



RIGSARKIVET

Rigsarkivets **guide** til
SPØRGESKEMA-
UNDERSØGELSER

”En ny undersøgelse viser” ... eller gør den?

**Hvad kan man egentlig konkludere
ud fra spørgeskemaundersøgelser?**

**I denne vejledning giver vi dig
værktøjer til at vurdere
spørgeskemaundersøgelser
styrker og svagheder**

HVAD ER EN SPØRGESKEMAUNDERSØGELSE?

Surveys, rundspørger, meningsmålinger... **Kært barn har mange navne.**

Fælles for dem er, at der er tale om *undersøgelser*, hvor man stiller en række på forhånd fastlagte spørgsmål til en gruppe svarpersoner. Vi bruger betegnelserne synonymt og kalder det med ét ord *spørgeskemaundersøgelser*.

Spørgeskemaundersøgelser består af to hovedelementer: *Indhold* i form af spørgsmål og svarmuligheder. Og *modtagere*, som også kaldes svarpersoner eller – med et fagudtryk – respondenter.

Indhold og modtagere skal være nøje afstemt i forhold til *formålet med undersøgelsen*. Et formål kan fx være at undersøge, om danskerne føler sig lykkelige, om gymnasieelever er ensomme, om danskernes holdning til legalisering af hash.

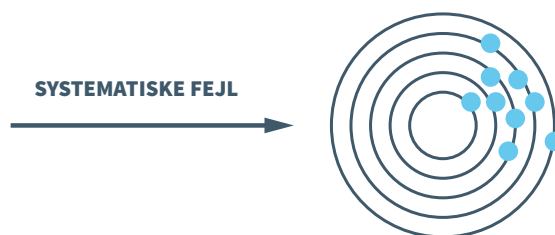
Indhold og modtagere skal også være nøje afstemt *i forhold til hinanden*. For hvad hjælper det at stille gode spørgsmål, hvis svarpersonerne ikke har den nødvendige viden, ikke har en holdning til emnet, ikke kan huske tilstrækkelig langt tilbage eller ikke ønsker at svare ærligt?

STIKPRØVER ELLER POPULATIONUNDERSØGELSER?

En stikprøve er en delmængde af en population. Spørgeskemaundersøgelser sendes typisk ud til en stikprøve, men indimellem laver man hele populationsundersøgelser. Populationer er fx de cirka 2.500 danske kommunalpolitikere og de 98 danske borgmestre. Vil man undersøge forhold i kommunerne, spørger man typisk alle 98 borgmestre eller alle kommunalpolitikere. Anderledes med populationen af voksne danskere, som er omtrent 4,6 millioner. Vil man vide, hvad voksne danskere mener, fx hvor lykkelige vi synes, vi er, så spørger man en stikprøve, typisk med 1.000 – 1.500 danskere.

SYSTEMATISKE MÅLEFEJL

Spørgeskemaundersøgelser har forskellige mulige fejlkilder. Man skelner grundlæggende mellem to typer fejl: Systematiske målefejl og tilfældige målefejl. Systematiske målefejl handler om undersøgelsens validitet, også kaldet gyldighed. Systematiske fejl betyder, at man måler forkert på en bestemt måde. Man rammer altså konsekvent og systematisk ”forbi skivens centrum”.



Men hvad vil det sige, at en undersøgelse er valid, dvs. har få systematiske målefejl? Vi opererer med tre slags validitet: Ekstern validitet, målingsvaliditet og intern validitet.

Ekstern validitet handler om undersøgelsens repræsentativitet, altså om resultaterne fra stikprøven kan generaliseres til at gælde hele populationen. Tjek om forudsætningerne for repræsentativitet er opfyldt: Er stikprøven udtrukket simpelt tilfældigt med lige adgang til hele populationen? Og er svarprocenten høj? Kan en frafaldsanalyse godtgøre, at der ikke er oplagte skævheder, altså at det ikke er en bestemt type personer, der har svaret?

Målingsvaliditet handler om at stille de rigtige spørgsmål – og stille dem metodisk korrekt. Hvordan måler man fx lykke? Stress? Ensomhed? Skal man måle det indirekte ved at stille spørgsmål til symptomer på stress, lykke og ensomhed? Eller skal man spørge direkte: "Føler du dig ensom?"

