67	923
2	100,0	3,26	8,70	0,36	0,36	87,32	1 380
3	100,0	2,64	10,54	0,14	1,25	85,44	721
4 ⁺	100,0	2,33	10,10	0,26	0,78	86,53	386

*) Med de som ikke ble truffet, er regnet de som var borte på arbeid, skole el. lignende, de som var flyttet og de som var bosatt el. på reise i utlandet for mer enn 3 mndr.

Tabellen viser at av kvinnene i utvalget med 0 levende fødsler, var det over 10 prosent som intervjuerne ikke kom i kontakt med. Prosentandelen i utvalget som intervjuerne ikke kom i kontakt med, avtar sterkt med antall levende fødsler. Tabellen viser videre at andelen personer i utvalget som nektet å delta i undersøkelsen varierer lite med antall levende fødsler. Det samme er tilfellet når det gjelder andelen personer i utvalget i frafallskategoriene "fysisk el. psykisk syke" og "annet". En kan da konkludere med at den viktigste årsaken til at gjennomsnittlig antall levende fødsler er så mye lavere i frafallet enn i resten av utvalget, er at kvinner er vanskeligere å treffe hjemme desto færre barn de har.

Tabell 5 viser gjennomsnittlig husholdningsinntekt for frafallet, de som svarte og hele utvalget i forskjellige aldersgrupper. Vi definerer i dette tilfellet husholdningsinntekt som summen av egeninntekt og ekte-mannens inntekt for de gifte kvinnene, og bare egeninntekt for de ugifte og tidligere gifte kvinnene.

Tabell 5. Husholdningsinntekt for frafallet, de som svarte og hele utvalget i forskjellige aldersgrupper. Gjennomsnitt i 1 000 kr.

Alder	Frafall	Personer som svarte	Hele utvalget
I a l t	58,2	67,7	66,0
18-24 år	27,2	34,0	32,7
25-34 år	61,2	75,8	73,5
35-44 år	86,2	88,9	88,4

Tabellen viser at husholdningsinntekten i gjennomsnitt er lavere for frafallet enn for resten av utvalget i alle aldersgruppene. Siden disse resultatene har mye til felles med resultatene fra Forbruksundersøkelsen 1973, er det igjen naturlig å foreta en oppdeling av utvalget etter husholdningstype. Tabell 6 viser gjennomsnittlig husholdningsinntekt for frafallet, de som svarte og hele utvalget i to forskjellige grupper for husholdningstype.

Tabell 6. Husholdningsinntekt for frafallet, de som svarte og hele utvalget i forskjellige grupper for ekteskapelig status. Gjennomsnitt i 1 000 kr.

Ekteskapelig status	Frafall	Personer som svarte	Hele utvalget
I a l t	58,2	67,7	66,0
Ugift/før gift	27,5	29,5	29,0
Gift	85,8	83,4	83,7

I likhet med Forbruksundersøkelsen 1973 ser vi at innenfor enkelte grupper for husholdningstype er det liten forskjell mellom gjennomsnittsinntekten for frafallet og resten av utvalget. Som i Forbruksundersøkelsen kan vi konkludere med at årsaken til at gjennomsnittsinntekten er så mye lavere for frafallet enn for resten av utvalget, er at frafallet er større blant personer fra små husholdninger enn fra store husholdninger. Som tidligere nevnt mener vi at dette igjen har sammenheng med at personer fra små husholdninger (el. personer uten barn) er vanskeligere å treffe hjemme enn personer fra store husholdninger (personer med barn).

2.4. Nektere sammenlignet med annet frafall

Som nevnt i innledningen har vi i Statistisk Sentralbyrå en hypotese som går ut på at for mange variables vedkommende er den skjevhet i resultatene som skyldes frafall på grunn av nekting, vesentlig mindre enn den skjevhet som skyldes frafall av andre årsaker. For å prøve å underbygge denne hypotesen har vi delt frafallet i de tre undersøkelsene vi til nå har studert resultatene fra, opp i grupper for frafallsårsak og sammenliknet nektere med andre.

I Valgundersøkelsen 1969 var valgdeltakelsen i hele utvalget, som vi så i avsnitt 2.1, 88 prosent, mens den var 71 prosent for frafallet. Blant de som nektet å delta i undersøkelsen, var valgdeltakelsen 81 prosent, mens den i kategorien for frafallsårsakene "ikke truffet hjemme" og "psykisk eller fysisk syk" var henholdsvis 65 og 55 prosent. Valgdeltakelsen blant "nekterne" lå altså mye nærmere valgdeltakelsen i hele utvalget enn den gjorde i de andre gruppene for frafallsårsak. Det betyr at skjevhetene i resultatene fra Valgundersøkelsen først og fremst skyldtes de som ikke ble truffet hjemme og de som falt fra på grunn av sykdom.

I avsnitt 2.2 og 2.3 så vi at gjennomsnittlig husholdningsinntekt var mye lavere for frafallet enn for hele utvalget i både Forbruksundersøkelsen 1973 og Fruktbarhetsundersøkelsen 1977. I tabell 7 kan vi sammenlikne gjennomsnittlig husholdningsinntekt for "nektere" med annet frafall og hele utvalget i de to ovenfor nevnte undersøkelsene.

Tabell 7. Husholdningsinntekt blant nektere, annet frafall og hele utvalget. Gjennomsnitt i 1 000 kr.

Undersøkelse	Nektere	Annet frafall	Hele utvalget
Forbruksundersøkelsen 1973	43,3	29,5	42,1
Fruktbarhetsundersøkelsen 1977 ..	71,6	40,7	66,0

I begge undersøkelsene ser vi at gjennomsnittlig husholdningsinntekt for nekterne er omtrent den samme som for hele utvalget, mens den er mye høyere enn for resten av frafallet.

