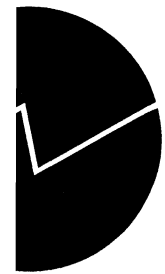


Anne Vedø

**Frafall i levekårspanelet
1980, 1983 og 1987**

Notater



1. INNLEDNING	2
2. DATA	3
3. KJØNN	8
3.1 FORDELING PÅ KJØNN OG FRAFALLSMØNSTER	8
3.2 FORDELING PÅ KJØNN OG DELTAKELSESTYPE	8
3.3 KONKLUSJON	10
4. ALDER	11
4.1 FORDELING PÅ ALDER OG FRAFALLSMØNSTER	11
4.2 FORDELING PÅ ALDER OG DELTAKELSESTYPE	12
4.3 KONKLUSJON	14
5. EKTESKAPELIG STATUS	15
5.1 FORDELING PÅ EKTESKAPELIG STATUS OG FRAFALLSMØNSTER	15
5.2 FORDELING PÅ EKTESKAPELIG STATUS OG DELTAKELSESTYPE	16
5.3 KONKLUSJON	17
6. UTDANNING	18
6.1 FORDELING PÅ UTDANNING OG FRAFALLSMØNSTER	18
6.2 FORDELING PÅ UTDANNING OG DELTAKELSESTYPE	19
6.3 KONKLUSJON	20
7. INNTEKT	21
7.1 FORDELING PÅ INNTEKT OG FRAFALLSMØNSTER	21
7.2 FORDELING PÅ INNTEKT OG DELTAKELSESTYPE	22
7.3 KONKLUSJON	23
8. BOSTED	24
8.1 FORDELING PÅ BOSTED OG FRAFALLSMØNSTER	24
8.2 FORDELING PÅ BOSTED OG DELTAKELSESTYPE	25
8.3 KONKLUSJON	26
9. PROBITANALYSE	27
REFERANSER	29
VEDLEGG	30
V.1 OM KONSTRUKSJON AV FILA PANEL 1 OG DEFINISJON AV NYE VARIABLE:	30
V.2 DEFINISJON AV DELTAKELSESVARIABLENE:	30
V.3 KJI-KVADRAT-TESTEN FOR UAVHENGIGHET:	32

1. Innledning

Et av problemene i utvalgsundersøkelser er frafallet, de uttrukne personene svarer ikke. Årsakene til frafall kan være mange. Noen nekter å svare av prinsipielle grunner, andre fordi de ikke har tid, noen er syke, og andre er bortreist. En kan velge å se bort fra frafallet, og basere tall og estimater på de dataene en har fått inn. Dersom personene i frafallet er omtrent som dem som har svart med hensyn på den størrelsen en undersøker, er dette ikke noe stort problem. Hvis det derimot er stor forskjell på disse to gruppene, vil en få dårlige estimater dersom frafallet ignoreres.

Det finnes mange metoder for å korrigere for frafall, f. eks. etterstratifisering, kalibrering av vektene og imputering av verdier i frafallsgruppen. I panelundersøkelser har en ofte tilgang til mer informasjon om personer som er frafall i en bestemt runde, fordi de kan ha svart i tidligere runder. For å få utnyttet denne informasjonen kreves det mer komplekse metoder for å korrigere for frafallet enn i andre undersøkelser.

I levekårsundersøkelsene i 80, 83, 87 og 91 er tabellene beregnet ut i fra nettoutvalget, uten hensyn til frafall. Det finnes imidlertid en tabell i hver undersøkelse som gjør det mulig å sammenligne bruttoutvalg, frafall, nettoutvalg og hele befolkningen med hensyn på kjønn, alder, ekteskapeleg status og landsdel. Disse tabellene bygger på hele levekårsutvalget, ikke bare panelet (se NOS B320 (1982), NOS B511 (1985), NOS B772 (1988)).

I dette notatet ser vi bare på panelet. Vi undersøker kjennemerkene kjønn, alder, ekteskapeleg status, utdanning, inntekt og bosted for å finne ut i hvilken grad de påvirker frafall.

2. Data

Levekårspanelet ble etablert i 1983. Det ble trukket på grunnlag av bruttoutvalget til Levekårsundersøkelsen 1980 og fulgt opp i 83, 87 og 91. Vi har altså data fra fire undersøkelser.

Dataene er organisert i tre filer for hver årgang: bruttoutvalgsfil, nettoutvalgsfil og frafallsfil.

- Bruttoutvalgsfilene inneholder bare noen få registeropplysninger om hver person. Personene er identifisert ved fødselsnummer.
- Nettoutvalgsfilene inneholder alle opplysningene fra spørreskjemaet. Personene er identifisert ved hjelp av et løpenummer.
- Frafallsfilene inneholder frafallsårsak for de personene som ikke har svart. Personene er identifisert ved fødselsnummer.

Ettersom det ikke finnes noen link mellom fødselsnummer og løpenummer, betyr dette at det på grunnlag av disse filene er umulig å koble sammen et fullstendig panel, dvs. et panel der vi for hver person og hvert år har frafallsårsak dersom personen er frafall og alle skjemaopplysninger dersom personen har svart.

Samtlige filer for 80, 83 og 87 er tilgjengelige, mens bruttoutvalg og frafall for 91 har vært umulig å oppdrive. På grunnlag av disse filene er det to måter å etablere et ufullstendig panel på:

Panel 1: Panel med frafallsopplysninger. Vi kan koble sammen bruttoutvalgsgfiler og frafallsfiler for 80, 83 og 87 etter fødselsnummer. Vi får da opplysninger om frafallsårsak i frafallet, men ingen opplysninger om de som har svart (utover det som står på bruttoutvalgsgfila). Vi får heller ingen data fra 91. Siden dette panelet inneholder fødselsnummer er det muligheter for å koble det opp mot registrene.

Panel 2: Panel med skjemaopplysninger: Vi kan koble sammen skjemafilene etter løpenummer. Vi får da data fra alle år, men vi mister de personene som aldri har svart, og vi får ingen opplysninger om frafallsårsak.

I denne analysen har vi valgt å koble opp Panel 1.

Vi har laget 4 nye *deltakelsesvariable* deltyp80-deltyp87 som tar verdiene «I» for intervjuet, «F» for frafall og «A» for avgang. I 83 og 87 ble panelet supplert med ca. 150 ungdommer for å holde aldersfordelingen i panelet riktig. Disse blir i det følgende omtalt som *tillegg*. For personer som kommer inn som tillegg senere år er hverken I, F eller A relevant, så disse får kode «tillegg83» eller «tillegg87». For noen personer gjelder at de ikke er IO, selv om de var med i panelet foregående år. Disse får kode «ikke IO». Med *deltakelsestypen* (eller *deltakelsen*) til en person et gitt år menes verdien til deltakelsesvariabelen dette året. I vedlegget finnes en mer detaljert beskrivelse av hvordan verdiene til deltakelsesvariablene er definert.

Sjekker vi Panel 1 mot opplysninger fra NOS-ene fra 83 og 87, finner vi at det er god overensstemmelse:

I NOS 83 side 16 finner vi følgende tall:

80\83	Intervjuet	Frafall	Sum
Intervjuet	1 409	263	1 672
Frafall	234	244	478
Sum	1 643	507	2 150

Konstrueres samme tabell fra Panel 1, finner vi

80\83	Intervjuet	Frafall	Sum
Intervjuet	1 410	263	1 673
Frafall	229	239	468
Sum	1 639	502	2 141

Det må sies å være god overensstemmelse mellom tabellene.

Vi kan gjøre tilsvarende sjekk med tall fra NOS 87.

Tall fra side 16 i NOS 87:

83\87	Intervjuet	Frafall	Sum
Intervjuet	1 411	263	1 674
Frafall	231	257	488
Sum	1 642	520	2 162

De samme tallene konstruert fra Panel 1:

83\87	Intervjuet	Frafall	Sum
Intervjuet	1 410	263	1 673
Frafall	225	250	475
Sum	1 635	513	2 148

Også her er det godt samsvar.

Frafallsmønsteret til en person beskrives av deltakelsestypen i 80, 83 og 87. En person som ble intervjuet i 80, var frafall i 83 og avgang i 87 har altså frafallsmønsteret IFA. Vi kan se på fordelingen av forskjellige frafallsmønstre i panel 1.

Frafallsmønstre for personer som deltok i 80, 83 og 87

Mønster	Antall	Prosent av 2175
III	1 146	52,7
IIF	182	8,4
IFI	145	6,7
IFF	90	4,1
FII	159	7,3
FIF	54	2,5
FFI	71	3,3
FFF	144	6,6
Sum	1 991	91,5

Prosentene refererer til panelet fra 80. De resterende 8,5 % av det opprinnelige panelet har enten ikke vært IO i 87 (154 personer, dvs. 7,1 %) eller de har et frafallsmønster som inneholder en avgang.

Under er en tabell over fordelingene på deltakelsestype de aktuelle årene.