Har du tjekket, at spørgsmålene i undersøgelsen ikke er ledende? At spørgsmålene er entydige? Og at de er afgrænsede i tid, fx "tænk tilbage på det seneste år" eller "tænk tilbage på den seneste uge"? Tror vi på, at svarpersonerne forstår spørgsmålet, har den nødvendige viden om emnet, er i stand til at genkalde sig den - og faktisk har en holdning, som de vil stå ved?

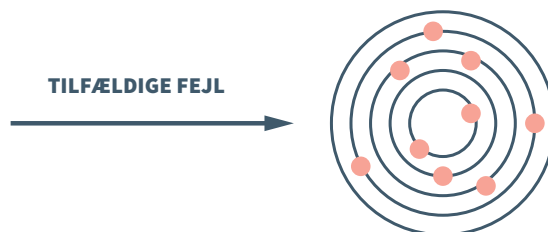
Intern validitet handler om mulige årsagssammenhænge. Fx om studiearbejde fører til stress. Eller om dårlig forældrekontakt fører til ensomhed. Tjek om forudsætningerne for at sandsynliggøre en årsagssammenhæng er opfyldt:

1. Er der gennemført en statistisk test, fx en Chi²-test eller en regressionsanalyse? Man skal ALTID teste for statistisk (talmæssig) sammenhæng.
2. Er der tale om en plausibel, meningsfuld forklaring? Det er ikke altid tilfældet. Man har fx påvist en stærk statistisk sammenhæng mellem nationers forbrug af chokolade og antal vundne nobelpriser. Men chokolade er – desværre – næppe en genvej nobelpriser.
3. Hvilken vej vender kausaliteten? Det er indimellem vanskeligt at svare på og følger ikke af en statistisk test. Er det fx dårlig forældrekontakt, der fører til ensomhed hos børn og unge? Eller er det i forvejen ensomme børn og unge, der har dårlig forældrekontakt?
4. Er der testet for 3. variabel, altså for falske sammenhænge? Måske skal den virkelige forklarende årsag til antal vundne nobelpriser findes i nationens generelle uddannelsesniveau. Denne mulige sammenhæng skal man teste statistisk for. Og sådan en test kan afsløre chokolade ⇒ nobelpriser som en falsk sammenhæng.

Hvis man skal minimere de systematiske målefejl og øge en undersøgelses validitet, skal man kunne sandsynliggøre, at undersøgelsen er repræsentativ; at spørgsmålene, der stilles, er relevante og stilles metodisk korrekt; at svarpersonerne er klædt på til at kunne svare; og at eventuelle årsagssammenhænge testes og i øvrigt er meningsfulde.

TILFÆLDIGE MÅLEFEJL

Tilfældige målefejl handler om undersøgelsens pålidelighed, også kaldet reliabilitet. De har den store fordel, at de kan estimeres som statistisk usikkerhed. Vi kan med andre ord sige noget kvalificeret om, hvor stort problemet med tilfældige målefejl er. Det kan man ikke på samme måde med systematiske målefejl. Tilfældige målefejl betyder, at man rammer forkert på en tilfældig, men ret forudsigelig måde. Man rammer altså inden for en bestemt afstand af skivens centrum.



Høj reliabilitet er kendetegnet ved, at undersøgelsen kan gentages igen og igen og med næsten samme resultat hver gang. Det betyder stor præcision og lille statistisk usikkerhed. Den statistiske usikkerhed udtrykkes typisk ved et konfidensniveau, som man på forhånd vælger, og et konfidensinterval, som man efterfølgende beregner. I de fleste sammenhænge vælger man et 95 procents konfidensniveau.

Hvis du læser en meningsmåling fra fx Epinion eller Gallup, beregner de også statistisk usikkerhed ud fra et 95 procents niveau. Med et 95 procents niveau tager man fejl 1 ud af 20 gange. Den risiko lever man fint med i de fleste målinger. Men ikke altid. I medicinske forsøg opererer man gerne med et 99 procents konfidensniveau. Ganske enkelt fordi man her ikke kan leve med at tage fejl hver 20. gang.

