

Kapittel 1

Confirmation Bias

- Bekreftelsesfeil er en bias hvor man heller ønsker å finne bevis for at ens egen hypotese stemmer, i stedet for å prøve å falsifisere den. Altså blir man mindre sensitiv for falsifiserende funn, og mer sensitiv for bekreftende funn / bevis. I praksis alle mennesker rammes av dette, men det er ikke positivt for forskning / å finne ny informasjon / bevis.

Science

- Vitenskap er aktivitet med mål om å produsere systematisk, reliabel kunnskap om omverdenen, og forklaringer på fenomenene vitenskapen beskriver. Å produsere vitenskap innebærer å følge den vitenskapelige metode, eller i hvert fall en type metode som resulterer i kunnskap som er systematisk og reliabel om omverdenen. Ulike felt innen vitenskapen er ofte svært forskjellige i framgangsmåten deres, og det er vesentlige forskjeller på hvordan for eksempel psykologisk forskning produseres, sammenliknet med geologisk forskning.

Empirical

- Empiri, eller empiriske bevis, er konkrete funn i naturen som svarer på empiriske spørsmål, det vil si spørsmål angående et eller annet konkret, funnet i naturen. Empiriske spørsmål kan besvares gjennom observasjon og analyse, i stedet for for eksempel tro eller ønske.

Basic research

- Grunnleggende forskning er forskning som gjøres med mål om å besvare empiriske eller teoretiske spørsmål, uten noen konkret agenda om applikasjon, før eller siden. Man ønsker altså å besvare grunnleggende spørsmål, som ikke nødvendigvis har noen, eller i hvert fall spesielt høy nytteverdi, men kanskje senere...

Applied research

- Applied forskning er forskning med et konkret mål om å løse et problem som eksisterer. Selv om denne forskningen også kan genereres ut ifra teorier er det i større grad praktiske problemer / spørsmål man ønsker å løse eller besvare.

Protoscience

- Protovitenskap er forskning helt på kanten av vitenskapelig forståelse på et gitt tidspunkt. Vanligvis følges vitenskapelig metode, men om claims ikke kan bevises empirisk kalles dette pseudovitenskap...

Scientific explanation

- En vitenskapelig forklaring er et svar på et vitenskapelig spørsmål som bygger på empirisk beviste fakta, og er altså ofte en forklaring på hvorfor eller hvordan et fenomen oppstår eller arter seg. "... Dette har altså en vitenskapelig forklaring..."

Parsimonious explanation

- En "parsimonious" forklaring er en forklaring på et fenomen som bygger på færrest mulige antagelser. Om det finnes flere mulige forklaringer på et fenomen vil de fleste foretrekke den som bygger på færrest antagelser, altså den mest "sparsommelige" forklaringen.

Commonsense explanation

- En sunn fornuft-forklaring er en forklaring som bygger på såkalt sunn fornuft, og er ikke det samme som en vitenskapelig forklaring. "Det er sunn fornuft at solen roterer rundt Jorden, vi står jo stille..."

Belief-based explanation

- En trosbasert forklaring er noe fundamentalt annet enn en vitenskapelig forklaring, da denne bygger på allerede etablert tro, sammenliknet med observert, analysert empiri. "Gud har skapt universet og alt i det", "Hva med evolusjonen og big bang teorien...?".

Pseudoexplanation

- En pseudoforklaring er en ikke-forklaring, det vil si en forklaring som egentlig ikke forklarer noe som helst. Pseudoforklaringer kan virke troverdige (aggressivt instinkt i boka s.23), men kan være sirkulære, kun et annet navn på samme fenomen osv.

Circular explanation / Tautology

- En sirkulær forklaring / tautologi er en type pseudoforklaring hvor forklaringen både brukes som forklaring, og bevis for egen forklaring. "Aggresjon skyldes et aggressivt instinkt, hvor aggresjonen skyldes instinktet, som beviser at aggresjonen skyldes instinkt (klønete formulert, grafikk s.23).

Method of authority

- Autoritetsmetoden er å oppsøke kilder til informasjon om et fenomen fra kilder som en anser å ha autoritet. "Jeg spurte presten om Jesus gikk på vannet, han sa ja." Hvilket svar man får avhenger av hvem eller hva man spør, og det er opp til ens subjektive vurdering av kildens autoritet hvor troverdig man tenker svaret er.

Rational method

- René Descartes mente at all kunnskap om universet kunne avledes fra resonnering, basert på grunnleggende selv-bevisende fenomener. Cogito ergo sum, og eventuelt andre selv-bevisende påstander kunne i praksis gi svar på alt annet. Skepsis er sentralt, altså skal man tvile på, eller ikke tro på, påstander som ikke er logisk beviselige. Dessverre er det vanskelig å finne korrekte svar på alt man lurer på kun gjennom resonnering, men likevel benyttes elementer av den rasjonelle metoden i vitenskap. I tillegg kan man risikere å få feil svar på ens spørsmål, om premissene for resonneringen er feil (som de kan være...).

Deductive reasoning

- Deduksjon er en prosess der man trekker kunnskap hentet fra et kjent faktum som er mer bredt, til å finne informasjon om andre gjerne mer spesifikke fenomener. "Det som tenker, må eksistere, jeg tenker, derfor er jeg". Brukes i forskning, men kombineres med testing basert på observert empiri.

Kapittel 2

Theory

- En teori er en hypotese som har blitt replikert og bevist i en såpass stor grad at den har oppnådd et nytt nivå av troverdighet, samtidig som den er troverdig eller "vitenskapelig akseptabel", er predikerende for ny, ukjent informasjon, og gjelder / er nyttig i et bredt system av vitenskapelig kunnskap. Eksempel: Teorien om evolusjon forklarer en rekke fenomener på en troverdig måte, som forskjellen på fugler på Galapagosøyene, samtidig som at den også kan bidra i å predikere utviklinger i evolusjon; De skilpaddene som ikke er motstandsdyktige mot klimaendringer og avfall i havet vil dø ut, mens de som er det vil overleve (absurd eksempel).

Hypothesis

- I den vitenskapelige metode er en hypotese en påstand (forklaring bak fenomen) som man opererer ut ifra, og prøver å teste. Hypoteser må være testbare og tentative (konkrete / fysisk eksisterende).

Law

- Et vitenskapelig bevist, kvantitativt forhold mellom to eller flere variabler, for eksempel loven om tyngdekraft, eller *matching law*, Richard Herrnsteins foreslåtte beskrivelse av hvordan duer fordelte tastetrykkene sine ifht to forskjellige forsterkningskjemaer (animal learning kontekst, s.37).

Model

- Kan være et synonym til teori, men brukes oftest om implementeringer av generelle teoretiske perspektiver i en gitt kontekst. Eksempel: Rescorla-Wagner-modellen om klassisk betinging og dens formalisering av mer generell teori om assosiative læring.

Mechanistic explanation

- En mekanistisk forklaring beskriver *hvordan* noe fungerer, altså mekanikken eller kjeden av hendelser som ligger bak et fenomen. Eksempel: Koffein gjør at du ikke blir trøtt fordi hormonet adenosin ikke blir tatt opp i hjernen siden koffein i stedet binder seg til adenosinreseptorene og derfor blokkerer hormonet.

Functional explanation

- En funksjonell forklaring tar utgangspunkt i et gitt fenomen og forklarer *hvorfor* fenomenet arter seg som det gjør. Eksempel: Overvektige kvinner er mindre attraktive fordi de forvirrer menneskehjernen til å tro at de er gravide, og derfor ikke er fruktbare.

Quantitative theory

- Kvantitative teorier er teorier hvor dens variabler og konstanter er definert ut fra matematiske formler og begreper. Eksempel= $E=MC^2$. Kvantitative teorier finnes også innenfor psykologi, for eksempel Norman Andersons "information integration theory" fra 1968 (s.43).

Kvalitativ teori

- Enhver teori som ikke defineres ut ifra matematiske begreper er en kvalitativ teori. Altså er en kvalitativ teori en teori som forklares med verbale begreper. Eksempel: Noam Chomskys "theory of language acquisition" fra 1965.

Descriptive theory

- Teorier har i praksis tre dimensjoner, nemlig graden av beskrivelse av fenomenet som teorien gir, altså om teorien beskriver eller også forklarer fenomenet. En deskriptiv teori tilhører det laveste nivået, og beskriver kun fenomenet og hvordan det arter seg, i stedet for å også forklare det.

Analogical theory

- En analogisk teori er en teori tilhørende det andre nivået av beskrivelse, der fenomenet også forklares gjennom analogier fra kjente fysiske systemer. Boka bruker en teori av Konrad Lorenz, "Lorenz's hydraulic model of motivation" fra 1950, s.46-47.

Fundamental theory

- En fundamental teori (begrep gitt av boka) tilhører det tredje og høyeste nivået av beskrivelse av forklaring som en teori kan gi, hvor det ikke benyttes noen analogier til

allerede kjente systemer, men at det heller brukes nye begreper spesifikke for teorien (på det gitte tidspunktet den kommer) for å beskrive systemet teorien har påvist.

Domain

- En teoris domene eller "scope", er i praksis området en teori belyser, altså hvor bredt en teori kan appliseres. Det er fordeler og ulemper med teorier som kan appliseres bredt og mindre bredt. En teori som dekker et smalere domene har større sannsynlighet for å beskrive (og forklare) systemene teorien gir uttrykk for på en bedre måte, men det kan være vanskelig å kombinere flere smale teorier og oppnå et helhetlig bilde, selv om begrepene som brukes er de samme.

Confirmational strategy

- En bekreftende strategi brukes vanligvis i startfasen til testing av teorier, for å undersøke om teorien har noen form for prediktiv evne. Om resultatet av bekreftende strategier er positivt, er teorien styrket, og motsatt. Her er det nyttig om det er flere forskjellige potensielle hendelser som kan oppstå, som beskrevet i boka, med 1/3 sjans for bekreftelse (om alt er tilfeldig), sammenliknet med en lavere sannsynlighet for positivt utfall om alt er tilfeldig.

Disconfirmational strategy

- I praksis det samme som å prøve å falsifisere en teori, der man heller har fokus på å finne ut om teorien har noen grad av prediktiv evne ved å prøve å motbevise den. Dette er viktig for å finne ut om en teori er gjeldende eller ikke (erroneous). Her er et positivt resultat at en teori avkreftes / falsifiseres.

Strong inference

- John Platt mente at den riktige framgangsmåten i forskning er å produsere flere testbare teorier bak et fenomen, og dermed teste disse opp mot hverandre med eksperimenter som avkrefter noen av dem, og støtter andre. "Strong inference" er i praksis at man trinnsvis prøver å avkrefte forskjellige teorier bak et gitt fenomen, og vurderer den siste gjenværende teorien som riktig eller sterkest.

