

Løsningsforslag metodeoppgaver

Viktig: Dette er til dels ufullstendige løsningsforslag. Dere må ikke sitere direkte fra disse forslagene til eksamen. Dette gjelder også forelesningsnotater – det er ikke vanlig å sitere fra upubliserte arbeider.

Oppgave 10

Kvalitative intervjuer kan følge mer eller mindre forhåndsbestemte temaer og spørsmål. Intervjuguiden for strukturerte intervjuer: klare spørsmålsformuleringer – for ustrukturerte intervjuer: kun bredt definerte temaer.

Ulike former for intervjuer kan plottes inn på en skala med det strukturerte og det ustrukturerte intervjuet som punkter på hver ende av skalaen.

1. Ustrukturerte eller åpne intervjuer:

Tema og spørsmål er i liten grad definert på forhånd. Tema er ofte bredt definert med fordypninger i det informanten finner interessant. Intervjuerens rolle er å fastslå hva samtalen skal dreie seg om men kommer i liten grad med direkte spørsmål. Denne intervjuformen krever at forskeren klarer å følge veldig nøye med under intervjuet. Både for å komme med oppfølgingsspørsmål og hvis intervjuet flyter ut og ikke lenger er relevant i forhold til problemstillingen. Fordi svarene fra ulike informanter kan bli svært forskjellige er ustrukturerte intervjuer vanskelige å organisere og analysere. Det åpne intervjuet kan også være noe mer rettet med forhåndsdefinerte spørsmålsområder.

2. Halvstrukturerte/semistrukturerte intervjuer: Intervjuet foregår på basis av en forholdsvis detaljert utarbeidet intervjuguide. Intervjuene er enklere å analysere fordi de samme spørsmålsområdene er dekket. Samtidig er intervjuformen mer fleksibel enn et strukturert intervju og gir store muligheter for å komme med oppfølgingsspørsmål.

(4) Strukturerte intervjuer: Spørsmålsområder med presise spørsmål. Spørsmålene stilles gjerne med utgangspunkt i et skjema. Dette gjør at svarene lettere kan fremstilles i et skjema og dermed gi gode muligheter for sammenlignende analyser. Intervjuet blir på den negative siden lett for strukturert: det kan lett skje at respondenten sitter inne med masse spennende informasjon som det fastlagte spørreskjemaet ikke klarer å dekke.

Oppgave 13

Observasjonsundersøkelser varierer i forhold til flere dimensjoner:

Lukket – åpent observasjonskontekst
Deltakende – ikke deltakende observasjon
Aktiv – passiv deltakende observasjon
Åpen – skjult forskningsaktivitet

1. Lukket eller åpent observasjonskontekst?

Nyhetsredaksjoner er typiske lukkede kontekster – krever spesiell adgang, både formell og uformell. Åpne observasjonskontekster hører kanskje til sjeldenhetene i medievitenskapen, men et offentlig tilgjengelig kinolokale er et eksempel

2. Deltakende eller ikke-deltakende observasjon

Ikke-deltakende observasjon typisk ustrukturerte og ofte skjult i den forstand at de observerte ikke vet om at det foregår en studie.

Eksempel: Lyn Loflands studie (1971) av hvordan fremmede på offentlige steder forholder seg til hverandre. Oppdaget en utbredt grooming – kontroll av hår/klær/utseende. Kanskje bør man også se på deltakende/ikke-deltakende som endepunkter på en skala, jfr. metodebokas innføring av begrepet tilstedeværende observatør (fordi forskere sjelden kan sies å være en del av nyhetsredaksjonen de studerer).

Metodeboka referer videre til Golds firedeling av observasjonsrollen avhengig av grad av deltaking:

- * fullstendig deltaker
- * deltakende observatør
- * observerende deltaker
- * fullstendig observatør

Firedelingen illustrer i hovedsak skalaen og gradene av deltaking. Inndelingene varierer dermed fra metodebok til metodebok

Et av de mest kjente eksemplene på deltakende observasjon: William F. Whytes studie av en gatehjørnegjeng i Boston på 30-tallet. Whyte fikk selvfølgelig ikke en rolle som vanlig gjengmedlem, men ble i større og større grad en del av observasjonskonteksten han studerte. Deltakende observasjon: tar del i hverdagen og situasjonene som oppstår i observasjonskonteksten – jfr. For eksempel Loflands studie over: i en deltakende rolle har du et forhold til de du studere – dele erfaringer, vise empati.

3. Aktiv eller passiv deltakende observasjon

Som deltakende observatør kan forskeren ha en mer eller mindre aktiv rolle. Whyte tok på seg rollen som passiv men deltakende observatør. Hvis han hadde forsøkt å påvirke gjengen eller konstruere situasjoner hadde han vært en aktiv deltakende observatør. I så tilfelle hadde studien nærmest vært et eksperiment eller felteksperiment. Dette krever at forskeren har stor kontroll og kan manipulere forholdene i miljøet.

4. Åpen eller skjult forskningsaktivitet

Hvorvidt forskeren observerer med åpne kort. Skjult observasjon – etisk problematisk. Vanligst og ofte også like greit å informere om forskningsaktiviteten og prosjektet. Jfr. Også metodebokas diskusjon av betydningen av å vise frem rolle som observatør (s. 113).

Oppgave 14

Ingen faste regler og oppskrifter for å analysere kvalitative data.

Analysen skal være forankret i teori og prosjektets problemstilling. Datamaterialet må vurderes i forhold til *hvor relevant det er for problemstillingene i prosjektet*. Dataene skal handle om problemstillingene og belyse teoretiske poenger. Også kvalitative studier må sikre at resultatene er valide.

Felles for all kvalitativ metode er at den forutsetter tolkning i hele forskningsprosessen. Utgangspunktet er subjektive virkeligheter, og det er ingen forutsetning at undersøkelsen

må finne objektive sannheter om disse virkelighetene. Målsetningen er å utvikle teori om de konkrete hverdagslige erfaringene som undersøkelsesobjektene måtte ha.

I hvert ledd er det forskerens vurderinger som avgjør resultatet. I kvantitativ metode pleier man å si at hvis studien er valid, vil andre forskere komme frem til akkurat de samme resultatene. Dette er prinsippet om etterprøvnbarhet. Kvalitative metoder gjør at det nesten er utenkelig at to forskere kommer frem til nøyaktig samme resultat. Nettopp fordi tolkning har så avgjørende betydning fremhever at kildekritikk også er viktig når analysen gjøres på eget materiale.

Hermeneutisk prosess: man har oftest et teoretisk utgangspunkt som påvirker tolkningen av dataene. De innsamlede dataene vil igjen påvirke forståelsen av det teoretiske utgangspunktet og forhåpentligvis en videre teoriutvikling.

For alle kvalitative data gjelder det at materialet må organiseres. Irrelevante elementer lukes ut, og materialet sorteres ofte i forhold til temaer eller dimensjoner. Hvis man først får en oversikt over de ulike delene i materialet forandres også ofte helhetsinntrykket. Kvalitativ metode er en hermeneutisk prosess. Dette betyr at forståelsen av delene gir en ny forståelse av helheten som igjen kan bety en for forståelse for delene.

Kriterier for god kvalitativ forskning:

- Forskningsprosjektet bør matche analyseobjektene meninger og tolkninger
- Teorien man utvikler bør ha en viss forklarende verdi for tilsvarende situasjoner
- Forskningsprosessen eksplisitt nevne alle prosedyrer og valg som er gjort.

Kvalitative metoder betyr ikke at forskeren kan gjøre som hun eller han vil, men det betyr at tolkninger er en viktig og uunngåelig del av bildet gjennom hele forskningsprosessen.

Vanlig fremgangsmåte for å analysere kvalitative intervjudata og andre kvalitative data:

(1) *Datareduksjon*: alt er ikke like interessant og relevant. Dataene kan sammenfattes i visse deler, og visse deler kan velges ut og velges bort ut fra teoretisk perspektiv.

(2) *Dimensjoner/kategorier*: Intervjudelene eller dataene kan struktureres langs konstruerte dimensjoner for å sammenfatte og sammenligne de ulike svarene. Forskeren benytter gjerne hjelpemodeller og matriser for sin egen del for å systematisere materialet.

(3) *Finne mønstre*: Hele hensikten med intervjuene er å finne mønstre hvis de eksisterer. Hvilke likheter og forskjeller finnes i svarmaterialene? En viss kvantifisering av materialet kan alltid være nyttig for å finne sammenhenger.

Oppgave 17

Tre vanlige måter å undersøke medievaner over tid: retrospektive spørsmål, panelundersøkelser eller tidsserieundersøkelser/gjentatte utvalg:

Ved å bruke *retrospektive spørsmål* kan en intervjuer ett utvalg på ett tidspunkt og stille spørsmål som dekker et tidsrom bakover i tid. Ulempen er at respondentenes svar sannsynligvis ikke er spesielt *reliable* rett og slett fordi det er vanskeligere å gjenfortelle noe som var enn å fortelle om noe som er. Hvis man tar høyde for problemene med en slik datainnsamlingsmetode er likevel retrospektive spørsmål ofte egnet.

