

Interne notater

STATISTISK SENTRALBYRÅ

80/11

9. mai 1980

OM PLANLEGGING AV
SPØRRESKJEMA OG DATAINNSAMLING
VED INTERVJU¹⁾

av

Gunvor Iversen

INNHold

	Side
1. Innledning	2
2. Gangen i en intervjuundersøkelse	2
3. Planlegging	3
3.1. Spesifisering av formålet	3
3.2. Konstruksjon av spørreskjema	4
3.2.1. Spørsmål som er vanskelige å stille	6
3.2.2. Spørsmål som er vanskelige å besvare ...	6
3.3. Holdningsmålinger	10
3.3.1. Holdninger målt ved hjelp av enkelt- spørsmål	10
3.3.2. Sammensatte skalaer	12
3.4. Prøveundersøkelse	14

INNHold (førts.)

	Side
4. Trekking/fastsetting av utvalg	17
4.1. Innledning	17
4.2. Ulike typer av utvalg	17
4.2.1. Ulike former for sannsynlighetsutvalg	17
4.2.2. Ulike typer av ikke-sannsynlighetsutvalg	18
4.2.3. Sammenlikning av de to hovedtypene av utvalg	18
4.3. Byråets utvalgsplan	18
4.4. Hvor stort bør utvalget være	19
5. Innsamlingsmetoder	20
5.1. Telefonintervju	20
5.2. Postale undersøkelser	20
5.3. Besøksintervju	21
5.3.1. Før intervjueren går ut i felten	21
5.3.2. Feltarbeidet	21
6. Feilkilder	22
6.1. Register- og utvalgsfeil	23
6.2. Målefeil	24
6.2.1. Egenskaper ved intervjueren	24
6.2.2. Tendensen til å svare sosialt ønskelig	24
6.2.3. Respondentens oppfattelse av spørsmålet	25
6.2.4. Andre feilkilder	25
6.3. Frafall	25
6.3.1. Årsaker til frafall	26
6.3.2. Reduksjon av effekten av frafall	27
7. Avsluttende bemerkninger	28
Litteratur	29

1) Notatet er med noen små forandringer opptrykk av et foredrag som ble holdt for forskere ved Transportøkonomisk Institutt på Hornsjø høyfjells-hotell 28. september 1979.

1. INNLEDNING

Hovedhensikten med dette notatet er å gi nye medarbeidere ved Underavdelingen for Intervjuundersøkelser, så vel som andre som er opptatt av spørreskjemakonstruksjon og intervjuteknikk, en innføring i de mer praktiske sidene ved slikt arbeid.

Det vil derfor ikke bli presentert noen nye synspunkter på spørreskjemakonstruksjon og intervjuteknikk, en innføring i de mer praktiske sidene ved slikt arbeid.

Det vil derfor ikke bli presentert noen nye synspunkter på spørreskjemakonstruksjon og intervjuteknikk. Notatet vil i stedet gi en oppsummering av en del erfaringer fra Underavdelingen, samt bygge på en del litteratur ellers. Litteraturhenvisninger vil ikke bli gitt underveis, men er listet opp bakerst. Jeg vil heller ikke si særlig mye om de teoretiske betraktninger og forskningsmessige erfaringer som ligger bak vår kunnskap om spørreskjemakonstruksjon, spørsmålsformuleringer og måleproblemer.

Interesserte lesere vil finne slike forskningsresultater blant annet i spesialtidsskrifter som Public Opinion Quarterly.

2. GANGEN I EN INTERVJUUNDERSØKELSE

En intervjuundersøkelse kan deles inn i følgende stadier:

1. Planlegging, herunder:
 - a) Spesifisering av formål
 - b) Teknisk planlegging (spørreskjemakonstruksjon)
 - c) Prøveundersøkelse
2. Fastsettelse av utvalg - type og størrelse
3. Datainnsamling, som også omfatter:
 - a) Innsamlingsmetoder
 - b) Feilkilder ved datainnsamlingen
4. Revisjon og koding
5. Dataoverføring og oppretting av materialet
6. Tabellkjøring og eventuelt analyse

I dette notatet skal jeg konsentrere meg om punktene 1-3, og la dem være en disposisjon for notatet. Men en må ha klart for seg at det å lage en intervjuundersøkelse må ses på som en helhet, hvor hver enkelt del har like mye å si for helheten, og hvor uklarheter framkommet på ett trinn i prosessen, vil få konsekvenser for de derpå følgende trinn. For eksempel må vi ha tenkt på hvordan materialet skal bearbeides når vi lager spørreskjemaet. Likedan må vi ta hensyn til revisjon og koding. Men altså, her vil jeg konsentrere meg om punktene 1-3.

3. PLANLEGGING

3.1. Spesifisering av formålet

Siktepunktet for arbeidet med å formulere og presisere formålet med en undersøkelse er å komme fram til problemstillinger som er så presise at de kan være en direkte veiledning for konstruksjonen av spørreskjemaet.

Dette kan bestå i tre faser som delvis overlapper hverandre.

Først må en oversette undersøkelsens generelle hensikt til et mer spesifisert sentralt siktemål. Det vil f.eks. være umulig å lage en undersøkelse som har det vage mål å finne ut noe om "måling av folks reisevaner". En må da bestemme seg for hensikten med å foreta en slik måling, og ut fra det snevre inn de opplysninger og forskningsopplegg som vil være brukbare.

Etter at dette er gjort må en prøve å finne fram til andre og mindre generelle temaer som henger sammen med hovedhensikten, og til slutt må en så komme fram til de spesielle opplysninger en trenger for å dekke de problemstillinger en er kommet fram til.

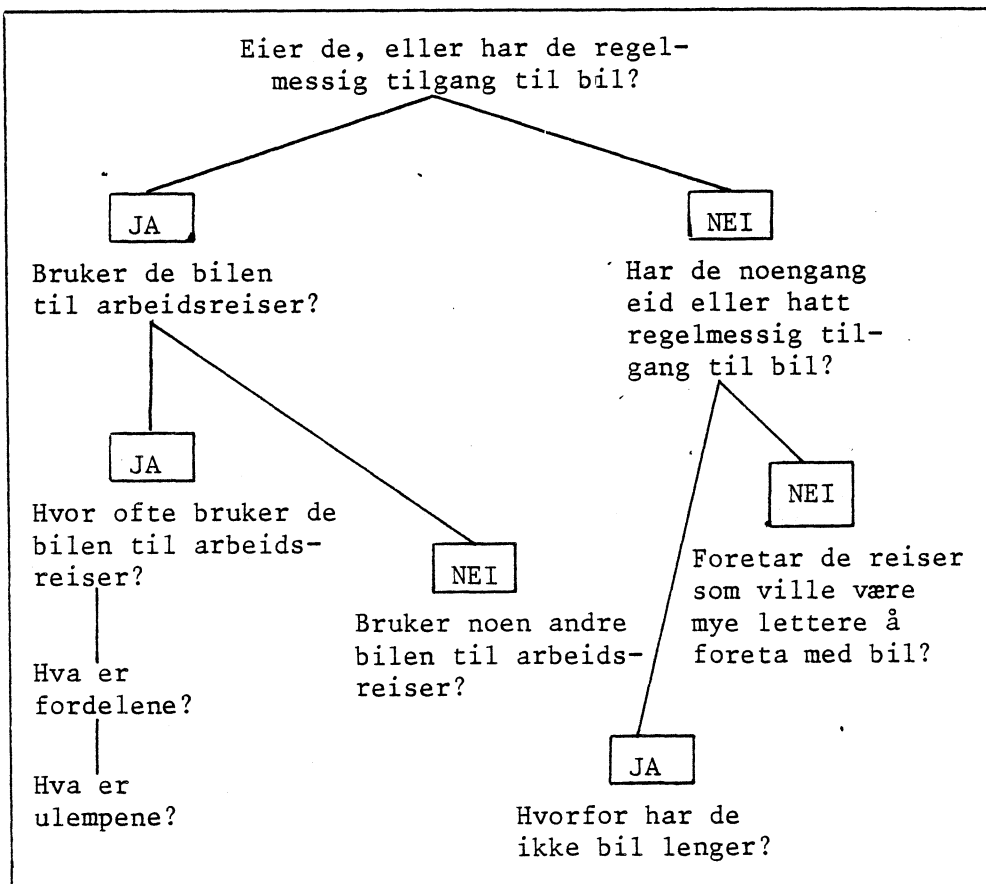
Et fjerde punkt som er svært viktig i denne spesifiseringen er også hvem en ønsker å stille de ulike spørsmål til, og hvordan disse gruppene skal defineres. Dette er et punkt som i utgangspunktet synes enkelt, men i virkeligheten krever det mye omtanke. Eksempler på slike grupper kan være husmødre, yrkesaktive, trygdede. Disse gruppene kan være mer eller mindre vanskelige å definere, f.eks. hva er en husmor? En kvinne som driver med husarbeid? Hvor mye husarbeid? Hva er husarbeid? osv., osv.