Hvis man skal minimere de tilfældige fejl og styrke en undersøgelses pålidelighed og præcision, er der én ting at gøre: Udvid stikprøven; spørg nogle flere!

Det optimale er en undersøgelse, hvor BÅDE de systematiske og tilfældige målefejl er minimeret. Illustreret her:



7 SPØRGSMÅL, DU SKAL STILLE TIL UNDERSØGELSER

1. Hvad er formålet med undersøgelsen?

Der er altid en grund til, at nogen gennemfører en spørgeskemaundersøgelse. Er formålet at sætte en bestemt dagsorden? At skabe en solohistorie? At etablere grundlag for en politisk beslutning? At sælge mere vaskepulver? At påvirke folkestemningen? Du bør altid overveje, hvem der er afsender, og hvad afsenderens formål er. Som i al anden kildekritik.

Afsenderens formål vil indimellem kunne aflæses i skævheder i undersøgelsen og dens konklusioner: Bærer det præg af, at overskrifter og konklusioner nærmest har været tænkt på forhånd? Er der oplagte spørgsmål, som ikke stilles? Er der væsentlige resultater, som ikke formidles eller fremhæves?

2. Er undersøgelsen repræsentativ?

Hvis resultaterne af en stikprøvebaseret undersøgelse skal kunne generaliseres, dvs. fortælle noget om en bredere gruppe, skal man kunne sandsynliggøre, at undersøgelsen er repræsentativ for denne bredere gruppe. Man skal lave en samlet vurdering af stikprøveudvælgelse, svarprocent, frafald og eventuel vægtning af svar.

Der er fx krav til måden, stikprøven er udtrukket på:

1. Man skal have adgang til hele populationen
2. Stikprøven skal udtrækkes simpelt tilfældigt eller stratificeret.

Ideen er, at alle i populationen skal have lige stor chance for at havne i stikprøven. Den første betingelse, adgang til hele populationen, er sjældent 100 procent opfyldt. Mange meningsmålingsinstitutter udtrækker i dag stikprøver enten helt eller delvist fra webpaneler, som i sagens natur ikke indeholder alle voksne danskere.

Det er også almindeligt, at undersøgelser har lave svarprocenter - ofte ned til 20-30 procent af de adspurgte. Det betyder risiko for systematiske forskelle mellem gruppen, der svarer, og dem, der ikke ønsker at deltage eller ikke træffes. En bortfaldsanalyse kan kaste lys over skævheder, og løsningen kan herefter fx være at vægte de indsamlede svar og på den måde opnå tilnærmelsesvis repræsentativitet.

I praksis har du ikke altid mulighed for at tjekke fremgangsmåden, hvis du læser en undersøgelse, som andre har lavet. Men tjek, hvad de skriver om udvælgelse, svarprocent og repræsentativitet. Du finder typisk oplysningerne i undersøgelsesrapportens metodeafsnit. Hvis undersøgelser kommer fra meningsmålingsinstitutter og forskningscentre, stoler man normalt på, at de er nogenlunde repræsentative – hvis de vel at mærke gør rede for det.

3. Hvad er den statistiske usikkerhed?

Statistisk usikkerhed er KUN relevant, hvis der er tale om en repræsentativ undersøgelse. Man skal aldrig beregne statistisk usikkerhed på ikke-repræsentative undersøgelser. Hvis undersøgelsen er repræsentativ, skal der beregnes statistisk usikkerhed på de resultater, du analyserer. Vælg den korrekte analysemetode:

- Overvej, om du skal beregne konfidensintervaller eller lave signifikanstest. Generelt giver konfidensintervaller mere information end signifikanstest.
- Overvej, om resultatet, som du skal beregne usikkerhed for, er en procentandel eller et gennemsnit. Det er to forskellige statistiske beregninger. En procentandel er fx, når 87 procent af danskerne erklærer sig ”meget tilfredse med livet”. Et gennemsnit er fx, når danskerne på en lykkeskala fra 0-10 i gennemsnit erklærer sig ”7,6 point lykkelige”.
- Overvej, om du skal beregne usikkerhed for et enkelt resultat eller for forskellen på to resultater. Forskellen på to resultater er fx, når en ny måling viser, at 87 procent af danskerne erklærer sig ”meget tilfredse med livet” mod 74 procent i en tidligere måling. Der er forskel på tallene 87 og 74, men forskellen er ikke nødvendigvis signifikant, når du tager højde for den statistiske usikkerhed.
- Overvej, om du skal lave en statistisk test for sammenhæng, fx en Chi²-test. Det kan fx være relevant, hvis du vil teste, om der er en sammenhæng mellem alder og tilfredshed med livet.

