

Frafallsproblemet øker

10

Stadig flere vegrer seg mot å svare når de blir bedt om å delta i helseundersøkelsen og andre surveyundersøkelser; i 1995 endte hvert fjerde besøk i et avslag. Folk som bor alene nekter oftest å delta. Frafallet har sannsynligvis ikke påvirket resultatene i vesentlig grad. Skal man redusere frafallet, må man oppsøke folk flere ganger.

Liv Belsby

Allerede på 1920- og 1930-tallet ble frafallsproblemer oppdaget i utvalgsundersøkelser. I 1951 peker Cochran på at frafall i personutvalg kan skape skjevhet siden de som ikke vil være med ofte atskiller seg fra resten av utvalget. For eksempel er "travle" yngre mennesker sjeldnere hjemme og mindre tilgjengelige for intervjuing enn mer etablerte barnefamilier.

Generell tendens

Generelt har frafall økt i spørreundersøkelser. Trolig henger dette sammen med at holdningen i samfunnet til spørreundersøkelser har endret seg over tid. På 1950- og 1960-tallet var de fleste positive til spørreundersøkelser. Denne tiden var preget av stor tro på samfunnsplanlegging. Motstanden økte på 1960- og 1970-tallet (Särndal mfl. 1992). Kanskje henger dette sammen med sterkere fokusering på individet og mindre tro på at et gode for fellesskapet nødvendigvis er et gode for den enkelte. Mange er dessuten vanskelig å få tak i eller har liten tid fordi de tilbringer mindre tid hjemme i dag enn tidligere. Dessuten konkurrerer nå mange spørreundersøkelser om tiden til intervjuobjektet.

Denne artikkelen handler om frafall i intervjudelen av Helseundersøkelsen 1995. Den postale delen, som inneholder de mest følsomme spørsmålene, har et større frafall enn intervjudelen.

Svarandelen i helseundersøkelsen har avtatt siden 1968

Svarprosentene i helseundersøkelsene gjenspeiler trenden mot lavere oppslutning om utvalgsundersøkelser; i 1968 var svarprosenten 94, i 1975 var den 89, i 1985 sank den til 79, og i 1995 var svarprosenten 75. I tillegg til endring i holdninger og hvor opptatt folk er, kan også det voksende antallet spørsmål ha påvirket svarprosenten.

I tabellen nedenfor presenteres frafall og oppgitte frafallsårsaker for henholdsvis Helseundersøkelsen 1975 og 1995.

Ønsker ikke å delta er den vanligste frafallsårsaken, og det er også her årsakene til det meste av frafallskningen i de siste 20 årene ligger.

Personer som bor alene, svarer sjeldnere

Hva karakteriserer personer som ofte svarer, eller vice versa de som

Tabell 1: Frafall og frafallsårsaker i Helseundersøkelsen 1975 og 1995

Årsak til frafall	Helseundersøkelsen 1975		Helseundersøkelsen 1995	
	Antall personer	Prosent av utvalget	Antall personer	Prosent av utvalget
Ønsker ikke å delta	684	5,50	2 061	15,09
Bortreist, eller borte pga. reiser, arbeid, studier e.l.	285	2,29	165	1,21
Andre årsaker til at personen ikke er å treffe	256	2,06	178	1,30
Andre årsaker til frafall	199	1,60	1 013	7,42
Totalt antall som ikke svarer	1 424	11,45	3 417	25,02
Totalt antall i utvalget	12 434		13 656	

ikke svarer? Opplysninger fra registre gir mulighet for å knytte informasjon til de personene som ikke er med i svardelen av utvalget. Trolig ligner frafallsmønsteret i helseundersøkelsene på frafallsmønsteret i andre person- og husholdningsundersøkelser i Statistisk sentralbyrå. Vi har derfor sett på slike frafallsanalyser. I en analyse av forbruksundersøkelsen fant Belsby (1995) at enpersonhusholdninger hadde større sannsynlighet for frafall enn husholdninger med flere personer. Personer bosatt i tettbygd strøk var også mindre villige til å være med i undersøkelsen enn de som var bosatt i griseværdte strøk. Ved en analyse av frafall i Levekårsundersøkelsen 1980, 1983 og 1987, fant Vedø (1997) bl.a. at enslige hadde lavere sannsynlighet for å svare enn andre. Bild mfl. (1997) fant at i Statistisk sentralbyrås undersøkelse om småbarnsfamiliers erfaringer med sykdom og helsetjenester, hadde personer som mottok

sosialhjelp og personer som var født i den 3. verden, større sannsynlighet for frafall enn andre. Videre fant de en tendens til at personer med lav inntekt og lav utdanning hadde større sannsynlighet for frafall.