Når vi så har spesifisert den generelle problemstillingen, mindre generelle problemstillinger, de opplysninger vi trenger, samt hvilke grupper vi ønsker å stille spørsmål til, kan vi gå over til selve skjema-konstruksjonen.

Et lite poeng før vi kommer så langt, er at når vi nå har klarlagt våre hensikter og hvilke grupper vi ønsker å stille spørsmål til, kan det være

hensiktsmessig å stille opp et gjennomstrømningskart over undersøkelsen, som sier hvilke grupper som skal ha ulike typer spørsmål.

Et veldig enkelt eksempel på et slikt gjennomstrømningskart er det følgende:



Og et råd til slutt her: Hvis en ikke klarer å lage et slikt gjennomstrømningskart før en går løs på skjemakonstruksjon, bør en i alle fall gjøre det etterpå for å sjekke at alle som skal ha et spørsmål har fått det, og at de som ikke skal ha det ledes utenom. Det vil spare en selv for mange ergrelser, og intervjuobjektene for mye irritasjon.

3.2 Konstruksjon av spørreskjema

Arbeidet med konstruksjon av spørreskjemaet som skal nyttes i en intervjuundersøkelse er en helt sentral del av planleggingsprosessen. Spørreskjemaet utgjør den konkrete operasjonalisering av de problemstillinger en ønsker å belyse, samtidig som utformingen av skjemaet har mye å si for intervjuernes arbeid på et senere tidspunkt. Også arbeidet med bearbeidingen av det innkomne materialet påvirkes sterkt av hvordan skjemaet er konstruert.

Kravene til det gode spørreskjema kan oppsummeres på følgende måte:

1. Skjemaet må sikre at vi får opplysninger som er nødvendige for å belyse våre problemstillinger.
2. Skjemaet må ikke være lengre enn nødvendig, dvs. det bør ikke inneholde andre spørsmål enn de som er viktige for å belyse problemstillingene.
3. Skjemaet må være utformet slik at intervjuerens arbeid med å skaffe og registrere de nødvendige opplysninger blir enklest mulig.
4. Skjemaet må være lagt opp slik at arbeidet med å utnytte resultatene kan foregå mest mulig effektivt.

Her skal jeg legge hovedvekten på enkelte sider ved spørsmålsformulering, da dette er det arbeidet som er mest tidkrevende.

Rent generelt kan en formulere følgende krav til det enkelte spørsmål:

- a) Spørsmålet skal så entydig som mulig presentere vår problemstilling for dem som skal svare.
- b) Spørsmålet skal ikke inneholde problemstillinger som stiller urealistiske krav til respondentens forkunnskaper, hukommelse eller fantasi.
- c) Spørsmålet skal ikke være ledende i den forstand at selve formuleringen påvirker svaret.
- d) Spørsmålet skal såvidt mulig gjøre svarene fra de forskjellige respondenter sammenlignbare.
- e) Spørsmålet skal såvidt mulig unngå å irritere, provosere eller støte den som skal svare.

Disse kravene kan selvsagt konkretiseres nærmere, og i det følgende skal jeg gi noen eksempler på typer spørsmål som det kan være vanskelig å stille, og i alle fall vanskelig å tolke svarene på.

Før vi går nærmere inn på selve spørsmålsformuleringen, kan det være grunn til å minne om en del tekniske ting ved skjemaet.

Et godt spørreskjema er noe mer enn en samling av gode eller brukbare enkeltspørsmål; det må være bygd opp slik at det motiverer respondenten og gir intervjueren brukbare arbeidsforhold. Hensynet til respondenten tilsier at det ikke skal være for langt, at spørsmålene må henge naturlig sammen, og at det ikke stilles så mange tilsynelatende like spørsmål om et emne at det virker som et kryssforhør. Det er også viktig at skjemaet er slik at respondenten til en viss grad skjønner hensikten med de spørsmålene som stilles.

Spørreskjemaet bør ikke være for langt. En time i gjennomsnitt regnes for den maksimale lengde på skjemaet i Byrået.

Rekkefølgen av spørsmålene er også av betydning. En skal i alle fall vokte seg for å stille opp spørsmål på en slik måte at respondenten får inntrykk av at hun eller han har svart på liknende spørsmål før. Dette kan virke masete, og være svært irriterende for respondenten.

ispedd med fremmedord og tekniske begreper. Ved formuleringen av spørsmål er det viktig at en er oppmerksom på denne tendensen.

b) Generaliseringer og abstraksjoner

Å lage en generell undersøkelse betyr ikke det samme som å forlange generaliserende svar på et eneste spørsmål, som f.eks.: Hvordan vil De stort sett si at bilen Deres fungerer? Her ber en om en endimensjonal skala for noe som opplagt består av flere dimensjoner. En får problemer med å svare på dette spørsmålet hvis en har en bil som har dårlig aksellerasjon, lavt bensinforbruk, litt dårlige bremseser, men ligger godt på veien.

Abstraksjoner er også vanskelige, og en må prøve å gjøre dem så konkrete som mulig gjennom å bryte dem ned i en serie med spørsmål.

c) Negative utsagn

Spørsmål eller utsagn som bruker nekteord er så å si umulig å forstå første gang en hører dem. Dette gjelder særlig når den som svarer skal si seg enig eller uenig i utsagnet.

Eks.: Hva vil det bety å svare uenig på følgende utsagn: "Barn som stjeler skal ikke straffes"? Problemer som dette oppstår fordi det ofte er ønskelig å blande positive og negative utsagn i holdningsmålinger. En må da finne positive måter å uttrykke det negative på, som f.eks.: "Barn som stjeler bør få en advarsel".

d) Spørsmål som fordrer kunnskaper

Studier av respondentens svaratferd viser at manglende kunnskaper ikke alltid hindrer folk i å svare på spørsmål. Verdien av svarene blir derimot svært tvilssom. Konsekvensen av dette må være at en så langt som mulig må unngå spørsmål som forutsetter at den som skal svare har kunnskaper ut over det helt elementære.

e) Hukommelsesspørsmål

Svært mange faktaspørsmål som blir stilt stiller krav til respondentens hukommelse. Stort sett kan glemsel påvirke svarene på følgende måte:

- i) Det mest vanlige er at folk glemmer hendelser som faktisk har skjedd. Sannsynligheten for at en hendelse skal glemmes øker jo lengre tid som går, og er størst for hendelser som respondenten synes er lite viktige.
- ii) Der det er spørsmål som gjelder hendelser som gjentas regelmessig, kan virkningen av glemselen bli at folk oppgir hva de vanligvis gjør, i stedet for hva de har gjort på et bestemt tidspunkt.
- iii) Respondentene vil ofte glemme når en bestemt hendelse har funnet

sted. Dersom spørsmålet gjelder hendelser innenfor en bestemt periode, kan vi ofte få en "telescoping effect" som består i at respondenten også oppgir hendelser som har foregått før perioden.

Den vanligste måten å hjelpe den som skal svare til å huske bedre, er ved å presentere de mulige hendelsene for vedkommende. Dette kan gjøres ved hjelp av detaljerte spørsmål eller ved hjelp av kort o.l.

Der det ikke er mulig å bruke et slikt opplegg med "aided recall", (hjelp til å huske) kan det ofte være en fordel å legge opp spørsmålet slik at IO får litt ekstra tid til å tenke seg om. Forsøk har vist at en får bedre svar på faktaspørsmål ved å erstatte en helt kort formulering med en som er litt lengre. Dette har også skjedd i de tilfelle der økingen i lengden ikke har gitt respondenten ekstra holdepunkter for hukommelsen.

f) Hypotetiske spørsmål

Dette er spørsmål der respondenten blir bedt om å svare på hvordan hun eller han vil reagere på en tenkt situasjon. Et eksempel på et hypotetisk spørsmål som er blitt benyttet i Byrået er: "Ønsker De å skifte til en annen jobb?" Slike spørsmål vil ofte stille for store krav til respondentens fantasi og selvforståelse, slik at verdien av resultatene blir nokså begrenset. Vi vet heller ikke hvilken ramme respondentene tolker spørsmålene ut fra. Noen vil svare ut fra idealsituasjonen, noen ut fra det de mener er realistisk. Svarene kan også reflektere tidligere erfaring og fremtidige forventninger og forhåpninger. Kort sagt, svarene blir vanskelige å tolke.

g) Ledende spørsmål

Det er allment anerkjent at en må unngå å bruke ledende spørsmål ved vanlige intervjuundersøkelser. Helt ytterliggående eksempler på slike spørsmål vil være:

Spør en bonde: "De liker sikkert ikke ordningen med sommertid?"