År	Intervju	Frafall	Frafall i ekte panel	Avgang	Antall IO
1980	77,3	21,8	21,57	0,8	2 175
1983	76,2	22,7	22,76	1,0	2 323
1987	76,2	23,1	23,50	0,7	2 373

Frafallsprosenten er svakt økende, 21,8 i 1980, 22,7 i 83 og 23,1 i 87. Her er frafallsprosenten hvert år regnet av det totale antall IO vedkommende år, altså med tillegget inkludert i 83 og 87. Hvis vi ønsker å undersøke tretthet i panelet er det mer rimelig å regne ut frafallsprosentene i det ekte panelet, altså blant de 2021 personene som har vært IO både i 80, 83 og 87. Disse er 21,57, 22,76 og 23,50. Forskjellen mellom 21,57 og 23,50 er signifikant¹. I tillegg må en ta hensyn til at i dette panelet er intervallene mellom intervjuene på tre og fire år. Det er altså ingen stor belastning. I de nye levekårsundersøkelsene vil intervallene bli på ett år. Det vil muligens kunne føre til mer merkbare tretthetsreaksjoner i panelet.

¹ For utregning, se vedlegget.

Vi kan sammenligne frafallet i panelet med frafallet i tverrsnittsundersøkelsene. Tabellen under viser frafallsprosentene i tverrsnittsundersøkelsene, panelet og det ekte panelet i 80, 83 og 87. I NOS-ene er avgangene fjernet fra bruttoutvalget. For å få sammenlignbare tall har vi også fjernet avganger fra panelets bruttoutvalg. Det ekte panelets bruttoutvalg er her de som har vært IO alle tre år, og aldri har vært avgang.

Frafallsprosentene i tverrsnittsundersøkelsene, panelet og det ekte panelet. I parentes bruttoutvalgets størrelse, etter at avganger er fjernet.

År	Tverrsnittsundersøkelse	Panel	Ekte panel
1980	23,5 (5 080)	22,0 (2 157)	21,5 (1 991)
1983	21,8 (5 023)	23,0 (2 299)	22,6 (1 991)
1987	21,8 (5 051)	23,2 (2 357)	23,6 (1 991)

Vi ser at det er noe større frafall i paneldelen enn i hele tverrsnittsutvalget.

I NOS-ene finnes tabeller der frafallet i tverrsnittsundersøkelsene er fordelt etter kjønn, alder og ekteskapelig status. Vi har regnet ut de tilsvarende tallene for paneldelen.

Frafallet i tverrsnittsutvalgene og panelutvalgene i 80, 83 og 87 fordelt etter kjønn, alder og ekteskapelig status. Tallene er oppgitt i prosent.

	1980		1983		1987	
	Tverrsnitt (1 195 pers)	Panel (475 pers)	Tverrsnitt (1 094 pers)	Panel (528 pers)	Tverrsnitt (1 099 pers)	Panel (548 pers)
KJØNN OG ALDER						
Menn						
16-24 år	51,8	50,1	49,0	46,2	52,4	52,9
25-44 år	9,5	11,0	9,3	8,0	10,1	8,4
45-66 år	20,3	17,7	18,0	18,2	19,3	21,4
67-79 år	16,0	17,5	15,9	15,2	16,1	16,6
Kvinner						
16-24 år	6,0	4,0	5,8	4,9	6,9	6,6
25-44 år	48,2	49,9	51,0	53,8	47,6	47,1
45-66 år	9,5	9,1	8,4	8,7	9,0	7,1
67-79 år	15,5	17,3	16,1	17,6	14,4	13,9
	15,6	16,4	16,7	17,4	15,1	13,9
	7,7	7,2	9,8	10,0	9,1	12,2
EKTESKAPELIG STATUS						
Gift	56,2	54,9	54,7	54,9	-	52,9
Ikke gift	43,8	45,1	45,3	45,1	-	47,1

Vi kjenner ikke frafallets fordeling på ekteskapelig status i tverrsnittsundersøkelsen i 87.

Det er ikke store forskjeller mellom frafallet i tverrsnittsundersøkelsene og frafallet i panelet med hensyn på alder, kjønn og ekteskapelig status.

Det finnes også tabeller over hvordan menn og kvinner i frafallet i tverrsnittsundersøkelsen fordeler seg på frafallsårsak. Vi har gruppert frafallsårsakene i to kategorier, nekting og annen årsak. I tabellen under sammenligner vi nekting i tverrsnittsutvalget og panelet.

Prosentandel nektere i frafallet blant menn og kvinner.

	1983		1987	
	Tverrsnitt (1 094)	Panel (528)	Tverrsnitt (1 099)	Panel (548)
Alle	59	55	56	59
Menn	53	48	53	57
Kvinner	64	61	59	62

Med nekting menes i denne tabellen at IO enten nekter selv eller at andre nekter for IO. Vi ser at kvinner nekter mer enn menn. Dette gjelder både i 83 og 87, og både tverrsnitt og paneldel. Tabellen viser også at den totale nekterandelen gikk ned fra 83 til 87 i tverrsnittsundersøkelsen, mens den gikk opp i paneldelen.

Om de etterfølgende kapitlene:

Vi vil undersøke om vi kan finne noen sammenheng mellom en del vanlige kjennemerker og frafall i levekårspanelet. I de påfølgende kapitlene tar vi for oss ett og ett kjennemerke. Først ser vi på sammenhengen mellom kjennemerket og *fracfallsmønster*, deretter ser vi på sammenhengen mellom kjennemerket og *deltakelse*. Til slutt skiller vi ut frafallet og ser på sammenhengen mellom kjennemerket og *fracfallsgrunn*. Vi opererer bare med to frafallsårsaker, nemlig nekting og andre årsaker. I disse tabellene er nektere definert som personer med registrert frafallsgrunn nekting. Alle de andre personene det har blitt registrert en frafallsgrunn for regnes som annet frafall. Personer som mangler frafallsgrunn er ikke tatt med i tabellene. I 1980 var det 146 personer som manglet frafallsgrunn, i 1983 kun 26 personer. Ingen manglet frafallsgrunn i 1987. Dette betyr at konklusjonene som angår frafallsgrunn i 1980 er usikre.

For hver tabell bruker vi en kji-kvadrat-test for å undersøke om en eventuell sammenheng er signifikant. Testen er beskrevet nærmere i vedlegget. I tabellene med fordelinger på deltakelsestype har vi satt opp P-verdien også etter at avgangene er fjernet fra tabellen. Dette er fordi vi først og fremst er interessert i å undersøke sammenhengen mellom et kjennemerke og *fracfall*. Når kji-kvadrat-testen viser en signifikant avhengighet mellom deltakelsestype og et kjennemerke er det derfor nyttig å vite om avgangene er utslagsgivende for signifikansen.

3. Kjønn

3.1 Fordeling på kjønn og frafallsmønster

Tabell 3.1 Menn og kvinner fordelt på frafallsmønster

Kjønn/ Mønster	Menn	Kvinner	Totalt antall
FFF	7,37	7,10	144
FFI	3,58	3,55	71
FIF	3,38	2,07	54
FII	8,29	7,69	159
IFF	4,30	4,73	90
IFI	6,45	8,09	145
IIF	10,24	8,09	182
III	56,40	58,68	1 146
Totalt antall	977	1 014	1 991

Det er ingen signifikant sammenheng (p-verdi lik 0,298) mellom kjønn og frafallsmønster.

Tabell 3.2 Betingede sannsynligheter for frafall i 87 med hensyn på kjønn og tidligere frafallsmønster

Kjønn/ Tidligere mønster	Menn	Kvinner	Totalt
FF	0,67	0,67	0,67
FI	0,29	0,21	0,25
IF	0,40	0,37	0,38
II	0,15	0,12	0,14

De betingede sannsynlighetene er omtrent like for kvinner og menn.

Vi kan slå sammen mønstrene i grupper etter antall intervjuer.

Tabell 3.3 Prosentvis fordeling på antall intervjuer for menn og kvinner

Ant. intervjuer/ Kjønn	0	1	2	3	Totalt antall
Menn	7,37	11,26	24,97	56,40	977
Kvinner	7,10	10,36	23,87	58,68	1 014
Totalt antall	144	215	486	1 146	1 991

Vi ser at prosentfordelingen er svært lik for menn og kvinner. Det er derfor ingen avhengighet mellom antall intervjuer og de intervjuedes kjønn.

3.2 Fordeling på kjønn og deltakelsestype

I dette avsnittet undersøker vi hver årgang av panelet for seg. Først ser vi på sammenhengen mellom kjønn og deltakelse, deretter på sammenhengen mellom kjønn og frafallsårsak.

3.2.1 1980

Tabell 3.4 Menn og kvinner fordelt på deltakelsestype

Deltakelse/ Kjønn	Intervju	Frafall	Avgang	Totalt antall
Menn	76,69	22,28	1,03	1 068
Kvinner	77,96	21,41	0,63	1 107
Totalt antall	1 682	475	18	2 175

P-verdi i kji-kvadrat-test: 0,511. Uten avganger: 0,586

Igjen ser vi svært lik prosentfordeling for menn og kvinner, dvs. kjønn har ingen betydning for frafallet. Det er flere menn enn kvinner blant avgangene (11 mot 7). Det er for få avganger til å si noe sikkert, men dette kan skyldes at menn dør tidligere enn kvinner.