Motivert av resultatene i disse analysene inkluderte vi følgende registervariabler i frafallsanalysen¹:

- antall personer i husholdningen

- alder til den eldste i husholdningen
- høyeste utdanning i husholdningen
- om noen i husholdningen mottok sosialhjelp
- overgangsstønad
- landbakgrunn etter verdensregioninndeling
- tettbygd/spredtbygd bosted
- den høyeste personinntekten i husholdningen²

Figur 2: Sannsynligheter for å være med i Helseundersøkelsen 1995 for forskjellige grupper¹

Husholdningsstørrelse²

- 1 person
- 2 personer
- 3 personer
- 4 personer
- 5 personer eller flere

Høyeste utdanning i familien

- 0-9 år
- 10-12 år
- 13-16 år
- 17 år og mer

Landbakgrunn³

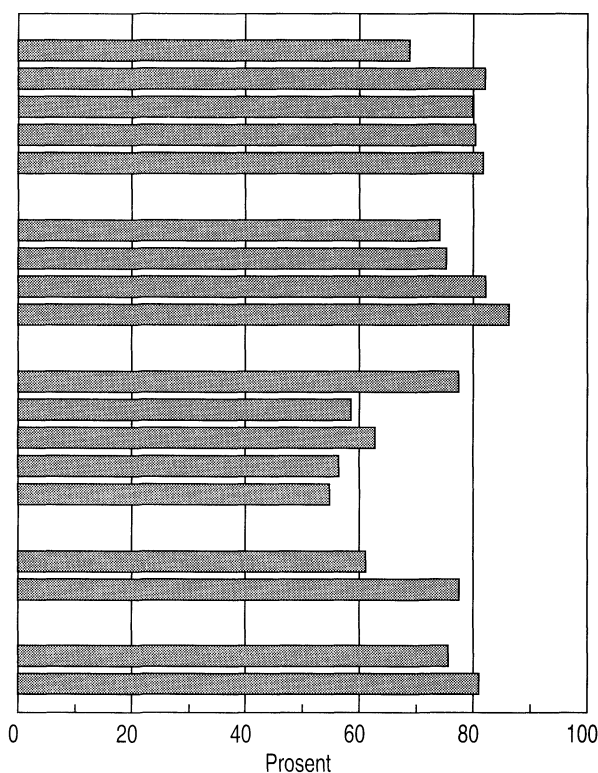
- Norge og Norden
- Vest-Europa ellers
- Øst-Europa
- Nord-Amerika og Oseania
- Sør-Amerika, Asia og Afrika

Mottok sosialhjelp

- Minst én i husholdningen
- Ingen i husholdningen

Bosted

- Tettbygd
- Spredtbygd



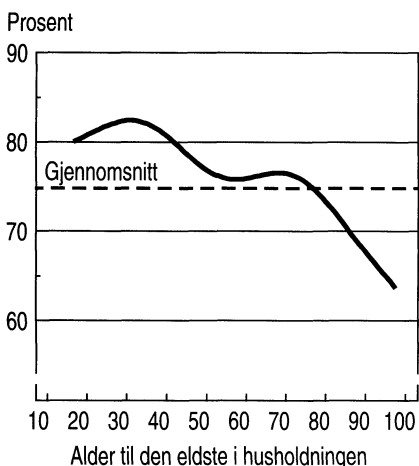
¹ Figuren viser den estimerte sannsynligheten for respons for de kategoriske forklaringsvariablene som ble funnet å være signifikante ved minst 1 prosents nivå i den logistiske modellen

² For de husholdningene hvor faktisk antall personer i husholdningen er ukjent, brukes familiestørrelse fra Det sentrale personregisteret

³ Landbakgrunn til en person er fødelandet til personen. Hvis barn har forskjellig fødeland fra foreldrene, blir landbakgrunn mors fødeland. For husholdninger med personer fra de forskjellige verdensregioner, brukes den verdensregionen som ligner kulturelt mest på vår norske, vest-europeiske kultur. Er for eksempel én i husholdningen fra Vest-Europa og de andre fra Afrika, vil husholdningen bli klassifisert som vesteuropeisk

Kilde: Helseundersøkelsen 1995

Figur 1: Sannsynlighet for å være med i Helseundersøkelsen 1995 for husholdninger, etter alder til den eldste i husholdningen



Kilde: Helseundersøkelsen 1995

De eldste vanskelige å nå

Figur 1 viser andelen av husholdningene som svarer ettersom hvor gammel den eldste i husholdningen er. Husholdninger hvor den eldste er omtrent 30 år har den største svarsannsynligheten. Deretter synker svarsannsynligheten fram til vel 50 år, for deretter å øke igjen. Rundt 70 år er det et ikke helt tydelig toppunkt igjen, før svarsannsynligheten avtar klart med økende alder.