Spør en eldre person: "De er vel enig i at fri barneoppdragelse er forkastelig?"

Slike eksempler kan vi finne mange av, og le av dem. Men det er andre, og mindre iøynefallende årsaker til at spørsmål blir ledende.

i) Dersom bare noen av de mulige svar på et spørsmål nevnes, vil respondentene ha en tendens til å velge ett av disse svaralternativene. For noen holdningsspørsmål hvor en ønsker at så mange som mulig skal gi uttrykk for et standpunkt, kan det være hensiktsmessig å ikke nevne den nøytrale midtkategorien. På den måten vil en tvinge folk til å ta et standpunkt. Med unntak av dette må regelen være at enten alle eller ingen svaralternativer nevnes.

ii) Spørsmål kan være formulert slik at svaralternativene ikke framtrer som likeverdige. Dette kan skje ved at det ene alternativet framtrer som særlig respektabelt eller normalt, f.eks. når det knyttes til personer eller institusjoner med høy prestisje. Ulikevekt mellom svaralternativene kan også komme av at det i spørsmålet bare argumenteres for et av ett av svaralternativene.

iii) Også bruken av positivt eller negativt ladete ord kan påvirke folks svar. Det er antakelig sannsynlig at formuleringene "gi hjelp" og "bruke penger" vil gi ulike svarfordelinger i en undersøkelse om U-hjelp. Det er også sannsynlig at følgende alternative spørsmålsformuleringer ville gi ulike svarfordelinger: "Er De enig i banksosialisering?" versus "Er De enig i at det offentlige skal ha representanter i forretningsbankenes styre?"

h) Flerdimensjonale spørsmål

En betingelse for at svarene fra de ulike respondentene skal være sammenlignbare er at de har oppfattet spørsmålet på samme måte. Spesielt for åpne spørsmål, dvs. spørsmål uten faste svaralternativ er det viktig at en gjør det klart hvilken dimensjon som dekkes gjennom spørsmålet.

Ek s.: "Hvorfor kjøpte De denne komfyren?" kan gi svar som: "Fordi jeg fikk et godt tilbud", "Fordi den gamle var gått i stykker", "Fordi en venn sa at dette merket var best" osv.

Det er videre viktig at hvert spørsmål bare inneholder én problemstilling; såkalte dobbeltløpete spørsmål skaper problemer både for respondenten og for den som skal tolke svarene. Slike dobbeltløpete spørsmål får en ofte når oppdragsgiveren eller statistikeren tar en sammenheng mellom to fenomener for gitt og inkluderer den i spørsmålet. For å ta et eksempel: Selv om spørsmålet "Hva ville du foretrekke, høyere lønn og sterk prisstigning eller lønnsstopp og redusert stigning i prisene?" ville bli oppfattet som en-dimensjonalt blant økonomene f.eks. i Statistisk Sentralbyrå, ville det skape problemer for alle dem som ønsket seg både mer lønn og stabile priser.

Som en avslutning på dette avsnittet vil jeg bare minne om at det som jo er av største betydning er hvordan den som svarer oppfatter spørsmålene. Dette må en bestandig ta hensyn til, og erfaringene på dette området kan summeres opp slik:

1. Hvis respondenten finner spørsmålet vanskelig å besvare, tolker hun det på en slik måte at det blir lettere å besvare.

2. Hvis et vidt uttrykk eller begrep anvendes i et spørsmål, vil det ofte bli tolket mindre vidt.
3. Respondenten forsøker å besvare hva hun oppfatter som spørsmålets mening eller ånd snarere enn å holde seg til den faktiske ordlyd.
4. Når et spørsmål inneholder mer enn noen få informasjonsbærende ord, er det særlig sårbart for misforståelser. Jo flere slike ord, jo større er sannsynligheten for misforståelser.
5. En respondent som bare hører en del av et spørsmål, vil ha en tendens til å rekonstruere spørsmålet på grunnlag av det hun har hørt.
6. Når et komplisert spørsmål følges av et tilbud om et enkelt valg av alternativer som f.eks. ja eller nei, vil respondenten legge mindre omtanke i arbeidet med å komme fram til et svar. Hvis hun blir bedt om å utarbeide sitt eget svar, vil respondenten tenke grundigere gjennom spørsmålet.

3.3 Holdningsmålinger

Med holdninger mener en gjerne forholdsvis varige mønstre i personers måte å oppfatte, vurdere og reagere på bestemte objekter. Problemet er imidlertid at det er en relativt svak sammenheng mellom slike holdninger og atferd, dessuten er det ofte vanskelig å finne fram til tilfredsstillende måter å måle holdninger på.

Et spesielt problem får vi når vi bruker påstander for å måle holdninger, nemlig enkelte respondenters tendens til å være enig med det intervjueren sier. Denne tendensen kalles vanligvis "acquiescence response set". Ved konstruksjon av holdningsskalaer kan en unngå denne effekten ved å formulere vekselvis positive og negative påstander, og summere svarene. Der en ser på reaksjonene på hver enkelt påstand, vil imidlertid effekten gjøre seg gjeldende.

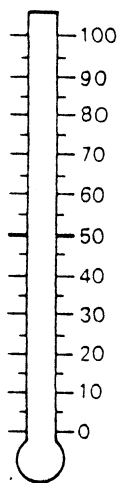
På tross av disse advarende ord, både hva angår verdien av holdningsspørsmål og vanskeligheter med hensyn til selve målingen, skal jeg likevel gå raskt gjennom en del måter å måle holdninger på. Først skal jeg ta for meg holdninger målt ved enkeltspørsmål, deretter holdninger målt ved hjelp av sammensatte skalaer.

3.3.1 Holdninger målt ved hjelp av enkeltspørsmål

a) Vurderingsskalaer

Disse kan være av to typer, grafiske og numeriske (kategoriserte).

Et eksempel på en grafisk vurderingsskala har vi i Institutt for samfunnsforsknings valgundersøkelse i 1969. Nedenfor er denne skalaen, som ble kalt et "sympati-termometer", gjengitt sammen med spørsmål og instruks til intervjueren.



Det er en rekke partier og andre organisasjoner her i landet som gjerne vil at man skal ta hensyn til deres krav og ønsker. Vi vil gjerne vite hvordan De ser på noen av disse gruppene. (VIS KORT 4 MERKET "TERMOMETER") På dette kortet finnes et termometer som vi kaller "sympati-termometer". Vi bruker det slik: Hvis De ikke har kjennskap til en gruppe, eller hverken føler sympati eller antipati for dens virksomhet, plasserer De den på midten - ved 50-graders-streken. Hvis De har sympati for en gruppe - eller liker dens virksomhet - så plasserer De gruppen mellom 50 og 100 grader. De plasserer den høyere dess mer sympati De har for den. Hvis De derimot ikke har noen sympati for gruppen, eller misliker dens virksomhet, så plasserer De den mellom 0 og 50 grader. Jo mindre sympati De har for den, dess nærmere 0 grader plasseres den.

Denne grafiske vurderingsskalaen angir både et nøytralt midtpunkt og bruker inndelingen i grader som referansepunkter for vurderingen. Dette er en bedre form for vurderingsskala enn der hvor bare ytterpunktene er angitt, og ikke noe mer.

Ved numeriske vurderingsskalaer har respondenten enda klarere holdpunkter for å plassere seg selv eller andre. En kan for eksempel be folk vurdere seg selv som meget interessert, noe interessert i forskjellige saksområder.

b) Sammenlikningsskalaer

Disse skiller seg fra vurderingsskalaer ved at en plasserer f.eks. påstander, saksområder o.l. ved å sammenlikne dem med andre påstander, saksområder o.l., i stedet for i forhold til en absolutt standard. En måte å gjøre dette på er å be den som skal svare å rangere de tingene en er interessert i vedkommendes holdning til. En kan for eksempel spørre respondentene om hvilket av ulike saksområder de er mest interessert i og hvilke de er minst interessert i. Dermed ville en få rangordnet saksområdene. En ville imidlertid ikke fått vite hvor interessert vedkommende var i det saksområdet hun var mest interessert i.