Kapittel 3

Informed consent

- At deltagere i forskning kan gi og gir informert samtykke er en essensiell del av å gjennomføre et etisk forsvarlig forskningseksperiment. Dette innebærer at deltagere informeres om de potensielle risikoene og "personlige kostnadene" deltagelse i eksperimentet innebærer, og at deltagerne har mulighet til å avgi samtykke eller nekte deltagelse avhengig av hva de selv ønsker. Det er viktig at informasjonen som gis er sann, selv om dette kan være en problemstilling ifbm eksperimenter der deltagere må "føres bak lyset" for at eksperimentet skal være valid. Å finne en balanse mellom informering av viktig og nyttig informasjon, men å samtidig ikke gi informasjon som kompromitterer eksperimentets nytte og kvalitet. Om det er en konflikt mellom disse må det gjøres en etisk beslutning, med fokus på ivaretagelse av deltagerens rettigheter og velvære.

Ill-gotten information

- Ill-gotten information er et begrep som refererer til funn gjort i forskning som ble ekstrahert ut ifra eksperimenter der deltagere eller andre kom til skade eller ble utsatt for etisk tvilsomme situasjoner. Ill-gotten information er altså informasjon hentet på tross av etiske retningslinjer, og bruken av denne kunnskapen kan ofte heller ikke forsvares da det finnes en risiko for at det skapes insentiver til å tilegne seg etisk uforsvarlig kunnskap.

Nuremberg Code

- Etter andre verdenskrig ble "Nuremberg Code" skapt som en samling av ti retningslinjer for etisk forsvarlig innsamling og produksjon av kunnskap. Nuremberg Code ble tatt inn i APA og HHS (departement i USA) sine etiske retningslinjer. Noen viktige punkter er at deltagelse i eksperimenter må være valgfritt, og at deltagere har rett til å vite detaljer rundt naturen, målet og lengden til et forskningsprosjekt de deltar i.

Declaration of Helsinki

- Helsinkierklæringen kom i 1964, og ble tatt i bruk av World Medical Association, hovedsakelig for bruk i medisinsk forskning, men også for samfunnsvitenskapene. Blant annet må forskningsprosjekter vurderes som etisk forsvarlige av eksterne, uavhengige grupper, og forskningen som gjøres må bygges på og samsvare med vitenskapelig aksepterte prinsipper og kunnskap. Alle retningslinjene i Helsinkierklæringen er implementert i APAs etiske retningslinjer.

Belmont Report

- Belmont-rapporten kom i 1979, og beskrev en rekke etiske prinsipper, blant annet "Respect for persons." (respekt for personers autonomi og rettigheter, og at personer med manglende evne til å ta informerte beslutninger må beskyttes), "Beneficence", altså at forskere ikke kan skade deltagere, og må maksimere fordelene ved forskningen kombinert med minimering av skade. "Justice" beskriver at deltagere og forskere skal dele kostnadene og de potensielle fordelene av forskningen, og at deltagere skal velges til forskning basert på nytte, og ikke på grunn av faktorer som tvang, at de er praktiske å bruke, eller tilhører en spesifikk minoritet.

Beneficence

- Definert i forrige punkt.

Respect for persons

- Definert i punkt før forrige punkt

Justice

- Definert i punkt "Belmont Report".

Ethical principles of psychologists and code of conduct 2002

- APAs etiske retningslinjer ble først påbegynt i 1947, og ble introdusert i 1953. Retningslinjene har gjennomgått et mylder av revisjoner og oppdateringer, senest i 2002. Dokumentet har blitt utbrodert i 2010 og 2016, med tydeliggjøring av protokoll om retningslinjene er i konflikt med lover eller organisasjonsbehov.

Institutional review board (IRB)

- Er en institusjon med formål å vurdere forskning som etisk forsvarlig, samt vurdering av risiko sammenliknet med potensielle fordeler av forskningen. IRB deler også risikoen for uheldige konsekvenser av forskning, da et prosjekt godkjent av institusjonen også står ansvarlig for disse. Derfor beskytter IRB også forskere mot økonomisk ruin om deltagere for eksempel skades under forskningen. Likevel er mange forskere lite glad i IRB, da de mener institusjonen har blokkert forskning som (fra forskerens perspektiv) er nyttig.

Institutional animal care and use committee (IACUC)

- IACUC er en institusjon med formål å sikre dyrs etiske behandling i forskning. Institusjonen krever for eksempel at det skal være til stede en veterinær med doktorgrad i veterinærmedisin samt erfaring, trening eller sertifisering i bruk av dyr i forskning.

Office of research integrity (ORI)

- ORI er en institusjon som har som mål å sikre at "research misconduct" ikke finner sted. Begrepet defineres som forskning der; fabrikasjon av data, manipulasjon av data, materialer, utstyr eller prosesser som ikke vises i forskningsloggen, eller plagiat, er til stede. ORI har også utformet rettigheter som skal beskytte "whistleblowers" fra negative konsekvenser av handlingen, da ORI betrakter dette som en essensiell del av kampen mot "research misconduct".

Kapittel 4

Empirical question

- Et spørsmål som kan besvares gjennom objektiv observasjon er et empirisk spørsmål. Altså må empiri benyttes til å besvare spørsmålet. Spørsmålet må baseres på definerte begreper, og kan derfor ikke være for bredt.

Operational definition

- En operasjonell definisjon er i praksis en konkretisering av hva relativt diffuse begreper betyr i en gitt kontekst, slik at disse kan måles med objektiv definisjon. Om du ønsker å måle depresjon, kan dette for eksempel gis en operasjonell definisjon lik "frekvens av depriment humør" som gjennom måling kan gi en skåre.

Literature review

- Å undersøke forskningslitteraturen i feltet man er interessert i å undersøke er nyttig av flere grunner.
 - Du kan oppdage om eller om ikke det eksperimentet man har planer om å gjennomføre allerede er gjort.
 - Om spørsmålet man ønsker å stille er besvart tilfredsstillende allerede (likner på første punkt).
 - Man kan hente inspirasjon for eventuelle forskningsdesign fra liknende eller relevante eksperimenter eller studier allerede gjennomgått.
 - Man kan få et inntrykk av hvilke spørsmål som ikke er "besvart", og dermed heller forsøke å besvare disse.

Primary source

- En primærkilde er en forskningsartikkel som inneholder tilstrekkelig informasjon til å kunne repliseres. Altså inneholder kilden detaljert informasjon om forskningens motivasjon, deltagere, utstyr / materiell, prosedyre, resultater og referanser. Kilder til forskningsinformasjon klassifiseres som primære eller sekundære.

Secondary source

- En sekundær kilde er en oppsummering, beskrivelse eller gjenfortelling av én eller flere primærkilder, og begrepet gjelder også metaanalyser. Bruk av sekundærkilder kan være nyttig for å få et overblikk over den relevante litteraturen, men det er nødvendig at man er obs på at forfatteren av sekundærkilden kan ha gjort feil eller tolkninger som ikke er riktige eller relevante for din forskning. Overdreven bruk av sekundærkilder er derfor ikke anbefalt, og om man skal bruke disse bør man prøve å forbedre disse gjennom lesning av andre relevante primærkilder, i stedet for å bare oppsummere en oppsummering. Typiske sekundærkilder kan være filmer, bøker, forelesninger, TV-programmer og magasiner / aviser.

Refereed journals

- "Bedømte journaler" er journaler hvor publisering av forskning innebærer gjennomgang fra vanligvis to eller flere "reviewers", og informasjon publisert i slike journaler er vanligvis mye mer troverdig enn forskning i såkalte "nonrefereed journals", selv om gjennomgangprosessen ikke nødvendigvis er en garanti.

Nonrefereed journals

- "Nonrefereed journals" er journaler hvor publisering av forskning ikke innebærer gjennomgang, men ofte heller kun økonomisk kompensasjon. Slike journaler har en tendens til å inneholde langt mindre troverdig eller pålitelig informasjon, altså forskning av lavere kvalitet. APA har verktøy for undersøkelse av kvalitet på journaler, "Journals in Psychology", og dette viser publiseringssraten til forskjellige journaler. Lavere enn 15% publiseringssrate antyder en journal av høy kvalitet.

Paper session

- Bøker og journaler kan ofte inneholde informasjon som er eldre, og en måte å få adgang til splitter ny informasjon kan være på såkalte "paper sessions" på forskningskonvensjoner. Her får forfattere av forskningsartikler anledning til å gi en 15-20 minutter lang presentasjon av forskningen sin. På denne måten kan man få tilgang på ny, mindre generell informasjon, men den korte presentasjonen nødvendiggjør som regel utelatelset av noe informasjon, ofte ifbm forskningens metodologi.

Poster session

- En "poster session" innebærer at en forsker har produsert en plakat som inneholder det meste av relevant informasjon fra forskningen sin, hengt den opp på en konvensjon, og står ved den klar til å besvare spørsmål og diskutere forskningen med de som har lyst til det. På denne måten kan man knytte kontakt til forskere på sitt eget felt, og hente nyttig informasjon fra en erfaren profesjonell forsker. Ulempen er den samme som med paper sessions, nemlig at besøk av konvensjoner ofte kan være dyrt (pga reisekostnader osv.).

Personal communications

- Personlig kommunikasjon refererer her til hvordan det faktisk er mulig å personlig nå ut til forskere og stille spørsmål eller på en annen måte åpne for dialog. Svar man får på denne måten kan være svært nyttige, og går under personlig kommunikasjon, sammenliknet med noen annen form for kommunikasjon.

PsycINFO

- PsycINFO er et verktøy for søk etter kilder innenfor psykologi tilbudt av APA, som dekker over 2500 journaler, i tillegg til dekning av diverse bøker, avhandlinger og andre relevante kilder innenfor psykologi. Her er ikke alle kilder nødvendigvis tilgjengelige for bruk uten videre.

PsycARTICLES

- PsycARTICLES er et annet verktøy APA tilbyr der kildene er direkte tilgjengelige for nedlastning, men med et mindre bredt spenn enn PsycINFO, da kildene hentes fra journaler tettere knyttet til APA, Canadian Psychological Association og et par andre utgivere. PsycINFO og PsycARTICLES er tilgjengelige med en rekke søkemotorer, men også APAs egen søkemotor PsycNET.

File drawer phenomenon

- En del forskningseksperimenter klarer ikke å oppnå positive resultater (funn som støtter hypotese) med alfa under 0,5 (statistisk signifikante resultater), og i disse tilfellene havner ofte disse forskningsartiklene i "filarkivet", uten at noen får vite noe mer om dem. Dette

resultater sannsynligvis i at forskning med negative resultater repeteres av andre forskere, noe som bruker opp ressurser forholdsvis unødvendig. Det har oppstått en bevegelse for å såkalt "preregistrerte" studier, slik at informasjon om forskningen er tilgjengelig enten resultatet er positivt eller negativt. Dette hindrer også teoretisk en del "research misconduct", ifm at studiene allerede er registrert i sin opprinnelige form, og ikke kan endres uforklarlig på samme måte som hvis den ikke var det.