Vi har tidligere hevdet at frafallet var spesielt stort blant personer fra små husholdninger i begge undersøkelsene. Når det gjelder Fruktbarhetsundersøkelsen, har vi i kap. 2.3 sett at det større frafallet blant mennesker fra små husholdninger enn blant andre først og fremst skyldtes at folk fra små husholdninger var vanskeligere å treffe hjemme enn andre. Det er naturlig å tro at det samme var tilfellet i Forbruksundersøkelsen. Ut fra dette er det ikke overraskende at gjennomsnittsinntekten for "annet frafall", som stort sett består av personer som ikke ble truffet hjemme, ligger så mye lavere enn for hele utvalget i begge undersøkelsene.

I tabell 4 har vi sett at andelen nektere i utvalget til Fruktbarhetsundersøkelsen varierer lite med antall levende fødsler. Dette viser at fordelingen av personer fra forskjellige husholdningstyper er omtrent den samme blant nekterne som i hele utvalget. Det er derfor ikke unaturlig at gjennomsnittsinntekten blant nekterne ligger nær gjennomsnittsinntekten i hele utvalget.

Til slutt i dette avsnittet skal vi presentere et resultat som ikke er overraskende ut fra tabell 4. Det er at gjennomsnittlig antall levende fødsler blant nekterne i Fruktbarhetsundersøkelsen ligger nær det tilsvarende tall i hele utvalget, og er mye høyere enn det tilsvarende tall i resten av frafallet. Gjennomsnittlig antall levende fødsler er 1,43 blant nekterne, 0,84 blant resten av frafallet og 1,50 i hele utvalget.

I de eksemplene vi har sett på, er den skjevheten i resultatene som skyldes frafall på grunn av nekting svært liten i forhold til den skjevhet som skyldes frafall av andre årsaker. De variablene vi har sett på, unntatt "valgdeltakelse", er sterkt korrelert med hva slags husholdning IO-ene kommer fra, som igjen er sterkt korrelert med sannsynligheten for å treffe folk hjemme. For variable som har svak sammenheng med husholdningstype, er det tvilsomt at nekterne "ligner" så mye mer på hele populasjonen enn resten av frafallet, som for de variable som vi har studert her.

3. Indirekte metoder for analyse av frafallseffekter

Det er sjeldent at informasjon om frafall finnes i eksisterende registre, og derfor blir vanligvis andre metoder brukt til å vurdere omfanget av frafallseffekter. Vi skal se eksempler på to slike metoder.

Den enkleste og mest vanlige metoden som er brukt, er å sammenlikne kjønns-, alders- og sivilstandsfordeling for frafallet med fordelingene i hele utvalget, og anta at sammenhengen mellom de nevnte variablene og variabler som vi er interessert i å studere, er den samme i frafallet som i resten av utvalget.

Den andre metoden vi skal se på, kan brukes i undersøkelser der datainnsamlingen foregår i flere faser. Metoden går ut på at en studerer utviklingen i resultatene fra fase til fase i datainnsamlingen, og på den måten skaffer seg innsikt i frafallseffektene. Eksemplene våre på denne metoden faller i to grupper. I den ene gruppen har vi undersøkelser der datainnsamlingen har foregått i én "postalfase" og én "intervjufase". I den andre gruppen har vi intervjuundersøkelser der vi kan dele datainnsamlingen inn i faser etter hvor mange besøk intervjuerne har måttet foreta for å oppnå intervju.

3.1. Sammenlikning av kjente fordelinger i frafallet med fordelinger i hele utvalget

Vi har tidligere sett at antall levende fødsler og størrelsen på husholdningsinntekten varierer sterkt med ekteskapelig status. Når en i en undersøkelse skal vurdere frafallseffektene på slike variable, kan det derfor gi god innsikt å sammenlikne fordelingen m.h.t. ekteskapelig status i frafallet med fordelingen i hele utvalget. Vi skal vise hvordan vi kan bruke vårt kjennskap til fordelingen m.h.t. ekteskapelig status i frafallet

i Fruktbarhetsundersøkelsen til å vurdere frafallseffektene på gjennomsnittlig antall levende fødsler og husholdningsinntekt.

Tabell 8 viser fordelingen m.h.t. ekteskapelig status for frafallet, de som svarte og hele utvalget i Fruktbarhetsundersøkelsen. Tabellen viser at ugifte er sterkt overrepresentert i frafallet. Det indikerer at de to variablene som vi er interessert i blir overestimert. (Vi har sett tidligere at dette er tilfellet.)

Tabell 8. Frafall, de som svarte og hele utvalget etter ekteskapelig status. Prosent

Ekteskapelig status	Frafall	Personer som svarte	Hele utvalget
I a l t	100	100	100
Ugift	40	24	27
Gift	53	71	68
Tidligere gift	6	5	5
Uoppgitt	1	0	0
Antall personer	910	4 137	5 047

For å få et enkelt inntrykk av frafallseffektene kan en lage en veid estimator der en justerer for skjevheten i fordelingen m.h.t. ekteskapelig status blant de som svarte, og sammenlikne denne estimatoren med det vanlige gjennomsnittet.

$$\text{La } \hat{S} = \bar{X}_s - \frac{1}{n} \sum_i n_i \bar{X}_{is},$$

der \bar{X}_{is} er gjennomsnittet av observasjoner fra gruppe i for ekteskapelig status, \bar{X}_s gjennomsnittet av observasjoner fra hele svardelen av utvalget, n_i antall personer i gruppe i for ekteskapelig status og n antall personer i hele utvalget. \hat{S} uttrykker effekten av skjevheten i fordelingen m.h.t. ekteskapelig status blant de som svarte i forhold til hele utvalget. I tabell 9 kan vi sammenlikne \hat{S} med den virkelige verdien, S , for skjevheten i henholdsvis gjennomsnittlig antall levende fødsler og husholdningsinntekt blant de som svarte, i forhold til hele utvalget.