En annen form for sammenlikningsskala er parvis sammenlikning. Her får respondentene i oppgave å rangordne f.eks. to og to påstander av gangen til alle kombinasjonsmuligheter er uttømt.

Ved parvis sammenlikning får vi bare en rangordning dersom den som svarer er konsistent i sin vurdering av de ulike objektene.

En tredje form for sammenlikningsspørsmål har vi der hvor respondenten bes om å rangere seg selv relativt til andre, for eksempel folk flest, andre på jobben, andre unge osv. Et eksempel vil være:

"Tror du at du selv er mer fornøyd eller mer misfornøyd med arbeidet ditt enn de andre du jobber sammen med, eller er du omtrent som gjennomsnittet?" Svarkategoriene vil da være: Mer fornøyd, mindre fornøyd, som gjennomsnittet, vet ikke.

3.3.2 Sammensatte skalaer

Jeg skal her ta for meg tre hovedtyper av slike skalaer, nemlig differensielle, summerte og kumulative skalaer.

a) Differensielle skalaer

Denne typen sammensatte holdningsskalaer er særlig knyttet til psykologen Thurstone's navn.

Som eksempel på en Thurstone-skala kan vi se på en som ble utviklet av Hinckley i 1932 for å måle holdning overfor negre i USA. Nedenfor er fire av de seksten påstandene som ble benyttet gjengitt.

- 8. I bunn og grunn er negeren og den hvite mann like. (10.3)
- 10. Ingen neger har den minste rett til å motsette seg eller i det hele tatt sette spørsmålstegn ved et ulovlig drap på en fra sin rase. (1.1)
- 11. Etter at negeren er utdannet til samme nivå som den hvite vil det fremdeles være en uovervinnelig kløft mellom dem. (4.2)
- 15. Det er mulig for den hvite og svarte rase å være medlemmer av den kristne familie uten å gifte seg med hverandre. (6.6)

Denne skalaen fra tredveårene illustrerer et av problemene ved en Thurstone-skala. De to påstandene som i følge dommerne som da ble brukt skulle representere noe i nærheten av en nøytral holdning til negre, vil i dag neppe oppfattes slik av særlig mange respondenter. Skalaverdienes gyldighet er med andre ord begrenset i tid og rom. Undersøkelser har vist at akkurat hvor på skalaen en påstand blir plassert avhenger av dommernes egne holdninger i det aktuelle spørsmålet. Derimot ser det ut til at en kan regne med at det stort sett vil være enighet om rangordningen av påstandene ut fra hvor negative/positive de er. Hinckley's skala kunne derfor trolig også i dag brukes til å rangordne personer ut fra hvor sterke rasefordommer de har, men avstanden mellom ulike skårer ville bety lite.

Et annet problem med en slik skala er at samme skåre kan oppnås på svært ulike måter. En respondent som sa seg enig i påstand 8 og 10 ville for eksempel få skåret 5.7, det vil si temmelig nær en som sa seg enig i 11 og 15. Men i det første tilfellet ville altså en nøytral skåre være et resultat av enighet med en påstand fra hver ekstrem av skalaen. Hvilken skåre en respondent får forteller derfor i seg sjøl lite om hvilke spørsmål han har sagt seg enig i (med mindre den tilsvarer den høyeste eller laveste skalaverdien som forekommer for noen påstand).

b) Summerte skalaer

En utforming av en summert skala som blir hyppig brukt til å måle holdninger, kalles en Likert-skala etter opphavsmannen. Som for Thurstone-skalaen er utgangspunktet en rekke påstander forskeren mener er relevante for holdningen som skal undersøkes. Men i motsetning til Thurstone-skalaen er alle påstandene klart positive eller negative i relasjon til objektet for holdningen.

Disse påstandene gis til et utvalg av respondenter som er representative for dem som det ferdige måleinstrumentet skal brukes på. Respondentene blir bedt om å angi for hver påstand om de er sterkt enig, enig, usikker, uenig eller sterkt uenig i påstandens innhold. Den enkelte respondents skåre kan beregnes ved at han får 5 poeng for hver gang han har erklært seg sterkt enig i en positiv påstand, 4 for hver enig osv., ned til 1 poeng for hver sterkt uenig. For negative påstander ordnes poengene motsatt vei, 5 for hver sterkt uenig osv., slik at alle svar som tyder på en positiv holdning til objektet gis de høyeste poengtallene. Respondentens skåre er lik summen av poengene han har oppnådd for hver enkelt påstand.

Neste skritt i utarbeidingen av en Likert-skala er å finne fram til hvilke påstander som skiller best mellom de respondentene som befinner seg på hvert sitt ytterpunkt på skalaen. Blant de påstandene som skiller best mellom respondentene på hvert ytterpunkt av holdningsvariabelen, velger en så ut det antallet påstander en ønsker til den endelige skalaen. Disse påstandene presenteres så for de egentlige respondentene i undersøkelsen, og den enkeltes skåre beregnes på samme måte som angitt ovenfor.

c) Kumulative skalaer

Som et eksempel på en kumulativ skala skal vi se på én Bogardus utviklet i 20-årene for å måte "sosial avstand" mellom respondenten og ulike etniske grupper. En oppgir navnet på en etnisk gruppe, og ber så om respondentens umiddelbare reaksjon på om han ville være villig til å ha ulike former for sosial kontakt med et "gjennomsnittsmedlem" av denne gruppen. Nedenfor er det angitt fire av de sju kontaktformene det blir spurt om i Bogardus-skalaen.

- A) Gi vedkommende adgang til den nærmeste familien gjennom ekteskap.
- B) Gi vedkommende adgang til data som nabo.
- C) Gi vedkommende adgang til landet som borger.
- D) Gi vedkommende adgang til landet som besøkende.

Disse skalaene er slik at i prinsippet er det bare ett svarmønster som fører til en bestemt skåre.

Skalaen er forsøkt konstruert slik at jo høyere på lista av spørsmål en får positive svar av respondenten, jo mindre sosial avstand ønsker han å holde til den aktuelle etniske gruppen. Jo høyere opp på lista vi kommer, jo færre er det som svarer positivt. Dette innebærer at om vi vet at en respondent har svart positivt på et bestemt spørsmål, skal han også ha svart positivt på spørsmålene lenger nede på lista.

Dermed er vi i stand til å slutte fra hvor mange spørsmål en respondent har besvart positivt til hvilke spørsmål dette er.

Erfaringene med å bruke slike sammensatte skalaer som de jeg har skissert her er svært begrenset i Norge, og for at slike skalaer skal kunne brukes, må det legges ned et stort arbeid både med hensyn til hvilke spørsmål/ utsagn som skal med, hvor mange som skal med osv. Det er også slik at skalaer konstruert i andre land, under andre historiske betingelser og under andre sosiale forhold ikke uten videre kan overføres direkte og brukes uten justeringer. Holdningsmålinger er med andre ord vanskelige å foreta, og som nevnt innledningsvis i dette avsnittet, kan verdien av det vi får ut av slike målinger også diskuteres.

3.3 Prøveundersøkelse

Vi forutsetter nå at vi har et ferdig spørreskjema, og før vi bruker det på vår endelige undersøkelsespopulasjon, den massen vi ønsker å undersøke, vil vi gjerne prøve skjemaet ut - det vil si vi foretar en prøveundersøkelse.

Det er vanskelig å planlegge en intervjuundersøkelse uten en god del kunnskap på forhånd om det emnet den skal dekke, de befolkningsgrupper den skal dekke, måten folk vil reagere på spørsmål på og til og med om hvilke svar de vil gi.

For eksempel, hvordan kan en være sikker på at et spørsmål er så meningsfylt for den gjennomsnittlige respondenten som det er for spørreskjemaeksperten? Og hvordan skal en bestemme seg for hvilke spørsmål som overhode skal stilles?

For å svare på slike spørsmål og for å få den kunnskapen som er nevnt foran, trenger vi både tester på foreløpige utgaver og prøve på den endelige versjonen.

Jeg skal her stort sett bare snakke om den siste formen for prøveundersøkelse, nemlig der hvor vi prøver ut en versjon av spørreskjemaet som ligger så nær opp til den endelige versjonen som overhode mulig.