Vi kan dele frafallet opp i nektete og andre.

Tabell 3.5 Menn og kvinner fordelt på frafallsgrunn

Frafallsgrunn/ Kjønn	Nektete	Annet fracfall	Totalt antall
Menn	57,14	42,86	161
Kvinner	69,64	30,36	168
Totalt antall	209	120	329

Sammenhengen mellom kjønn og frafallsgrunn er signifikant (p-verdi lik 0,019). Vi ser at kvinner nekter mer enn menn i denne undersøkelsen.

3.2.2 1983

Tabell 3.6 Menn og kvinner fordelt på deltakelsestype

Deltakelse/ Kjønn	Intervju	Frafall	Avgang	Totalt antall
Menn	77,47	21,31	1,22	1 145
Kvinner	75,04	24,11	0,85	1 178
Totalt antall	1 771	528	24	2 323

Det er heller ikke signifikant sammenheng mellom kjønn og frafall i 83 (p-verdi lik 0,199, uten avganger 0,118).

Tabell 3.7 Menn og kvinner fordelt på frafallsgrunn

Frafallsgrunn/ Kjønn	Nektete	Annet fracfall	Totalt
Menn	46,98	53,02	232
Kvinner	60,00	40,00	270
Totalt antall	271	231	502

Vi får samme resultat som for 80, kvinner nekter mer enn menn.

3.2.3 1987

Tabell 3.8 Menn og kvinner fordelt på deltakelsestype

Deltakelse/ kjønn	Intervju	Frafall	Avgang	Totalt antall
Menn	75,02	24,56	0,42	1 181
Kvinner	77,43	21,64	0,92	1 192
Totalt antall	1 809	548	16	2 373

Fortsatt har ikke kjønn noen stor betydning for frafallssannsynligheten (p-verdi 0,09, uten avganger 0,106).

Tabell 3.9 Menn og kvinner fordelt på frafallsgrunn

Frafallsgrunn/ Kjønn	Nektete	Annet frafall	Totalt antall
Menn	52,76	47,24	290
Kvinner	58,91	41,09	258
Totalt antall	305	243	548

Fortsatt nekter kvinner mer enn menn, men utslaget er ikke signifikant (p-verdi lik 0,148).

3.3 Konklusjon

Det synes ikke være noen forskjell mellom kvinner og menn med hensyn til frafallsmønsteret som helhet, eller med hensyn til sannsynligheten for å falle fra i et gitt år. I 80 og 83 fant vi imidlertid at kjønn hadde betydning for frafallsårsak, idet kvinner nektet mer enn menn disse årene. I 87 derimot var denne sammenhengen ikke lenger signifikant selv om tendensen gikk i samme retning.

4. Alder

Det er ikke urimelig å tenke seg at alder og frafall henger sammen. For eksempel kan en forestille seg at unge mennesker er lite hjemme, og derfor vanskelige å treffe.

4.1 Fordeling på alder og frafallsmønster

Tabell 4.1 Aldersgrupper fordelt på frafallsmønster

Alder i 80/ Mønster	16-24	25-44	45-66	67-80	Totalt antall	Gj.sn. alder
FFF	6,63	5,27	9,73	7,43	144	44,8
FFI	5,48	3,64	2,58	3,38	71	37,4
FIF	4,03	2,38	2,29	3,38	54	41,3
FII	10,37	8,41	6,72	6,08	159	39,2
IFF	5,19	4,27	4,43	4,73	90	41,4
IFI	9,80	7,03	6,44	6,76	145	39,3
IIF	9,51	8,91	9,16	9,46	182	41,7
III	48,99	60,10	58,66	58,78	1 146	42,3
Totalt antall	347	797	699	148	1 991	41,7

Uavhengighetstesten viser en signifikant (p-verdi lik 0,038) sammenheng mellom alder og frafallsmønster. I aldersgruppen 16-24 år er det færre som ble intervjuet alle tre år enn i de andre aldersgruppene. I gruppen 25-44 år er det få som er frafall alle år, og mange som er intervjuet alle år. 45-66-åringene har størst sannsynlighet for frafall alle tre år. Høyeste gjennomsnittsalder finner vi i gruppen med frafallsmønster FFF, laveste i gruppen med mønster FFI.

Tabell 4.2 Sannsynlighetene for frafall gitt alder i 87 og tidligere frafallsmønster.

Alder i 87/ Mønster	16-24	25-44	45-66	67-80	Totalt	Gj.sn. alder
FF	1,00	0,54	0,72	0,82	0,67	49,4
FI	0,17	0,22	0,27	0,32	0,25	46,7
IF	0,33	0,37	0,36	0,46	0,38	47,1
II	0,29	0,13	0,14	0,14	0,14	49,3

Tallene i kolonnen under 16-24 år er svært upålitelige, da bare 38 personer i det ekte panelet var i denne aldersgruppen i 87. I gruppen med to frafall varierer sannsynligheten for enda et frafall mye med alder. I de tre andre gruppene ser ikke alder ut til å ha fullt så stor betydning.

Vi slår sammen frafallsmønstrene etter antall intervjuer.

Tabell 4.3 Aldersgrupper fordelt på antall intervjuer

Alder i 80/ Ant. intervjuer	16-24	25-44	45-66	67-80	Totalt antall	Gj.sn. alder
0	6,63	5,27	9,73	7,43	144	44,8
1	14,70	10,29	9,30	11,49	215	40,0
2	29,68	24,34	22,32	22,30	486	40,2
3	48,99	60,10	58,66	58,78	1 146	42,3
Totalt antall	148	699	797	347	1 991	41,7

Her ser vi enda klarere forskjellen i frafallsmønster for de fire aldersgruppene (p-verdi 0,001).

4.2 Fordeling på alder og deltakelsestype

4.2.1 1980

Tabell 4.4 Aldersgrupper fordelt på deltakelsestype

Alder/ Deltakelse	16-24	25-44	45-66	67-80	Totalt antall	Gj.sn. alder
Intervju	72,45	79,17	77,82	76,98	1 682	43,5
Frafall	26,17	20,34	21,64	21,03	475	42,8
Avgang	1,38	0,49	0,54	1,98	18	45,2
Totalt antall	363	816	744	252	2 175	43,4

Tabellen viser en forskjell mellom den yngste aldersgruppen og de andre (p-verdi lik 0,035).

Vi kan se om det er noen forskjeller på aldersgruppene i forhold til frafallsgrunn.

Tabell 4.5 Aldersgrupper fordelt på frafallsgrunn

Alder/ Frafallsgrunn	16-24	25-44	45-66	67-80	Totalt antall	Gj.sn. alder
Nektene	45,76	62,17	75,38	54,05	209	46,3
Annet frafall	54,24	37,86	24,62	45,95	120	40,9
Totalt antall	59	103	130	37	329	44,4

Her er det store forskjeller (p-verdi lik 0,00057). Vi ser at nektning er en mye vanligere frafallsårsak blant 45-66-åringene enn blant de andre, og spesielt sammenlignet med de yngste, 16-24 år.

4.2.2 1983

Tabell 4.6 Aldersgrupper fordelt på deltakelsestype

Alder/ Deltakelse	16-24	25-44	45-66	67-80	Totalt antall	Gj.sn. alder
Intervju	75,34	77,97	75,75	73,82	1 771	44,3
Frafall	23,85	21,58	23,31	23,24	528	44,8
Avgang	0,81	0,46	0,95	2,94	24	57,4
Totalt antall	369	876	738	340	2 323	44,6

Det er ikke noen stor forskjell i deltagelse mellom de fire gruppene (p-verdi 0,011, uten avganger 0,712), og mindre utslag enn for 1980.

Vi deler opp frafallet i nektere og andre.

Tabell 4.7 Aldersgrupper fordelt på frafallsgrunn

Alder/ Frafallsgrunn	16-24	25-44	45-66	67-80	Totalt antall	Gj.sn. alder
Nektene	34,52	49,16	67,07	58,67	271	47,8
Annet frafall	65,48	50,84	32,93	41,33	231	41,4
Totalt antall	84	179	164	75	502	44,9

Tendensen her er den samme som for 1980 (p-verdi 0,0000077). Det er relativt sett flest nektene i aldersgruppen 45-66 år og klart færrest nektene i aldersgruppen 16-24 år.

4.2.3 1987

Tabell 4.8 Aldersgrupper fordelt på deltakelsestype

Alder/ Deltakelse	16-24	25-44	45-66	67-80	Totalt antall	Gj.sn. alder
Intervju	78,01	78,02	75,93	70,71	1 809	44,1
Frafall	21,74	21,42	23,79	27,18	548	45,7
Avgang	0,26	0,55	0,28	2,11	16	55,2
Totalt antall	391	901	702	379	2 373	44,5

Mønsteret ser annerledes ut i 1987. Det er en tendens til større frafall for de eldste (p-verdi lik 0,002, uten avganger 0,097).