Tabell 9. Estimerte og virkelige skjevheter i gjennomsnittlig antall levende fødsler og husholdningsinntekt blant de som svarte, i forhold til hele utvalget

Variabel	\hat{S}	S
Ant. levende fødsler ..	0,06	0,07
Husholdningsinntekt ...	1 700 kr.	1 700 kr.

Vi ser at \hat{S} i begge tilfellene ligger svært nær S. Grunnen til at estimeringen av skjevhetene går så bra her, er at både husholdningsinntekt og antall levende fødsler er høyt korrelert med ekteskapelig status.

3.2. Oppfølging med intervjuere i postundersøkelser

I undersøkelser i Byrået der datainnsamlingen foregår pr. post er det ofte slik at en sender ut intervjuere etter at den ordinære datainnsamlingen er ferdig for å intervju personer som ikke har svart i posten. Ved å sammenligne resultatene fra de to fasene av innsamlingen kan en skaffe seg innsikt om frafallseffektene.

Vi skal presentere resultatene fra to undersøkelser der datainnsamlingen har foregått både postalt og ved intervjuing. Den ene undersøkelsen er "Ungdomsundersøkelsen 1975", og den andre undersøkelsen er "Prøveundersøkelsen for folke- og bolig tellingen 1980".

3.2.1. Resultater fra Ungdomsundersøkelsen 1975

Undersøkelsen, som hadde som formål å belyse de problemer som ungdommen har til skole og arbeidsliv, ble gjennomført som en postundersøkelse. 2 uker etter at skjemaene var blitt sendt ut til hele utvalget, ble det sendt purring med skjema vedlagt. Den delen av utvalget som ikke svarte på de to første henvendelsene, ble delt i to deler. Den ene delen ble oppsøkt av intervjuere, mens den andre delen ble purret for 2. gang.

Av 8 818 personer i utvalget sendte 63 prosent inn besvarelse på det først utsendte skjema, 15 prosent svarte på skjemaet fra 1. purring, 3 prosent på skjemaet fra 2. purring, mens 7 prosent ble intervjuet. Drøyt 12 prosent fikk en ikke noe svar fra. Tabell 10 viser de som svarte, frafallet og hele utvalget etter kjønn, alder, kommunetype og region.

Tabell 10. Svarerne, utvalget og frafallet etter kjønn, alder, kommune-
type og region. Prosent

	Svarerne	Utvalget	Frafall
Kjønn	100	100	100
Menn	51	51	53
Kvinner	49	49	47
Alder			
17 år	33	33	29
18 "	33	34	36
19 "	33	34	35
Kommunetype			
Landbrukskommuner	4	4	3
Mindre sentrale, blandede landbruks- og industrikommuner	9	10	11
Sentrale, blandede landbruks- og industrikommuner	6	6	6
Fiskerikommuner	4	4	2
Mindre sentrale industrikommuner	5	5	6
Sentrale industrikommuner	14	14	15
Særlig sentrale, blandede tjeneste- ytings- og industrikommuner unntatt Oslo, Bergen og Trondheim	15	15	15
Øvrige blandede tjenesteytings- og industrikommuner	20	21	23
Andre kommuner	5	5	4
Oslo	8	8	7
Bergen	5	6	6
Trondheim	3	3	3
Region			
Oslo	8	8	7
Akershus	9	9	10
Østfold, Vestfold	11	11	12
Hedmark, Oppland	9	9	10
Buskerud, Telemark	9	9	10
Agder, Rogaland	12	12	12
Hordaland, Sogn og Fjordane	14	14	13
Møre og Romsdal	7	7	6
Trøndelag	16	15	14
Nordland	4	4	4
Troms og Finnmark	2	2	2
Tallet på personer	7 706	8 818	1 112

Sjøl om fordelingen av svarerne på kjennemerkene i tabellen ikke skiller seg stort fra fordelingen av utvalget, trenger det ikke å være sam- svar mellom fordelingene på andre kjennemerker.

Fordelingen på viktigste gjøremål av dem som svarte på forskjellige trinn i undersøkelsen kan tyde på at frafallet er svært skjevt med hensyn til viktigste gjøremål.

Tabell 11. Personer i grupper for hvilket trinn i undersøkelsen det ble svart på skjemaet, etter viktigste gjøremål

	I alt	Lønnet arbeid	På skole	Annet	Tallet på personer som svarte
A l l e	100	29	64	7	7 706
Utfylt først tilsendte skjema	100	25	70	5	5 575
Utfylt skjema for 1.purring	100	35	57	9	1 299
Utfylt skjema for 2.purring	100	44	39	16	221
Intervjuet	100	45	35	20	611

Resultatene i tabellen indikerer at andelen med lønnet arbeid blir underestimert, og at andelen som går på skole blir overestimert.

3.2.2. Resultater fra Prøveundersøkelsen for Folke- og bolig telling 1980

Denne undersøkelsen, som foregikk i 1978, kaller vi heretter for P-78. Personene som var med i utvalget fikk først tilsendt skjema i posten som de skulle fylle ut og returnere. Etter at postundersøkelsen var ferdig, ble det satt i gang en kvalitetskontrollundersøkelse (KU-78). Opplegget for den var at en intervjuet hele utvalget til prøveundersøkelsen for å kunne vurdere kvaliteten på svarene i postundersøkelsen. Heldal (1980).