Vi kan imidlertid også bruke prøveundersøkelsen som et hjelpemiddel på et tidligere trinn, f.eks. for å finne ut hvilke spørsmål av et helt batteri som best måler det vi ønsker å måle, og som vi deretter bruker i hovedundersøkelsen.

Prøveundersøkelsen kan også brukes som en selvstendig undersøkelse.

Den typen prøveundersøkelser jeg skal snakke om kan ses på som en generalprøve og skal gi veiledning om følgende forhold:

- a) Hvor godt definert og registrert den undersøkelsespopulasjonen er som en ønsker å trekke utvalg fra. Dette gjelder kanskje særlig hvis en bruker registre som en ikke vet nok om mht. hvor fullstendige de er.
- b) Variasjonen innen befolkningen med hensyn til det emne som skal undersøkes. Dette vil være av betydning når utvalgsstørrelsen skal bestemmes. Selv om prøveundersøkelsene ofte vil være vel små til dette bruk, kan de likevel være av verdi hvis det ikke finnes annen tilgjengelig statistikk å vurdere ut fra.
- c) Frafallet som kan forventes. Det sannsynlige antall nektere og ikke-kontakter kan beregnes dersom en ikke på forhånd har et godt grunnlag for å vite noe om dette. Effektiviteten i ulike metoder for å redusere frafallet kan sammenlignes. Ulike undersøkelsesopplegg kan vurderes når det gjelder å påvirke frafallets størrelse - f.eks. valg mellom postalintervju og personlig intervju.
- d) Brukbarheten av datainnsamlingsmetoden. Alternative metoder for datainnsamling er tilgjengelige, f.eks. observasjon, postalintervju, personlig intervju, gruppeintervju, og en kan trenge data om de ulike metoders relative kostnader, nøyaktighet osv. for å gjøre et fornuftig valg av innsamlingsmetode.
- Prøven kan også vise om intervjuerne gjør en effektiv jobb, om for mye press blir lagt på dem eller på respondentene, og om skjemaene bør sjekkes nærmere.
- e) Om skjemaet ser ut til å fungere. Dette er uten tvil den viktigste funksjonen for en prøveundersøkelse. Det er flere ting en må passe på.
- hvor lett det er å behandle skjemaet i felten,
 - hvor effektivt skjemaet er satt opp,
 - hvor klare definisjonene er, og selvsagt:
 - hvor gode spørsmålene er.

En må her sjekke om formuleringene er enkle, klare og frie for tekniske uttrykk. En må også se etter tegn på om noen misforstår spørsmålene eller ikke er nok informert til å kunne gi fornuftige svar, eller om de f.eks. av ulike grunner gir unøyaktige svar. En slik grunn kan være "social desirability", dvs. at respondenten svarer ut fra det vedkommende anser som sosialt ønskelig i stedet for ut fra faktiske forhold. En må også se etter om det er ting som tyder på at det legges alt for mye press på folks hukommelse.

Det er ikke lett å gi regler for å gjenkjenne slike spørsmål som er skissert her, men enkelte tips kan gis.

Noen ganger er den sanne fordelingen av svar kjent. Større avvik fra denne kjente fordelingen er da en advarsel om at noe er galt, enten med utvalget eller med spørsmålet.

Hvis en ikke er sikker på hvilke formuleringer som fungerer best, kan en prøve ut alternative formuleringer på tilfeldige deler av utvalget.

For stor opphopning av svar i en ende av fordelingen kan være et tegn på et ledende spørsmål, eller det kan bety at folk svarer på en stereotyp måte.

Et stort antall "vet ikke"-svar er ofte (men ikke nødvendigvis) et dårlig tegn. Det kan tyde på et uklart spørsmål, et som bruker uvanlige ord, eller et som går utenom respondentens erfaringsområde.

Hvis mange nekter å svare på et bestemt spørsmål, kan dette være et tegn på at spørsmålet kan være delikat eller sårende. En kan da kanskje endre rekkefølgen slik at dette delikate/sårende spørsmålet blir introdusert på en bedre måte, eller en bør kanskje også vurdere å kutte spørsmålet helt ut.

Kontakten med intervjuerne er svært viktig i denne fasen. Hver intervjuer bør eller skal rapportere om hvordan intervjuingen har gått, hvilke vanskeligheter en har støtt på underveis, osv.

Den som planlegger undersøkelsen bør gjerne også selv gå ut i felten; enten ved å intervjuer selv, følge med en intervjuer, eller begge deler.

- f) En bør også bruke prøveundersøkelsen til å sjekke hvor effektiv instruksjonen er og hvor godt intervjuerne mestrer skjemaet.
- g) Sjekke de kodene som er valgt på prekodete spørsmål, samt fastlegge svarkategoriene for spørsmål som har vært "åpne" i prøven, men som en ønsker å ha prekodet i hovedundersøkelsen.

Uten en prøveundersøkelse er det ofte vanskelig å finne fram til de relevante svaralternativene.

- h) Varigheten av og kostnadene forbundet med hovedundersøkelsen vil en også ha et bedre grunnlag for å vurdere hvis det er foretatt en prøve.

Prøven er altså den siste mulighet vi har til å finne ut hvor effektiv hovedundersøkelsen vil bli, og om det er noen vits i å følge opp de problemstillingene vi hadde i utgangspunktet.

4. TREKKING/FASTSETTING AV UTVALG

4.1 Innledning

På et tidlig tidspunkt i planleggingsprosessen er det bare nødvendig med en omtrentlig definisjon av den undersøkelsesbefolkning en ønsker å dekke. Men når formålet er klart og skjemaet konstruert, må en bestemme seg for hvilke befolkningsgrupper en vil dekke.

I Byrået f.eks. er undersøkelsesbefolkningen ofte personer mellom 16 og 74 år som ikke er i institusjoner. Men det er klart at dette kan variere, og en må ha undersøkelsesbefolkningen klart definert før en begynner å trekke utvalg.

4.2 Ulike typer av utvalg

Jeg skal bare her gå kort inn på typer av utvalg, men jeg skal skissere opp de to hovedtypene av utvalg, nemlig sannsynlighetsutvalg og ikke-sannsynlighetsutvalg. Jeg skal også si noen ord om fordeler og ulemper ved de to hovedtypene.

4.2.1 Ulike former for sannsynlighetsutvalg

Enkel tilfeldig eller systematisk utvelging der enhetene trekkes direkte fra universlista.

Stratifisert utvelging: enhetene ordnes i strata ut fra bestemte egenskaper (stratifiseringsvariablene). Fra hvert stratum trekkes et bestemt antall enheter til utvalget ved enkel tilfeldig eller systematisk utvelging.

Klyngeutvelging: enhetene grupperes sammen i klynger. Et visst antall klynger trekkes ut. Utvalget kan bestå av alle enheter i disse klyngene, eller det kan foretas utvelging innenfor de uttrukne klyngene igjen. På hvert trinn i utvelgingen brukes enkel tilfeldig eller systematisk utvelging, eventuelt med stratifisering.

Stratifisert utvelging, klyngeutvelging og enkel tilfeldig eller systematisk utvelging er ikke gjensidig utelukkende opplegg for sannsynlighetsutvelging. Enhver sannsynlighetsutvelging skjer ved at enheter trekkes ut ved enkel tilfeldig eller systematisk utvelging. I noen tilfeller deles enhetene først inn i strata, i andre grupperes de sammen i klynger. Ved spesielt kompliserte utvelgingsopplegg kan alle framgangsmåtene kombineres. Vi kan for eksempel ha to-trinns klyngeutvelging med stratifisering på ett eller begge trinn og bruk av enkel tilfeldig og systematisk utvelging på hvert sitt trinn.

Kravet til et sannsynlighetsutvalg er altså at alle enheter på universlista (i universet) skal ha en kjent sjanse for å komme med i utvalget.

4.2.2 Ulike typer av ikke-sannsynlighetsutvalg

Disse utvalgene er altså slik at ikke alle enheter har en kjent sjanse for å komme med i utvalget.

Eksempler på slike utvalg er skjønnsmessig utvelgning, utvelgning ved selvseleksjon, slumpemessig utvelgning, kvoteutvelgning.

4.2.3 Sammenligning av de to hovedtypene av utvalg

Fordelene med sannsynlighetsutvalg er at vi kan generalisere de resultater vi har funnet i utvalget til også å gjelde hele den befolkningen utvalget er trukket fra. Problemene med denne utvelgingsmåten ligger blant annet i at dette er en svært kostbar måte å foreta datainnsamling på. Det er bestemte personer som er trukket ut, og det er disse personene som skal intervjues. Ved flyttinger f.eks. kan en imidlertid lage regler for hvordan en skal finne erstatning for uttrukne personer.