Tabell 4.9 Aldersgrupper fordelt på frafallsgrunn

Alder/ Frafallsgrunn	16-24	25-44	45-66	67-80	Totalt antall	Gj.sn. alder
Nektene	38,82	51,30	65,87	61,17	305	48,0
Annet frafall	61,18	48,70	34,13	38,83	243	42,7
Totalt antall	85	193	167	103	548	45,7

Fremdeles nekter 16-24-åringene minst og 45-66-åringene mest (p-verdi 0,00021).

4.3 Konklusjon

Vi fant en sammenheng mellom frafallsmønster og alder. Aldersgruppen 16-24 år er den gruppen der det er vanskeligst å oppnå intervju alle tre år. 45-66-åringene har oftere frafall alle tre år enn andre aldersgrupper. Dette medvirker til at gruppen med frafallsmønster FFF har høyere gjennomsnittsalder enn de andre frafallsmønster-gruppene.

Da vi analyserte dataene år for år fant vi ikke sterk sammenheng mellom alder og frafall. Derimot fant vi at gruppen 45-66 år nektet mye mer enn gruppen 16-24 år. Dette gjaldt alle tre år.

5. Ekteskapelig status

5.1 Fordeling på ekteskapelig status og frafallsmønster

Ekteskapelig status kan endre seg med tid. Vi har tilgang til ekteskapelig status i 83 og 87, og har valgt å bruke ekteskapelig status i 83 i tabell 5.1 og tabell 5.3.

Tabell 5.1 Gifte og ikke gifte fordelt på frafallsmønster

Sivilstand/ Mønster	Gift	Ikke gift	Totalt antall
FFF	6,90	7,87	144
FFI	2,84	4,96	71
FIF	2,15	3,79	54
FII	6,51	10,79	159
IFF	4,60	4,37	90
IFI	6,13	9,48	145
IIF	8,20	10,93	182
III	62,68	47,81	1 146
Totalt antall	1 305	686	1 991

Vi ser at gifte har mye høyere sannsynlighet (p-verdi 0,000) for å svare tre ganger enn ugifte.

Tabell 5.2 Sannsynlighetene for frafall i 87 gitt ekteskapelig status i 87 og tidligere frafallsmønster

Status i 87/ Mønster i 80/83	Gift	Ikke gift	Totalt
FF	0,69	0,64	0,67
FI	0,21	0,33	0,25
IF	0,40	0,36	0,38
II	0,11	0,20	0,14

Hovedinntrykket er at gitt frafall i 83 er det mindre frafall blant ikke gifte enn gifte, mens gitt at de svarer i 83 er det langt høyere frafall blant dem som ikke er gift.

Vi slår sammen frafallsmønstrene i grupper etter antall intervjuer.

Tabell 5.3 Gifte og ikke gifte fordelt på antall intervjuer

Ant. intervjuer/ Sivilstand	0	1	2	3	Totalt antall
Gift	6,90	9,58	20,84	62,68	1 305
Ikke gift	7,87	13,12	31,20	47,81	686
Totalt antall	144	215	486	1 146	1 991

Denne tabellen viser enda klarere at gifte har klart mindre frafall enn ikke gifte (p-verdi lik 0,000).

5.2 Fordeling på ekteskapelig status og deltakesestype

5.2.1 1980

I 1980 har vi bare opplysninger om ekteskapelig status for frafallet, og kan derfor ikke gjennomføre noen analyse.

5.2.2 1983

Tabell 5.4 Gifte og ikke gifte fordelt på deltakesestype

Deltakelse/ Sivilstand	Intervju	Frafall	Avgang	Totalt antall
Gift	78,87	20,77	0,36	1 396
Ikke gift	72,28	25,67	2,05	927
Totalt antall	1 771	528	24	2 323

Sannsynligheten for frafall er høyere blant dem som ikke er gift enn blant gifte (p-verdi lik 0,000, uten avganger 0,003).

Tabell 5.5 Gifte og ikke gifte fordelt på frafallsgrunn

Frafallsgrunn/ Sivilstand	Nektere	Annet frafall	Totalt antall
Gift	62,23	37,77	278
Ikke gift	43,75	56,25	224
Totalt antall	271	231	502

Gifte har oftere nekting som frafallsgrunn enn ikke gifte (p-verdi lik 0,000), men det kan skyldes at ikke gifte er vanskeligere å treffe.

5.2.3 1987

Tabell 5.6 Gifte og ikke gifte fordelt på deltakesestype

Deltakelse/ Sivilstand	Intervju	Frafall	Avgang	Totalt antall
Gift	78,70	20,94	0,36	1 385
Ikke gift	72,77	26,11	1,11	988
Totalt antall	1 809	548	16	2 373

Mønsteret er helt identisk med 1983 (p-verdi lik 0,0008, uten avganger 0,002).

Tabell 5.7 Gifte og ikke gifte fordelt etter frafallsgrunn

Frafallsgrunn/ Sivilstand	Nektere	Annet frafall	Totalt antall
Gift	65,86	34,14	290
Ikke gift	44,19	55,81	258
Totalt antall	305	243	548

Vi ser at de gifte fortsatt nekter mer enn dem som ikke er gift (p-verdi lik 0,000).

5.3 Konklusjon

Sannsynligheten for frafall er høyere for ikke gifte enn for gifte. Dette henger antageligvis sammen med at gifte er mer hjemme enn ikke gifte, og dermed lettere å få tak i. Blant dem som er frafall er andelen nektere mye høyere blant de gifte enn de andre. Dette fenomenet er vanskelig å forklare. Det virker ikke rimelig å anta at det er flere prinsipielle nektere blant gifte enn blant ikke gifte. En mulig forklaring er at vi har «skjult nekting», dvs. at en del av dem som blir registrert som «annet frafall» fordi de ikke er hjemme, ville ha nektet dersom de hadde vært hjemme. Siden de gifte som nevnt er lettere å treffe hjemme, vil vi få mindre skjult nekting blant gifte enn blant ikke gifte.

6. Utdanning

6.1 Fordeling på utdanning og frafallsmønster

Tabell 6.1 Utdanningsgrupper fordelt på frafallsmønster

Utdanning 87/ Mønster	0-9 år	10-12 år	13-16 år	Over 16 år	Totalt antall
FFF	8,47	7,37	4,45	1,69	140
FFI	3,78	3,05	4,45	5,08	70
FIF	3,03	2,21	3,77	0,00	52
FII	7,26	8,53	7,88	6,78	156
IFF	5,75	4,11	3,42	1,69	88
IFI	7,11	7,89	6,16	8,47	145
IIF	11,35	8,32	7,88	3,39	179
III	53,25	58,53	61,99	72,88	1 132
Totalt antall	661	950	292	59	1 962

Det klareste utslaget er at andelen som er med alle tre år øker betydelig med lengden på utdanningen.

Tabell 6.2 Sannsynlighetene for frafall i 87 gitt utdanning i 87 og tidligere frafallsmønster

Utdanning 87/ Tidligere mønster	0-9 år	10-12 år	13-16 år	Over 16 år	Totalt
FF	0,69	0,71	0,50	0,25	0,67
FI	0,29	0,21	0,32	0,00	0,25
IF	0,45	0,34	0,36	0,17	0,38
II	0,18	0,12	0,11	0,04	0,14

De fire første kolonnene er basert på de 1962 personene i det ekte panelet som har utdanningskode i 87. I kolonnen for utdanning over 16 år er tallene svært usikre. Bare 4 personer i denne gruppen var frafall i 87. Det ser ut til at frafallet i 87 avtar betydelig med lengden på utdanning uavhengig av mønsteret for 80/83.

Tabell 6.3 Utdanningsgrupper fordelt på antall intervjuer

Utdanning 87/ Ant. intervjuer	0-9 år	10-12 år	13-16 år	Over 16 år	Totalt antall
0	8,47	7,37	4,45	1,69	140
1	12,56	9,37	11,64	6,78	210
2	25,72	24,74	21,92	18,64	480
3	53,25	58,53	61,99	72,88	1 132
Totalt antall	661	950	292	59	1 962

Vi ser at sannsynligheten for tre intervjuer øker med utdanningsnivå (p-verdi 0,020), noe vi allerede så fra tabell 6.1.

6.2 Fordeling på utdanning og deltakelsestype

6.2.1 1980

Tabell 6.4 Utdanningsgrupper fordelt på deltakelsestype

Utdanning / Deltakelse	0-9 år	10-12 år	13-16 år	Over 16 år	Totalt antall
Intervju	76,74	78,08	78,10	82,69	1659
Frafall	22,06	21,43	21,49	17,31	461
Avgang	1,20	0,50	0,41	0,00	16
Totalt antall	834	1008	242	52	2136

39 personer mangler utdanningskode i 80.

Det er ikke signifikant sammenheng mellom deltakelse og utdanningsnivå dette året (p-verdi lik 0,583, uten avganger 0,850).

Tabell 6.5 Utdanningsgrupper fordelt på frafallsårsak

Utdanning / Frafallsgrunn	0-9 år	10-12 år	13-16 år	Over 16 år	Totalt antall
Nektere	69,53	60,26	64,52	71,43	208
Annet frafall	30,47	39,74	35,48	28,57	114
Totalt	128	156	31	7	322

14 personer i frafallet mangler utdanningskode i 80.