Peer review

- Peer review er en prosess gjort vanligvis av to ulike "eksperter", vanligvis i forbindelse med "refereed journal"-publikasjon. Årsaken til at ordet "ekspert" er i hermetegn er at prosessen har vist seg å ha en del problematiske aspekter, deriblant at ekspertene som skal gjennomgå forskningen ofte ikke er eksperter, eller i hvert fall på det spesifikke feltet som de gjennomgår. I tillegg til dette er det også en del effekter av kjønn, anerkjennelse av institusjonen forskningen er gjort ifm med, og "peer-review"-ernes egne biases. Det har blitt gjort forskning som antyder at det kan være relativt tilfeldig om en studie blir godkjent for publikasjon eller ikke, og dette skyldes ofte ekspertenes inkompetens. Til tross for de mange problemene med peer-review-prosessen er det likevel kanskje et bra tiltak for kvalitetssikring, da det i hvert fall er større sjans for at en studie er av god kvalitet om den er godkjent, selv om bedre studier kan unnlates å bli godkjent i flere tilfeller.

Kapittel 5

Causal relationships

- Kausale forhold kan kun påvises eksperimentelt, det vil si der forskere kan påvirke uavhengige variabler fritt. Korrelasjon og kausasjon er ikke det samme, og kausalitet mellom variabler påvises når det er påvist et konkret årsaksforhold mellom dem, og ikke bare at de forekommer sammen.

Correlational relationship

- Korrelasjonelle relasjoner kan påvises i situasjoner hvor ulike variabler forekommer sammen eller samtidig, uten at man kan si noe om årsak. Typisk gjøres korrelasjonell forskning gjennom måling av variabler slik de arter seg i "naturen", og ikke i en laboratoriesetting.

Correlational research

- Forskning på korrelasjon tilhører kategorien noneksperimentell forskning, og er forskning hvor man ønsker å observere om to eller flere variabler kovarierer, dvs at det finnes en sammenheng mellom hvordan de to opptrer. Om det er mye eller lite av den variabelen, er det sannsynligvis mye eller lite av den andre, avhengig av hvordan korrelasjonen ser ut. I korrelasjonell forskning ønsker man å observere variabler slik de arter seg naturlig, og ikke manipulere dem i det hele tatt, i motsetning til eksperimentell forskning.

Third-variable problem

- I noneksperimentell forskning, og forså vidt også i eksperimentell forskning (selv om målet er å unngå dette), finnes det situasjoner hvor det er en tredje, ukjent variabel som forårsaker sammenhengen mellom de observerte variabelenes relasjon. Eksperimentell forskning innebærer at man har som mål å identifisere og deretter fjerne og / eller minimere tredje variabelers påvirkning på eksperimentets resultater. Tredjevariabler / konfundering er hovedårsaken til at man ikke kan påvise kausalitet med korrelasjonell forskning, fordi man ikke kan kontrollere disse på noen måte på grunn av forskningens natur.

Directionality problem

- Hvilken retning en kausalitet går, dvs hvilken variabel som har skylden i at den andre endres, er vanskelig eller umulig å vurdere ut ifra korrelasjonell forskning, og dette er nok en grunn til at kausalitet ikke kan påvises med korrelasjonell forskning.

Experimental research

- Eksperimentell forskning er forskning der man bevisst og aktivt manipulerer en uavhengig variabel i forhold til en eller flere avhengige variabler. Her er det også essensielt at potensielle tredjevariabler er identifisert og fjernet eller minimert (under kontroll), for å kunne påvise kausalitet mellom variabler. Her kan også kausalitetens retning bestemmes, da den uavhengige variabelen er påvist forårsakende hvis den avhengige variabelen endres når den uavhengige variabelen endres. Eksperimentet kan også endres for å finne kausalitetens retning for ytterligere undersøkelser.

Independent variable

- En uavhengig variabel er en variabel i et eksperiment som kontrolleres og manipuleres av eksperimentøren, og variabelen er altså uavhengig fra andre variabler, som for eksempel deltakers atferd (avhengig variabel). Man kan ha flere uavhengige variabler, og hver av dem må ha minst to ulike grader av "effekt", ergo "2 timer søvn vs. 8 timer søvn".

Treatments

- Treatments er situasjonene assosiert med graden av den uavhengige variabelen. En treatment er altså i praksis hva den uavhengige variabelen har bestemt at den avhengige variabelen skal "utsettes for".

Dependent variable

- Den avhengige variabelen er en variabel som eksperimentøren ikke kan manipulere (annet enn eventuelt mer indirekte gjennom direkte manipulasjon av den uavhengige variabelen), for eksempel en deltakers atferd eller prestasjon.

Experimental group

- Den eksperimentelle gruppen er den gruppen som blir utsatt for treatment, sammenliknet med kontrollgruppen som ikke mottar treatment. Altså er den eksperimentelle gruppen de deltagere som opptrer som avhengig variabel (altså deres atferd eller hva enn det er som måles...) og blir utsatt for den uavhengige variabelen / variablene.

Control group

- Kontrollgruppen blir i motsetning til den eksperimentelle gruppen ikke utsatt for treatment, og gruppen eksisterer for å gi et tydeligere bilde av om det var de(n) uavhengige variabelen(e) (treatment) som førte til endring (hvis det var endring), eller om det skyldtes en eller flere tredjevariabler. I dette tilfellet ville man sett liknende endring i begge grupper, i motsetning til bare den som mottok treatment.

Extraneous variables

- Eksperimentell forskning innebærer kontroll over "extraneous" variabler, det vil si variabler som kan påvirke og ugyldiggjøre påvisningen av relasjon mellom uavhengig og avhengig variabel(er). Disse variablene kan typisk oppstå om det er vesentlige forskjeller mellom deltagere, eller at de behandles annerledes eller utsettes for ulike forhold generelt.

Random assignment

- Random assignment, eller randomisering er et statistisk grep som benyttes for å minimere effekten av "extraneous" variabler, da man håper at de individuelle forskjellene mellom

deltagere vil jevne seg ut, og dermed gjøre det mulig å vite om treatment faktisk hadde effekt eller ikke.

Demonstration

- En demonstrasjon er til forskjell fra et sant eksperiment en situasjon hvor kun én enkelt gruppe blir utsatt for én spesifikk treatment. Altså mangler det en uavhengig variabel, da ingen variabler blir manipulert over flere nivåer, og nytteverdien forsvinner i stor grad da det ikke kan påvises kausasjon. I praksis viser man bare at resultatet av en gitt treatment kombinert med en gitt gruppe gir *dette* resultatet, uten at man aner om treatmenten hadde effekt eller ikke.

Internal validity

- Intern validitet referer i et eksperimentelt design til at det utelukkende var den uavhengige variabelen som gav endring i den avhengige, og at eksperimentet altså tester hypotesen. I et korrelasjonsdesign beskriver begrepet om endringer i kriterievariabelen utelukkende skyldtes endringer i prediktorvariabelen, og ikke endringer i andre "extraneous" variabler. Om intern validitet ikke er tilstrekkelig, vil det oppstå såkalte rivalhypoteser, eller hypoteser som også kan forklare funn, sammenliknet med hypotesen man begynte forskningen med. Liste over trusler mot intern validitet s. 159

Confounding

- Konfundering oppstår når såkalte konfunderende variabler oppstår og åpner opp for muligheten til at observert effekt skyldes andre variabler enn de spesifisert i eksperimentets design. Konfundering kan være en hovedårsak til at en studie mangler intern validitet, i hvert fall om disse gir forklaringer som er like gode som hypotesen (rivalhypoteser), men konfundering er ikke alltid A) mulig å unngå, og B) en alvorlig trussel for intern validitet. Hvis identifiserte konfunderende variabler ikke kan gi den observerte effekten, eller heller fører til forminskning av effekten, kan den interne validiteten likevel være noenlunde akseptabel. Dette er heldig fordi det ofte er umulig å totalt fjerne alle (potensielt) konfunderende variabler. Konfunderende variabler forekommer i korrelasjonell og eksperimentell forskning, men er vanligvis et større problem i korrelasjonell forskning.

External validity

- Ekstern validitet referer til om hvorvidt et eksperiments funn kan generaliseres til omverdenen, eller om det kunstige lab-miljøet skapt i søken etter intern validitet, gjør at ekstern validitet synker. Enkelte mener at ekstern validitet alltid bør prioriteres, da forskning som kun viser hva som teoretisk sett "vil skje" under visse kunstige omstendigheter har begrenset nytteverdi. Det er altså en dikotomi mellom intern og ekstern validitet, og hva som bør prioriteres er som regel opp til forskerne å vurdere, spesielt med henblikk på forsøkets mål. Liste over trusler mot ekstern validitet s. 162, for eksempel reaktivitet, "multiple treatment interference" osv.

Simulation

- Simulering er en teknikk som brukes i laboratoriet for å gi funn bedre ekstern validitet, samtidig som man har kontroll over "extraneous" variabler. Simulering kan gjøres med VR-teknologi, eller fysisk simulering av omverdenen inne i labben, og disse er ofte nyttige i situasjoner hvor etiske problemstillinger hindrer feltstudier osv. Praktiske problemstillinger gjør også simulasjon ofte attraktivt, da simulering kan være billigere, mindre tidskrevende og enklere å få til enn feltstudier, samtidig som man opprettholder kontroll over variabler. Det er som regel i simulasjonens beste interesse at situasjonen eller fenomenet som simuleres fremstår og generelt er mest mulig virkelighetsnært og realistisk. Derfor krever simulasjon nøye undersøkelser av fenomenet / situasjonen som man ønsker å simulere.

Kapittel 6

Reliability

- Reliabilitet er i hvilken grad resultater funnet kan repliseres, det vil si hvilken grad av variabilitet mellom resultater som oppstår når replikasjoner gjøres. Det finnes flere typer reliabilitet, for eksempel "reliability of measure", altså en målings reliabilitet vurdert ut ifra variabiliteten mellom målinger når flere gjøres. Dette vises ofte med en +/- x-verdi av målinger, dvs hvor mye under eller over en måling andre målinger kan vise. "Reliability of population estimates" er et mål på hvor tilsvarende en populasjon et sample vil være, og vises med en vurdering av "margin of error" (feilmargin), for eksempel +/- 5% (eksempel i boka s. 175).

Test-retest reliability

- Målinger av psykologiske konstrukter som intelligens eller angst kan ofte fremstå som ureliable da konstruktene som måles har en tendens til å endre seg over tid. Likevel benyttes ofte en test-retest-strategi for å finne en tests reliabilitet, i hvert fall ifm de mer stabile konstruktene som intelligens. Konseptet er at man administrerer en test som gir en måling, og senere administreres den nøyaktig samme testen på nytt. Et høyere samsvar mellom målingene indikerer høyere reliabilitet og motsatt. Ulempen her er at reliabiliteten kan fremstå kunstig høy om deltagere husker forrige test og svarene som da ble gitt, og i tillegg kan deltageren som testes endre seg fra første til andre test gjennomføres, og gi annerledes svar. Dette er på mange måter et tegn på bedre reliabilitet, men testens reliabilitet vil fremstå lavere.