Husk: En statistisk test er en nødvendig forudsætning for overhovedet at tale om årsagssammenhænge. Og beregning af statistisk usikkerhed er en nødvendig forudsætning for overhovedet at tale om udviklinger og forskelle, fx for at kunne sige, om noget er gået frem eller tilbage.

4. Hvordan er spørgsmålene formuleret?

God spørgeteknik er helt afgørende for at få troværdige svar. Man kan populært sagt få folk til at svare næsten hvad som helst. Vær opmærksom på, om spørgsmålene er afgrænsede i tid, hypotetiske, socialt følsomme, ledende eller balancerede, entydige og forståelige for svarpersonerne. Derudover skal du overveje, om svarpersonerne faktisk har den nødvendige viden, har holdninger til emnet og må formodes at ville svare ærligt på spørgsmålet.

Man må gerne stille hypotetiske spørgsmål. Og det sker hele tiden. Fx ”Hvilket parti ville du stemme på, hvis der var folketingsvalg i morgen?” Men hypotetiske spørgsmål er gratis at svare på. Og svarene kan afspejle meget andet, end hvad man faktisk *ville* stemme – fx at man ønsker at demonstrere utilfredshed med de etablerede politiske partier.

Man må også gerne stille socialt følsomme spørgsmål, altså spørgsmål, som kan opleves som følsomme at svare ærligt på. Fx fordi emnet er forbundet med status, prestige eller politisk korrekthed. Også det sker hele tiden. Fx når man undersøger alkoholvaner, sex, godgørenhed, motion, sort arbejde og kostvaner. Men man risikerer, at folk underdriver, overdriver eller direkte lyver.

Kan du identificere problemerne med spørgsmålsformuleringerne herunder?

(Du kan finde vores bud på, hvad der er galt med formuleringerne på side 11)

- Oplever du – eller har du på et tidspunkt oplevet – at blive udsat for mobning af en medstuderende?
- Har du kendskab til, at en medstuderende er blevet udsat for mobning?
- Er du enig i, at mobning kan have alvorlige konsekvenser for den, det går ud over?
- Ville du gribe ind, hvis du oplevede, at en medstuderende blev udsat for mobning?
- Har du selv mobbet en medstuderende?
- Oplever du mobning, chikane eller anden grænseoverskridende adfærd?
- Er du enig eller uenig i kommissoriet bag regeringens nyligt formulerede mobbekodeks, som skal implementeres på alle landets gymnasier fra årsskiftet?

5. Er svarmulighederne velvalgte?

Tjek at svarmulighederne i undersøgelsen er balancerede, udtømmende, gensidigt udelukkende og veldefinerede. Forkerte og uhensigtsmæssige svarmuligheder kan give systematiske fejl og dermed problemer med målingsvaliditeten.

1. Svarmuligheder skal være balancerede

Ubalance: Vil du sige, at du er lykkeligere end nogensinde, meget lykkelig, ret lykkelig, ikke særlig lykkelig?

Balance: Vil du sige, at du er meget lykkelig, ret lykkelig, ikke særlig lykkelig, slet ikke lykkelig?