Det er ikke signifikant avhengighet mellom utdanningsnivå og frafallsgrunn (p-verdi lik 0,425).

6.2.2 1983

Det finnes ikke noe register over høyeste fullførte utdanning i 83.

6.2.3 1987

Tabell 6.6 Utdanningsgrupper fordelt på deltakelsestype

Utdanning / Deltakelse	0-9 år	10-12 år	13-16 år	Over 16 år	Totalt antall
Intervju	72,10	77,79	79,75	91,67	1 788
Frafall	26,73	21,88	20,25	6,67	534
Avgang	1,17	0,34	0,00	1,67	14
Totalt antall	767	1 193	316	60	2 336

37 personer manglet utdanningskode i 87.

Her ser vi at frafallssannsynligheten synker med økende utdanning (p-verdi lik 0,00019, uten avganger 0,00047). Blant dem med over 16 års utdanning er den helt nede i 6,67 %.

Tabell 6.7 Utdanningsgrupper fordelt på frafallsgrunn

Utdanning / Frafallsgrunn	0-9 år	10-12 år	13-16 år	Over 16 år	Totalt antall
Nektere	57,07	56,70	50,00	75,00	300
Annet frafall	42,93	43,30	50,00	25,00	234
Totalt antall	205	261	64	4	534

14 personer i frafallet i 87 mangler utdanningskode for 87.

Heller ikke dette året finner vi avhengighet mellom utdanningsnivå og frafallsgrunn (p-verdi lik 0,645).

6.3 Konklusjon

Sannsynligheten for tre intervjuer øker med økende utdanning. I 87 fant vi at frafallssannsynligheten sank med økende utdanning. Utdanningsnivået hadde ikke betydning for frafallsårsak.

7. Inntekt

I tabellene under er inntektsgrupperingen valgt slik at det blir omtrent like mange personer i hver inntektsgruppe. Inntektene er oppgitt i kroner.

7.1 Fordeling på inntekt og frafallsmønster

Vi har valgt å bruke inntekten i 87 i tabellene i dette avsnittet, ettersom 232 personer i det ekte panelet ikke står i ligningsregisteret for 83.

Tabell 7.1 Inntektsgrupper fordelt på frafallsmønster

Inntekt 87/ Mønster	Ingen inntekt	100- 55 000	55 000- 125 000	125 000- 175 000	Over 175 000	Totalt antall	Gj.sn. inntekt
FFF	10,51	8,21	7,18	4,87	5,35	144	80 500
FFI	2,63	2,61	3,74	4,87	3,79	71	117 069
FIF	4,04	1,49	3,16	1,39	2,90	54	95 863
FII	7,68	7,09	6,32	9,28	8,91	159	111 242
IFF	6,26	4,85	4,60	3,48	3,34	90	85 944
IFI	6,06	8,58	9,20	7,89	5,79	145	101 886
IIF	11,52	6,72	7,18	10,21	8,46	182	105 425
III	51,31	60,45	58,62	58,00	61,47	1 146	109 029
Totalt antall	495	268	348	431	449	1 991	105 179

P-verdi: 0,005

Tabell 7.2 Betingede sannsynligheter for frafall gitt inntekt i 87 og tidligere frafallsmønster

Inntekt 87/ Tidligere mønster	Ingen inntekt	100- 55 000	55 000- 125 000	125 000- 175 000	Over 175 000	Totalt	Gj.sn. inntekt
FF	0,80	0,76	0,66	0,50	0,59	0,67	92 576
FI	0,34	0,17	0,33	0,13	0,25	0,25	107 343
IF	0,51	0,36	0,33	0,31	0,37	0,38	95 781
II	0,18	0,10	0,11	0,15	0,12	0,14	108 535

Tabell 7.3 Inntektsgrupper fordelt på antall intervjuer

Inntekt 87/ Ant. intervjuer	Ingen inntekt	100- 55 000	55 000- 125 000	125 000- 175 000	Over 175 000	Totalt antall	Gj.sn. inntekt
0	10,51	8,21	7,18	4,87	5,35	144	80 500
1	12,93	8,96	11,49	9,74	10,02	215	98 714
2	25,25	22,39	22,70	27,38	23,16	486	106 272
3	51,31	60,45	58,62	58,00	61,47	1 146	109 029
Totalt antall	495	268	348	431	449	1 991	105 179

Vi ser at gjennomsnittsinntekten øker med antall intervjuer. Den viktigste forskjellen går mellom dem uten inntekt og resten. Mens kun halvparten av de uten inntekt deltar alle tre ganger så er det 6 av 10 blant de andre som deltar alle tre årene og det er liten forskjell mellom inntektsgruppene (p-verdi lik 0,016).

7.2 Fordeling på inntekt og deltakelsestype

7.2.1 1980

Tabell 7.4 Inntektsgrupper fordelt på deltakelsestype

Inntekt / Deltakelse	Ingen inntekt	100-30 000	30 000-70 000	70 000-95 000	Over 95 000	Totalt antall	Gj.sn. inntekt
Intervju	74,30	78,55	76,98	79,70	79,42	1 678	51 242
Frafall	23,09	21,20	22,80	20,30	20,58	470	49 964
Avgang	2,61	0,25	0,23	0,00	0,00	16	3 931
Totalt antall	537	401	443	404	379	2 164	50 614

11 personer stod ikke i ligningsregisteret for 1980.

Når vi ser bort fra avgangene, finner vi ingen avhengighet mellom deltakelse og inntekt i 1980 (p-verdi lik 0,676).

Tabell 7.5 Inntektsgrupper fordelt på frafallsårsak

Inntekt / Frafallsårsak	Ingen inntekt	100-30 000	30 000-70 000	70 000-95 000	Over 95 000	Totalt antall	Gj.sn. inntekt
Nektere	60,71	58,62	70,15	71,88	54,72	207	50 443
Annet frafall	39,29	41,38	29,85	28,13	45,28	119	53 221
Totalt antall	84	58	67	64	53	326	51 457

5 personer i frafallet stod ikke i ligningsregisteret.

Det er heller ikke sterk sammenheng mellom inntekt og frafallsgrunn dette året (p-verdi lik 0,210).

7.2.2 1983

Tabell 7.6 Inntektsgrupper fordelt på deltakelsestype

Inntekt / Deltakelse	Ingen inntekt	100-40 000	40 000-90 000	90 000-125 000	Over 125 000	Totalt antall	Gj.sn. inntekt
Intervju	70,07	76,60	78,62	79,14	78,26	1 506	78 265
Frafall	28,52	22,66	21,15	20,86	21,50	441	71 887
Avgang	1,41	0,74	0,23	0,00	0,24	9	27 011
Totalt antall	284	406	435	417	414	1 956	76 591

367 personer stod ikke i ligningsregisteret for 83. Dette er 16 % av IO-ene.

Vi får lavere p-verdi enn i 80, men sammenhengen mellom deltakelse og inntekt er fremdeles ikke sterkt signifikant når vi tar vekk avgangene (p-verdi lik 0,096). Dessuten mangler vi inntektsopplysninger for en relativt stor del av dataene. Sett i lys av den tilsvarende tabellen for 1987, aner vi likevel en tendens til høyere frafall i gruppen uten inntekt. Vi ser også at gjennomsnittsinntekten er lavere blant frafallet enn blant de intervjuede.

Tabell 7.7 Inntektsgrupper fordelt på frafallsårsak

Inntekt / Frafallsårsak	Ingen inntekt	100- 40 000	40 000- 90 000	90 000- 125 000	Over 125 000	Totalt antall	Gj.sn. inntekt
Nektere	57,89	55,56	60,92	47,06	54,32	231	68 956
Annet frafall	42,11	44,44	39,08	52,94	45,68	188	73 659
Totalt antall	76	90	87	85	81	419	71 067

87 personer i frafallet stod ikke i ligningsregisteret for 83. Dette er 16 % av frafallet.

Her ser vi heller ingen sammenheng mellom inntekt og frafallsgrunn (p-verdi lik 0,451).

7.2.3 1987

Tabell 7.8 Inntektsgrupper fordelt på deltakelsestype

Inntekt / Deltakelse	Ingen inntekt	100- 55 000	55 000- 125 000	125 000- 175 000	Over 175 000	Totalt antall	Gj.sn. inntekt
Intervju	66,23	77,66	78,34	80,43	80,09	1809	101 324
Frafall	31,52	21,69	21,44	19,57	19,91	548	86 690
Avgang	2,25	0,65	0,22	0,00	0,00	16	5 319
Totalt antall	533	461	457	465	457	2373	97 297

Her får vi en klar sammenheng mellom inntekt og deltakelse (p-verdi lik 0,00000, uten avganger 0,00000). Vi ser at frafallssannsynligheten synker med økende inntekt, men fortsatt er det de uten inntekt som skiller seg ut.