Parallel-forms reliability

- Parallel-forms reliability er en strategi for å øke reliabilitet ved at to relativt forskjellige tester som skal måle det samme konstruktet gis, i stedet for to identiske. På denne måten skal man i teorien unngå et problem med test-retest-strategien, at deltagere husker hva som ble svart først. Problemene med dette er at de to parallelle testene kan variere i grad av parallellitet, som vil påvirke reliabiliteten negativt om dette er vesentlig. I tillegg gjenstår problemstillingen med at konstruktets verdier kan være forskjellige fra test til test, som får testene til å fremstå mindre reliable enn de egentlig er.

Split-half reliability

- Her administreres de to parallelle testene samtidig (på gruppenivå, dvs at test 1 og test 2 fordeles jevnt ulikt, og følges på med av den testen som ikke er tatt). På denne måten har ikke deltageren rukket å endre seg særlig mye (ingenting), noe som eliminerer denne problemstillingen, men spørsmålet om de to testene faktisk er parallelle gjenstår.

Accuracy

- Nøyaktighet referer til i hvilken grad en måling vil samsvare med det den måler, noe som kan begrenses av manglende presisjon av måling. Ofte indikeres nøyaktighet med en +/- (enhet) av det den skal måle. En baderomsvekt skal måle vekten til en kalibrert 25 kg vektskive, og den viser (etter gjentakende målinger) et avvik på 1 kg. Vekten har en nøyaktighet på +/- 1 kg / 4%. Ofte ønsker man å finne et måleinstruments gjennomsnittlige nøyaktighet, og om målingen i snitt er høyere eller lavere enn det som er riktig, har måleinstrumentet en "bias". Problemstillingen innenfor målingen av psykologiske konstrukter er her at det ikke finnes en standard for konstruktet, og man måler derfor heller ut ifra et gjennomsnitt fra befolkningen. Snittet for befolkningen på testen er 15, du scorer 5, og du har færre tegn på dette konstruktet enn snittpersonen.

Face validity

- Face validity er om hvorvidt en test overfladisk fremstår som å måle det man ønsker å måle. Dessverre er dette et svakt mål for validitet, da tester som ikke har særlig høy grad av face validity, kan måle det man ønsker å måle helt fint. Samtidig kan face validity spille en rolle for subjekter som testes, da det kan ha betydning om subjekter oppfatter å bli testet for det de skal testes for, da de kan oppleve testen som unyttig under testing.

Content validity

- Dette begrepet beskriver om hvorvidt en tests innhold samsvarer med det som skulle testes. Om en test på skolen reflekterer hva som ble undervist om og står i pensum, har testen innholdsvaliditet.

Criterion-related validity

- Kriterievaliditet beskriver hvor godt en testscore reflekterer et individs verdi / prestasjon i forhold til et spesifikt kriterie. Kriterievaliditet har to subtyper, concurrent validity og predicitive validity.

Concurrent validity

- "Concurrent validity" (samtidig validitet?) reflekterer hvor godt en tests resultater vil samsvare med en annen tests resultater om de to testene administreres samtidig og skal måle det samme. Slik kan tester måles ut ifra andre tester, for eksempel nye tester vs anerkjente tester.

Predictive validity

- Prediktiv validitet er et begrep som brukes for å beskrive om hvorvidt en test og dens resultater kan brukes til å predikere fremtidige resultater i ett spesifikt kriterie. Eksempel: SAT-scorer predikerer prestasjon i universitetet senere.

Construct validity

- Begrepet brukes ifm målinger av psykologiske konstrukter, og beskriver om hvorvidt en tests resultater vil samsvare med atferden man forbinder med konstruktet. Eksempel: Du scorer lavt på en intelligenstest, og om testen har god konstruktvaliditet er det forventet at din atferd vil være indikativ for lav intelligens.

Nominal scale

- En nominell skala er en skala som ikke måler noe kvantitativt, men i stedet eksisterer som en skala for ulike kvalitative fenomener. Altså er en nominell skala en skala for variabler som ikke er forskjellige i antall, men type. Typisk kan en nominell skala brukes for måling av deltageres politiske ståsted, uten at de ulike ståstedene rangeres som høyere eller lavere. Dette gjelder selv om valgmulighetene kan representeres med tall. Her gir det altså ingen mening å prøve å regne ut noe som helst matematisk ut fra disse arbitrære tilknytningene mellom tall og valgmulighet.

Ordinal scale

- Ordinale skalaer benyttes når ulike grader av en variabel kan plasseres langs en skala for eksempel fra lav, moderat, høy. Selv om skalaen per definisjon er kvantitativ er det ingen kjent konstant matematisk avstand mellom de ulike gradene skalaen angir, noe som gjør det totalt meningsløst å prøve å bruke matematiske operasjoner til å ytterlig forenkle svar på slike tester. Om lav = 1, moderat = 2 og høy = 3, sier det oss lite om befolkningens svar i snitt var 2,5.

Interval scale

- En intervallskala er en skala hvor nullpunktet ikke indikerer totalt fravær av faktoren skalaen indikerer grad eller forekomst av. En typisk intervallskala kan være temperaturskalaen celsius. Her er null kun plassert av praktiske årsaker (frysepunkt vann, normalt trykk), og indikerer ikke et totalt fravær av termisk energi. Dette betyr også at det er absurd å hevde at det er "dobbel så varmt" når det er 30°C, som når det er 15°C. Et eksempel på en intervallskala innenfor psykologi kan være IQ-scorer; det er absurd å hevde at en med iq på 140 er "dobbel så smart" som en med iq på 70, og det er slett ikke gitt at en iq lik 0 indikerer et totalt fravær av intelligens.

Ratio scale

- En "ratio scale" er en skala hvor nullpunktet indikerer totalt fravær av variabelen som måles. Kelvin-skalaen er et eksempel på en "ratio scale", da 0 kelvin indikerer totalt fravær av termisk energi. Man kan altså (teoretisk sett, jeg er ikke fysiker) hevde at det er dobbelt så varmt når det er 270 kelvin som når det er 135. Et eksempel på ratio scales i psykologiske eksperimenter kan være antall items husket i en hukommelsestest, der 0 betyr at null items ble husket, mens 30 betyr at 30 items ble husket (for eksempel).

Range effects

- "Rekkevidde-effekter" finnes i to typer, tak- og gulv-effekter, og disse gjør seg gjeldende om målinger i praksis er for "lette" eller "vanskelige", da man ikke vil klare å detektere noen forskjeller mellom deltagerne om alle scorer toppscore, eller bunnscore, som gitt av for lette vs for vanskelige målinger. Altså oppstår gulveffekter når en variabel har sin laveste mulige verdi (du vet ikke om verdien "kunne vært lavere"), eller høyest mulige verdi (verdien kunne ha vært høyere). Range effects er skumle da disse kan resultere i at man må redesigne hele eksperimentet om de viser seg å være svært påvirkende.

Behavioral measure

- Behavioral measures er avhengige mål (mål på avhengige variabler) som benyttes til måling av avhengige, atferdsrelaterte variabler. Mål på avhengige variabler som benyttes er frekvens (frekvensen av observert atferd i et spesifisert tidsrom), "latency" (forsinkelse), som kan brukes som mål på hvor lang tid det tok fra treatment ble innført, til deltager reagerte atferdsmessig. Et tredje avhengig mål på atferd kan være "number of errors" (antall feil), en deltager gjør i en situasjon hvor det finnes definerte korrekt atferd, altså atferd som avviker fra dette. Dessverre kan ikke "behavioral measures" måle de underliggende årsakene som ligger bak atferden, men fungerer heller til å kartlegge selve atferden.

Physiological measure

- Fysiologiske mål brukes i måling av fysiologiske faktorer som hjerterate, søvnstadiet eller hjerneaktivitet gjennom for eksempel EEG, PET eller fMRI. I psykologi kan fysiologiske mål være nyttige i undersøkelse av fysiologiske forhold som kan ligge i grunn for eller indikere psykologiske tilstander, fenomener eller for eksempel læring. Her må det trekkes inferenser fra fysiologiske mål til psykologiske variabler, og her ligger det en risiko for at det trekkes feil inferenser. Verktøy for måling av avhengige fysiologiske variabler hos mennesker er noninvasive, men det finnes invasive verktøy, for eksempel implantater av måleinstrumenter osv. Dette gjøres hovedsakelig ifb med studier på dyr, eller mer spesifikt (som regel) rotter eller mus.

Self-report measure

- Et tredje mål på avhengige variabler knyttet til atferd innebærer selvrapportering, dvs at deltagerne selv bes å vurdere variabler knyttet til deres egne emosjoner, tanker eller atferd. Dette gjøres ofte med Likert skalaer, altså at det legges fram en påstand som deltager vurderer om de er enige eller uenige med på en skala fra 1-5 (sterkt uenig til sterkt enig).

Eksempel: "Jeg blir ofte lei meg om kveldene", om deltager opplever at dette dekker dem svært godt, kan de svare 5, om ikke, kan de svare 1, eller noe i mellom. Self-report measures kan ha nedsider som at validiteten synker om det finnes motivasjon for deltagere til å svare uærlig eller at de husker feil i lys av leste selvrapporingsmål.

Q-sort methodology

- Q-sort metodologi innebærer at man etablerer "målekategorier" (for eksempel "liker veldig godt" og "liker svært lite"), og at forskjellige items plasseres inne i disse målekategoriene. Etter at et sett items har blitt plassert i forskjellige målekategorier kan resultatet analyseres og undersøkes for mønstre, for eksempel om deltagere som liker Ferrari svært godt, ofte ikke er glad i Mini. Analysen kan gjøres med spesielle Q-sort korrelasjonelle- og faktoranalyse-teknikker. Data produsert med Q-sort metodologi kan også settes inn i standard analyse av varians for å undersøke main effects og interaksjoner mellom variabler.

Implicit measure

- Implisitte mål er populære innen sosialpsykologi for å unnsnippe problemet med selvrapporering hvor deltagere (sannsynligvis) unnlater / er motiverte til å unnlate å svare på politisk ukorrekte (sosialt avvikende) måter. Implisitte mål er ikke under direkte bevisst kontroll, og brukes der man mistenker at deltagere ikke ville besvart riktig (på grunn av sosialt press, eller liknende), men i stedet måles emosjonelle reaksjoner for eksempel til bilder av medlemmer av andre etniske grupper osv.