Vi kaller de personene som svarte i begge undersøkelsene for nettoutvalget. Vi skal her sammenlikne data fra nettoutvalget med data fra alle som svarte i P-78. (Filosofien er at de som svarte i P-78, men som var frafall i KU-78, har mer felles med frafallet enn resten av de som svarte i P-78.) Tabell 12 viser fordelingen m.h.t. kjønn, alder og ekteskapelig status for nettoutvalget, alle som svarte i P-78 og hele utvalget.

I tabellen ser vi at fordelingene m.h.t. de tre variablene er mer lik fordelingene i hele utvalget blant alle som svarte i P-78 enn i nettoutvalget. Da det ikke er urimelig å tro at dette er tilfellet også for variablene eier-/leieforhold og leilighetens areal, kan en ut fra tabell 13 trekke konklusjoner om frafallseffektene på disse to variablene.

Tabell 12. Bruttoutvalg, besvarte personskjema (P-78) og nettoutvalg etter kjønn, alder og ekteskadelig status. Prosent

	Hele utvalget	De som svarte i P-78	De som svarte i P-78 og KU-78
I a l t	100	100	100
Kjønn			
Menn	47,8	48,0	47,1
Kvinner	52,2	52,0	52,9
Alder			
16-29 år	27,6	26,7	24,4
30-49 år	30,3	33,4	35,0
50-69 år	29,1	29,1	29,7
70 år eller mer	13,1	10,7	10,9
Ekteskadelig status			
Ugift	24,3	21,6	18,8
Gift	61,7	68,1	71,8
Enke/enkemann	9,1	7,1	6,4
Skilt	3,5	2,4	2,3
Separert	1,4	0,8	0,8
Antall personer	3 729	2 408	1 988
Prosent	100	64,6	53,3

Tabell 13. De som svarte i P-78 og de som svarte i både P-78 og KU-78 etter eier-/leieforhold og etter leilighetens areal. Prosent

	De som svarte i P-78	De som svarte i P-78 og KU-78
I a l t	100	100
Eier-/leieforhold		
Selveie	59,3	61,1
Borettslag, aksjeselskap	17,7	17,7
Arbeidsforhold	5,7	5,3
Tidsbegrenset leiekontrakt ...	2,7	1,7
Annet	9,6	9,1
Uoppgitt	5,1	5,2
Leilighetens areal		
Under 40 m ²	7,1	5,7
40- 59 m ²	12,4	10,2
60- 79 m ²	17,4	17,8
80- 99 m ²	25,9	27,2
100-129 m ²	18,6	19,6
130 m ² eller mer	14,4	15,5
Uoppgitt	4,4	3,8

Tabell 13 viser fordelingene m.h.t. eier-/leieforhold og leilighetens areal for alle som svarte i P-78 og i nettutvalget. Vi ser at både andelen selveiere og andelen personer med store leiligheter er lavere blant alle som svarte i P-78, enn i nettutvalget. Det indikerer at begge disse andelene blir overestimert. Tilsvarende indikerer tabellen at andelen personer med små leiligheter blir underestimert.

3.3. Analyse av data fra gjenbesøk

I de fleste intervjuundersøkelser kan en dele opp utvalget i grupper etter antall besøk som må foretas før det blir oppnådd intervju. Ved å studere utviklingen i resultatene fra 1. besøk til 2. besøk osv kan en danne seg et bilde av frafallseffektene. Denne metoden for analyse av frafallseffekter er blitt formalisert ved en frafallsmodell i Thomsen og Siring (1980). Metoden har den fordel, som vi skal se i neste avsnitt, at den foruten å kunne gi et bilde av frafallseffekter ved estimering av

gjennomsnitt i populasjonen, også kan gi et bilde av frafallseffekter når målet er å estimere forholdet mellom to eller flere variable.

Tabell 14 viser gjennomsnittlig antall levende fødsler pr. IO i Fruktbarhetsundersøkelsen etter hvilket besøk intervjueren oppnådde intervju i.

Tabell 14. Gjennomsnittlig antall levende fødsler pr. IO etter antall besøk

	Antall besøk							
	1	2	3	4	5	6	7	8
Gjennomsnittlig antall levende fødsler	1,84	1,56	1,47	1,30	1,28	1,08	0,85	0,75

Vi ser at gjennomsnittet avtar med økende besøksnr. Det indikerer at kvinner er vanskeligere å treffe hjemme desto færre barn de har, og at kvinnene i frafallet har færre barn i gjennomsnitt enn de som svarte i utvalget. Dette samsvarer helt med resultatene fra kap. 2.3.

Et liknende eksempel har vi fra Boforholdsundersøkelsen 1973. Tabell 15 viser fordelingen av husholdningenes boligtype fordelt etter hvilket besøk intervjueren oppnådde intervju i.