Tabell 7.9 Inntektsgrupper fordelt på frafallsårsak

Inntekt / Frafallsårsak	Ingen inntekt	100- 55 000	55 000- 125 000	125 000- 175 000	Over 175 000	Totalt antall	Gj.sn. inntekt
Nektere	54,17	52,00	59,18	60,44	53,85	305	87 407
Annet frafall	45,83	48,00	40,82	39,56	46,15	243	85 790
Totalt antall	168	100	98	91	91	548	86 690

Det er fremdeles ingen sammenheng mellom inntekt og frafallsgrunn (p-verdi lik 0,708).

7.3 Konklusjon

Det var signifikant avhengighet mellom frafallsmønster og inntekt. Personer helt uten inntekt ble sjeldnere intervjuet tre ganger enn andre. I 80 og 83 fikk vi ikke signifikant sammenheng mellom inntekt og deltagelse, men vi kunne likevel se en tendens til høyere frafall blant dem uten inntekt. I 87 var det helt klart at frafallssannsynligheten sank med økende inntekt. Vi så ingen sammenheng mellom inntekt og frafallsgrunn noen av årene.

8. Bosted

Det har tidligere vist seg at det er større frafall i byer enn på landet. Dette kan henge sammen med at personer bosatt i større byer er mindre hjemme enn de som bor på landet eller i mindre byer. I dette kapittelet er «storby» definert som Oslo, Bergen, Trondhjem, Kristiansand, Stavanger og Tromsø.

8.1 Fordeling på bosted og frafallsmønster

Tabell 8.1 Personer bosatt i og utenfor storby fordelt på frafallsmønster

Bosted 83/ Mønster	Storby	Resten	Totalt antall
FFF	9,78	6,35	144
FFI	3,52	3,58	71
FIF	3,91	2,30	54
FII	8,61	7,77	159
IFF	5,68	4,12	90
IFI	6,46	7,57	145
IIF	7,83	9,59	182
III	54,21	58,72	1 146
Totalt antall	511	1480	1 991

Sammenhengen mellom bosted og frafallsmønster er signifikant (p-verdi lik 0,030). Personer bosatt i storby har oftere et frafallsmønster med mange frafall enn personer utenfor storby.

Tabell 8.2 Betingede sannsynligheter for frafall gitt bosted i 87 og tidligere frafallsmønster

Bosted 87/ Tidligere mønster	Storby	Resten	Totalt
FF	0,72	0,67	0,67
FI	0,32	0,23	0,25
IF	0,48	0,35	0,38
II	0,13	0,14	0,14

Når en person har svart i både 80 og 83 er frafallssannsynligheten uavhengig av bosted. For de tre andre mønstrene er frafallssannsynligheten høyest for dem som bor i storby. Den største forskjellen ser vi blant dem som ble intervjuet i 80 og var frafall i 83.

Tabell 8.3 Personer bosatt i og utenfor storby fordelt på antall intervjuer

Ant. intervjuer/ Bosted	0	1	2	3	Totalt antall
Storby	9,78	13,11	22,90	54,21	511
Resten	6,35	10,00	24,93	58,72	1 480
Totalt antall	144	215	486	1146	1 991

P-verdi: 0,009

Vi ser at de som bor i storby oftere svarer ingen eller en gang enn resten.

8.2 Fordeling på bosted og deltakelsestype

8.2.1 1980

Vi har ikke tilgang på opplysninger om bosted i 1980.

8.2.2 1983

Tabell 8.4 Personer bosatt i og utenfor storby fordelt på deltakelsestype

Deltakelse/ Bosted	Intervju	Frafall	Avgang	Totalt antall
Storby	72,41	26,09	1,51	598
Resten	77,57	21,57	0,87	1 725
Totalt antall	1 771	528	24	2 323

Det er signifikant sammenheng mellom deltagelse og bosted dette året (p-verdi lik 0,026, uten avganger 0,019). Det er høyere frafall i storbyene enn ellers i landet.

Tabell 8.5 Personer bosatt i og utenfor storby fordelt på frafallsårsak

Frafallsgrunn/ Bosted	Nektere	Annet frafall	Totalt antall
Storby	49,33	50,67	150
Resten	55,97	44,03	352
Totalt antall	271	231	502

Bostedets betydning for frafallsårsak er ikke signifikant (p-verdi lik 0,172).

8.2.3 1987

Tabell 8.6 Personer bosatt i og utenfor storby fordelt på deltakelsestype

Deltakelse/ Bosted	Intervju	Frafall	Avgang	Totalt antall
Storby	72,58	26,34	1,08	558
Resten	77,36	22,09	0,55	1 815
Totalt antall	1 809	548	16	2 373

Også dette året er sammenhengen mellom deltagelse og bosted signifikant (p-verdi lik 0,042, uten avganger 0,032). Tabellen er nesten identisk med tilsvarende tabell for 83.

Tabell 8.7 Personer bosatt i og utenfor storby fordelt på frafallsårsak

Frafallsgrunn/ Bosted	Nektere	Annet frafall	Totalt antall
Storby	45,58	54,42	147
Resten	59,35	40,65	401
Totalt antall	305	243	548

Dette året er det signifikant avhengighet mellom bosted og frafallsgrunn (p-verdi lik 0,004). Personer bosatt utenfor storbyene har oftere nekting som frafallsårsak enn personer som bor i storbyene.

8.3 Konklusjon

Vi fant sterk sammenheng mellom frafall og bosted både i 83 og 87. Personer som bor i storbyene falt oftere fra enn personer som bor utenfor. Nekting var en vanligere frafallsårsak blant folk ellers enn blant folk i storbyene, både i 83 og 87, men det var bare i 87 at sammenhengen var signifikant.

9. Probitanalyse

I de foregående kapitlene har vi sett på hvordan variablene kjønn, alder, ekteskapelig status, utdanning, inntekt og bosted hver for seg påvirker sannsynligheten for frafall. Vi så også på hvordan disse variablene hang sammen med frafallsårsak. I dette kapitlet undersøker vi hvordan alle variablene sammen påvirker frafallssannsynlighet og frafallsårsak. Vi bruker følgende probit-modell: (se for eksempel McCullagh and Nelder: Generalized Linear Models)

$$\Phi^{-1}(p) = \alpha + \sum_{i=1}^8 x_i^T \beta_i$$

der

Φ er den normale kumulative fordelingsfunksjonen

p er sannsynligheten for frafall (evt. sannsynligheten for å være nektek gitt at en er i frafallet)

$x_1^T = (x_{11}, x_{12})$ er indikatorvariable for kjønn

$x_2^T = (x_{21}, x_{22}, x_{23}, x_{24})$ er indikatorvariable for aldersgruppe (samme gruppering som tidligere)

$x_3^T = (x_{31}, x_{32})$ er indikatorvariable for ekteskapelig status (gift, ikke gift)

$x_4^T = (x_{41}, x_{42}, x_{43}, x_{44})$ er indikatorvariable for utdanningsgruppe (samme gruppering som tidligere)

$x_5^T = (x_{51}, x_{52}, x_{53}, x_{54}, x_{55})$ er indikatorvariable for inntekt (samme gruppering som tidligere)

$x_6^T = (x_{61}, x_{62})$ er indikatorvariable for bosted

$x_7^T = (x_{71}, x_{72})$ er indikatorvariable for frafall i 80 (evt. nekting i 80). Denne brukes bare i analysen for 83.

$x_8^T = (x_{81}, x_{82}, x_{83}, x_{84})$ er indikatorvariable for frafallmønster i 80 og 83 (evt. mønster for nekting). Denne brukes bare i analysen for 87.

I disse analysene tar vi bare med personene i det ekte panelet, dvs. de som har vært IO både i 80, 83 og 87. Vi ser først hvordan forklaringsvariablene påvirker deltagelse.

Tabell 9.1 P-verdier for forklaringsvariablene når avhengig variabel er deltagelse

År/ Variabel	1980	1983	1987
Kjønn	0,1158	0,1114	0,0008
Alder	0,2230	0,2203	0,2522
Ekteskapelig status	Ikke tilgj	0,0437	0,0111
Utdanning	0,5688	Ikke tilgj.	0,0495
Inntekt	0,4516	0,0294	0,0005
Bosted	Ikke tilgj.	0,4773	0,1478
Frafall i 80	Uaktuelt	0,0001	Uaktuelt
Tidligere mønster	Uaktuelt	Uaktuelt	0,0001

Tallene er basert på 1 980 observasjoner i 80, 1 775 observasjoner i 83 og 1 988 observasjoner i 87.

I 1980 er det ingen av variablene som påvirker sannsynligheten for deltagelse. I 83 er ekteskapelig status, inntekt og frafall i 80 signifikante, og i 87 har både kjønn, ekteskapelig status, utdanning, inntekt og tidligere frafallsmønster signifikant betydning for deltagelse. Dette kan tolkes som en

paneleffekt. Først etter at personene har vært med en tid, begynner de forskjellige gruppene å utvikle forskjellig svarsansynlighet. En annen mulighet er at denne utviklingen er et fenomen i tiden, dvs. at vi ville ha sett det samme selv om undersøkelsen ikke hadde vært en panelundersøkelse.

Vi bruker så den samme modellen til å undersøke hvordan variablene henger sammen med nekting.