Forutseende forskere setter selvfølgelig i gang med *panelundersøkelser*. Man intervjuer da de samme respondentene på ulike tidspunkter. I så tilfelle er det virkelig mulig å studere utviklingen over tid. Problemet er selvfølgelig at panelundersøkelser er kostbare og nødvendigvis strekker seg over lang tid. I tillegg vil man sannsynligvis oppleve at endel av respondentene faller bort underveis.

Ved en *tidsserieundersøkelse* eller *gjentatte utvalg* (læreboka) blir ulike sammenliknbare utvalg intervjuet om de samme spørsmålene på ulike tidspunkter. En slik undersøkelse kan selvfølgelig ikke gi svar på individuelle endringer (slik panelundersøkelsen kan) men kan gi et godt bilde av bruttoendringer som har funnet sted. Det største problemet er derimot hvorvidt de ulike utvalgene lar seg sammenlikne. Selv om utvalgene er trukket ut som et *sannsynlighetsutvalg* kan de være problematiske å sammenlikne.

Oppgave 21

Hvorvidt det er hensiktsmessig å formulere åpne (ingen svaralternativer/verdier er utarbeidet på forhånd) eller lukkede spørsmål (faste svaralternativer/verdier) i en spørreskjemaundersøkelse avhenger av flere forhold. I utgangspunktet ønsker man at svarene lett lar seg sammenligne og at svarene kan fremstilles i tabeller for senere utregning av statistiske mål. Derfor er det oftest en overvekt av lukkede spørsmål.

Hvis *utfallsrommet* for et spørsmål er oversiktlig og begrenset brukes oftest lukkede spørsmål. Utfallsrommet er alle mulige og rimelig samt forskjellige svar som kan tenkes til et spørsmål. Dermed henger utformingen av spørsmål også sammen med hvor mye vi vet om det vi spør. Åpne spørsmål brukes dermed når en vil undersøke respondentens kunnskapsbredde og fantasi – her er det vanskelig å kjenne utfallsrommet.

a) Hvis hensikten er å hente inn bred og grundig informasjon om et fenomen kan det være gunstig med åpne spørsmål – sannsynligvis kjenner man i slike tilfeller ikke utfallsrommet.

b) Dersom alle respondentene har grundige kunnskaper om det de spørres om, kan det i visse tilfeller være bedre med lukkede spørsmål og i visse tilfeller riktigere med åpne spørsmål. En risiko ved lukkede spørsmål er at respondenter egentlig ikke har noen formening om svar, men krysser av et av svarene litt tilfeldig. Når respondentene har grundige kunnskaper er det lite sannsynlig at svarene blir tilfeldige, og dermed kan det argumenteres for at lukkede spørsmål er gunstig, siden det gjør det enklere å sammenligne svarene. Men samtidig kan det være ekstra vanskelig å kjenne utfallsrommet og da vil det være fornuftig å eventuelt supplere med åpne spørsmål.

c) Hvis respondentene har lite kunnskap om et fenomen kan lukkede spørsmål føre til at respondentene velger svaralternativer litt tilfeldig. Samtidig er spørreskjemaundersøkelser sjelden kunnskapstester, men man bør være obs for at lukkede spørsmål i visse tilfeller ikke nødvendigvis fører til valide resultater.

d) Hvis respondentene er lite motiverte, vil som oftest lukkede spørsmål være å foretrekke, fordi de krever mindre av respondenten.

e) Hvis man ikke kjenner respondentene, er det også vanskelig å kjenne utfallsrommet. Man kjenner selvfølgelig ikke 1000 respondenter, men kan kjenner til omtrentlig hva som karakteriserer norske studenter i større grad enn man kjenner til pakistanske familier. I en

tenkt undersøkelse av TV-mønstre blant pakistanske familier kan det dermed være en god ide å hente litt ekspertkunnskap før man utformer ferdige svaralternativer.

Oppgave 23

1. *Eksplorativ forskningsdesign:* Kartlegge et område der det tidligere ikke er drevet systematisk forskning. Man *utforsker* da et nytt område for å bli kjent med undersøkelsesobjektet og problemfeltet. Fordi lite er fastsatt på forhånd er utforskningen også fleksibel – gir mulighet til å ta hensyn til uventede forhold. Ved eksplorativ forskning har man gjerne kun problemstillinger formulert som presiserte temaer. Kan fungere som en del av et større opplegg med målsetning om å komme frem til en mer presis problemstilling.

2. *Deskriptiv eller beskrivende forskningsdesign:* Man har i utgangspunktet klare forestillinger om strukturer og sammenhenger. Teori og gjerne hypoteser å arbeide ut fra > et mer strukturert forskningsopplegg – man trenger ikke å undersøke alt, fordi man har en ide om hvordan det vil være strategisk å gjennomføre studien. Problemstilling utformet som spørsmålsstillinger eller hypoteser.

3. *Kausal forskningsdesign:* Man har i utgangspunktet forestillinger om årsakssammenhenger. Viktig å kontrollere for tidsrekkefølge og for bakenforliggende faktorer. Problemstilling i form av hypoteser som angir årsakssammenhenger mellom variabler/fenomener. Målsetning å påvise årsaker til mønstre.

Oppgave 24

Mulige problemstillinger ut fra artiklene "Fryktens mediepedagogikk" av Alex Iversen og "Hvorfor blir barn drapsmenn" av Ingjerd Skagestad:

Den mest åpenbare hypotesen man finner eksplisitt i artikkelen "Hvorfor blir barn drapsmenn" er: *Det er høy sammenheng mellom vold på fjernsyn og voldelig atferd.*

Andre problemstillinger:

Barn som spiller voldelige videospill kan senke terskelen for å utføre tilsvarende handlinger i virkeligheten.

Artikkelen "Fryktens mediepedagogikk" peker på de komplekse kontekstene mediebruken foregår i. Her er det vanskeligere å finne eksplisitt uttalte hypoteser, men det er naturlig å problematisere mediebruken:

Hvordan forholder barn seg til vold i mediene?

En slik problemstilling vektlegger at mediebruk er en aktiv og tolkende prosess. Dette er hovedpoenget til Alex Iversen.

For videre å undersøke i hvilken grad barn klarer å skille mellom ulike typer vold – for eksempel mellom fiksjonsvold og fakta:

Skiller barn mellom ulike typer vold i mediene og i så fall på hvilke måte?

Oppgave 25

Når man utarbeider problemstillinger, er det nødvendig å vurdere hvorvidt det faktisk er mulig å finne svar på den ved hjelp av ulike metodiske tilnæringsmåter. Hypotesene som er utarbeidet på bakgrunn av Skagestads artikkel stiller store krav til undersøkelsen, og

typiske studier ville vært eksperimenter og observasjon. Slike studier har vært gjennomført – gjerne der mediebruken tas ut av sin kontekst.

De to siste problemstillingene: *Hvordan forholder barn seg til vold i mediene? Og: Skiller barn mellom ulike typer vold i mediene og i så fall på hvilke måte?* er det mulig å få svar på gjennom kvalitative intervjuer med barn om deres faktiske mediebruk.

[I en eksamensbesvarelse bør du i slike oppgaver videre gjøre rede for hvordan du ville lagt opp studien, gjennomført innsamlingen av data – i dette tilfellet gjennomført intervjuene – og hva du bør passe på i analysen av dataene.]

Oppgave 28

Ofta gjengis spørreundersøkelser i avisene uten at feilmarginer oppgis. Feilmarginer angir det intervallet for utvalgsresultatet som vi kan regne med som nesten sikkert at det riktige tallet for universet ligger. Dette gjelder ved *sannsynlighetsfordeling*. Universtallet ligger innenfor feilmarginen for utvalgsresultatet.

I SSB tobakksbruksundersøkelse publisert i Aftenposten 2. februar gjengis hovedresultatene uten å vise til feilmarginene. Vanligvis oppgis resultater fra spørreundersøkelser som sikrere enn de egentlig er. Spesielt små endringer trenger ikke å reflektere endringer i universet. Dette ser man når man inkluderer feilmarginene.

Størrelsen på feilmarginen avhenger av utvalgets størrelse og av resultatet i utvalget. Jo nærmere resultatet ligger 50% desto større feilmargin må vi regne med, og jo større utvalget er jo mindre feilmarginer må vi bruke (se tabell 9.3 på side 258 i metodeboka eller side 327 i Hellevik).