Tabell 15. Husholdningenes boligtype etter nummer på besøket. Prosent

Besøk nr.	Antall IO som blir besøkt	Våningshus	Enebolig	Rekkehus	2-4 mannsbolig	Blokk	Annet
1	1 535	24,3	39,3	6,6	12,4	13,2	4,2
2	910	12,5	37,3	9,8	16,5	18,5	5,5
3	314	7,0	37,6	8,3	19,4	23,6	4,1
4	82	1,2	30,5	6,1	26,8	30,5	4,9
5	25	4,0	28,0	12,0	16,0	32,0	8,0
6	7	0,0	42,9	14,3	14,3	28,6	0,0
7	1	0,0	0,0	100,0	0,0	0,0	0,0
8	3	0,0	33,3	0,0	0,0	33,3	33,3
Uoppgitt	29	20,7	51,7	3,4	17,2	6,9	0,0
T o t a l ..	100,0	17,8	38,2	7,8	14,9	16,6	4,6

Tabellen viser altså at i første besøk svarer 24,3 prosent at de bor i våningshus. Etter å ha gjenbesøkt de som ikke var tilstede i første besøk, synker denne andelen til 17,8 prosent. Det indikerer at den nevnte andelen er større blant de som svarte enn i hele utvalget. Tilsvarende indikerer resultatene i tabellen at andelene av husholdninger som bor i 2-4-manns boliger og blokker er lavere blant de som ble intervjuet enn i hele utvalget.

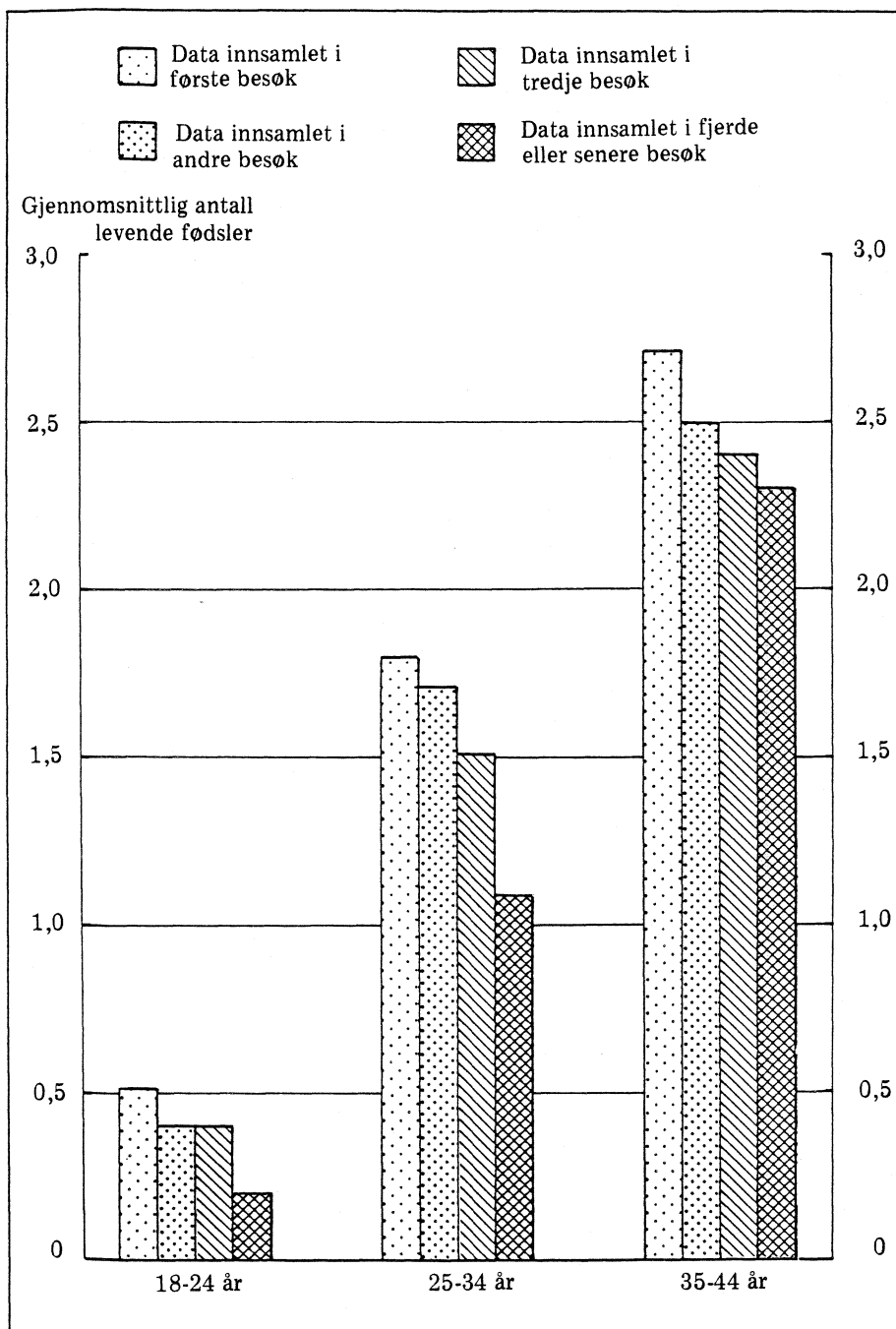
3.4. Effekter av frafall på multivariabel analyse

Det eksisterer svært få eksempler der frafallseffekten på forholdet mellom to eller flere variable er blitt studert. Thomsen og Siring (1980). I dette avsnittet skal vi se på forholdet mellom fruktbarhet (målt ved antall levende fødsler) og alder¹⁾ og utdanning i Fruktbarhetsundersøkelsen. For å studere effekten av frafall skal vi foreta separat analyse av data innsamlet i 1. besøk, 2. besøk, 3. besøk osv.

I fig. 1 er det gjort et forsøk på å studere effekten av frafall på forholdet mellom alder og fruktbarhet. Innenfor hver aldersgruppe ser en at fruktbarheten avtar med økende antall besøk, mens differansen i fruktbarhet mellom aldersgruppene synes å være uavhengig av antall besøk. Det finner vi oppmuntrende siden det indikerer at frafallseffekten ved analyse av forholdet mellom alder og fruktbarhet er mindre enn ved analyse av bare fruktbarhet.