Tabell 9.2 P-verdier for forklaringsvariablene når avhengig variabel er frafallsgrunn

År/ Variabel	1980	1983	1987
Kjønn	0,0365	0,6036	0,0526
Alder	0,0463	0,9376	0,1069
Ekteskapelig status	0,0024	0,6622	0,4763
Utdanning	0,1447	Ikke tilgj.	0,3802
Inntekt	0,1584	0,3560	0,4859
Bosted	0,0155	0,0464	0,0002
Nekter i 80	Uaktuelt	0,0001	Uaktuelt
Tidligere mønster	Uaktuelt	Uaktuelt	0,0083

Med tidligere mønster menes her tidligere mønster for nekting. Variablen tar de fire verdiene NN, N*, *N, og **, der N står for nekting og * står for alt annet (dvs. frafall med annen grunn enn nekting, intervju eller avgang). Første posisjon representerer 1980, annen posisjon 1983.

Personer som mangler verdier for frafallsgrunn eller en forklaringsvariabel er ikke tatt med. Tallene er basert på 294 observasjoner i 80, 144 i 83 og 116 i 87.

I 1980 har både kjønn, alder, ekteskapelig status og bosted betydning for frafallsgrunn, mens i 83 og 87 er det bare bosted som slår ut. Ettersom vi har utelatt alle observasjoner som mangler verdi for en forklaringsvariabel eller avhengig variabel, er tallene basert på relativt få observasjoner. Vi kan derfor ikke trekke for sterke slutninger.

Noen av resultatene fra probit-analysen virker tilsynelatende inkonsistente med resultatene fra tabellene i tidligere kapitler. For eksempel fant vi i kapittel 8 at bosted hadde signifikant sammenheng med deltagelse både i 83 og 87, mens dette ikke var tilfelle i probit-analysen. Disse forskjellene kan oppstå av flere årsaker. P-verdien i krysstabellen over bosted og deltagelse uttrykker hvor sterk sammenheng det er mellom deltagelse og bosted *alene*. I probit-analysen uttrykker p-verdien for bosted hvor stor forklaringskraft bosted har på deltagelse *etter* at de andre variablene i analysen har forklart sitt. En sterk forklaringsvariabel som «tidligere frafall» kan lett slå ut andre variable som er korrelert med tidligere frafall. Dessuten brukte vi i probit-analysen bare ekte panel med alle variable tilstede. I tabellene i de foregående kapitlene tok vi utgangspunkt i alle IO vedkommende år, og vi trengte bare å utelate observasjoner som manglet verdi for det kjennemerket tabellen angikk. Dette merkes spesielt godt i probit-analysen over frafallsgrunn, ettersom svært mange manglet frafallsgrunn i 1980. I tillegg bruker vi i probit-analysen en modellantakelse som vi ikke bruker i kji-kvadrat-testen.

Referanser

Statistisk Sentralbyrå (1982): *Levekårsundersøkelsen 1980*. Norges offisielle statistikk B 320.

Statistisk Sentralbyrå (1985): *Levekårsundersøkelsen 1983*. Norges offisielle statistikk B 511.

Statistisk Sentralbyrå (1988): *Levekårsundersøkelsen 1987*. Norges offisielle statistikk B 772.

P. McCullagh and J.A. Nelder (1989): *Generalized Linear Models*. Chapman and Hall.

Vedlegg

V.1 Om konstruksjon av fila Panel 1 og definisjon av nye variable:

1. Frafallsfilene fra 80, 83 og 87 og utvalgsfilene fra 83 og 87 linkes etter fødselsnummer.

2. Definerer nye variable:

FAAR: Fødselsår tatt fra fødselsnummeret

PANEL80: Levekårsfilene inneholder alle IO-ene i levekårsundersøkelsen, ikke bare panelet. Panelet var ikke planlagt i 80, så det finnes ingen panelmerker på 80-filene. Panelet ble trukket i 83, fra bruttoutvalget i 80. På utvalgsfila for 83 er det en variabel PANEL83 som markerer hvilke personer som er med i panelet. På utvalgsfila for 83 er det også en variabel MERKE83 som viser hvilke personer som var IO i 83 (utvalgsfilene inneholder av administrative årsaker flere personer enn dem som skal intervjues). Av ukjente årsaker er det 210 personer som ifølge PANEL83 skulle være med i panelet, men som ikke har IO-merke for 83. Disse 210 personene har heller ikke noe panelmerke på utvalgsfila for 87, og ikke en eneste en av dem befinner seg på frafallsfila for 80 (da vi ikke har utvalgsfila for 80 er det umulig å sjekke om noen av dem befinner seg der). Vi regner derfor med at de ikke har vært IO i 80 og at det er en feiltagelse at de har fått panelkode i 83. PANEL80 skal være en 0-1 variabel som er 1 hvis og bare hvis personen var IO i 80 og ble trukket ut til å være med i panelet i 83. Vi setter derfor:

PANEL80=1 hvis PANEL83=P og MERKE83=X (2174 personer), PANEL80=0 ellers.

TILLEGG83: Ifølge NOS fra 83 ble panelet supplert med 150 16-18-åringer i 83 for å opprettholde aldersfordelingen. Ser vi på aldersfordelingen til dem som har panelkode i 83, oppdager vi lett at ingen er født etter 63. Dette betyr at tillegget ikke har fått noen panelkode. På utvalgsfila for 87 er det derimot en variabel PANEL87 som har verdien EP (ekte panel) for personer som har vært med i panelet før 87. Vi kan derfor sile ut tillegget i 83 ved å sette:

TILLEGG83=1 hvis PANEL87=EP og FAAR=64, 65 eller 66 (148 personer), TILLEGG83=0 ellers.

TILLEGG87: Tillegget i 87 er lett å finne, ettersom variabelen PANEL87 på utvalgsfila for 87 markerer uekte panel. Vi setter:

TILLEGG87=1 hvis PANEL87=UP, TILLEGG87=0 ellers.

3. Panelet defineres nå som de personer der PANEL80=1 eller TILLEGG83=1 eller TILLEGG87=1. Dette utgjør 2527 personer.

V.2 Definisjon av deltakelsesvariablene:

DELTYP80:

1. IO

A. Intervjuede

Hvis PANEL80=1 og personen ikke befinner seg på frafallsfila for 80 settes DELTYP80 lik I.

B. Frafall og avgang

Dette er personene på frafallsfila. Her har vi to variable å gå ut fra, AVGANG80 og FRAFALLS-GRUNN80. Det er mye missing begge steder, men likevel greier disse variablene å bli inkonsistente.

19 personer har AVGANG80 lik 1. 6 av disse har frafallsgrunn langvarig sykdom i 80, 1 er på arbeid e.l., 1 nekter, 3 har kode 08 som ikke finnes på skjemaet, 7 har kode 12 «annet, spesifiser» og en er missing. Vi vil definere avgang som utenfor populasjon, og i så fall må sykdom, nekting og arbeid regnes som frafall. Vi regner med at avgangsvariabelen er mer pålitelig enn frafallsgrunnen siden avgang er fylt ut i byrået og frafallsgrunn av intervjuer. Vi definerer derfor:

Hvis AVGANG80=1 settes DELTYP80=A, ellers settes DELTYP80=F.

2. Ikke IO

Dette gjelder tillegget i 83 og 87.

Hvis TILLEGG83=1 settes DELTYP80 = tillegg83.

Hvis TILLEGG87=1 settes DELTYP80 = tillegg87.

DELTYP83:

1. IO

A. Intervjuede

Bruker samme prinsipp som i 80, de IO-ene som ikke er på frafallsfila anses som intervjuet. Hvis MERKE83=X og personen ikke er på frafallsfila for 83 settes DELTYP83 lik I.

B. Frafall og avgang

Hvis personen er på frafallsfila for 83 må vi skille mellom frafall og avgang. Frafallsfila fra 83 inneholder ingen variabel avgang83, så her må vi holde oss til frafallsgrunn 83. Det er 26 personer på frafallsfila for 83 som mangler frafallsgrunn.

Hvis FRAFALLSGRUNN 83 er 80 (død), 90 (flyttet til utlandet), 94 (bor på institusjon) eller 95 (annen avgang) settes DELTYP83 lik A, ellers settes den lik F.

2. Ikke IO

Dette gjelder tillegget i 87.

Hvis TILLEGG87=1 settes DELTYP83= tillegg87.

DELTYP87:

1. IO

A. Intervjuede

Hvis PANEL87=EP eller PANEL87=UP og personen ikke er på frafallsfila for 87 settes DELTYP87 lik I. I teorien skulle hele panelet oppfylle enten PANEL87=EP eller PANEL87=UP, men i praksis ser det ut til at endel personer har falt ut av IO-listen i 87, ettersom de har blank panelkode.

B. Frafall og avgang

Hvis frafallsgrunn i 87 er 80, 90, 94 eller 95 (avgangsgrunnene) settes DELTYP87 lik A. Ellers settes DELTYP87 lik F.

Dette fører til at de med frafallsgrunn 60 (Mangler opplysninger fra intervjuer) og 65 (Kostnader, mangler intervjuere o.l.) blir klassifisert som frafall.