Utvalget i SSBs tobakksbruksundersøkelse er stort: 5000 personer. Dermed blir også feilmarginen mindre: rundt 1%. Aftenpostens overskrift 'Færre røykere i 2003' holder likevel stikk siden nedgangen i andel røykere har gått ned fra 29,4% i 2002 til 26,3% i 2003. Slutningen avisen har trukket er både i overskrift og brødtekst er derfor rimelig. Derimot referer Aftenposten til at det også har vært en nedgang i andelen mannlige røykere og blant andelen unge røykere. Når Aftenposten skriver dette, tar de vitterlig ikke med hele bildet. Utvalget innenfor de enkelte gruppene er for små, til at man med sikkerhet kan si at nedgangen innenfor utvalget også gjelder for populasjonen.

På SSBs websider kan vi lese mer om undersøkelsen, og blant annet lese at hvis resultatene brytes ned i forhold til bakgrunnsvariabler som alder og kjønn, er ikke endringene alltid signifikante. Dette skyldes at utvalgsgruppene blir for små, og feilmarginene følgelig så store at vi ikke med sikkerhet kan generalisere til å gjelde hele universet. Men totalt sett er altså endringene for andel ikke-røykere i Norge signifikante.

Oppgave 30

- a) Spredningsmål sier noe om hvor mye enhetenes verdier i en fordeling varierer. Preges fordelingen av ekstreme verdier eller har de fleste enhetene verdier nær gjennomsnittet? Variasjonsbredden påvirkes sterkt av ekstreme verdier. Variasjonsbredden er lik differensen mellom høyeste og laveste observerte verdi. Dermed vil en enkelt enhet med sterkt avvikende verdi føre til en høy variasjonsbredde selv om alle andre enheter har verdier samlet om gjennomsnittet.

Også standardavviket blir sterkt påvirket av ekstreme verdier. Dette skyldes utregningen av standardavviket, der man kvadrerer hver enhets avvik fra gjennomsnittet.

- b) Modalprosenten er prosentandel av enhetene som har den vanligste verdien. Jo større modalprosent, desto mindre spredning i fordelingen. Modalprosentene er dermed lite påvirket av ekstreme verdier for enkelte enheter. Modalprosenten er derimot uheldig å bruke hvis man skal sammenligne spredningen i ulike frekvensfordelinger: jo færre verdier en variabel har, dess høyere er minimum for modalprosenten.

Oppgave 31

Sannsynlighetsutvalg og ikke-sannsynlighetsutvalg tilsvare det Metodeboka kaller *tilfeldig* og *ikke-tilfeldig* uvalg.

Sannsynlighets- eller tilfeldig utvalg: *Et utvalg der alle medlemmene av populasjonen har en kjent sannsynlighet for å komme med i utvalget.*

1. Enkelt tilfeldig utvalg:

et utvalg der alle medlemmene i populasjonen har *lik* sannsynlighet til å komme i utvalget. Alle enhetene i populasjonen blir nummerert. Ved hjelp av en *tilfeldighetsmekanisme* trekkes det ut numre – disse enhetene tas med i utvalget. Hvor stor sjansen til å bli trukket ut er avhenger av utvalgets størrelse i forhold til populasjonen. Skal man trekke et utvalg på 1000 av en populasjon på 100 000 blir *utvalgsbrøken* $1000/100\ 000$: en hundredel. Utvalget vil da altså være 1 promille av den totale populasjonen.

2. Systematisk utvalg:

Ved systematisk utvalg har alle enheter samme sjanse til å bli plukket ut men ikke alle kombinasjoner av enheter. I systematiske utvalg kommer nemlig enheter som står med en bestemt avstand til hverandre med i utvalget. Skal man for eksempel trekke et utvalg på 1000 av en populasjon på 100 000 tar en hver 100 enhet på lista. Startpunktet kan trekkes tilfeldig ved å velge ett tall mellom 1 og 100. Hvis startpunktet for eksempel trekkes til å bli enhet nummer 57, vil utvalget bestå av enhet nummer 57, 157, 257, 357... 99 957. Systematisk utvelgning forutsetter at enhetenes rekkefølge på populasjonslista er tilfeldig, altså at det ikke er periodisitet i materialet.

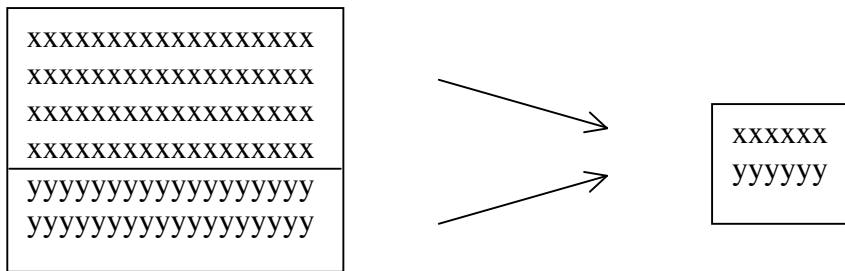
3. Klyngeutvalg

Her skjer utvalget i to runder. Først deles populasjonen inn i *primære samplingsområder* for eksempel på basis av geografi. Deretter trekkes et utvalg av disse klyngene ved hjelp av enkel tilfeldig eller systematisk utvelgning. Enhetene trekkes fra de utvalgte klyngene ved hjelp av enkel tilfeldig eller systematisk utvelgning.

4. Stratifisert utvalg [*Metodeboka skiller ikke mellom disproporsjonale og proporsjonale stratifiserte utvalg: dette er altså ikke relevant for eksamen*].

Innebærer at små grupper systematisk blir overrepresentert i utvalget. Enhetene har altså ikke lik sannsynlighet til å bli valgt ut, men alle enhetene har en *kjent* sannsynlighet til å bli valgt u. Enhetene grupperes sammen i ulike kategorier (strata) på basis av deres verdi på en eller flere variabler. Utvalget kan gjøres proporsjonalt ved å la utvalgsbrøken være lik for de ulike strata: man sikrer at enheter fra hvert strata kommer med i utvalget men det

tallmessige forholdet mellom de ulike strata i utvalget tilsvarer det en har i universet. Utvalget kan gjøres *disproporsjonalt* ved å la utvalgsbrøken være forskjellig for ulike strata – en gruppe blir dermed overrepresentert i utvalget



Oppgave 33: [Oppgaven går utover metodeboka – ikke eksamensrelevant].

Ved et proporsjonalt stratifisert utvalg deles enhetene sammen i kategorier på samme måte som over, men når antallet i utvalget skal fastsettes skal det tallmessige forholdet mellom de ulike strata i utvalget tilsvare det en har i universet. Så lenge denne sannsynligheten er *kjent* er det snakk om sannsynlighetsutvalg. Utvalgsbrøken for de ulike strataene er lik: det tallmessige forholdet mellom de ulike strata i utvalget tilsvarer det en har i universet.

Konkret:

I universet:

600 hunder

400 katter

Det skal trekkes ut 10% i et utvalg: 60 hunder og 40 katter.

Oppgave 35

Utvalgsbrøken er forholdet mellom antall enheter som skal være med i utvalget dividert på antall enheter i universet. Skal utvalget være på 1000 enheter og hele Norges befolkning er universet blir utvalgsbrøken $1000/4\ 500\ 000$, eller $1/4500$. I et proporsjonalt stratifisert utvalg er utvalgsbrøkene for de forskjellige strataene like, men hensikten er altså at det skal være det samme forholdet i utvalget som i universet.

Løsningsforslag oppgave 36

Deduktiv forskning	Induktiv forskning
<ul style="list-style-type: none"> Hypotetisk-deduktiv metode: utvikle hypoteser og teorier som testes mot empiri Betydelig kunnskap og teori eksisterer på området. Forskningsprosess bærer preg av å finne frem til empiriske forhold som kan belyse spørsmål utarbeidet i teorien Deduksjon: konklusjonen 	<ul style="list-style-type: none"> Forskningen har bakgrunn i mer eller mindre tilfeldige iakttagelser Ut fra disse iakttagelsen forsøker forskeren å arbeide seg frem til mer generelle forestillinger om fenomenet. Konklusjonen går utover informasjonen som finnes i premissene > Konklusjonen er <i>sannsynligvis sann</i>.

avledes av sanne premisser. Krever gyldige/logiske resonnementer <ul style="list-style-type: none"> • Modeller som er logisk avledet fra en overordnet teori 	<ul style="list-style-type: none"> • Teoretisk generalisering • Modeller som kan være en syntese av flere empiriske observasjoner • Fra empiri til teori • Starte med datainnsamling • ”Tabula rasa”
--	---

Oppgave 37

Enhetene i en studie er studieobjektene som undersøkes.

- Nyhetsstilbudet på NRK og TV2 har delvis blitt mer mangfoldig siden TV2 startet sine TV-sendinger i 1992.* Enhetene her er NRK og TV2
- Grupper med høy utdanning integreres bedre i det digitale samfunnet enn grupper med lav utdanning.* Her er enheten mennesker. Høy og lav utdanning er dikotomiserte verdier på variabelen utdanning.
- De siste 20-30 årene har det skjedd en kommersialisering av norske medier.* Enheten her er norske medier.