1) Det som her kalles "alderseffekt" er en kombinasjon av alders-, kohort- og periodeeffekt.

FIGUR 1. Gjennomsnittlig antall levende fødsler i ulike grupper for alder og antall besøk

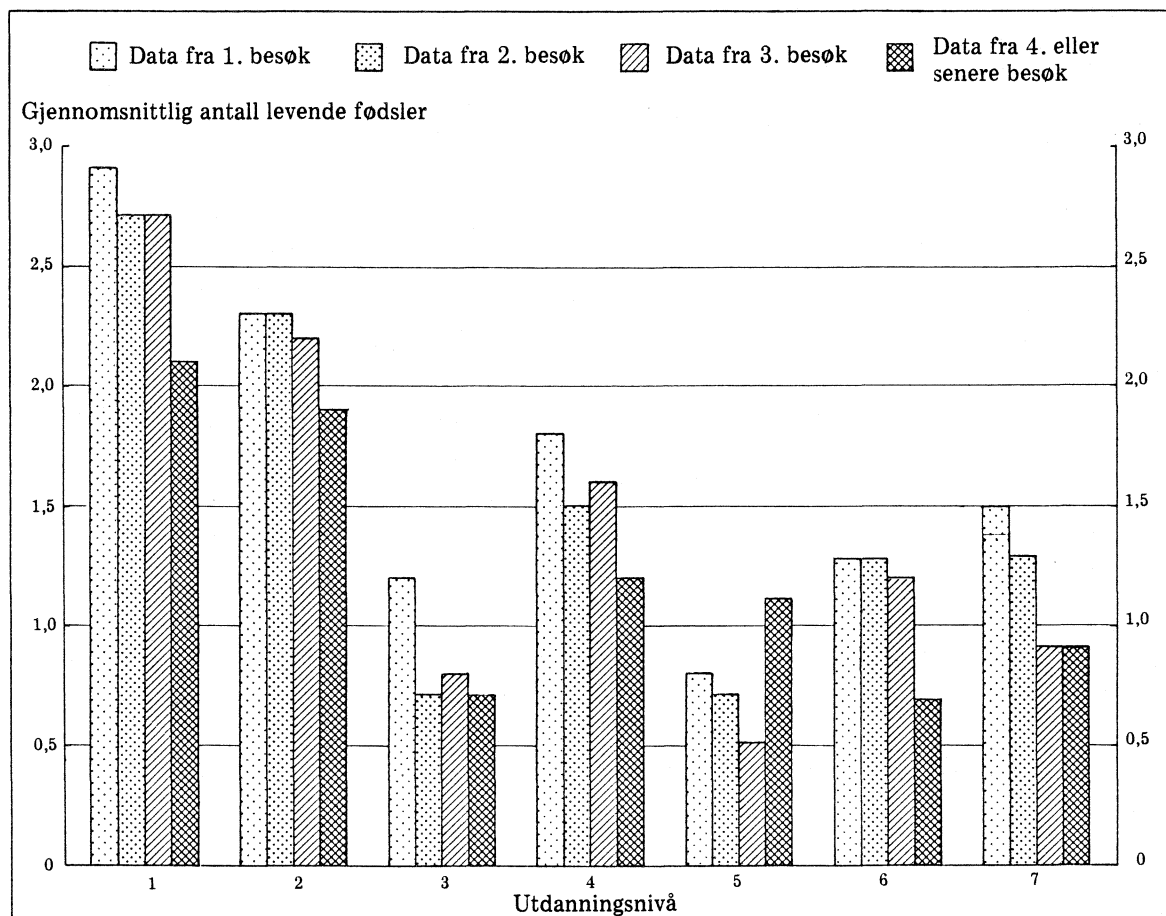


I fig. 2 er det gjort en tilsvarende studie av forholdet mellom fruktbarhet og utdanning. De som ble intervjuet i Fruktbarhetsundersøkelsen, ble klassifisert til en av følgende 7 grupper for høyeste fullførte utdanning:

- Gruppe 1: 7-årig folkeskole
 " 2: Framhaldskole
 " 3: Ungdomskolenivå
 " 4: Gymnasnivå I
 " 5: Artium
 " 6: Gymnasnivå II
 " 7: Universitet, høyskole

Fig. 2 indikerer at også analyse av sammenhengen mellom fruktbarhet og utdanning blir mindre påvirket av frafall enn analyse av bare fruktbarhet.

FIGUR 2. Gjennomsnittlig antall levende fødsler i ulike grupper for utdanning og antall besøk



Vi vil nå vise hvordan en kan studere den simultane sammenhengen mellom fruktbarhet, alder og utdanning, og se hvordan denne analysen blir påvirket av frafall. Den teknikken vi skal bruke er standardisering som foreslått i Pullum (1978).

Før vi begynner på analysen, skal vi gi en kort beskrivelse av modellen og estimeringsmetodene som er brukt. Vi er altså interessert i sammenhengen mellom fruktbarhet, Y (målt som antall levende fødsler) og forklaringsvariablene utdanning, X_1 , og alder X_2 . Vi antar at sammenhengen er lineær ved passende bruk av skalaer for de 3 variable:

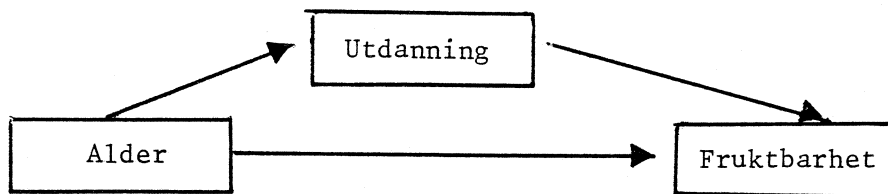
$$(3.1) \quad Y = b_1 X_1 + b_2 X_2 + C_0 + \text{residual}$$

Mellom alder og utdanning antar vi også en lineær sammenheng:

$$(3.2) \quad X_1 = b_{12} X_2 + C_1 + \text{residual}$$

Ligningene (3.1) og (3.2) definerer nå en kausal kjede. Under våre antagelser kan koeffisientene alle bli estimert ved minste kvadraters metode.