Implicit association test (IAT)

- IAT (implicit association test), brukes for å for eksempel måle "fordommer" (prejudice). Her måles deltageres fordommer gjennom analyse av deres svar på assosiasjonsoppgaver hvor en liste med ord kan knyttes raskt til for eksempel ulike etniske grupper. Her skjer assosiasjonen såpass fort at deltagere (forhåpentligvis) ikke klarer å la være å "svare ærlig".

Demand characteristics

- Demand characteristics er "cues" som deltagere av eksperimenter får ved et uhell fra forskeren eller forsøkets miljø, som har med forsøkets formål å gjøre eller deltagerens forventede atferd. Det finnes ulike demand characteristics, og dessverre hender det at deltagere blir villedet av disse, slik at de agerer på visse måter, sammenliknet med hvordan de "ville ha agert", som heller ikke samsvarer med forsøkets faktiske formål. Dette kalles performance cues (Adair).

Role attitude cues

- Role attitude cues er signaler som deltagere kan hente ut ifra et eksperiment som signaliserer at det er forventet at deltageren skal endre holdning (attitude) for å tilpasse seg eksperimentet. Allerede eksisterende "attitudes" / holdninger til eksperimentet kan finnes hos deltagere, deriblant den kooperative, defensive og negative holdningen. Deltagere med kooperativ holdning har blitt vist å kunne fås til å gjøre repetitive, tullete og kjedelige oppgaver i timevis, kun fra et behov for å blidgjøre forskeren. Deltagere med defensive / bekymrede holdninger deltar i eksperimentet allerede bekymret for hvordan deltagelse vil forvolde dem noen form for skade. En slik bekymret holdning har blitt vist å kunne endre et eksperiments utfall, spesielt ift "compliance and attitude change" (overholdelse av regler og holdningsendringer). Negative holdninger til eksperimentet hos deltagere er mindre vanlig i dag da det ikke er lov å tvinge folk til å delta, men før var dette mer vanlig. Da deltok ofte deltagere med interesse i å ødelegge forskningen, på grunn av aggresjon rettet mot eksperimentet.

Experimenter bias

- Experimenter bias oppstår når en eksperimentørs atferd påvirker resultatene oppnådd under eksperimentering, og som vist i situasjonen med "clever Hans" kan disse oppstå helt utilsiktet. Experimenter bias fungerer som en konfunderende variabel, da man umulig (om ikke tiltak settes inn) kan vite om effektene observert skyldes treatment eller forskerens utilsiktede påvirkning.

Expectancy effects

- Forventningseffekter oppstår når forskeres forventninger påvirker deltageres atferd under eksperimentering, og er en kilde til konfundering som både truer intern og ekstern validitet. Forventningseffekter er effektene av experimenter bias, og kan oppstå på subtile måter, som at forventninger til positiv resultater i en gruppe kan gjøre at administratoren av uavhengige variabler administrerer variabelen på en overdrevent bra måte (sammenliknet med andre grupper der dårlige resultater er forventet). I tillegg kan forventningseffekter oppstå når det ligger i forskeres beste resultater at deres hypotese styrkes, slik at forskere uten at det er med vilje kommuniserer hva de forventer av atferd til deltagere.

Single-blind technique

- Single-blind teknikken bygger på at forskere ikke vet hvilke grupper deltagere har blitt delt inn i, noe som senker sannsynligheten for at experimenter bias spiller inn. Single-blind teknikken er en "blind technique".

Double-blind technique

- Double blind technique innebærer at verken deltager eller forsker vet hvilken treatment deltagere har blitt utsatt for. På denne måten minimerer man både experimenter bias og deltageres forventninger til effekt (som kan skape placebo).

Pilot study

- En pilotstudie er en variant av en studie i mindre skala, ment å teste / etablere prosedyrer, materiell (utstyr) og parametere som skal brukes i fullskala studier. Ofte er pilotstudier eksempler på studier hvor "noe gikk galt", slik at studien måtte endres eller forkastes, mens andre ganger er pilotstudier designet fra bunnen av for sitt formål som pilotstudie. Pilotstudier kan være nyttige for å forbedre og klargjøre forskningsdesign før de gjøres i fullskala, slik at eventuelle problemer kan detekteres før mer ressurser iverksettes. En ulempe er at pilotstudier også krever ressurser, ift penger, tid og innblanding fra for eksempel IACUCer (om dyr involveres). Om en pilotstudie ikke gjøres, og en fullskalastudie heller iverksettes, men viser seg å ha problemer, kan denne virke som en pilotstudie før man prøver på nytt.

Manipulation check

- En manipulation check gjøres for å sikre at subjektive faktorer, for eksempel opplevelse av tiltrekning til en person, kan sies å samsvare med en forskers opplevelse. Om du vil teste om forfatteres attraktivitet påvirker hvordan tekstene deres blir bedømt, kan det altså være greit å vite om deltagerne er enige i at de attraktive og mindre attraktive er attraktive eller mindre tiltrekkende.

Kapittel 7

Reliability

- Reliabilitet er i hvilken grad resultater funnet kan repliseres, det vil si hvilken grad av variabilitet mellom resultater som oppstår når replikasjoner gjøres. Det finnes flere typer reliabilitet, for eksempel "reliability of measure", altså en målings reliabilitet vurdert ut ifra

variabiliteten mellom målinger når flere gjøres. Dette vises ofte med en +/- x-verdi av målinger, dvs hvor mye under eller over en måling andre målinger kan vise. "Reliability of population estimates" er et mål på hvor tilsvarende en populasjon et sample vil være, og vises med en vurdering av "margin of error" (feilmargin), for eksempel +/- 5% (eksempel i boka s. 175).

Test-retest reliability

- Målinger av psykologiske konstrukter som intelligens eller angst kan ofte fremstå som ureliable da konstruktene som måles har en tendens til å endre seg over tid. Likevel benyttes ofte en test-retest-strategi for å finne en tests reliabilitet, i hvert fall ifm de mer stabile konstruktene som intelligens. Konseptet er at man administrerer en test som gir en måling, og senere administreres den nøyaktig samme testen på nytt. Et høyere samsvar mellom målingene indikerer høyere reliabilitet og motsatt. Ulempen her er at reliabiliteten kan fremstå kunstig høy om deltagere husker forrige test og svarene som da ble gitt, og i tillegg kan deltageren som testes endre seg fra første til andre test gjennomføres, og gi annerledes svar. Dette er på mange måter et tegn på bedre reliabilitet, men testens reliabilitet vil fremstå lavere.

Parallel-forms reliability

- Parallel-forms reliability er en strategi for å øke reliabilitet ved at to relativt forskjellige tester som skal måle det samme konstruktet gis, i stedet for to identiske. På denne måten skal man i teorien unngå et problem med test-retest-strategien, at deltagere husker hva som ble svart først. Problemene med dette er at de to parallelle testene kan variere i grad av parallellitet, som vil påvirke reliabiliteten negativt om dette er vesentlig. I tillegg gjenstår problemstillingen med at konstruktets verdier kan være forskjellige fra test til test, som får testene til å fremstå mindre reliable enn de egentlig er.

Split-half reliability

- Her administreres de to parallelle testene samtidig (på gruppenivå, dvs at test 1 og test 2 fordeles jevnt ulikt, og følges på med av den testen som ikke er tatt). På denne måten har ikke deltageren rullet å endre seg særlig mye (ingenting), noe som eliminerer denne problemstillingen, men spørsmålet om de to testene faktisk er parallelle gjenstår.

Accuracy

- Nøyaktighet referer til i hvilken grad en måling vil samsvare med det den måler, noe som kan begrenses av manglende presisjon av måling. Ofte indikeres nøyaktighet med en +/- (enhet) av det den skal måle. En baderomsvekt skal måle vekten til en kalibrert 25 kg vektskive, og den viser (etter gjentakende målinger) et avvik på 1 kg. Vekten har en nøyaktighet på +/- 1 kg / 4%. Ofte ønsker man å finne et måleinstruments gjennomsnittlige nøyaktighet, og om målingen i snitt er høyere eller lavere enn det som er riktig, har måleinstrumentet en "bias". Problemstillingen innenfor målingen av psykologiske konstrukter er her at det ikke finnes en standard for konstruktet, og man måler derfor heller ut ifra et gjennomsnitt fra befolkningen. Snittet for befolkningen på testen er 15, du scorer 5, og du har færre tegn på dette konstruktet enn snittpersonen.

Face validity

- Face validity er om hvorvidt en test overfladisk fremstår som å måle det man ønsker å måle. Dessverre er dette et svakt mål for validitet, da tester som ikke har særlig høy grad av face validity, kan måle det man ønsker å måle helt fint. Samtidig kan face validity spille en rolle for subjekter som testes, da det kan ha betydning om subjekter oppfatter å bli testet for det de skal testes for, da de kan oppleve testen som unyttig under testing.

Content validity

- Dette begrepet beskriver om hvorvidt en tests innhold samsvarer med det som skulle testes. Om en test på skolen reflekterer hva som ble undervist om og står i pensum, har testen innholdsvaliditet.

Criterion-related validity

- Kriterievaliditet beskriver hvor godt en testscore reflekterer et individs verdi / prestasjon i forhold til et spesifikt kriterie. Kriterievaliditet har to subtyper, concurrent validity og predictive validity.

Concurrent validity

- "Concurrent validity" (samtidig validitet?) reflekterer hvor godt en tests resultater vil samsvare med en annen tests resultater om de to testene administreres samtidig og skal måle det samme. Slik kan tester måles ut ifra andre tester, for eksempel nye tester vs anerkjente tester.

Predictive validity

- Prediktiv validitet er et begrep som brukes for å beskrive om hvorvidt en test og dens resultater kan brukes til å predikere fremtidige resultater i ett spesifikt kriterie. Eksempel: SAT-scorer predikerer prestasjon i universitetet senere.

Construct validity

- Begrepet brukes ifm målinger av psykologiske konstrukter, og beskriver om hvorvidt en tests resultater vil samsvare med atferden man forbinder med konstruktet. Eksempel: Du scorer lavt på en intelligens test, og om testen har god konstruktvaliditet er det forventet at din atferd vil være indikativ for lav intelligens.

Nominal scale

- En nominell skala er en skala som ikke måler noe kvantitativt, men i stedet eksisterer som en skala for ulike kvalitative fenomener. Altså er en nominell skala en skala for variabler som ikke er forskjellige i antall, men type. Typisk kan en nominell skala brukes for måling av deltageres politiske ståsted, uten at de ulike ståstedene rangeres som høyere eller lavere. Dette gjelder selv om valgmulighetene kan representeres med tall. Her gir det altså ingen mening å prøve å regne ut noe som helst matematisk ut fra disse arbitrære tilknytningene mellom tall og valgmulighet.

Ordinal scale

- Ordinale skalaer benyttes når ulike grader av en variabel kan plasseres langs en skala for eksempel fra lav, moderat, høy. Selv om skalaen per definisjon er kvantitativ er det ingen kjent konstant matematisk avstand mellom de ulike gradene skalaen angir, noe som gjør det totalt meningsløst å prøve å bruke matematiske operasjoner til å ytterlig forenkle svar på slike tester. Om lav = 1, moderat = 2 og høy = 3, sier det oss lite om befolkningens svar i snitt var 2,5.