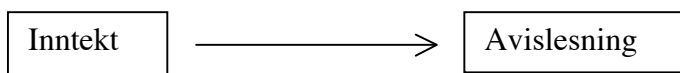
Oppgave 39

Korrelasjon mellom to variabler innebærer kun at det er en statistisk sammenheng og sier i seg selv ikke noe om hvorvidt det er en *årsakssammenheng* mellom variablene. En korrelasjon mellom to variabler kan for eksempel skyldes andre variabler enn de vi har tatt med i en antatt årsaksmodell.

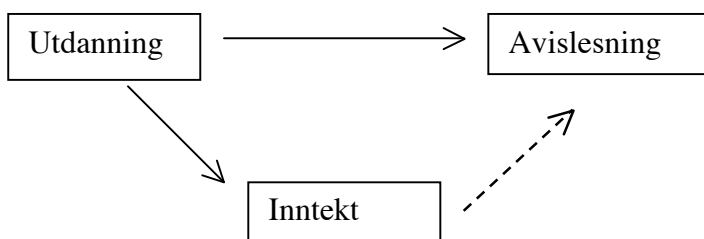
For å påvise at det virkelig er en årsakssammenheng mellom to variabler, må følgende forhold fastslås:

1. Det må foreligge en statistisk sammenheng mellom årsaksvariabelen og effektvariabelen
2. Det må være en bestemt tidsrekkefølge mellom variablene. Årsak kommer før effekt.
3. Vi må vise at sammenhengen ikke er *spuriøs*. Det må ikke være en bakenforliggende faktor som er årsak til både årsaksvariabelen og effektvariabelen.

Vi kan for eksempel tenke oss at vi har funnet en statistisk sammenheng mellom inntekt og avislesning. Høy inntekt gir høy verdi på variabelen avislesning:



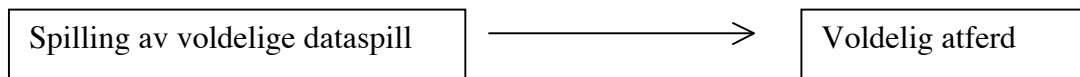
Så viser det seg at sammenhengen mellom inntekt og avislesning blir svekket ved kontroll for utdanning. Utdanning er en bakenforliggende variabel som i dette tenkte tilfellet påvirker både inntektsvariabelen og avislesning.



Dette er et tenkt eksempel. Det er sannsynlig at også inntekt har en effekt på avislesning i seg selv. Hovedpoeng: selv om vi har funnet en statistisk sammenheng, kan det tenkes flere bakenforliggende variabler som egentlig forklarer sammenhengen.

Oppgave 40

En spuriøs sammenheng er en statistisk sammenheng der den uavhengige variabelen ikke forklarer variasjonen i den avhengige variabelen. En bakenforliggende uavhengig variabel er årsak til både den uavhengige (ofte antatte årsaks) variabelen og den avhengige effektvariabelen.



Finner man en sammenheng mellom voldelige dataspill og voldelig atferd, er det naturlig å kontrollere for en rekke bakenforliggende variabler: psykisk helse, familieforhold, sosiale forhold, interesser.

Oppgave 41

Operasjonelle definisjoner angir hvordan et fenomen skal måles. I motsetning angir teoretiske definisjoner mer innholdsmessige kjennetegn ved fenomenet. Å operasjonalisere begreper vil ofte også være nødvendig for å synliggjøre alle relevante aspekter som begrepet dekker. Hvilke *kjennetegn*, aspekter eller dimensjoner rommer begrepet og hvordan skal vi empirisk bestemme om et fenomen omfattes av et begrep og dermed hvorvidt det er relevant for studien. Vi forsøker med andre ord å gjøre begrepet mer konkret. En operasjonaliserings hensiktsmessighet vurderes i forhold til det den skal brukes til. En *operasjonell definisjon* er dermed ofte en detaljert beskrivelse av de nødvendige prosedyrene for å ta stilling til om et empirisk fenomen faller inn under begrepet.

I en spørreskjemaundersøkelse fungerer ofte spørsmålene som operasjonaliseringer av viktige begreper og problemstillinger. Begreper som inngår i teorier og hypoteser blir operasjonalisert gjennom spørsmålene. Et begrep må ofte måles med flere spørsmål for å fange opp ulike aspekter. Også når man skal operasjonalisere i form av å utarbeide en rekke spørsmål som til sammen kan si noe om et fenomen, er første ledd å tenke gjennom hvilke kjennetegn og dimensjoner dette fenomenet omfatter. Høy validitet sikres ved at det er et godt samsvar mellom det empiriske og det teoretiske nivået.

- a) *Medieinstitusjon*: omfatter både organisasjoner og strukturer samt praksiser, normer og konvensjoner for hvordan man gjør ting. Dobbel rolle som materielle institusjoner og symbolproduserende enheter. Få tak i både *produksjons-* og *publiseringsdelen* og institusjonsdelen > forsøksvis operasjonell definisjon: medieinstitusjoner er bedrifter med ansatte og ledelse som produserer og eller distribuerer medieinnhold.
- b) *Nyhet*: I en innholdsanalyse av et utvalg norske aviser, vil en operasjonalisering av nyhet avklare både for forskeren og leseren hvordan kartleggingen bør foregå og

hvordan den har foregått. Hva er en nyhet i forhold til en feature eller kommentar?
 Sub-dimensjoner: Aktualitet, produseres raskt, nærhet i tid og rom, uventede hendelser, dramatiske hendelser, referanser til elite-nasjoner eller elite-personer > forsøksvis operasjonell definisjon: nyhet: avisartikkel som referer til eller beskriver en aktuell hendelse eller utviklingen i en sak [her begrenses nyhet til å gjelde nyheter i aviser].

- c) *Datamediert kommunikasjon*: Forsøksvis operasjonell definisjon: kommunikasjon formidlet gjennom datamaskiner: e-post, chat, messenger. Mobiltelefonen som minidatamaskin – hensiktsmessig å inkludere sms, mms, telefonsamtaler.
- d) *Komedie*: hvordan skal vi vurdere om en film er en komedie og ikke drama, krim eller thriller? Benytte oss av filmens eget sjangerstempel? > forsøksvis operasjonell definisjon: spillefilmer merket komedie eller komedie i kombinasjon med andre sjangrer.
- e) *Populærkultur*: Når er noe populærkultur og når er det finkultur eller subkultur? > forsøksvis operasjonell definisjon: Populærkultur referer her til masseproduserte bøker, filmer, tv-programmer, nettsider, musikkutgivelser, magasiner, radioprogrammer, (videospill?).
- f) *Kunst*: Når er noe kunst, når er det kreativ reklame, kan tegneserier være kunst? Intensjon å uttrykke noe. > forsøksvis operasjonell definisjon: Kunst referer her til originale visuelle, auditive eller audio-visuelle verker som er laget av enkeltindivider, mindre grupper.

[Kritikk: det er veldig vanskelig å operasjonalisere begreper ute av kontekst. Med en overordnet problemstilling blir det ofte greiere å konkretisere hva studien dreier seg om. Dere får ikke denne oppgaven eller tilsvarende til eksamen!]

Oppgave 42

Inntekt kan muligens måles som en variabel på nominalt nivå hvis man dikotomiserer variabelen og opererer med verdiene *ja* eller *nei*.

Eksempel på variabel: *Har du fast inntekt?*

Eksempel på verdier: *Ja – nei*

Eller: løsningsforslag fra metodegruppe 2 [Bedre enn det over]:

Inntekt som variabel på nominal nivå:

Eksempel på verdier: trygd, lønn, studielån

Inntekt målt som variabel på ordinal nivå:

Eksempel på verdier: *Lav – middels – høy*

Inntekt målt som variabel på forholdstallsnivå:

Respondentene oppgir inntekt i tall: eksempel: 250 000 kroner

Oppgave 44

a) Enhetene i undersøkelsen er personer i Norge

b) Det er to variabler i tabellen: Uavhengig variabel: alder. Denne variabelen er på forholdstallsnivå. Den andre variabelen må vel egentlig være noe sånt som: *Hvor mange ganger har du vært på kino siste 12 måneder*. Denne variabelen er også på forholdstallsnivå. Alternativt kan spørsmålet ha vært stilt på følgende måte: *Har du vært på kino i løpet av de siste 12 månedene?* Dette ville vært en variabel på nominalnivå med verdiene *ja* og *nei*.

c) [Dere får ikke hypoteseprøving til eksamen, men for de som er interessert... Husk at selv om dere ikke skal regne ut forkastningsormåder etc. kan det være relevant å skulle forklare hva f. eks. hypoteseprøving brukes til]

Når man trekker et utvalg vil ikke det gi et helt korrekt bilde av populasjonen. Derfor legger vi til et visst slingringsmonn for hvor store variasjoner i utvalgets karakteristika man kan tillate seg – statistisk bestemmer vi en grense for dette slingringsmonnet: denne grensen kaller vi forkastningsområdet til nullhypotesen. Vi begynner med å formulere en nullhypotese og en alternativ hypotese:

Formuler nullhypotese og alternativ hypotese

H_0 : Aldersgruppen 67-79 år går ikke mer på kino enn for elleve år siden. $P(\text{kino}) \leq 0,10$

H_1 : Aldersgruppen 67-79 år går mer på kino enn for elleve år siden. $P(\text{kino}) > 0,10$

[Legg merke til hvordan andel som har vært på kino ikke blir oppgitt i prosent. Dette for at utregningen under skal bli enklere.]