Husholdninger med én person har en svarsannsynlighet på rundt 69 prosent. Husholdninger med fra to til fem eller flere personer har svarsannsynlighet rundt 80 prosent. En stor andel av enpersonhusholdningene er eldre kvinner over 80 år. Videre ser vi at husholdninger hvor personen(e) med høyest utdanning har utdanning på universitets- eller høgskolenivå har en svarprosent på omtrent 86. Mens husholdninger hvor høyeste utdanning er på grunnskolenivå, ligger under 70 prosent.

Husholdninger hvor ingen er norsk-fødte har lavere svarsannsynlighet, rundt 55-60 prosent, enn husholdninger med minst én norsk person. De som får sosialhjelp har lavere svarsannsynlighet enn de som ikke får. Husholdninger i spredtbygde strøk har større svarsannsynlighet enn de i tettbygde strøk.

Økonomien i husholdningen, målt med den høyeste personinntekten, påvirker ikke svarsannsynligheten noe særlig. Gruppen som mottar overgangsstønad, det vil si enslige foreldre, svarte omtrent like ofte som andre. Følgelig er disse to variablene ikke med i modellen for å beregne svarsannsynligheter.

Vi brukte en statistisk test som viser hvilken forklaringsvariabel som ale-ne betyr mest i modellen. Den gav følgende rekkefølge med den viktigste forklaringsvariabelen først: Antall personer i husholdningen, alder til den eldste i husholdningen, høyeste utdanning i husholdningen, landbakgrunn etter verdensregion-

inndeling, om noen i husholdningen mottok sosialhjelp og tettbygdt/spredtbygdt bosted.

Hvordan påvirker frafallet tall fra helseundersøkelsen?

Siden det er overhyppighet av noen kjennetegn blant de som ikke svarer, vil det kunne skape skjevhet i materialet. For å få et anslag på hvor stor skjevhet frafallet skaper, beregner vi anslag på noen typer sykkelighet med og uten korrigering for frafall.

Vi bruker en metode som stammer fra Politz og Simmons (1949). Vi bruker to versjoner av modellen for svarsannsynligheter for å justere data³. Tabell 2 viser estimert andel syke uten frafallsjustering, og justert for frafall basert på disse to modellene.

Tallene i tabellen tyder heldigvis på at frafallet bare i liten grad påvirker sykkelighetstallene for hele befolkningen. Det kan tenkes at andre typer sykkelighet eller undergrupper eller geografiske områder påvirkes mer. De som ikke svarer ser ut til å være litt mer syke enn de som er med i undersøkelsen.

Metoden for å justere for frafall forutsetter homogenitet i grupper med lik verdi av forklaringsvariablene i frafallsmodellen. Eller med andre ord; gitt utdanningsnivået, landbakgrunn i husholdningen, om noen i husholdningen mottar sosialhjelp eller ikke, og om husholdningen bor i tettbygdt eller spredtbygdt strøk, eventuelt også alder til eldste i husholdningen og husholdningens størrelse, forutsettes sannsynligheten for sykdom å være den samme for de som svarer og ikke svarer. For å undersøke om denne forutsetningen holder, måtte personer som ikke svarer bli oppsøkt, og på en

Tabell 2: Tall for sykkelighet for Helseundersøkelsen 1995 beregnet uten og med to alternative modeller for å korrigere for frafall



Sykdom	Ingen justering for frafall	Modell for svarsannsynlighet	
		Inkluderer høyeste utdanning i husholdningen, landbakgrunn, sosialhjelp og tettbygdt/spredtbygdt bosted	Inkluderer alder til eldste i husholdningen, antall personer i husholdningen, høyeste utdanning i husholdningen, landbakgrunn, sosialhjelp og tettbygdt/spredtbygdt bosted
Varig sykdom	57,24	57,46	58,04
Nervøse lidelser	6,77	6,87	7,01
Hjerte- og karsykdommer	14,35	14,43	14,92
Åndedrettssykdommer	17,44	17,54	17,50
Sykdommer i underhuden	4,46	4,48	4,50
Skjelett- og muskelsykdommer	24,79	25,34	24,93

Kilde: Helseundersøkelsen 1995

eller annen måte bli overtalt til å bli med i helseundersøkelsen.