Interval scale

- En intervallskala er en skala hvor nullpunktet ikke indikerer totalt fravær av faktoren skalaen indikerer grad eller forekomst av. En typisk intervallskala kan være temperaturskalen celsius. Her er null kun plassert av praktiske årsaker (frysepunkt vann, normalt trykk), og indikerer ikke et totalt fravær av termisk energi. Dette betyr også at det er absurd å hevde at det er "dobbeltså varmt" når det er 30°C, som når det er 15°C. Et eksempel på en intervallskala innenfor psykologi kan være IQ-scorer; det er absurd å hevde at en med iq på 140 er

"dobbel så smart" som en med iq på 70, og det er slett ikke gitt at en iq lik 0 indikerer et totalt fravær av intelligens.

Ratio scale

- En "ratio scale" er en skala hvor nullpunktet indikerer totalt fravær av variabelen som måles. Kelvin-skalaen er et eksempel på en "ratio scale", da 0 kelvin indikerer totalt fravær av termisk energi. Man kan altså (teoretisk sett, jeg er ikke fysiker) hevde at det er dobbelt så varmt når det er 270 kelvin som når det er 135. Et eksempel på ratio scales i psykologiske eksperimenter kan være antall items husket i en hukommelsestest, der 0 betyr at null items ble husket, mens 30 betyr at 30 items ble husket (for eksempel).

Range effects

- "Rekkevidde-effekter" finnes i to typer, tak- og gulv-effekter, og disse gjør seg gjeldende om målinger i praksis er for "lette" eller "vanskelige", da man ikke vil klare å detektere noen forskjeller mellom deltagere om alle scorer toppscore, eller bunnscore, som gitt av for lette vs for vanskelige målinger. Altså oppstår gulveffekter når en variabel har sin laveste mulige verdi (du vet ikke om verdien "kunne vært lavere"), eller høyest mulige verdi (verdien kunne ha vært høyere). Range effects er skumle da disse kan resultere i at man må redesigne hele eksperiment om de viser seg å være svært påvirkende.

Behavioral measure

- Behavioral measures er avhengige mål (mål på avhengige variabler) som benyttes til måling av avhengige, atferdsrelaterte variabler. Mål på avhengige variabler som benyttes er frekvens (frekvensen av observert atferd i et spesifisert tidsrom), "latency" (forsinkelse), som kan brukes som mål på hvor lang tid det tok fra treatment ble innført, til deltager reagerte atferdsmessig. Et tredje avhengig mål på atferd kan være "number of errors" (antall feil), en deltager gjør i en situasjon hvor det finnes definerte korrekt atferd, altså atferd som avviker fra dette. Dessverre kan ikke "behavioral measures" måle de underliggende årsakene som ligger bak atferden, men fungerer heller til å kartlegge selve atferden.

Physiological measure

- Fysiologiske mål brukes i måling av fysiologiske faktorer som hjerterate, søvnstadiet eller hjerneaktivitet gjennom for eksempel EEG, PET eller fMRI. I psykologi kan fysiologiske mål være nyttige i undersøkelse av fysiologiske forhold som kan ligge i grunn for eller indikere psykologiske tilstander, fenomener eller for eksempel læring. Her må det trekkes inferenser fra fysiologiske mål til psykologiske variabler, og her ligger det en risiko for at det trekkes feil inferenser. Verktøy for måling av avhengige fysiologiske variabler hos mennesker er noninvasive, men det finnes invasive verktøy, for eksempel implantater av måleinstrumenter osv. Dette gjøres hovedsakelig ifb med studier på dyr, eller mer spesifikt (som regel) rotter eller mus.

Self-report measure

- Et tredje mål på avhengige variabler knyttet til atferd innebærer selvrapporing, dvs at deltagere selv bes å vurdere variabler knyttet til deres egne emosjoner, tanker eller atferd. Dette gjøres ofte med Likert skalaer, altså at det legges fram en påstand som deltager vurderer om de er enige eller uenige med på en skala fra 1-5 (sterkt uenig til sterkt enig). Eksempel: "Jeg blir ofte lei meg om kveldene", om deltager opplever at dette dekker dem svært godt, kan de svare 5, om ikke, kan de svare 1, eller noe i mellom. Self-report measures kan ha nedsider som at validiteten synker om det finnes motivasjon for deltagere til å svare uærlig eller at de husker feil i lys av leste selvrapporingsmål.

Q-sort methodology

- Q-sort metodologi innebærer at man etablerer "målekategorier" (for eksempel "liker veldig godt" og "liker svært lite"), og at forskjellige items plasseres inne i disse målekategoriene. Etter at et sett items har blitt plassert i forskjellige målekategorier kan resultatet analyseres og undersøkes for mønstre, for eksempel om deltagere som liker Ferrari svært godt, ofte ikke er glad i Mini. Analysen kan gjøres med spesielle Q-sort korrelasjonelle- og faktoranalyse-teknikker. Data produsert med Q-sort metodologi kan også settes inn i standard analyse av varians for å undersøke main effects og interaksjoner mellom variabler.

Implicit measure

- Implisitte mål er populære innen sosialpsykologi for å unnsnippe problemet med selvrapporing hvor deltagere (sannsynligvis) unnlater / er motiverte til å unnlate å svare på politisk ukorrekte (sosialt avvikende) måter. Implisitte mål er ikke under direkte bevisst kontroll, og brukes der man mistenker at deltagere ikke ville besvart riktig (på grunn av sosialt press, eller liknende), men i stedet måles emosjonelle reaksjoner for eksempel til bilder av medlemmer av andre etniske grupper osv.

Implicit association test (IAT)

- IAT (implicit association test), brukes for å for eksempel måle "fordommer" (prejudice). Her måles deltageres fordommer gjennom analyse av deres svar på assosiasjonsoppgaver hvor en liste med ord kan knyttes raskt til for eksempel ulike etniske grupper. Her skjer assosiasjonen såpass fort at deltagere (forhåpentligvis) ikke klarer å la være å "svare ærlig".

Demand characteristics

- Demand characteristics er "cues" som deltagere av eksperimenter får ved et uhell fra forskeren eller forsøkets miljø, som har med forsøkets formål å gjøre eller deltagerens forventede atferd. Det finnes ulike demand characteristics, og dessverre hender det at deltagere blir villedet av disse, slik at de agerer på visse måter, sammenliknet med hvordan de "ville ha agert", som heller ikke samsvarer med forsøkets faktiske formål. Dette kalles performance cues (Adair).

Role attitude cues

- Role attitude cues er signaler som deltagere kan hente ut ifra et eksperiment som signaliserer at det er forventet at deltageren skal endre holdning (attitude) for å tilpasse seg eksperimentet. Allerede eksisterende "attitudes" / holdninger til eksperimentet kan finnes hos deltagere, deriblant den kooperative, defensive og negative holdningen. Deltagere med kooperativ holdning har blitt vist å kunne fås til å gjøre repetitive, tullete og kjedelige oppgaver i timevis, kun fra et behov for å blidgjøre forskeren. Deltagere med defensive / bekymrede holdninger deltar i eksperimentet allerede bekymret for hvordan deltagelse vil forvolde dem noen form for skade. En slik bekymret holdning har blitt vist å kunne endre et eksperiments utfall, spesielt ift "compliance and attitude change" (overholdelse av regler og holdningsendringer). Negative holdninger til eksperimentet hos deltagere er mindre vanlig i dag da det ikke er lov å tvinge folk til å delta, men før var dette mer vanlig. Da deltok ofte deltagere med interesse i å ødelegge forskningen, på grunn av aggresjon rettet mot eksperimentet.

Experimenter bias

- Experimenter bias oppstår når en eksperimentørs atferd påvirker resultatene oppnådd under eksperimentering, og som vist i situasjonen med "clever Hans" kan disse oppstå helt utilsiktet. Experimenter bias fungerer som en konfunderende variabel, da man umulig (om ikke tiltak settes inn) kan vite om effektene observert skyldes treatment eller forskerens utilsiktede påvirkning.

Expectancy effects

- Forventningseffekter oppstår når forskeres forventninger påvirker deltageres atferd under eksperimentering, og er en kilde til konfundering som både truer intern og ekstern validitet. Forventningseffekter er effektene av eksperimenter bias, og kan oppstå på subtile måter, som at forventninger til positiv resultater i en gruppe kan gjøre at administratoren av uavhengige variabler administrerer variabelen på en overdrevent bra måte (sammenliknet med andre grupper der dårlige resultater er forventet). I tillegg kan forventningseffekter oppstå når det ligger i forskeres beste resultater at deres hypotese styrkes, slik at forskere uten at det er med vilje kommuniserer hva de forventer av atferd til deltagere.

Single-blind technique

- Single-blind teknikken bygger på at forskere ikke vet hvilke grupper deltagere har blitt delt inn i, noe som senker sannsynligheten for at eksperimenter bias spiller inn. Single-blind teknikken er en "blind technique".

Double-blind technique

- Double blind technique innebærer at verken deltager eller forsker vet hvilken treatment deltagere har blitt utsatt for. På denne måten minimerer man både eksperimenter bias og deltageres forventninger til effekt (som kan skape placebo).

Pilot study

- En pilotstudie er en variant av en studie i mindre skala, ment å teste / etablere prosedyrer, materiell (utstyr) og parametere som skal brukes i fullskala studier. Ofte er pilotstudier eksempler på studier hvor "noe gikk galt", slik at studien måtte endres eller forkastes, mens andre ganger er pilotstudier designet fra bunnen av for sitt formål som pilotstudie. Pilotstudier kan være nyttige for å forbedre og klargjøre forskningsdesign før de gjøres i fullskala, slik at eventuelle problemer kan detekteres før mer ressurser iverksettes. En ulempe er at pilotstudier også krever ressurser, ift penger, tid og innblanding fra for eksempel IACUCer (om dyr involveres). Om en pilotstudie ikke gjøres, og en fullskalastudie heller iverksettes, men viser seg å ha problemer, kan denne virke som en pilotstudie før man prøver på nytt.

Manipulation check

- En manipulation check gjøres for å sikre at subjektive faktorer, for eksempel opplevelse av tiltrekning til en person, kan sies å samsvare med en forskers opplevelse. Om du vil teste om forfatteres attraktivitet påvirker hvordan tekstene deres blir bedømt, kan det altså være greit å vite om deltagerne er enige i at de attraktive og mindre attraktive er attraktive eller mindre tiltrekkende.

Kapittel 9

Open-Ended items

- Open-ended items er en type spørsmål som kan stilles i spørreundersøkelser, altså spørsmål der deltagere kan svare med egne ord, sammenliknet med å kun huke av en boks eller liknende. Fordelen her er at deltagere kan komme med svar som forskeren ikke hadde regnet med å få, men formatet gjør det samtidig vanskelig å oppsummere dataen som har blitt samlet inn, på grunn av større volum av tekst osv. Unngå fristelsen til å kun tolke resultater, sammenliknet med oppsummering.