2. Ikke IO

Hvis PANEL87 hverken er EP eller UP settes DELTYP87= 'ikke IO'. Dette gjelder 154 personer.

V.3 Kji-kvadrat-testen for uavhengighet:

Kji-kvadrat-testen tester avhengighet mellom to variable. Anta at vi ønsker å teste om det er avhengighet mellom de to variablene X og Y , der X er delt inn i gruppene 1 til R og Y er delt inn i gruppene 1 til C . Vi kan sette opp følgende to-veis-tabell:

X Y	1	...	j	...	R	Totalt
1	n_{11}		n_{1j}		n_{1R}	n_{1o}
M.						
i	n_{i1}		n_{ij}		n_{iR}	n_{io}
M.						
C	n_{C1}		n_{Cj}		n_{CR}	n_{Co}
Totalt	n_{o1}		n_{oj}		n_{oR}	n

Her betegner n_{ij} antall observasjoner i X -gruppe i og Y -gruppe j .

Vi vil teste H_0 : X og Y er uavhengige

La $m_{ij} = n_{i0}n_{0j} / n$. Under H_0 er

$$Q = \frac{\sum_i \sum_j (n_{ij} - m_{ij})^2}{m_{ij}}$$

kji-kvadrat-fordelt med $(R-1)(C-1)$ frihetsgrader. La q være tallet man får når man setter inn dataene i formelen for Q . P-verdien er sannsynligheten for at $Q \geq q$ under H_0 .

Fotnote 1:

Det er 2021 personer som var IO i både 80, 83 og 87. 25 av disse var avganger enten i 80 eller 87 eller begge deler. For å forenkle regningen under har vi valgt å se bort fra disse personene. Siden det dreier seg om såpass få personer har dette ingen betydning for resultatet.

La sannsynligheten for at et IO skal bli frafall i hhv. 80 og 87 være p_{80} og p_{87} . Vi vil teste

$$H_0: p_{87} - p_{80} = 0 \text{ mot } H_1: p_{87} - p_{80} > 0$$

Antall IO i ekte panel = $n = 1996$

Antall frafall i 80 i ekte panel = $Y_{80} = 430$

Antall frafall i 87 i ekte panel = $Y_{87} = 473$

Vi kan tenke oss en multinomisk forsøksrekke med 4 mulige utfall:

II: Intervju i både 80 og 87. La dette ha sannsynlighet π_{II} .

IF: Intervju i 80, frafall i 87. La dette ha sannsynlighet π_{IF} .

FI: Frafall i 80, intervju i 87. La dette ha sannsynlighet π_{FI} .

FF: Frafall i både 80 og 87. La dette ha sannsynlighet π_{FF} .

Under har vi satt opp en tabell med disse sannsynlighetene.

87/ 80	I	F	Sum
I	π_{II}	π_{IF}	π_{I0}
F	π_{FI}	π_{FF}	π_{F0}
Sum	π_{0I}	π_{0F}	1

Her er $\pi_{F0} = p_{80}$ og $\pi_{0F} = p_{87}$.

Fra dataene får vi følgende opplysninger:

87/ 80	I	F	Sum
I	n_{II}	n_{IF}	n_{I0}
F	n_{FI}	n_{FF}	n_{F0}
Sum	n_{0I}	n_{0F}	n

Her betegner n_{II} antall personer som har blitt intervjuet både i 80 og 87, n_{IF} er antall personer som ble intervjuet i 80 men var frafall i 87 osv.

Vi ser at $Y_{80} = n_{F0}$ og $Y_{87} = n_{0F}$.

Vi har at

$$Z = \frac{Y_{87} - Y_{80}}{\hat{S}(Y_{87} - Y_{80})}$$

der $\hat{S}(Y_{87} - Y_{80})$ er det estimerte standardavviket til $Y_{87} - Y_{80}$, er tilnærmet standard normalfordelt, dvs. med forventning lik 0 og standardavvik lik 1 under H_0 .

For å finne en estimator for standardavviket til $Y_{87} - Y_{80}$ regner vi først ut variansen.

$$V(Y_{87} - Y_{80}) = V(Y_{87}) + V(Y_{80}) - 2Cov(Y_{80}, Y_{87})$$

Etter litt regning finner vi at

$$Cov(Y_{80}, Y_{87}) = n\pi_{FF} - np_{87}p_{80}$$

Dermed er

$$V(Y_{87} - Y_{80}) = np_{87}(1 - p_{87}) + np_{80}(1 - p_{80}) - 2(n\pi_{FF} - np_{87}p_{80})$$

Under H_0 er $p_{80} = p_{87} = p$. Setter vi inn dette i ligningen over får vi at

$$V(Y_{87} - Y_{80}) = 2n\pi_{IF}$$

Under H_0 er $\pi_{IF} = \pi_{FI}$, så vi estimerer variansen til $Y_{87} - Y_{80}$ med

$$\hat{V}(Y_{87} - Y_{80}) = 2n \frac{n_{IF} + n_{FI}}{2n} = n_{IF} + n_{FI}$$

Vi ender dermed opp med testobservatoren

$$Z = \frac{Y_{87} - Y_{80}}{\sqrt{n_{IF} + n_{FI}}}$$

Fra dataene får vi følgende tabell

87/ 80	I	F	Sum
I	1292	274	1566
F	231	199	430
Sum	1523	473	1996

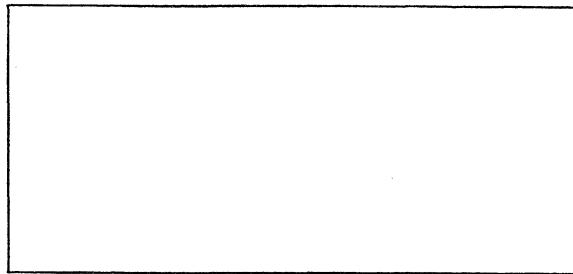
$$Z = \frac{473 - 430}{\sqrt{274 + 231}} = 1,91$$

Dette har en p-verdi på 0,028.

De sist utgitte publikasjonene i serien Notater

- 97/11 A. Barstad: Frihetens århundre? Levekår i Norge i et 100-årsperspektiv. 37s.
- 97/12 G. Sparby: Inntekts- og formuesundersøkelsen 1992: Dokumentasjon. 101s.
- 97/13 V. Pedersen: Inntekts- og formuesundersøkelsen 1993: Dokumentasjon. 94s.
- 97/14 V. Pedersen: Inntekts- og formuesundersøkelsen 1994: Dokumentasjon. 93s.
- 97/15 Metodevalg og kostnader ved etablering og drift av et boligregister. 29s.
- 97/16 K. Vassenden: Innvandrersstatistikkprosjektet: Styringsgruppas evaluering. 34s.
- 97/17 E. Heilund, A.M.K. Holmøy, A.C. Steen og B.L. Western: Beregning av vektorer til inntekts- og formuesundersøkelsene 1994. 30s.
- 97/18 S. Størmer: Sammenhengen mellom sysselsetting i AKU og registrene: Måling av kvalitet og prediksjon. 36s.
- 97/19 H.M. Teigum: Verdiundersøkelsen 1996: Dokumentasjonsrapport. 84s.
- 97/20 T. Ouren og T. Vik: Prosjektrapport: Voksenopplæringsprosjektet 1995-1996. 24s.
- 97/21 S. Hansen og T. Skoglund: Beregning av sysselsetting og lønn i nasjonalregnskapet. 36s.
- 97/22 H. Lövkvist: Standardiserte rater - en metodebeskrivelse med eksempler fra dødsårsaksstatistikken. 45s.
- 97/23 T. Vik og T. Ouren: Voksenopplæringsprosjektet: Dokumentasjon av databasen og rutiner. 34s.
- 97/24 A.S. Andersen: Gerix-data: Gir de grunnlag for å vurdere inntektssystemet for kommunene? 58s.
- 97/25 Ø. Amundrud og I. Tuveng: Utredning av utvalgsplan for sentral sykefraværstatistikk. 36s.
- 97/26 J. Lyngstad: Innvanderers demografi og levekår. 38s.
- 97/27 L. Rogstad, P. Schønning, M.V. Dysterud og S. Homstvedt: Arealstatistikk i Norge: Resultater fra en brukerundersøkelse. 30s.
- 97/28 H.N. Næsheim og I. Tuveng: Muligheter for å få yrkesdata i registerbaserte statistikker. 37s.
- 97/29 J. Rodriguez: Sesongjustering i praksis - en innføring: April 1997. 71s.
- 97/30 K-G. Lindquist: Database for energiintensive næringer: Tall fra industristatistikken. 17s.

Notater



B Returadresse:
Statistisk sentralbyrå
Postboks 8131 Dep.
N-0033 Oslo

Statistisk sentralbyrå

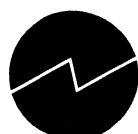
Oslo:
Postboks 8131 Dep.
0033 Oslo

Telefon: 22 86 45 00
Telefaks: 22 86 49 73

Kongsvinger:
Postboks 1260
2201 Kongsvinger

Telefon: 62 88 50 00
Telefaks: 62 88 50 30

ISSN 0806-3745



Statistisk sentralbyrå
Statistics Norway