Modellen kan uttrykkes ved følgende diagram:

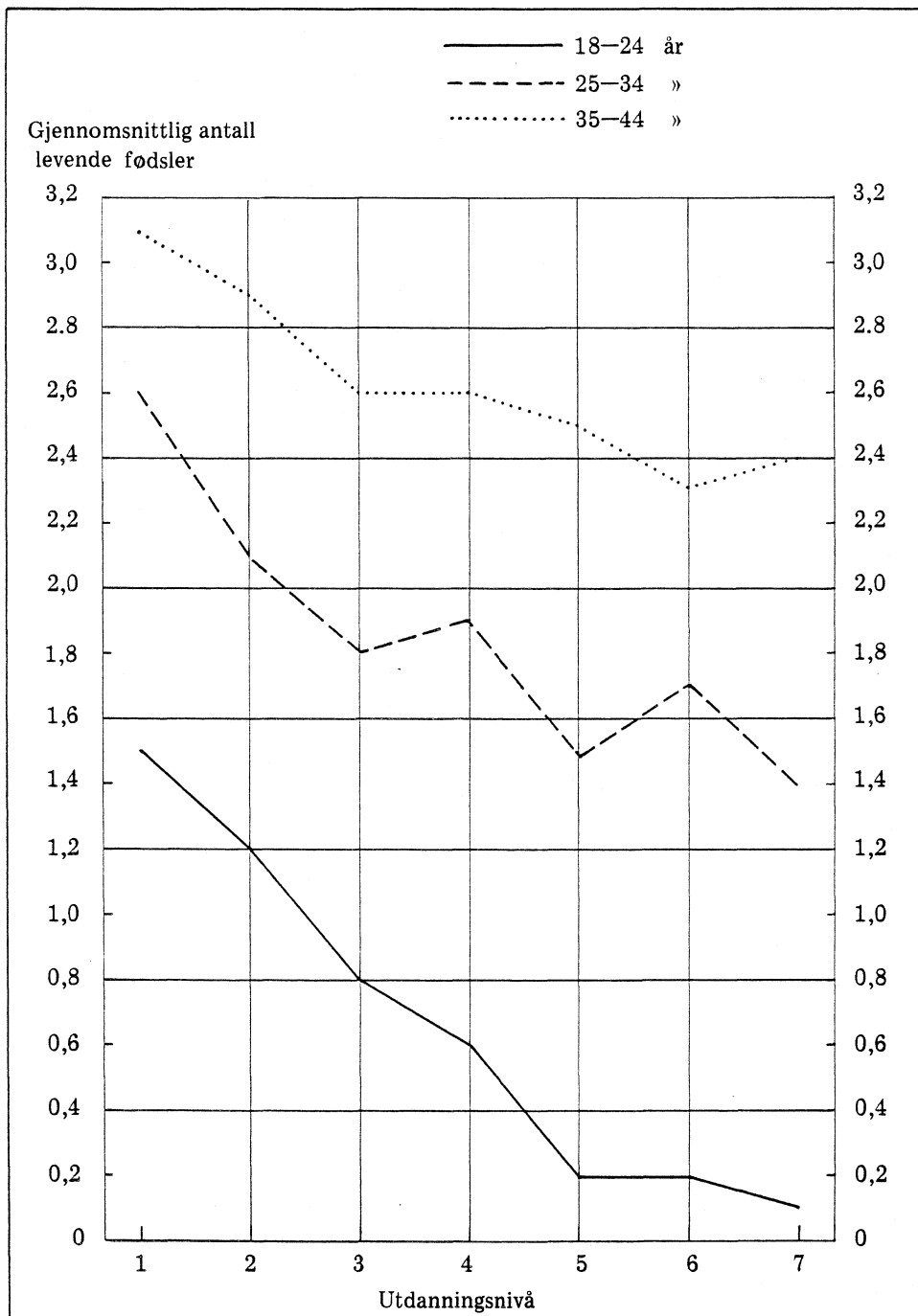


I Fruktbarhetsundersøkelsen er begge forklaringsvariablene målt som ordinale variable, og de er derfor transformert til dummy-variable. Dette vil ha liten betydning for analysen.

For å få en idé om hvilken effekt frafall har på en slik analyse, skal vi estimere nettoeffekten av utdanning på fruktbarhet med data innsamlet i henholdsvis 1. besøk, 2. besøk, 3. besøk og 4. og senere besøk.

Det første trinnet i analysen er å sjekke om det er samspill mellom alder og utdanning i deres effekt på den uavhengige variabelen, fruktbarhet. Fig. 3 viser grafisk hvordan fruktbarheten avtar med økende utdanningsnivå i alle tre aldersgruppene. Skjønt linjene ikke er helt parallelle, synes samspillet mellom alder og utdanning i deres effekter på fruktbarheten å være neglisjerbart, slik at det er akseptabelt å bruke standardisering her. (En formell test burde bli anvendt, men vi hopper over det her.)

FIGUR 3. Gjennomsnittlig antall levende fødsler etter utdanning i ulike aldersgrupper



Tabell 16 viser gjennomsnittlig antall levende fødsler i grupper for alder og utdanning.