Restricted-items (closed-ended items)

- Restricted items-spørreundersøkelser baserer seg på at svaralternativene allerede er formulert, og kun innebærer avhuking fra respondenter. Dette muliggjør enklere analyse og oppsummering av resultater enn open-ended items-spørreundersøkelser, på bekostning av mengden informasjon som kan hentes. I tillegg løper man risikoen at man unnlater å nevne et svaralternativ i undersøkelsen som burde være tilstede. Dillmann et al. (2014) nevner fire viktige ting å være obs på ved utforming av slike spørreundersøkelser (s.297).

Partially open-ended items

- Delvis open-ended items-spørreundersøkelser bygger på at det er preformulerte svaralternativ som respondent kun behøver å krysse av, som i restricted-items-spørreundersøkelser. Forskjellen er at det legges til en "annen"-kategori (other), hvor respondenter kan svare med egne ord. Dette minimerer sannsynligheten for at det mangler dekkende svaralternativ, samtidig som at en slik spørreundersøkelse stort sett beholder fordelene til en restricted-items-spørreundersøkelse.

Rating scale

- En variasjon av restricted-items-formatet benytter såkalte "rating scales" ("rangeringsskalaer"?), hvor svaralternativene i praksis er plassert langs en skala, fra et punkt til et annet, ideelt sett med mellom 7 og 10 ankerpunkter som respondenten kan krysse av på. Det er viktig å bruke korrekte eller passende begreper på skalane (se s.300).

Mail survey

- En mail survey ("brevundersøkelse") er en spørreundersøkelse som distribueres via posten, da dette ofte er en praktisk og ukomplisert måte å distribuere undersøkelser på. Dessverre har metoden et problem kalt "nonresponse bias", altså at svært mange av mottakerne av undersøkelsen ikke svarer på den.

Nonresponse bias

- Nonresponse bias er kort fortalt at store deler av mottakere av brevundersøkelser tenderer til å ikke svare på disse undersøkelsene. Dette kan gjøre resultatene fra undersøkelsene mindre generaliserbare, spesielt om det er overordnede faktorer som hindrer disse mottakerne å svare. Det finnes måter å minske nonresponse bias, for eksempel ved å kontakte respondenter flere ganger på flere måter (fire er ideelt ifølge Dillmann). Det øker også responsraten å sende med 1 dollar i undersøkelsen.

Internet survey

- Internetspørreundersøkelser er blitt populære, da dette er en effektiv måte å nå store antall folk, med relativ enkelhet. Ulemper er at internettet har et skjevt utvalg av (den amerikanske) befolkningen representert, med minoritets- alder- og utdanningsstatus som viktige faktorer. Andre ulemper er at utvikling og vedlikehold av en nettside hvor undersøkelser kan huses er ressurskrevende, og dette er nødvendig med mer komplekse spørreundersøkelser (sammenliknet med epostundersøkelser).

Telephone survey

- En telefonundersøkelse er en spørreundersøkelse som distribueres over telefon. Deltagere i slike undersøkelser har endret seg vesentlig ifb med endring av folks telefontyper, da landlinetelefoner og deres utstrakte bruk gjorde mediet langt mer attraktivt, sammenliknet med smarttelefoner. Dette skyldes at smarttelefoner er vanskeligere å nå på grunn av inaktive nummere, "call-blocking" osv. Dette er problematisk fordi brukere av smarttelefoner ser ut til å være vesentlig forskjellige fra brukere av fasttelefoner, og om overdrevent mange respondenter benytter fasttelefoner kan dette påvirke generaliserbarheten til undersøkelsens funn negativt.

Face-to-face interview

- Face-to-face-intervjuet bygger på at respondenter svarer på undersøkelser med intervjueren fysisk tilstede. Intervjuet kan gjøres strukturert med bruk av faste skjemaer som følges og fylles ut, eller ustrukturerte intervjuer hvor intervjueren har en anelse om hva som skal spørres om, men at det ikke finnes samme struktur som i det strukturerte alternativet. Ustrukturerte intervjuer kan åpne for mer fleksibilitet (i forhold til svar) på bekostning av analyserbarhet og oppsummerbarhet, og fordelene og ulempene er motsatte for det strukturerte intervjuet. En ulempe med teknikken er at svar og deres kvalitet kan påvirkes av intervjuerens implisitte menneskelige dynamikk, sammenliknet med en nettside eller et brev. Tilstedeværelsen av et levende menneske som intervjuer deltagere kan også minske sannsynligheten for ærlige besvarelser på sjenerende spørsmål, for eksempel i forbindelse med sex eller liknende.

Mixed-mode survey

- Mixed-mode-surveys bygger i motsetning til "unimethod"-undersøkelser på at undersøkelser distribueres med flere enn én enkelt metode. For eksempel kan en gitt undersøkelse både distribueres via internett og post, og dette vil sannsynligvis øke responsraten. Dessverre kan negative mixed-mode effekter oppstå som beskrevet av Lugtig et al. (2011).

Probability sampling

- Probability sampling (sannsynlighetsutvalg?) er en kategori av samplingteknikker hvor man har en gitt, ideelt lik på tvers av individer, sannsynlighet for å trekke ut enkeltindivider fra en populasjon til et utvalg.

Representative sample

- Et representativt utvalg er et sample med høy grad av generaliserbarhet til utvalget det er ment å representere, det er altså et sample som representerer populasjonen det er trukket ut fra på en god måte. Dette er altså det motsatte av et "biased sample".

Biased sample

- Et biased sample er et sample som ikke representerer populasjonen som det er trukket ut fra eller ment å representere. Eksempelet i boka fra det amerikanske presidentvalget i 1936 er illustrativt (s.314). Kort fortalt var literary digests sample ikke representativt fordi det var ment å representere alle amerikanere med stemmerett, men i stedet kun representerte bil- og telefoneiere, en privilegert klasse som ikke delte den jevne amerikaners politiske holdninger eller partitilknytning. En måte å unngå biased sampling kan være å unngå biased kilder (bil- og telefoneiere i 1936), eller i det minste kombinere denne kilden med flere, mindre biased kilder, som "voter registration lists".

Simple random sampling

- Simple random sampling (randomisering) er en teknikk for å forbedre generaliserbarheten fra et sample til en populasjon, ved at utvalget man bruker har blitt utvalgt helt tilfeldig fra populasjonen. Dette kan gjøres med "random number generators", eller for eksempel tilfeldig "dialing" (telefonundersøkelser). Simple random sampling er ikke en garanti for representativitet, men sannsynligheten for at samplet er skjevt utvalgt synker med økende størrelser på utvalg. Dette vil også øke ressursbehovet for forskningen.

Stratified sampling

- Stratifisert randomisert sampling er en metode for økning av generaliserbarhet, hvor man deler populasjonen inn i ulike segmenter (for eksempel yrkesgrupper, etnisitet eller alder), og deretter velger tilfeldige deltagere fra hvert segment (strata).

Proportionate sampling

- Proporsjonal sampling er den mest populære metoden for sampling (Kish, 1965), og innebærer at samplingen gjøres proporsjonal med populasjonen forøvrig, altså avhenger størrelsen på strataene samlet på proporsjonen av befolkningen strataet representerer. Eksempel fra boka: Populasjonen består av 5000 mennesker, 3000 hvite, 1500 svarte og 500 latinamerikanske, og samplet på 500 mennesker har tilsvarende demografiske proporsjoner (sample med 1200 personer består av 720 hvite, 120 latinamerikanere og 360 svarte). Proporsjonal og stratifisert sampling kan gjøres etter at utvalget er hentet, ved at man tilfeldig henter ut det antallet av hvert strata man ønsker fra respondenter til undersøkelsen.

Systematic sampling

- Systematisk sampling brukes ofte i forbindelse med stratifisert sampling, og er en teknikk for randomisering. Her hentes hver "kth", for eksempel hver femte, fjerde, tredje osv person i en liste av personer, med tilfeldig start.

Cluster sampling

- Cluster sampling er en randomiseringsteknikk der man henter en tilfeldig sammenhengende masse av deltagere fra en større populasjon, eller flere større strata av en populasjon. Dette er tids- og ressursbesvarende, men kan begrense utvalgets representativitet, om clusterets deltagere likner mye på hverandre, men ikke resten av populasjonen eller strataet av populasjonen.

Multistage sampling

- Multistage sampling er en variant av cluster sampling hvor man først finner større cluster, som man deretter velger deltagere fra tilfeldig, og så henter man tilfeldige elementer fra denne gruppen, sammenliknet med at elementene hentes først. Metoden kan kombineres med stratifiseringsprosedyrer for å sikre et representativt utvalg.

Sampling error

- Sampling error er den graden utvalget er annerledes fra populasjonen det er ment å representere. Dette kan skyldes at utvalget er for lite (større utvalg gir lavere sampling error). Sampling error estimeres som en feilmargin, gitt med +/- x %, der x er avviket man forventer å kunne regne med mellom utvalg og populasjon. Man ønsker et mest mulig økonomisk sample, dvs et sample som er stort nok til å sikre høy reliabilitet, men ikke fler. Dette skyldes at økning i samplestørrelse og dets fordeler minsker jo mer utvalget økes. Det finnes derfor formler for utregning av nødvendig samplestørrelse med et gitt "confidence level" og ønsket feilmargin:

$$n' = (z * z)P' \frac{1 - P'}{ME * ME}$$

n' = nødvendig antall deltagere

Z = confidence level i prosent

P' = estimert proporsjon (om ikke kjent, et konservativt valg er 0,5)

ME = feilmargin

Kapittel 10

Between-subjects design

- Between-subjects design er én av tre typer eksperimentelle design, sammen med within- og single-subject design. I et between-subjects design administreres hver treatment til forskjellige grupper. For eksempel treatment til gruppe 1, treatment 2 til gruppe 2, og gruppe 3 er kontrollgruppe.

Within-subjects design

- I et within-subjects design administreres hver treatment til samme gruppe på forskjellige tidspunkt. Her benyttes gjennomsnitt av de ulike deltagerne i gruppens atferd ifb med måling av effekt osv.

Single-subject design

- I single subject design administreres hver treatment til hver deltager (den ene i hver gruppe) over tid, hvor hver enkelt deltager fungerer som sin egen kontrollgruppe (i likhet med within-subjects design, men på individnivå). Tanken er at det er lettere å detekttere (også manglende) effekt av treatment ved å undersøke hvert enkeltindivids respons.