2. Svarmuligheder skal være udtømmende

Ikke udtømmende: Hvor meget motionerer du ugentligt? 1 time, 2 timer, 3 timer, 4 timer

Udtømmende: Hvor meget motionerer du ugentligt? Jeg motionerer ikke, mindre end 1 time, 1-2 timer, 3-4 timer, mere end 4 timer

3. Svarmuligheder skal være gensidigt udelukkende

Ikke gensidigt udelukkende: Hvor mange nyhedsindslag ser du på en typisk dag? 0-1, 1-2, 2-3, 3 eller flere

Gensidigt udelukkende: Hvor mange nyhedsindslag ser du på en typisk dag? 0, 1, 2, 3 eller flere

4. Svarmuligheder skal være veldefinerede

Ikke veldefinerede: Lider du af søvnbesvær? Altid, ofte, indimellem, af og til, sjældent, aldrig

Veldefinerede: Lider du af søvnbesvær? Dagligt, flere gange om ugen, ugentligt, sjældnere, aldrig

6. Er der sammenhæng mellem spørgsmål og konklusioner?

Har man faktisk undersøgt det, der konkluderes? Konklusioner skal referere entydigt til de spørgsmål, der er stillet. Det lyder selvindlysende, men er ikke altid tilfældet. I en undersøgelse stilles spørgsmålet: ”Hvilken af regeringens ministre, mener du, klarer sig bedst?” Kan undersøgelsen bruges til at konkludere, hvilken af regeringens ministre der er mindst populær?

Ofte sammenligner man en aktuel undersøgelse med tidligere gennemførte undersøgelser. Men er spørgsmålene formuleret på samme måde? I en undersøgelse stilles spørgsmålet: ”Hvor tit læser du noget af følgende: Romaner, historier og eventyr?” I en opfølgende undersøgelse ændres formuleringen en smule: ”Hvor tit læser/hører du noget af følgende: Romaner, historier og eventyr?” Er de to undersøgelser sammenlignelige?

7. Refererer undersøgelsen sine resultater neutralt og troværdigt?

Slås svarmuligheder sammen i formidlingen? Man ser ofte, at svarmulighederne ”Helt enig” og ”delvist enig” eller ”helt tilfreds” og ”delvist tilfreds” slås sammen i formidlingen. Det er en problematisk praksis, som kan være med til at skjule, at der faktisk er uenighed eller utilfredshed.

Er der ikke-signifikante forskelle, som alligevel fremhæves? Hvis beregning af statistisk usikkerhed viser, at forskelle eller udviklinger ikke er signifikante, skal de ikke formidles som forskelle eller udviklinger. En undersøgelse viser, at rød blok ville få 50,4 procent af stemmerne ved et folketingsvalg i morgen. Den korrekte fortolkning er ”dødt løb mellem rød og blå blok”.

Hvilke resultater fremhæves? Det er et subjektivt valg! En undersøgelse viser, at 75 procent af danskerne bekymrer sig om de globale klimaforandringer. Hvad er nyheden? At 75 procent bekymrer sig, eller at 25 procent af danskerne er ubekymrede? Den væsentlige historie findes ikke altid i de største tal og andele.

Er der resultater, som helt negligeres i formidlingen? Måske er der resultater, som harmonerer dårligt med undersøgelsens formål. Så er der risiko for, at de – bevidst eller ubevidst – nedtones eller helt negligeres. Hold øje med, hvad der *ikke* formidles.

Praktiseres cherry picking? Cherry picking er fx, når man fremhæver enkeltstående data, som bekræfter ens egen opfattelse, og samtidig lukker øjnene for store datamængder, som peger i en anden retning. Et eksempel: 9 ud af 10 samtidige meningsmålinger viser, at tilslutningen til de politiske partier er uændret. Men en tiende måling viser, at det største parti står til at miste 12 mandater. Lukker du øjnene for de 9 øvrige målinger og vælger kun at fokusere på den afvigende? Det er cherry picking – og det skal man holde sig fra.

ORDFORKLARINGER

Kausalitet

Betyder årsagssammenhæng, altså om noget påvirker noget andet. Man taler her om uafhængige og afhængige variable.

Konfidensinterval

Et beregnet interval, som lægger sig omkring en værdi fra stikprøven, når man generaliserer fra stikprøve til populationen. Inden for det interval finder man med en given sandsynlighed (ofte 95%) det sande svar i populationen. Udtryk for den statistiske usikkerhed.

Konfidensniveau

Et tal (typisk 95%) som fortæller, hvor sikre vi er på, at vores generalisering er korrekt, altså at populationens sande svar er indeholdt i vores beregnede konfidensinterval.

Population

Alle individer eller enheder, som deler ét eller flere karakteristika, og som stikprøven ideelt set udtrækkes fra. En population kan fx være alle borgmestre, alle danskere, alle idrætsforeninger, alle tømrerlærlinge.