Denne analysen tyder på at en frafallsprosent rundt 25 ikke svekker kvaliteten noe særlig på helseundersøkelsen. Likevel ville det naturligvis være ønskelig å øke deltakelsen. I den amerikanske National Health Interview Survey lyktes det å holde svarprosenten på 95 prosent i perioden 1985 til 1989 (Couper 1996). I den samme perioden økte gjennomsnittlig antall forsøk på å komme i kontakt med intervjuobjektet fra 2,56 til 2,95. Kanskje burde man for den neste helseundersøkelsen også anstrenge seg mer for å øke svarprosenten?

1. Modellen estimeres på husholdningsnivå, eller mer presist; hvis minst én person i husholdningen er med i undersøkelsen, får responsvariabelen verdien 1, ellers får den verdien 0. Variablene for alder og inntekt blir brukt som kontinuerlige variabler i modellen. De andre forklaringsvariablene kodes som kategoriske variabler. Det er vanlig å formulere denne typen sammenheng med en logistisk link.

I tradisjonelle logistiske modeller må eventuelle transformasjoner av forklaringsvariablene defineres før estimeringen. For eksempel kan en ha en formening om at alder inngår med både 1. og 2. grads ledd. Ofte er det en ulempe å måtte formulere funksjonsformen på forhånd fordi en ikke har nok kunnskap om hva som er god transformasjon. Ved å velge en "tilfeldig" transformasjon kan en risikere å ikke finne en god tilpasning til data. Moderne regresjonsmetoder (Venables og Ripley 1994) beregner en ikke-parametrisk funksjon ved "spline"-metoder. Fordelen er at man slipper å transformere basert på mer eller mindre godt funderte hypoteser om funksjonsformen. Den estimerte sammenhengen mellom alder og frafall viser at alder bør inngå som et fjerdegrads polynom i modellen.

2. Personinntekt er disponibel inntekt, som i selvangivelsen, pluss summen av utlignet skatt.

3. Metoden går ut på å justere de opprinnelige vektene i helseundersøkelsen, basert på

inverse trekkesannsynligheter, med den inverse svarsannsynligheten. Personer over 80 år er overrepresentert i helseundersøkelsen. Derfor baserer vi beregningen av svarsannsynligheten på to forskjellige frafallsmodeller. Den ene modellen inkluderer alle forklaringsvariablene som påvirket frafallet signifikant, se figurene 1 og 2. I den andre modellen er ikke alder eller antall personer i husholdningen med. Antall personer i husholdningen er sterkt korrelert med alder idet eldre mennesker ofte bor alene. Derfor ble også denne variabelen fjernet fra modellen.

Litteratur

Belsby, Liv (1995): *Forbruksundersøkelsen. Vektmetoder, frafallskorrigeringer og intervjuer-effekt*, Notater 95/18, Statistisk sentralbyrå.

Bild, Hanna, Jon Erik Finnvald, Kari Kveim Lie, Rannveig Nordhagen og Arnfinn Schalm (1997): *Brukererfaringer med forebyggende helsetjenester*, Under utgivelse i serien Rapporter, Statistisk sentralbyrå.

Cochran, William Gemmell (1951): "Modern methods in the sampling of human populations", *American Journal of Public Health*, 41, 647-653.

Couper, Mick P. (1996): *Theoretical and Practical Aspects of Survey Nonresponse*, Metodekonferanse på SCB, Stockholm september 1996.

Politz, A. og W. Simmons (1949): "An attempt to get not-at-homes into the sample without callbacks", *Journal of the American Statistical Association* 44, 9-31.

Särndal, Carl-Erik, Bengt Swenson and Jan Wretman (1992): *Model Assisted Survey Sampling*, Springer Series, Springer-Verlag.

Vedø, Anne (1997): *Frafall i leveårspanelet 80, 83 og 87*, Notater 97/31, Statistisk sentralbyrå.

Venables, Bill N. og Brian D. Ripley (1994): *Modern Applied Statistics with S-plus*, Springer-Verlag.

Liv Belsby er førstekonsulent i Statistisk sentralbyrå, Seksjon for statistiske metoder og standarder.