Et annet problem, og mer alvorlig, ligger i at vi alltid vil få et visst frafall ved slike utvalgsmetoder, og hvis frafallet er stort og/eller skjevt fordelt, vil dette kunne ødelegge mulighetene for generaliseringer, noe som jo i utgangspunktet var den viktigste fordelene med denne måten å trekke utvalg på. Jeg skal komme nærmere inn på frafallet noe senere.

Ikke-sannsynlighetsutvalg blir billigere enn sannsynlighetsutvalg, og hvis vi ikke trenger generalisering, eller ikke mener at generalisering er viktig på det nåværende tidspunkt, f.eks. i pretester av spørreskjema, kan en godt bruke slike utvalg.

Imidlertid er sannsynlighetsutvalg det som oftest blir brukt i samfunnsvitenskapene, og jeg vil derfor heretter stort sett snakke om slike utvalg.

4.3 Byråets utvalgsplan

Det kan nå være på sin plass å gi et grovt bilde av den utvalgsplan som blir brukt i Byrådet. Planen er ellers utførlig beskrevet annetsteds i en bok av Ib Thomsen (se litteraturlisten). Her skal jeg derfor bare gi en kort skisse.

Av kostnadsgrunner er det nødvendig å velge ut enhetene slik at vi får en viss geografisk konsentrasjon i oppgaveinnhenting. I stedet for å be om opplysninger fra noen ganske få personer eller husholdninger i hver eneste kommune i Norge, spør vi heller noen flere enheter i et færre antall kommuner. På den måten får vi en mindre intervjuerstab å rekruttere og lære opp, og intervjuerne behøver ikke reise så langt mellom hver respondent.

Utplukkingen av de stedene som skal tjene som basis for oppgaveinnhenting er skjedd i samsvar med statistisk-vitenskapelige metoder. Vi har først tatt med alle byer som har en folkemengde på ca. 30 000 og mer. Deretter har vi delt resten av landet inn i 79 deler som blant annet har det til felles at folkemengden er omtrent like stor i hver del. Disse "delene" kaller vi strata. Fra hvert av de 79 strata har vi så trukket ett utvalgsområde. Hvert utvalgsområde kan bestå av én eller flere kommuner. Sammen med de nevnte byene har vi 102 utvalgsområder som til sammen omfatter 131 av landets 454 kommuner og har vel 60 prosent av landets folkemengde.

Det er da fra disse 102 utvalgsområdene med de 131 kommunene vi skal trekke våre tilfeldige utvalg av husholdninger eller personer, utvalg som til sammen vil representere "et Norge i miniatyr"; det vil si det som er typisk for Norge med hensyn til næringsliv, bosettingsstruktur og inntektsforhold.

Utvalgsområdenes geografiske omfang og antall observasjoner i hvert område gjør at vi må ha flere intervjuere i hvert utvalgsområde.

4.4 Hvor stort bør utvalget være?

Det er tre forhold som spiller en avgjørende rolle for hvor stort utvalg en trenger.

- a) Den nøyaktighet en ønsker på resultatene.
- b) Hvor ensartet befolkningen er.
- c) Antall oppsplittinger av utvalget en ønsker å gjøre.

Ad. a) I praksis er det som regel svært vanskelig på forhånd å si mye om hvor nøyaktige resultater en ønsker. Dette henger sammen med at en svært sjelden kan gi tall for konsekvensene av å basere beslutninger på grunnlag av unøyaktige estimatorer. Det er likevel viktig at en forsøker å komme fram til visse uttrykk for hvor gode resultater en ønsker.

Ad. b) Det sier seg selv at en trenger et relativt lite utvalg i tilfeller hvor det er liten variasjon i befolkningen med hensyn til de variablene vi ønsker å måle. I det ekstreme tilfellet hvor alle elementene har samme verdi på den variabelen en ønsker å studere, trenger en bare en verdi (en respondent) for å anslå gjennomsnittlig verdi i befolkningen.

Ad. c) Hvis en i en landsomfattende undersøkelse i tillegg til tall for hele landet også ønsker tall for deler av populasjonen, må en ta hensyn til dette ved valg av utvalgsstørrelse. I praksis er det ofte slik at det er ønsket om oppsplitting av materialet som setter de største krav til utvalgsstørrelsen. Når det i Byrådet skal gis et raskt, foreløpig anslag på utvalgs-

størrelse, finner en ofte ut hvilke oppsplittinger av materialet en ønsker seg, deretter regnes utvalget ut på en slik måte at vi er sikre på å få minst 40 personer innenfor hver oppsplitting. Ettersom formålene med undersøkelsene blir mer og mer klargjorte, kan en så justere denne utvalgsstørrelsen. Men det er helt klart at det er umulig å anslå utvalgsstørrelsen før formålet med undersøkelsen er klargjort.

5. INNSAMLINGSMETODER

Vi har nå et spørreskjema og et utvalg, og vi skal nå samle inn våre opplysninger.

Jeg skal her ta for meg tre mulige måter å gjøre dette på, nemlig telefonintervjuing, postalt intervju og besøksintervju.

5.1 Telefonintervju

Der hvor telefontettheten tilsier det, kan det være hensiktsmessig å ta telefonen i bruk under datainnsamlingen. I Byrået blir telefonintervju brukt i Arbeidskraftundersøkelsene, og telefonen blir også brukt til forhåndsavtale om intervju. Det er imidlertid tvilsomt om telefonintervju bør benyttes ved intervjuer som tar lang tid.

Det må imidlertid advares sterkt mot å trekke utvalg på grunnlag av folk som har telefon, dvs. via telefonkatalogen. Telefontettheten er i Norge i dag på ca. 60 prosent, og telefonabonnenter er ikke et tilfeldig utvalg av Norges befolkning. Dobbelt ille blir det hvis en bruker telefonkatalogen til å finne ut personer som skal intervjues. Tenk bare etter hvor mange kvinner som står oppført i katalogen!

Men når dette er sagt, la meg slå fast at telefonen kan være et nyttig hjelpemiddel for å få utført datainnsamlingen på en smidig måte.

5.2 Postale undersøkelser

Mange forhold ligger godt til rette for å bruke postinnsamling. Vi har et velorganisert postvesen, og i alle fall Byråets register har navn og adresser.

Fordelen med postundersøkelser er at det er lett å sende spørreskjema til et stort antall respondenter. Ulempene ved postundersøkelser består kanskje først og fremst i at svarprosenten ofte kan bli lavere enn ved besøksintervju, og at spørreskjemaet ikke bør være for langt og at det må være

enkelt satt opp uten alt for mange spørsmål som bare skal stilles til enkelte respondenter og ikke til alle.

Det bør imidlertid nevnes at vi i Byrådet har hatt gode erfaringer med å kombinere postalundersøkelser og besøksintervju. En har da latt intervjuere oppsøke de personer som ikke innen en viss tid har sendt tilbake spørreskjemaene de var blitt tilsendt.

5.3 Besøksintervju

5.3.1 Før intervjuerne går ut i felten

Dersom slike intervju skal bli vellykkede, forutsettes det både en grundig utvelgning av intervjuere og at de har fått en grundig opplæring. Dette gjelder såvel generelt som for den enkelte undersøkelse. Dette betyr da blant annet at intervjuerne blir grundig orientert om undersøkelsens formål, hvorfor den gjennomføres, hvem som står bak undersøkelsen, osv.

Vi må imidlertid være litt forsiktige her, slik at ikke hypoteser og forventede sammenhenger blir for klart formulert for intervjuerne. En kan nemlig i slike tilfelle risikere at intervjuerne ubevisst tolker svar i en slik retning at forventningene blir innfridd / hypotesene bekreftet.

Før vi sender intervjuerne ut i felten, er det også ønskelig (i Byrådet nødvendig) at de som er trukket ut til å delta i en undersøkelse får brev om dette. Brevet bør inneholde en kort redegjørelse for undersøkelsen, samt understreke hvor viktig det er at den som er trukket ut faktisk deltar i undersøkelsen. Anonymiteten til respondenten bør også påpekes.

5.3.2 Feltarbeidet

Hovedpoenget for intervjueren i felten er jo å få gjennomført intervjuet, men her kan det oppstå en del problemer. I det følgende skal jeg bare kort skissere en del typiske situasjoner, reaksjoner fra respondenten, og strategier for likevel å få foretatt intervjuet. Det må understrekes at i opplæringen av intervjuere vil det å lære å takle slike situasjoner være av stor betydning.