Reliabilitet

Betyder pålidelighed. Fagudtryk som omhandler en undersøgelses tilfældige målefejl. Kan estimeres som statistisk usikkerhed.

Signifikanstest

Tester ved hjælp af hypotesetest, om en forskel på to andele eller gennemsnit fra stikprøver er signifikant, når man vil generalisere til populationen.

Simpel tilfældig udvælgelse


Ved simpel tilfældig udvælgelse har alle i populationen lige stor chance for at havne i stikprøven, efter samme princip som når man trækker numre i banko eller lodder i en tombola.

Stratificeret udvælgelse

Ved stratificeret udvælgelse opdeler man populationen i lag som fx køn, alder og uddannelse og udtrækker folk fra hvert lag tilfældigt og proportionelt, så stikprøven matcher populationen på hvert lag.

Validitet

Betyder gyldighed. Fagudtryk som omhandler en undersøgelses systematiske målefejl. Vi opererer med tre slags validitet: Intern, ekstern og målingsvaliditet.



Lav dine egne analyser
på spørgeskemadata fra
DDA.DK/SIMPLE-SEARCH

SPØRGSMÅLSFORMULERINGER: HVORFOR PROBLEMATISKE?

På side 8 præsenteres eksempler på problematiske spørgsmålsformuleringer. Her kan du læse, hvad der kan siges at være galt med de pågældende formuleringer.

Generelt

Begrebet ”mobning” defineres ikke. Mobning er i de fleste afskygninger et socialt følsomt emne at spørge ind til.

Oplever du – eller har du på et tidspunkt oplevet – at blive udsat for mobning af en medstuderende?

Største problem: Spørgsmålet mangler afgrænsning i tid. Afspejler ja-svar et aktuelt eller historisk problem? Vi aner det ikke. Derudover et socialt følsomt emne.

Har du kendskab til at en medstuderende er blevet udsat for mobning?

Hvad betyder ”kendskab til”? Risiko for, at mange svarpersoner refererer til samme case. Samlet set svagt belæg. Risiko for, at problemet blæses kunstigt højt op.

Er du enig i, at mobning kan have alvorlige konsekvenser for den, det går ud over?

Ledende spørgsmålsformulering af to grunde: 1) Ledende indledning (Er du enig i.); 2) Ladet udtryk (alvorlige konsekvenser).

Ville du gribe ind, hvis du oplevede, at en medstuderende blev udsat for mobning?

Hypotetisk spørgsmål. Gratis at svare på. Risiko for, at svaret ikke afspejler, hvordan man faktisk ville handle i en virkelig situation.

Har du selv mobbet en medstuderende?

Socialt meget følsomt spørgsmål. Oplagt risiko for, at folk underrapporterer – bevidst eller ubevidst. Mangler også tidsafgrænsning.

Oplever du mobning, chikane eller anden grænseoverskridende adfærd?

Flere spørgsmål i ét. En uskik – spørg kun om én ting ad gangen. Hvad svarer man i givet fald ja til? Mobning? Chikane? Anden grænseoverskridende adfærd? Svært at konkludere præcist på svarene.

Er du enig eller uenig i kommissoriet bag regeringens nyligt formulerede mobbekodeks, som skal implementeres på alle landets gymnasier fra årsskiftet?

Vanskeligt spørgsmål at forstå meningen med. Stiller store krav til svarpersonen. Forventer vi, at svarpersonerne forstår formuleringen? Har viden om det omtalte kommissorium? Og har en holdning til det? Risiko for, at svarpersoner opfinder en holdning til lejligheden i stedet for at svare ”ved ikke” eller ”har ikke taget stilling”.

Rigsarkivet stiller Danmarks største samling af spørgeskemabaserede forskningsdata frit til rådighed for forskere og studerende

Med disse data kan du lave dine egne analyser af
samfundsforhold, politik, sundhed og meget andet.

Du kan bestille data på
dda.dk/simple-search

Yderligere oplysninger ser du på
sa.dk/data

Tekst til guide: Kresten Roland Johansen
Layout: Lind Graphics
Udgivet 2018