Testen vil være ensidig, fordi vi sier større/mindre enn, ikke lik/forskjellig. Det holder med andre ord at vi finner at sannsynligheten for å få et gitt resultat i utvalget er større enn 0,10 for at vi skal kunne forkaste nullhypotesen.

Før vi kan finne *sannsynlighetsfordelingen* burde vi ha visst hvor mange av de 1856 spurte var i aldersgruppen. Det sier ikke tabellen noe om. Vi antar at med fem alderskategorier, var 20% av utvalget i aldersgruppen: $N = 371$ (20% av 1856).

Finn sannsynlighetsfordelingen for utvalgsresultatet gitt at nullhypotesen er sann:

Vi skal regne ut hvor mange av 371 tilfeldige mennesker i aldersgruppen som går på kino per år, gitt at det er sant at 10% av aldersgruppen 67-79 år går på kino hvert år.

$$M = N \cdot p$$

M betyr gjennomsnitt. N betyr hvor mange som er spurt. P betyr sannsynligheten for å ha vært på kino:

$$M = 371 \cdot 0,10 = 37,1$$

[Vi kunne selvfølgelig regnet ut gjennomsnitt som $M = 371 \cdot 10\% / 100 = 37,1$. Men 0,1 er jo nettopp tallet vi får når vi deler 10/100.]

Videre må vi regne ut standardavviket. Denne måten å regne ut standardavvik på har vi lov til å bruke når antall enheter i utvalget er større enn 30, når $N \cdot p > 10$ og når $N \cdot (1-p) > 10$.

$$s = \sqrt{N \cdot p \cdot (1-p)} \quad [\sqrt{\cdot}: \text{kvadratroten av}]$$

$$s = \sqrt{371 \cdot 0,1 \cdot (1 - 0,1)}$$

$$s = \sqrt{371 \cdot 0,1 \cdot 0,9}$$

$$s = \sqrt{33,39} \approx 5,77$$

Her velges et signifikansnivå på 1%. Signifikansnivået angir sjansen for at hypotesetesten vår skal resultere i at vi forkaster en sann nullhypotese. Det er altså sannsynligheten for å forkaste en sann nullhypotese.

Nå må vi finne *forkastningsområdet* som svarer til et signifikansnivå på 1%. Forkastningsområdet består av de utfallene som er mest usannsynlige dersom nullhypotesen er sann. Nå må vi konsultere en Z-tabell for å finne ut grensen for forkastningsområdet av H_0 . Tabell 9.2 i Metodeboka viser z-verdier. I tabellen ser vi et *signifikansnivå* på 1% i en ensidig test gir en z-verdi på 2,33. Z representerer hvor mange standardavvik over gjennomsnittet kritisk verdi ligger, eller hvor mange prosent av *normalfordelingen* som ligger over verdien

$M + Z * s$. Dette er altså formelen for den kritiske verdien:

$$M + Z * s = 37,1 + 2,33 * 5,77 \approx 37,1 + 13,4441 \approx 50,54$$

Vår kritiske verdi er altså 50,54. Hvis 50,54 eller flere av 371 i aldersgruppen går på kino i løpet av et år, kan vi forkaste nullhypotesen og anta at flere i aldersgruppen 67-79 år går på kino enn for elleve år siden.

Undersøke om utvalgsresultatet ligger i forkastningsområdet eller ikke:

I utvalget var 20% av aldersgruppen på kino: $371 * 0,2 = 74,2$

Utvalgsresultatet ligger i forkastningsområdet til nullhypotesen siden 74,2 er mer enn 50,54.

Konklusjon

Vi forkaster nullhypotesen og beholder H_1 og antar at flere i aldersgruppen 67-79 år går på kino nå enn for elleve år siden.

Oppgave 47

a)

Kjønn	Kvinne		Mann		Totalt
	Høy	Lav	Høy	Lav	
NRK 1	180	100	200	120	600
Ikke NRK 1	95	125	50	130	400
I alt	275	225	250	250	1000

For å regne ut den bivariate sammenhengen mellom alder og har sett NRK1 uttrykt som prosentdifferanse må vi først sette tallene inn i en ny tabell:

Absolutte tall

		Alder	
		HØY	LAV
Har sett NRK1	JA	180 + 200	100 + 120
	NEI	95 + 50	125 + 130
Totalt		525	475

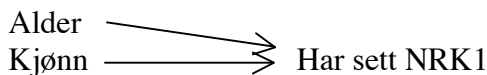
Nå må vi regne de absolutte tallene til prosenter: altså hvor mange av de med høy alder har sett NRK1, hvor mange har ikke sett NRK1, samt hvor mange av de med lav alder har sett NRK1, hvor mange har ikke sett NRK1.

Relative tall:

		Alder		
		HØY	LAV	%-diff.
Har sett NRK1	JA	72,4%	46,3%	26,1
	NEI	27,6%	53,7%	-26,1
Totalt		100%	100%	

Prosentdifferansen mellom alder og har sett NRK1 er 26,1. Dette gir et tall på styrken i sammenhengen mellom alder og andel som har sett NRK1.

b) Den opprinnelige tabellen gir informasjon om sammenhengen mellom 3 variabler: kjønn, alder og har sett NRK1. Da er det interessant å se på hvordan de to uavhengige variablene virker samtidig.



Også i dette tilfellet må vi først prosentueres tabellen:

Kjønn	Kvinne		Mann		Totalt
	Høy	Lav	Høy	Lav	
Alder					
NRK1	65%	44%	80%	48%	60% (n= 600)
Ikke NRK1	35%	56%	20%	52%	40% (n= 400)
Totalt	100% n= 275	100% n= 225	100% n= 250	100% n= 250	100% N= 1000

Vi forenkler tabellen og ser kun på tallene for de som har sett NRK1:

Kjønn	Kvinne		Mann	
	Høy	Lav	Høy	Lav
NRK1	65%	44%	80%	48%

Nå kan vi regne ut effekten av hver av de to uavhengige variablene på den avhengige variabelen *har sett NRK1*:

Effekten av alder når vi kontrollerer for kjønn:

$$E_{\text{alder}} = \frac{(65-44) + (80-48)}{2}$$

$$E_{\text{alder}} = \frac{21 + 32}{2} = \frac{53}{2} = 26,5$$

Effekten av kjønn når vi kontrollerer for alder:

$$E_{\text{kjønn}} = \frac{(65 - 80) + (44 - 48)}{2}$$

$$E_{\text{kjønn}} = \frac{-15 - 4}{2} = \frac{-19}{2} = -9,5$$

Tallene sier noe om styrken i sammenhengen. Vi ser at effekten av alder er sterkere enn effekten av kjønn. Høy alder gir større sannsynlighet for å ha sett NRK1. Vi ante allerede

at alder har mer å si for å forklare om enhetene har sett NRK1. Men vi ser at også kjønn har betydning. At effekten for kjønn er $-9,5$ betyr at flere menn enn kvinner ser NRK1. Vi får et minus-tall fordi vi regnet ut $65 - 80$ og ikke $80 - 65$). Regner du ut effekten av kjønn andre veien ($80-65$ og $65-48-44$) blir resultatet $9,5$. Det blir det samme hvilken vei du regner, men du må passe på å være konsekvent for begge prosentdifferansene og deretter tolke resultatet riktig. Resultatene våre betyr altså at sjansene for å ha sett NRK1 øker hvis vi går fra lav alder til høy alder (effekten av alder er $26,5$) og hvis vi går fra kvinner til menn (effekten av kjønn er $9,5$).

Oppgave 50

a) Spørreskjemaer er strukturerte med faste spørsmål og gjerne faste svaralternativer. Spørreskjemaene benyttes fordi det gjerne er snakk om mange respondenter – gjerne 1000 – og de som intervjues skal svare på de samme spørsmålene. Fordi de fleste spørsmålene har faste svaralternativer kan svarene telles og sammenlignes. Det er på basis av slike undersøkelser at det er meningsfullt å regne ut sentraltendensmål som gjennomsnitt, median og modus.