Tabell 16. Gjennomsnittlig antall levende fødsler i grupper for alder og utdanning

Alder	I alt	Utdanningsnivå							Antall personer
		1	2	3	4	5	6	7	
I a l t	1,8	3,0	2,3	1,2	1,8	0,9	1,4	1,7	1 478
18-24 år	0,5	1,5	1,2	0,8	0,6	0,2	0,2	0,1	339
25-34 "	1,9	2,6	2,1	1,8	2,0	1,3	1,7	1,5	667
35-44 "	2,7	3,2	2,9	2,6	2,5	2,5	2,3	2,5	472
.....		121	189	98	678	97	94	201	

Vi definerer de totale effektene til en variabel som differansene mellom de marginale gjennomsnittene og det totale gjennomsnitt. For utdanning finner vi fra tabell 16 følgende marginale gjennomsnitt og totale effekter:

	Utdanningsnivå						
	1	2	3	4	5	6	7
Gjennomsnittlig antall levende fødsler	3,0	2,3	1,2	1,8	0,9	1,4	1,7
Totaleffekter av utdanning	1,2	0,5	-0,6	0,0	-0,9	-0,4	-0,1

Kvinnene i utdanningskategori 1 har altså i gjennomsnitt 1,2 flere levende fødsler enn i gjennomsnitt totalt.

Tilsvarende for alderskategoriene får vi følgende resultater:

	Alder		
	18-24 år	25-34 år	35-44 år
Gjennomsnittlig antall levende fødsler	0,5	1,9	2,7
Totaleffekter av alder ..	-1,3	0,1	0,9

Totaleffektene til variablene er i dette tilfellet (som i så mange andre) misvisende. I vår modell består totaleffekten av utdanning av to effekter, nemlig nettoeffekten av utdanning, og den indirekte effekten av alder på utdanning og fruktbarhet.

En mulig måte å estimere nettoeffekten av utdanning på er å holde aldersfordelingen "konstant" ved standardisering. (Denne teknikken er identisk med stianalyse med dummy-variable.) Det vil si at en anvender den totale aldersfordelingen innen hver utdanningsgruppe.

Det standardiserte gjennomsnittet i den første utdanningskategorien, er for eks.: $(1,5 \cdot 339 + 2,6 \cdot 667 + 3,2 \cdot 472)/1\ 478$. De standardiserte gjennomsnitt av antall levende fødsler er i de forskjellige utdanningsgruppene følgende:

	Utdanningsnivå						
	1	2	3	4	5	6	7
Gjennomsnittlig antall levende fødsler standardisert for alder	2,5	2,1	1,8	1,8	1,4	1,5	1,5

Det veide gjennomsnitt av de standardiserte gjennomsnittene er 1,7. Differansene mellom de standardiserte gjennomsnittene og det veide gjennomsnittet kan bli tolket som nettoeffektene av de forskjellige utdanningsnivåene på gjennomsnittlig antall levende fødsler (Pullum (1978)). Differansen er følgende

	Utdanningsnivå						
	1	2	3	4	5	6	7
Nettoeffekt av utdanning	0,7	0,3	0,0	0,0	-0,4	-0,3	-0,3

Vårt mål er ikke å gi en kritisk tolkning av disse nettoeffektene. Det som vi er interessert i, er å finne ut hvilke effekter frafall kan ha på en slik analyse. For å studere dette problemet, gjentar vi analysen med data innsamlet i henholdsvis 2. besøk, 3. besøk og 4. og senere besøk. De estimerte total- og nettoeffektene av utdanning er følgende:

Besøk nr.		Utdanningsnivå						
		1	2	3	4	5	6	7
2	Totaleffekt av utdanning	1,1	0,8	-0,9	-0,1	-0,8	-0,2	-0,2
	Nettoeffekt av utdanning	0,6	0,5	-0,4	0,0	-0,5	0,0	-0,4
3	Totaleffekt av utdanning	1,2	0,8	-0,7	0,1	-0,9	-0,2	-0,5
	Nettoeffekt av utdanning	0,4	0,6	0,1	0,1	-0,5	-0,1	-0,5
4. og senere	Totaleffekt av utdanning	1,0	0,8	-0,3	0,0	-0,1	-0,3	0,0
	Nettoeffekt av utdanning	-0,1	0,5	-0,1	0,0	0,3	-0,4	-0,2

Resultatene synes å indikere at frafallet har liten eller ingen systematisk effekt på estimeringen av nettoeffekten av utdanning.

Analysen som vi har gjort i dette avsnittet, har tatt sikte på å besvare spørsmålet om frafallet har noen effekt på forholdet mellom alder, utdanning og fruktbarhet. Resultatene indikerer at frafallet har liten eller ingen systematisk effekt ved analyse av forholdet mellom de tre variablene.

4. Referanser

- Heldal, J. (1980): Prøveundersøkelse i Folke- og bolig tellingen 1980. Rapport fra kvalitetskontrollundersøkelsen. Statistisk Sentralbyrå INo 80/3.
- Herigstad, H. (1978): Korrigering for fråfall ved forbruksundersøkinga. Statistisk Sentralbyrå. Stensil.
- Pullum, T.W. (1978): Standardization. World Fertility Survey. Technical Bulletin, No 31 Tech. 597.
- Thomsen, I. (1971): On the Effect of Non-Response in the Norwegian Election Survey 1969. Statistisk Tidsskrift, 3.
- Thomsen, I. og Siring, E.: On the Causes and Effects of Non-Response, Norwegian Experiences. Statistisk Sentralbyrå. Artikler nr. 121.