- a) Respondenten har "ingen interesse" for saken:

Her kan samfunnsnytte og eventuell egennytte for respondenten trekkes fram. En kan også forklare at en ikke har noe med offentlige myndigheter å gjøre, og en kan også understreke taushetsplikten.

- b) Den som skal svare sier at hun eller han ikke er "representativ".

En kan da forklare at det ikke har noen mening å snakke om at den

enkelte respondent er representativ, men at det er viktig for representativiteten for undersøkelsen at alle som er trukket ut deltar.

c) Frykt for vanskelige spørsmål.

En kan da stille noen av de første spørsmålene i skjemaet for på den måten (forhåpentligvis) å vise respondenten at dette ikke er så farlig/vanskelig.

d) Frykt for ukjent rolle hos respondenten.

En kan da bruke samme taktikk som over.

e) Respondenten har ikke tid.

En kan da gjøre en av to ting:

- 1) si at intervjuet ikke tar lang tid (dersom det faktisk ikke gjør det - aldri ellers),
- 2) avtale nytt tidspunkt.

f) Respondenten har ingen tro på undersøkelsen.

En må på en pen måte forklare at denne påstanden ikke holder.

g) Er jeg pliktig til å medvirke?

En må da si fra at medvirkning er frivillig, men en kan kanskje nevne konsekvensen av frafall, og kanskje også understreke nytten av at undersøkelsen blir gjennomført.

h) Alvorlig sykdom eller dødsfall i respondentens nærmeste familie.

Det er i seg selv grunn god nok til at intervju ikke foretas.

Som vi ser, er det mange problemer som kan oppstå i forbindelse med det å oppnå et intervju, og det burde også klart ha framgått at det er svært viktig å gi intervjuerne en grundig innføring i den enkelte undersøkelse for dermed å motivere dem til å få tatt så mange intervju som mulig.

6. FEILKILDER

Jeg har nå behandlet planlegging av en undersøkelse, utvalgstreking, innsamling av data. I hver av disse fasene må en regne med at det foreligger kilder til feil. Til slutt i dette notatet skal jeg derfor gå gjennom de viktigste av disse feilkildene.

Feilene kan inndeles i 3 typer:

1. Register- og utvalgsfeil
2. Målefeil
3. Frafall

6.1. Register- og utvalgsfeil

Når en trekker enheter til en statistisk undersøkelse, er målet at alle enheter i en populasjon skal ha en kjent sannsynlighet for å komme med i utvalget. Ved en totaltelling ønsker en utvalgssannsynlighet lik 1. På grunn av manglende opplysninger om populasjonen, f.eks. dårlige registre, krever det både fantasi og mye arbeid å nå dette målet. En kan si at et register er perfekt dersom alle analyseenheter foreligger i registeret bare en gang og med gode identifikasjoner. Slike registre er sjeldne, og det typiske er at en må utbedre alvorlige svakheter før en trekker utvalg fra et register. Det foreligger et virvar av eksempler på problemer som oppstår ved trekking av utvalg fra register. Nedenfor skal gis en systematisk beskrivelse som dekker de fleste registerproblemer som kan oppstå. Feilene inndeles i fire typer:

- Type 1. Registeret mangler elementer. Med dette menes at populasjonen inneholder elementer som ikke forekommer i registeret.
- Type 2. Elementene forekommer i klynger i registeret. Dette er tilfellet når en ønsker å trekke personer, men bare har adgang til et register over adresser.
- Type 3. Fremmedelementer i registeret. Det vi si at registeret inneholder elementer som ikke tilhører den populasjonen en ønsker å lage en undersøkelse i.
- Type 4. Dubletter. Med dette menes at det samme element forekommer flere steder i registeret.

Før en tar et register i bruk til trekking av utvalg, bør en sikre seg å få oversikt over hvilke av de fire feiltyper som forekommer i det aktuelle tilfelle. Etter at en har kartlagt de feiltyper som foreligger, kan en velge å se bort fra dem. Dette kan en f.eks. gjøre i tilfeller hvor hyppigheten av feil er liten. Hvis en f.eks. vet at det mangler elementer i registeret, men at det bare gjelder noen ganske få elementer, kan en velge å se bort fra denne feilen. Tilsvarende for de øvrige feiltyper.

6.2. Målefeil

Målefeil kan være både systematiske og usystematiske. De usystematiske målefeilene trenger vi ikke bekymre oss alt for mye om. Når vi tar gjennomsnittet over flere observasjoner, vil slike feil ofte oppveie hverandre, slik at effekten f.eks. på et gjennomsnitt vil bli liten.

Derimot kan det være grunn til å ta for seg en del årsaker til systematiske målefeil.

6.2.1. Egenskaper ved intervjueren

Det har lenge vært kjent at personlige kjennetegn ved intervjueren, så som kjønn, alder, utdanning og sosiale status, kan virke inn på de svar vedkommende får. For eksempel er det vist at veteraner fra den andre verdenskrig rapporterte flere medaljer til eldre kvinnelige intervjuere enn til andre.

Egenskaper ved intervjueren kan altså påvirke respondenten ved at hun tillegger intervjueren bestemte forventninger, og at svarene går i retning av å stemme overens med disse forventningene.

Også intervjuerens holdninger kan, via de antipatier eller sympatier som vekkes hos den som skal svare, virke inn på resultatet. Dette avhenger selvsagt mye av hvor strukturert spørreskjemaet er.

En har også funnet at svarene avhenger mye av de forventningene som intervjueren har til respondenten. Svarene på spørsmål tidlig i intervjuet kan gi intervjueren et inntrykk av respondentens holdning, og dette kan ha innflytelse på intervjuerens tolkning av svar på spørsmål seinere i intervjuet.

Dette at svarene kan variere fra intervjuer til intervjuer kaller vi intervjuerarians. Når det gjelder å beregne slik intervjuarians, har vi blant annet i Byrået det problemet at en bestemt intervjuers arbeid stort sett er konsentrert innen et bestemt geografisk område. Det blir derfor vanskelig å skille en intervjuereffekt fra en geografisk effekt.

6.2.2. Tendens til å svare sosialt ønskelig

For mange spørsmål er svaralternativene ulike med hensyn til hvor akseptable eller ønskelige de er ut fra sosiale normer. På spørsmålet: "Snyter De på skatten?" er åpenbart det sosialt ønskelige svaret: "Nei", og en må derfor anta at det velges oftere enn de faktiske forhold skulle tilsi. En forventer at respondenten snarere svarer det som betraktes som mest ønskelig i stedet for det som best karakteriserer henne eller ham, dersom det er forskjell mellom disse to alternativene. Denne svartendensen har vært undersøkt i forbindelse med holdningsmålinger og personlighetstester, og det ser ut til å

være en betydelig variasjon i denne effekten. Betydningen av slike effekter er naturligvis avhengig av hvor klare normene er for hva som er ønskelig.

6.2.3. Respondentens oppfattelse av spørsmålet

Forekomsten av svarsavvik avhenger også av skjemaets form og innhold. Av erfaring vet en at kvaliteten på svarene avhenger av hvor klar, vesentlig eller relevant problemstillingen er for respondenten. Stort sett regner en med at spørsmål om konkrete atferdsaspekter vil være mindre utsatt for svarsavvik enn spørsmål om meninger, holdninger, ønsker osv.

6.2.4. Andre feilkilder

Respondenten kan gi feil svar fordi han mangler nødvendige kunnskaper, eller fordi han rett og slett ikke husker det riktige svaret.

6.3. Frafall

Et av de største problemer i forbindelse med en utvalgsundersøkelse er frafallet, det vil si de personer en ikke får kontakt med, men som er trukket ut til undersøkelsen.

Det er jo slik at en uten alt for store vansker kan trekke sannsynlighetsutvalg, og idéen er da, som vi har vært inne på før, at en da kan beregne hvor nøyaktige resultatene for utvalget er, og dermed kan en også generalisere fra utvalget til hele den befolkningen utvalget er trukket fra.

I praksis er det, som også nevnt tidligere, et stort problem her, og det er frafallet.

Frafallet kan i verste fall ødelegge våre generaliseringsmuligheter, eller i alle fall gjøre generaliseringene mer usikre.

I en godt planlagt undersøkelse er det imidlertid vanligvis mulig å holde frafallet på et noenlunde rimelig nivå. Jeg har f.eks. vært inne på betydningen av intervjuernes dyktighet og opplæring i å takle situasjoner hvor respondenten vegrer seg og ved utsending av brev om undersøkelsen, og ved å avtale tid for intervjuet på forhånd.