Det er ulike måter å gjennomføre en spørreskjemaundersøkelse, og det er vanlig å skille mellom en intervjuundersøkelse og enquete. En *intervjuundersøkelse* kan enten foregå ved at intervjueren oppsøker respondentene eller – det vanligste – at intervjuet foregår over telefon. *Enquete* er en spørreskjemaundersøkelse der spørsmål og svar avgis skriftlig. Fordelen med sistnevnte er at det er lettere å få svar på litt ømtålige temaer fordi forholdet mellom de som står bak undersøkelsen og respondentene er helt anonyme.

b) Uavhengige variabler i Fridas undersøkelse: Alder, Kjønn

Avhengig variabel i Fridas undersøkelse: Holdning til Internett

Avhengig variabel er det fenomenet vi vil forklare eller studere variasjoner i. Uavhengige variabler er de faktorene som vi antar påvirker eller forklarer fordelingen på den avhengige variabelen. I Fridas tilfelle er både alder og kjønn bakgrunnsvariabler – forholdsvis konstante over tid. Den avhengige variabelen er en holdningsvariabel. Holdninger er egenskaper som kan endres gradvis.

c) Som oftest vil både intervjuer og enqueter ha en overvekt av lukkede spørsmål. Dette betyr at det allerede er utarbeidet svaralternativer. Årsaken er selvfølgelig at dette gjøre det mye enklere å sammenlikne svarene og utarbeide ulike mål. Et problem dette kan føre til: det er ofte helt umulig å finne rett svarkategori – ingenting passer perfekt og svarene blir bare tilnærmet riktig. Åpne spørsmål brukes derimot også og spesielt når forskeren ikke har mulighet til å ramse opp alle svaralternativene på forhånd. Hvis man for eksempel ønsker å undersøke hvilke programmer et utvalg av respondenter har sett den siste uka, er det like greit å la svaralternativet stå åpent.

Både spørsmålsformuleringene og utarbeidelsen av svaralternativer er veldig viktig. Spørsmålene må for det første være enkelt formulert og *entydige*. Hvis ulike respondenter tolker spørsmålet på ulike måter, vil ikke resultatet være spesielt pålitelig. Spørsmålene kan derfor gjerne være spesifikke og konkrete. Svaralternativene må videre være både uttømmende og ikke overlappende.

Om man utarbeider åpne eller lukkede spørsmål henger også sammen med i hvilken grad man tror man kjenner *utfallsrommet* til et spørsmål (jf. løsningsforslag til oppgave 21)

d) I hvilken grad er folk positivt eller negativt innstilt til Internett?

Når en slik problemstilling skal operasjonaliseres er det viktig å tenke gjennom hvordan folk kan være positivt eller negativt innstilt til Internett. Er det her snakk om informasjonssikkerhet? Er det snakk om teknologiske barrierer som eventuelle barrierer som påvirker folks holdning til Internett. Siden problemstillingen er lite spesifikk, velger jeg her å legge opp til variabler og verdier som til sammen kan gi svar på folks erfaringer med Internett, eventuelle problemer de har med å orientere seg og se hvilke funksjoner nettet kan ha, om de er opptatt av informasjonssikkerhet, om de stoler på informasjonen de finner på nett og om de føler teknologien setter grenser for bruken.

Variabel 1:

Bruker du Internett?

- Ja, ofte
- Ja, av og til
- Lite
- Aldri

Her er det relevant å be de som svarer *aldri* svare på andre spørsmål enn de som bruker nettet:

Variabel 2 for aldri-brukerne

Hvorfor har du aldri brukt Internett?

- Har aldri hatt tilgang
- Ser ikke behovet
- Har ikke mulighet til å sette meg inn i ny teknologi
- Fordi informasjonssikkerheten ikke er god nok
- Andre grunner: Utdyp

De som har svart bekreftende på spørsmål 1 svarer på en rekke andre spørsmål for å få informasjon om den konkrete Internettbruken:

Variabel 2 for brukere:

Hva bruker du Internett mest til?

- Informasjon/nyheter
- Underholdning
- Spill
- Bank, bestillings- og betalingstjenester
- Kommunikasjon: e-post, chat, messenger

[I en eksamensbesvarelse ville det vært lurt å komme med noen flere eksempler og vise hvordan man utarbeider spørsmål som til sammen kan gi svar på hvorvidt respondentene er positivt eller negativt innstilt til Internett].

Oppgave 52

[a) Reliabilitet og validitet går jeg ikke gjennom enda en gang, se metodeboka side 39 og legg spesielt merke til figur 2.1]

b) Hvis vi skal generalisere resultatene fra et utvalg til hele populasjonen må:

- utvalget være enheter som er en del av populasjonen (du kan ikke inkludere svensker i et utvalg hvis du vil si noe om nordmenns TV-vaner)

- må alle enhetene i populasjonen ha en *kjent* sannsynlighet (større enn 0) til å bli trukket ut. Ingen i populasjonen skal i utgangspunktet være avskåret fra å komme med.

Enkelt tilfeldig utvalg, systematisk utvalg, klyngeutvalg og stratifisert utvalg tilfredstiller disse kravene (se løsningsforslag 31)

c)

Kjønn	Kvinne		Mann	
Alder	Under 35 år	35 år og over	Under 35 år	35 år og over
Andel som bruker Internett daglig	35%	19%	47%	28%

Umiddelbart ser vi at flere menn og flere yngre bruker Internett daglig. Vi har altså to uavhengige variabler: kjønn og alder og en avhengig variabel, *Internettbruk*.

Vi er interessert i å finne ut mer om styrken på sammenhengene mellom alder og Internettbruk og kjønn og Internettbruk. Hvilken variabel har mest å si for om enhetene bruker Internett daglig.

Effekten av alder kontrollert for kjønn:

$$E_{\text{alder}} = \frac{(35 - 19) + (47 - 28)}{2} = \frac{16 + 19}{2} = 17,5$$

Effekten av kjønn kontrollert for alder

$$E_{\text{kjønn}} = \frac{(35 - 47) + (19 - 28)}{2} = \frac{-12 - 9}{2} = -10,5$$

Vi ser at sammenhengen mellom alder og daglig Internettbruk er sterkere enn sammenhengen mellom kjønn og daglig Internettbruk. Hvis vi går fra den eldre alderskategorien til den yngre øker sannsynligheten for daglig Internettbruk (effekten av alder er 17,5). Hvis vi går fra kvinne til mann øker også sannsynligheten for daglig Internettbruk, men ikke i like stor grad som for alder (effekten av kjønn er 10,5).

Oppgave 55

Gjennomsnitt: 55 minutter, median: 30 minutter, modus: 0 minutter.

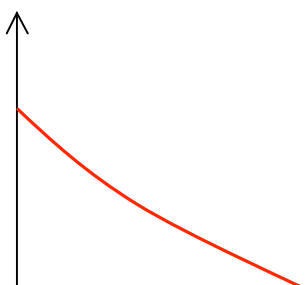
Gjennomsnitt: legge sammen, dividere på antall enheter.

Modus: Verdien som forekommer oftest. Høyest frekvens.

Medianen: verdien som deler en ordnet/rangert fordeling i to.

Når forskjellene mellom de ulike sentraltendensmålene er så store er det grunn til å tro at fordelingen av enheter i undersøkelsen er skjev. Gjennomsnittet påvirkes av ekstreme verdier, og her er det sannsynligvis enkelte enheter som lytter svært mye mer på radio enn flertallet.

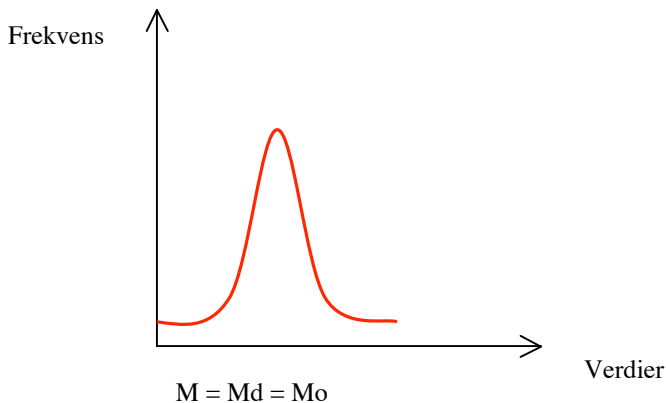
Frekvens



Fordelingen er sannsynligvis skjev slik grafen viser

$$Mo > Md > M$$

Dersom gjennomsnitt, median og modus hadde vært tilnærmet de samme ville vi sagt at fordelingen var symmetrisk:



Oppgave 57

- Uavhengig variabel: bosted. Det er naturlig å anta at bosted kommer før i tid enn bruk av Internett (jfr. Metodeboka om bakgrunnsvariabler, personlighetsvariabler og holdningsvariabler). Tilbudet om Internett-tilknytning (f. eks. via ADSL) avhenger blant annet av folketall.
- Bostedsvariabelen er satt opp som ordinalnivå: det er mulig å rangere fra spredtbygd til ulike grader av tettbygd. Variabelen internettbruk er dikotomisert, det vil si den er gitt to verdier. Spørsmålet om målenivå er da ikke relevant, og vi står fritt til å benytte statistiske teknikker uansett hvilke målenivåer variablene i utgangspunktet forutsetter.
- Tabellen viser at det er en klar sammenheng mellom bosted og andel som bruker Internett daglig. Høy verdi på bostadsvariabelen gir høy verdi (definert som 'bruker Internett daglig') på Internett-variabelen.
- Først regner jeg ut i absolutte tall hvor mange som bruker og ikke bruker Internett daglig for alle verdier på bostedsvariabelen. Deretter dikotomiserer jeg variabelen bosted ved å slå sammen de to laveste verdiene og de to høyeste verdiene:

Sammenhengen mellom bosted og andel som har brukt Internett en gjennomsnittsdag 2002.

bosted	Bruker ikke Internett daglig	Bruker Internett daglig	Sum	(N)
Lav (under 20 000)	744	327	100%	1071
Høy (over 20 000)	452	318	100%	770
Sum	65%	35%	100%	1841

Nå regner jeg ut hvor mange prosent de absolutte tallene utgjør av for henholdsvis lav og høy verdi på variabelen bosted og for andel som bruker Internett daglig (704/1071), (327/1071) osv: nå kan jeg regne ut prosentdifferansene ved å trekke fra høy minus lav verdi for variabelen bruker Internett daglig. I utregningen av prosentdifferansene må man passe på å regne ut prosentdifferansen riktig retning. Tabellen er satt opp motsatt av det vanligste (se oppgaveløsning oppgave 2 side 212):

Sammenhengen mellom bosted og andel som har brukt Internett en gjennomsnittsdag 2002.

bosted	Bruker ikke Internett daglig	Bruker Internett daglig	Sum	(N)
Lav (under 20 000)	69%	31%	100%	1071
Høy (over 20 000)	59%	41%	100%	770
% differanse	10%	-10%	100%	1841

Den utregnede prosentdifferansen sier hvor sterk sammenhengen er mellom den uavhengige variabelen bosted og den avhengige variabelen Internettbruk.

Prosentdifferansen kan variere mellom 0 og 100. Vår prosentdifferanse på 10% innebærer dermed at det er en sammenheng mellom tettbygde strøk og andel som bruker Internett daglig, men denne sammenhengen er ikke veldig sterk.

e) Gamma beregnes ved at vi undersøker om enhetene er ordnet på samme måte langs begge variablene. For hvert par av enheter ser vi på om enheten som har høyest verdi på den ene variabelen også har høyest verdi på den andre. Gamma kan si oss om en sammenheng er monoton – det vil si at den stiger hele tiden eller synker hele tiden. Gamma er et standardisert mål, det vil si at det varierer mellom -1 via 0 til 1 . At Gamma er $0,22$ betyr at det er en sammenheng mellom bostad og andel som bruker Internett og at denne sammenhengen er positiv. Det er altså en tendens at høy verdi på variabelen bosted gir høy verdi på den avhengige variabelen.

f) Det er sannsynlig at også andre uavhengige variabler har betydning for verdiene på den avhengige variabelen: kjønn, utdanning, inntekt, yrke,

g) (Se også oppgaveløsning 52 b) Enkelt tilfeldig utvalg gir gode muligheter til å generalisere resultatene fra utvalg til populasjon. Under tilfeldige utvalg (gjerne kalt systematiske utvalg i andre metodebøker) regnes gjerne enkelt tilfeldig utvalg, systematisk utvalg, klyngeutvalg og stratifisert utvalg. Alle gir muligheter til å generalisere resultatene, men ved enkelt tilfeldig utvalg blir feilmarginene minst [se tabell side 258]. Feilmarginene vi må bruke hvis vi skal generalisere fra utvalg til populasjon avhenger av utvalgsstørrelsen, sikkerhetsnivå (med hvor stor sikkerhet kan vi si at resultatene gjelder for populasjonen) og resultatet i utvalget: jo mer heterogent utvalg (med resultater nærmere 50%), jo større feilmargin. I dette tilfellet må vi regne med en feilmargin på rundt 2%. Det vil si at hvis vi vil generalisere resultatet fra utvalget til populasjonen, vil andel som bruker Internett daglig i spredtbygde strøk med 95% sikkerhet ligge mellom 29% og 33%. Andelen som bruker Internett daglig i tettbygde strøk vil med 95% sikkerhet ligge mellom 39% og 43%.

Oppgave 58

Andel radiolyttere en gjennomsnittsdag, 2002 (konstruerte tall)

Kjønn	Kvinne		Mann		Totalt
	Høy	Lav	Høy	Lav	
Radio	426	100	458	103	1087

Ikke radio	262	150	236	135	783
I alt	688	250	694	238	1870

a) **Bivariat sammenheng mellom kjønn og radiolytting:** Jeg legger først sammen resultatene for høy og lav alder for hhv. kvinner og menn:

radiolyttere en gjennomsnittsdag, 2002 (konstruerte tall)

Kjønn	Kvinne		Mann		Totalt
	Høy	Lav	Høy	Lav	
Radio	426 +100		458+103		1087
Ikke radio	262+150		236+135		783
I alt	688+250		694+238		1870

Så regner jeg ut hvor mange prosent av kvinnene som hørte på radio og hvor mange prosent av mennene som hørte på radio (jeg regner altså ikke ut prosent av totalen men av kvinner og menn for seg). Nå kan jeg også regne ut prosentdifferansen:

Andel radiolyttere en gjennomsnittsdag, 2002 (konstruerte tall)

Kjønn	Kvinne	Mann	% -diff
Radio	56% (526/938 x100)	60%	-4%
Ikke radio	44%	40%	4%
I alt	100% (n=938)	100% (n=932)	N=1870

Prosentdifferansen mellom kjønn og radiolytting er på -4%. Dette viser at det er en svak sammenheng mellom kjønn og radiolytting (0%: ingen sammenheng; +/-100% maksimal statistisk sammenheng). Prosentdifferansen viser at sannsynligheten for å høre på radio er høyere for menn enn for kvinner.

Bivariat sammenheng mellom alder og radiolytting: Jeg legger først sammen resultatene for hhv. kvinner og menn med høy alder og kvinner og menn med lav alder:

radiolyttere en gjennomsnittsdag, 2002 (konstruerte tall)

Alder	Høy	Lav	Totalt
Radio	426+458	100+103	1087
Ikke radio	262+236	150+135	783
I alt	688+694	250+238	1870

Så regner jeg ut hvor mange prosent av de med høy alder som hørte på radio, og hvor mange prosent av de med lav alder som hørte på radio. Nå kan jeg også regne ut prosentdifferansen:

Andel radiolyttere en gjennomsnittsdag, 2002 (konstruerte tall)

Alder	Høy	Lav	% -diff
Radio	64%	42%	22%
Ikke radio	36%	58%	-22%
I alt	100% (n=1382)	100% (n=488)	N=1870

Prosentdifferansen mellom alder og radiolytting er 22%. Dette betyr at det er en sterkere sammenheng mellom alder og radiolytting enn det er mellom kjønn og radiolytting. 22% er en forholdsvis sterk sammenheng. Prosentdifferansen viser at sannsynligheten for å høre på radio er høyere hos de med høy alder sammenlignet med de med lav alder.

b) Effekten av hver av de uavhengige variablene på den avhengige variabelen:

Først regner jeg ut tabellen i prosent. Også nå prosentuerer jeg med basis i kolonnetotalen (dvs. for kvinner med høy alder; for kvinner med lav alder; for menn med høy alder; for menn med lav alder):

Andel radiolyttere en gjennomsnittsdag, 2002 (konstruerte tall)

Kjønn	Kvinne		Mann		
Alder	Høy	Lav	Høy	Lav	Totalt
Radio	62%	40%	66%	43%	58%
Ikke radio	38%	60%	34%	57%	42%
I alt	100% (n=688)	100% (n=250)	100% (n=694)	100% (n=238)	N=1870

Når jeg skal regne ut effektene av hver av de uavhengige variablene tar jeg utgangspunkt i de som har hørt på radio:

Effekten av alder kontrollert for kjønn (Jeg holder verdien for kjønn konstant i utregningen av hver del sammenheng: (høy alder – lav alder for kvinner)+(høy alder –lav alder for menn)):

$$E_{\text{alder}} = \frac{(62-40) + (66-43)}{2}$$

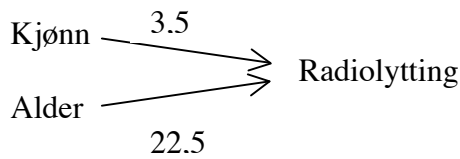
$$E_{\text{alder}} = \frac{45}{2} = 22,5$$

Effekten av kjønn kontrollert for alder (Jeg holder verdien for alder konstant i utregningen av hver del sammenheng: (kvinne – menn for høy alder)+(kvinne –menn for lav alder)):

$$E_{\text{kjønn}} = \frac{(62-66) + (40-43)}{2}$$

$$E_{\text{kjønn}} = \frac{-7}{2} = -3,5$$

Effektene sier noe om styrken i sammenhengen mellom de uavhengige variablene og den avhengige variabelen. Sammenhengen mellom alder og radiolytting er sterkere enn sammenhengen mellom kjønn og radiolytting: alder betyr mer for radiolytting enn kjønn. Variabelen kjønn har liten betydning. At effekten av kjønn er –3,5 indikerer at flere menn enn kvinner hører på radio. Resultatet av den multivariate analysen viser at sannsynligheten for å høre på radio (avhengig variabel) øker fra lav til høy alder (22,5) og fra kvinner til menn (3,5).



Med utregning av effekt kan man kontrollere om de uavhengige variablene påvirker hverandre. Når vi får tilnærmet likt resultat mellom utregningen av effekt og utregningen av prosentdifferanser betyr det de uavhengige variablene ikke påvirker hverandre. Effekten av de uavhengige variablene blir tilnærmet de samme som prosentdifferansene i oppgave a fordi variablene det ble kontrollert for i oppgave b hadde liten innvirkning på resultatet. Det betyr at kjønn har liten betydning i forhold til hvor mye de med høy alder hører på radioen. Prosentdifferansen for kjønn var 4% mens effekten for kjønn kontrollert for alder var 3,5. Dette betyr at alder har en viss innvirkning på effekten av kjønn (om enn en liten innvirkning).

Oppgave 2 side 212/213

Variabler med verdier (kategorier)

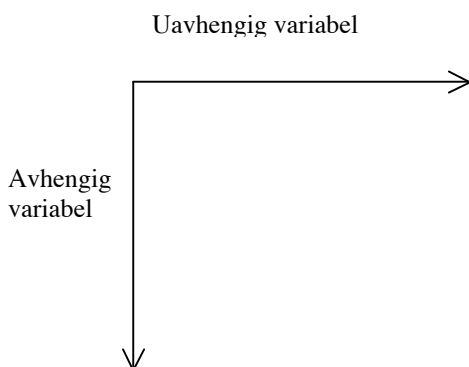
Uavhengige: Utdanning. Variabelen er dikotomisert med verdiene høy og lav

Avhengig: antall kinobesøk. Også denne variabelen er dikotomisert med verdiene høy og lav

Valget av avhengig og uavhengig variabel er naturlig. Det blir forholdsvis søkt å snakke om at for hyppige kinobesøk går på bekostning av utdanning.

Prosentdifferanse

Tabellen er satt opp omvendt. Vanligvis setter man opp:



Derfor må vi passe på å regne ut prosentdifferansen vertikalt (og ikke horisontalt slik eksemplene i boken viser ellers).

$$62-31 = 31$$

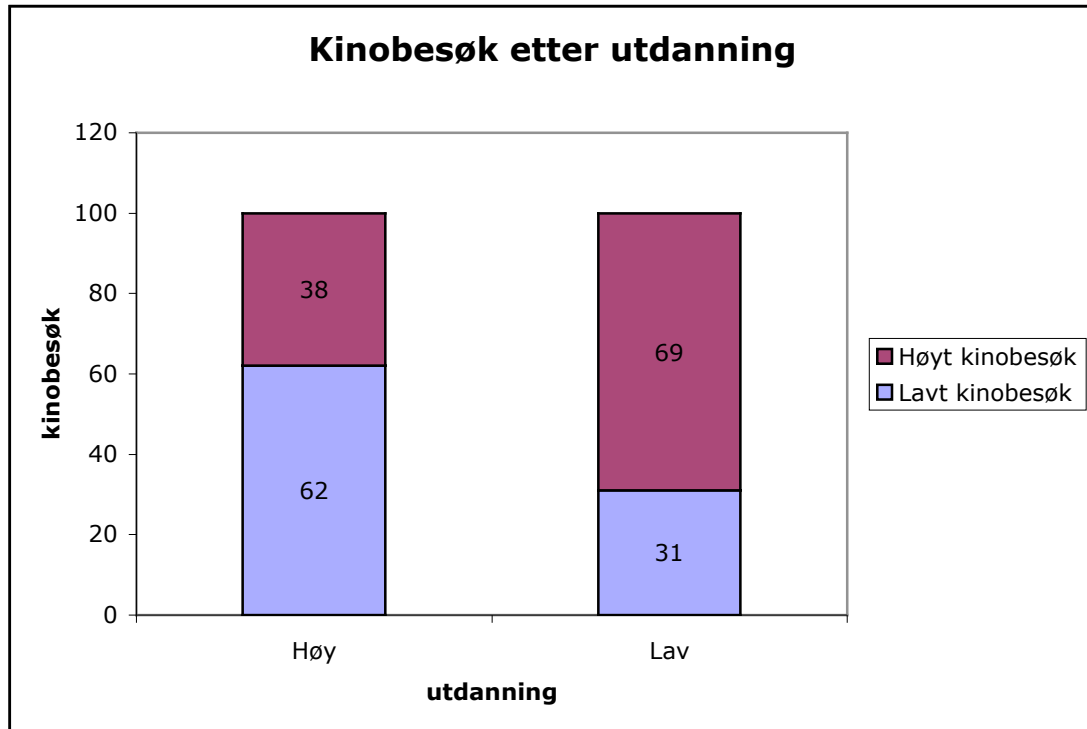
$$38-69 = -31$$

Dette tallet er et mål på styrken i sammenhengen mellom utdanning og kinobesøk. Sammenhengen er forholdsvis sterk (prosentdifferansen kan variere mellom 0 og 100) og

innebærer at sjansene for å ha høy verdi på antall kinobesøk øker hvis vi går fra høy til lav verdi på utdanningsvariabelen.

Søylediagram som viser innholdet

Jeg velger å fremstille tabellen med todelte søyler som til sammen er på 100%. En søyle for de med høy utdanning og en søyle for de med lav utdanning:



Spuriøs sammenheng?

Hvis en sammenheng er spuriøs er det ikke den uavhengige variabelen årsaksvariabelen, men det finnes en bakenforliggende faktor som er årsak til både årsaksvariabelen og effektvariabelen. Hvis jeg skulle undersøke om denne statistiske sammenhengen mellom utdanning og kinobesøk var spuriøs, ville jeg i det minste undersøkt sammenhengen mellom alder og kinobesøk og bosted og kinobesøk.

Oppgave 3 side 213

Hvem leser VG

	Oslo/Akershus	Resten
Lav utdanning	45%	17%
Høy utdanning	23%	27%

Når vi skal regne ut effekten av hver av de uavhengige variablene på den avhengige må vi regne ut prosentdifferensene for hver uavhengig variabel og dele på 2.

Effekten av bosted:

$$E_{\text{bosted}} = \frac{(45-17)}{2} + \frac{(23-27)}{2} = \frac{28}{2} - \frac{4}{2} = 12$$

Effekten av utdanning:

$$E_{\text{utd.}} = \frac{(45-23)}{2} + \frac{(17-27)}{2} = \frac{22}{2} - \frac{10}{2} = 6$$

Når vi skal tolke resultatene må vi passe på hvordan vi regnet ut effektene. I dette tilfellet regnet vi ut effektene av bosted som Oslo – (minus) resten av landet. Når resultatet ble 12 betyr dette at *bosted Oslo/Akershus har en positiv effekt på lesing av VG*. Tilsvarende for effekten av utdanning som ble regnet ut som lav utdanning – (minus) høy utdanning. Effekten på 6 betyr at *lav utdanning har en positiv effekt på lesing av VG*. Det ser ut til at bosted har høyere effekt for lesing enn utdanning. Men selv om vi regner ut prediksjonseffektene bør vi alltid tolke selve tabellen. I tabellen ser vi at det er de med lav utdanning som bor i Oslo/Akershus som trekker opp bosteds betydning. Gjennomsnittet for de to prosentdifferansene for utdanning er 6, men dette tallet er basert på to veldig ulike prosentdifferanser på 22 og – 10. Effekten av utdanning er altså svært ulik i Oslo og i resten av landet. I tabellen ser vi også klart at utdanning har ulik effekt i Oslo/Akershus i forhold til resten av landet. Så når vi gjennom prediksjonseffektene konkluderer med at lav utdanning har en positiv effekt på lesing av VG kamouflerer dette at i resten av landet er det faktisk høy utdanning som har en positiv effekt på lesing av VG.