En kan imidlertid også foreta grove beregninger på hvilke effekter skjevhet på grunn av frafallet vil ha på resultatene.

Hvis en har trukket et utvalg og en har fått en svarprosent på f.eks. 80, kan en ikke uten videre slå seg til ro med dette. Hvis de 20 prosentene som ikke svarte var lik de som svarte, på alle relevante kjennetegn, trengte en ikke bekymre seg om noe, bortsett kanskje fra reduksjonen i antall.

Erfaringen viser imidlertid at frafallet er forskjellig fra resten, og en kan i alle fall ikke gå ut fra at det ikke er det. Et eksempel fra Fruktbarhetsundersøkelsen i Byrået skulle vise dette. Jeg har her en tabell over gjennomsnittlig antall levendefødte barn pr. kvinne fordelt etter antall besøk. Vi ser at gjennomsnittlig antall levendefødte barn synker ettersom antall besøk øker. Dette betyr at hvis vi hadde stoppet intervjuingen uten å foreta gjenbesøk, ville vårt utvalg ha en overrepresentasjon av kvinner med mange barn.

Gjennomsnittlig antall levendefødte barn pr. kvinne etter antall besøk

	Antall besøk								
	Total	1	2	3	4	5	6	7	8
Gjennomsnittlig antall levendefødte barn pr. kvinne		1.84	1.56	1.47	1.30	1.28	1.08	0.85	0.75

6.3.1. Årsaker til frafall

For å kunne vurdere frafallets effekt på viktige resultater i en undersøkelse, er det vanlig å inndele dette etter årsaken til frafallet. For dette formål er det utarbeidd et prekodet frafallsskjema til bruk under intervjuernes feltarbeid. Skjemaet inneholder 10 frafallsårsaker, men det er vanlig at en før publisering slår sammen til noen få årsaker. Den klassifikasjon som publiseres varierer fra undersøkelse til undersøkelse, men det nedenstående er blitt en temmelig standardisert inndeling i mange typer undersøkelser, kanskje først og fremst intervjuundersøkelser.

- 1) Nekter
- 2) Ikke å treffe
- 3) Sykdom
- 4) Annen årsak

N e k t e r

Ved nesten alle statistiske undersøkelser vil en ha en gruppe av personer som nekter å gi de ønskede opplysninger. Dette skyldes i det vesentlige to faktorer, egenskaper ved respondenten og den intervjuteknikk som brukes.

Det er viktig å være oppmerksom på at disse på ingen måte er umulig å påvirke. En må likevel regne med at selv om en gjør store anstrengelser for å

redusere gruppen av nektere, vil det alltid bli igjen en hard kjerne, som det ikke er mulig å få tak i uten bruk av eventuelle lovbestemmelser.

I k k e å t r e f f e

Mens nektere er spredt utover hele utvalget, viser det seg at folk i byene er vanskeligere å treffe enn folk i mindre tettbygde strøk. Dessuten er det lettere å treffe en eller annen i en husholdning enn et bestemt medlem av en husholdning.

Intervjutidspunktet har også mye å si. Vanligvis er kvelden den beste tiden når en ønsker å få tak i andre enn husmødre og eldre. Forarbeid som f.eks. forhåndsavtale eller brev sendt IO på forhånd kan redusere frafallet. Sist, men ikke minst, vil gjenbesøk ofte øke svarprosenten betraktelig.

S y k d o m

Sykdom hos IO selv eller i nærmeste familie er ofte årsak til frafall. Årsaken til å skille ut denne kategorien er at den har en helt spesiell virkning på resultatene i mange undersøkelser, fordi sykdom har en viktig innflytelse på mange andre kjennetegn for personen.

Det kan være verd å merke seg at fordelingen på de ulike frafallskategorier ser ut til ikke å endre seg, selv om frafallet som sådan ser ut til å øke. Dette kan f.eks. bety at folk ikke er mindre villige til å svare på spørsmål enn før, dvs. nekter mer, men at folk er blitt vanskeligere å få tak i, er mer borte fra boligen sin på grunn av ulike typer aktiviteter. Selve økningen i frafallet representerer imidlertid et svært alvorlig problem i Byråets utvalgsundersøkelser.

6.2.2 Reduksjon av effekten av frafall

Hvis en skal kunne korrigere for frafallet, må en ha noen opplysninger om frafallet. I Byrådet får vi vanligvis opplysninger om kjønn, alder, ekteskapelig status, bosted.

Enhver undersøkelse har også et frafallsskjema som fylles ut, slik at det kan foretas analyser på frafallet.

En bør også foreta sammenligninger mellom det netto-utvalget en får og den befolkningen utvalget er trukket fra, dersom det foreligger kjente tall for denne befolkningen. Slike opplysninger kan f.eks. være kjønn og alder. Slik kan en få et begrep om omfanget av og retningen av skjevheter i utvalget på grunn av frafall.

Slike sammenligninger kan aldri bli noen forsikring om fravær av skjevheter, men kan gi en forsikring om at det endelige resultatet ikke er alt for ille.

En er heller ikke så ille ute hvis skjevheter i utvalget ikke har sammenheng med de problemstillinger som en vil undersøke. Men hvis en f.eks. skal undersøke forekomst av trafikkstøy og en finner ut at de som ikke har svart bor i svært trafikkbelastede områder, må en gjøre noe for å rette opp skjevhetene, dersom resultatene skal bli av noen verdi.

Selv med store anstrengelser for å redusere størrelsen på frafall, vil det alltid ha en viss størrelse etter at innsamlingen er avsluttet. I visse undersøkelser kan frafallet være så stort at en må gjøre noe for å redusere effekten av det. En kan da veie de innsamlede data. Den mest vanlige metoden består i å inndelegge utvalget i grupper etter noen få kjennemerker, og deretter veie de forskjellige gruppegjennomsnitt med en faktor som er omvendt proporsjonal med frafallsprosenten innen gruppen.

Det er imidlertid ikke uten videre gitt at en klarer å redusere effekten av frafall gjennom veiing. Ved veiing forutsetter en jo at frafallet og de som svarer også er like på andre kjennetegn enn de som en bruker ved veiingen. Dette kan en ikke uten videre gå ut fra.

På grunnlag av de analyser av frafall og av veiing som er foretatt, har en derfor i Byrået kommet fram til at det ser ut til å være mer fruktbart å arbeide for å redusere størrelsen på frafallet enn å utvikle metoder for å motvirke virkningene av frafallet.

7. AVSLUTTENDE BEMERKNINGER

Jeg har her forsøkt å gi en oversikt over en del forhold en må ta hensyn til når en skal gjennomføre en intervjuundersøkelse. Hensikten har blant annet vært å vise at de ulike deler av en undersøkelse henger nøye sammen, og at feil og uklarheter som oppstår på et punkt i planleggingen "henger med" gjennom hele undersøkelsen. En kan for eksempel ikke regne med at en kan redusere virkningene av et dårlig spesifisert formål gjennom å stille mange enkeltspørsmål som i seg selv er gode nok.

En annen hensikt har vært å vise at planlegging og gjennomføring av en intervjuundersøkelse er en tidkrevende jobb som krever stor arbeidsinnsats, både fra de som planlegger undersøkelsen og fra de som står for den endelige datainnsamlingen. Endelig har jeg hatt som siktepunkt at dette notatet skal kunne gi praktisk veiledning for folk som selv skal i gang med å planlegge spørreskjema og gjennomføre datainnsamling via intervju.

LITTERATUR

- Hellevik, Ottar: Forskningsmetode i sosiologi og statsvitenskap. Oslo 1971.
- Hoynville, G., Jowell, R.: Survey Research Practice, London 1978.
- Høst, Sigurd: Ja-effekten (Acquiescence response set) og bruken av påstander i intervjuundersøkelser. ANO, IO 75/34, Statistisk Sentralbyrå 1975.
- Moser, C.A. and Kalton, G.: Survey Methods in Social Investigations, London 1973.
- Payne, S.L.: The Art of Asking Questions, New Jersey 1951.
- Thomsen, Ib & Høst, S.: Rutiner ved intervjukontoret. ANO IB 73/1, Statistisk Sentralbyrå 1973.
- Thomsen, Ib: Prinsipper og metoder for Statistisk Sentralbyrås utvalgsundersøkelser. SØS nr. 33, Oslo 1977.