Error variance

- Error variance utydeliggjør graden av effekt treatment har, da det finnes forskjeller på hver enkelt deltager i en gruppe, som gjør at man vanskelig kan se om endringer skyldes manipulasjon av uavhengig variabel, eller kun disse forskjellene. Disse forskjellene skyldes extraneous variabler, som man i praksis ikke kan unnslippe. Reduksjon av error variance er en av hovedårsakene til at eksperimentell forskning ofte gjøres i laboratorium, da man på denne måten kan minimere kilder til extraneous variabler som oppstår fra miljøet. Error variance kan også minskes ved å øke effektiviteten til uavhengig variabel, som i eksempelet med THC, kan denne dosen økes. Fullstendig fjerning av error variance går ikke, men man kan også komme nærmere ved å randomisere tildelingen av gruppe mellom deltagere. Her håper man at individuelle forskjeller mellom grupper vil tilsvare hverandre i kontroll vs treatment condition.

Randomized two-group design

- Et randomisert togruppedesign er et design som består av to grupper som deltagere har blitt tildelt med lik sannsynlighet, og er et av de enklere forskningsdesignene man kan tenke seg. Her har uavhengig variabel to nivåer, treatment og kontroll, og man kan her finne effekter av treatment sammenliknet med kontroll på en relativt ukomplisert måte (om effekten finnes...). Tydelige resultater kan hentes med relativt få deltagere, og designet er altså svært økonomisk. Dessverre er informasjonen man kan hente om effekten av uavhengig variabel relativt begrenset, man får kun vite om den gitte graden av treatment hadde effekt eller ikke, og for eksempel ikke noe om hvilken grad av treatment som behøves. Man får altså ikke særlig mye informasjon om hvordan uavhengig variabel påvirker avhengig variabel (funksjonen). Se s.337-338 om Gold's eksperiment.

Parametric design

- Et parametrisk design er et design hvor uavhengig variabel manipuleres kvantitativt på tvers av conditions, for eksempel i Guéguens undersøkelse av 19-åringens tiltrekning avhengig av høyden på hælene, flat, 5cm eller 9 cm.

Nonparametric design

- Et nonparametrisk design er et design hvor uavhengig variabel manipuleres kvalitativt på tvers av grupper, dvs at uavhengig variabel ikke endres i grad, men av type. Jeg antar at man kan argumentere for at Guéguens eksperiment var både nonparametrisk og parametrisk avhengig av om man regner flate sko som en sko uten hæler, eller som "0cm"-hæl. Uansett var forsøket et eksempel på et randomisert multigruppedesign. S.339.

Multiple control group design

- Multiple control group designet er en variasjon av single-factor multigroup design som benyttes om kun én kontrollgruppe ikke er tilstrekkelig for å utelukke alternative forklaringer av funn. For eksempel kan hver enkelt condition ha en egen kontrollgruppe.

Matched-groups design

- Matched-group designs benyttes der man har grunn til å tro at individuelle forskjeller har store påvirkninger på uavhengig variabel, for eksempel ifb med reaksjonstid. Rent konkret går prosessen ut på at man henter et utvalg av en populasjon, hvor hver enkelt deltager blir pretestet og gitt en måling på en viss avhengig variabel. Deretter deles de med tilsvarende mål inn i par, der den ene i paret tilfeldig havner i kontrollgruppe, mens den andre blir utsatt for treatment. På denne måten kan man være trygg på at gjennomsnittet for hver gruppe på avhengig variabel er lik, slik at effekten av treatment kan sees tydeligere. Altså forminsker man error variance, om det faktisk er en sammenheng mellom karakteristikken deltagerne har blitt paret på grunn av, og den avhengige variabelen vi vil undersøke om uavhengig variabel har en effekt på. Det er essensielt at denne sammenhengen finnes, da effekten av paringen vil bli negativ for eksperimentets kvalitet i stedet, da de slutningsstatistiske verktøyene som kan benyttes for slike eksperimenter er dårligere enn i andre eksperimenter.

Matched-pairs design

- Matched-pairs design er matched-groups-ekvivalenten til randomiserte togruppedesign, slik som i eksempelet med reaksjonstidseksempelet på s.340. Her har man altså to grupper, hentet fra de to nivåene av uavhengig variabel, som gir en kontroll, og en treatment condition. De samme fordelene og ulempene av randomiserte togruppedesign gjelder her og matched-groups design, med lavt ressursbehov, begrenset informasjon (ingen om funksjon), og at effekt er tydelig sammenliknet med kontroll.

Carryover effects

- Carryovereffekter er et alvorlig problem med within-subjects design, der deltagere utsettes for ulike nivåer av uavhengig variabel. På grunn av dette kan effekter av å ha blitt utsatt for en condition bære med seg følger til neste condition, typisk i situasjoner hvor for eksempel tester av ulike former for læring foregår. Om Tom i condition 1 lærer å lese, og deretter blir utsatt for en annen læringstreatment i condition 2, kan effektene av at han har lært å lese i condition 1 videreføres, og gi overdreven høy effekt i condition 2 (for eksempel). "Carryover effects occur when exposure to a treatment alter the behavior observed in a subsequent treatment" (s.345). Carryover kan arte seg forskjellig og av forskjellige grunner, for eksempel "learning", "fatigue", "habituation" osv.

Counterbalancing

- Counterbalancing er et tiltak som brukes for å minimere effekten av carryover på den observerte effekten mellom uavhengig og avhengig variabel. Complete counterbalancing skjer ved at hver enkelt deltager utsettes for en unik rekkefølge av treatments, tildelt tilfeldig. Komplette counterbalancing (motbalansering?) kan være ressurskrevende og jo flere treatments, desto (enormt mye) flere rekkefølger kreves. Et forsøk med tre treatments krever 6 subjekter ($3 \times 2 \times 1 = 6$), mens et forsøk med fire treatments krever 24 subjekter ($4 \times 3 \times 2 \times 1$). Altså blir komplette counterbalancing raskt ekstremt ressurskrevende jo flere treatments som inngår. Partial counterbalancing (delvis motbalansering?) innebærer at man velger ut noen rekkefølger som distribueres tilfeldig på subjekter, noe som ikke nødvendigvis vil balansere ut carryovereffekter, men sannsynligvis komme nær i å gjøre det. Det er også viktig å merke seg at for at counterbalancing generelt skal fungere, må carryovereffektgraden være lik eller tilnærmet lik på tvers av rekkefølger.

Factorial design

- Et faktorielt design er et ressursbesparende alternativ om man ønsker å teste flere uavhengige variabler på en avhengig variabel. Her refereres hver variabel til som en faktor, og det må finnes minst to, med minst to nivåer av hver faktor. Et faktorielt design med to faktorer A og B, der A har 2 nivåer og B har tre nivåer, kalles et 2×3 faktorielt design (Nivåer av A * Nivåer av B). I dette eksemplet finnes to faktorer, men om det fantes en tredje faktor C med 4 nivåer ville vi hatt et $2 \times 3 \times 4$ faktorielt design.

Main effect

- I et faktorielt design kan man både finne main effects og interaksjoner. Main effects er effekten en uavhengig variabel har på den avhengige, og vil i et rutenett tilhørende et faktorielt design være gjennomsnittet av to nivåer eller to treatmenttyper. Illustrasjon (figure 10-7) s.358.

Interaction

- En interaksjon oppstår når et gitt nivå av en uavhengig variabel interagerer med en annen uavhengig variabel og / eller motsatt. I eksempelet s.358 ser vi at effektene av "fear type" er uproporsjonalt store når "fear intensity" er høy, sammenliknet med lav. Dette er en interaksjon mellom de to uavhengige variablene.

Simple main effect

- Simple main effects er så vidt jeg har forstått hver enkelt linje (om effekter grafes) sin hovedeffekt, for eksempel "fear intensity" sin egen hovedeffekt, eller "fear type" sin egen hovedeffekt. Altså er en "simple main effect" en uavhengig variabels effekt, på et gitt nivå av en annen uavhengig variabels effekt. Om linjene i en slik graf ikke er parallelle, kan det være en interaksjon tilstede. Se s.359, figur 10-9.

Higher-order factorial design

- Et higher-order factorial design er et faktorielt design som bygger på mer enn to uavhengige variabler. Nyten av slike design vil raskt bli begrenset da det vil skapes såpass mange grupper, med et såpass stort behov for deltagere per gruppe (minst fem), at et slikt design slutter å være økonomisk. I tillegg blir det desto mer komplisert å undersøke interaksjoner, men om man i større grad er ute etter å undersøke main effects, kan man øke studiens økonomiske fordeler ved å prioritere deltagere til gruppene der main effects kan vises. I design med tre faktorer, vil "simple effects" bestå av toveisinteraksjoner av de andre to faktorene. Se s.371-361.

Kapittel 11

Kapittel 11

Mixed design

- Et "mixed design" (split-plot) er et faktorielt forskningsdesign som kombinerer karakteristika fra between- og within-subjects design. Et blandet design grafes ofte ut i fire (eller flere) ruter hvor sammenfallet mellom ulike nivåer av en treatment, samt typen treatment vises. Eksempel: Narrative evidence kontra statistical evidence vannrett, med personlig relevans loddrett, (høy vs. Lav), (Zhang, 2014). Fordeler: Flere uavhengige variabler undersøkes, gir resultater for flere kombinasjoner av variabler.

The nested design

- Design som kombinerer within- og between-subjects design, gjennom nivådeling av within-subjects faktorer, (f. Eks. A1, A2 osv), med flere grader av between-subjects-faktorer B (B1, B2, osv) "nestet" under hver A-variant. Hvert nivå av A har flere nivåer av B "nestet under seg", slik at man kan undersøke flere interaksjoner. Fordel: Mer praktisk / økonomisk enn faktoriale design. Ulemper: På grunn av nestingen blir noen interaksjoner ikke undersøkt.

Covariate

- Det legges til en ekstra korrelerende variabel til den avhengige variabelen for å kontrollere for å kontrollere effekten av den uavhengige variabelen. En slik Covariate brukes altså da for hver deltager i eksperimentet. Eks. Tommy og Erik deltar i en studie av kunnskap om japansk kultur. Det tenkes at det finnes en korrelerende variabel, holdninger til playstation vs. Xbox-debatten. Tommy hater playstation, og Erik elsker playstation. Likevel skårer de to likt på den uavhengige variabelen, men effekten er større for Tommy som hater playstation.

Quasi-independent variable

- En kvasi-uavhengig variabel er en variabel som behandles som uavhengig i en korrelasjonsstudie, men som altså ikke helt er det. Typisk brukes disse i tilfeller der det skjer en lov- eller annen endring i forhold til en avhengig variabel. Her kan man få interessante data om endringen av et forhold og dens påvirkning på andre forhold i et større / samfunnsperspektiv. Eksempel: Det innføres en reklamekampanje om bruk av sikkerhetsbelter. Synker trafikkødeligheten? Det utføres målinger av dødelighet før og etter reklamekampanjen.

Quasi-experimental design

- Kvasiekperimentelt design er en type forskningsdesign hvor man benytter en kvasiavhengig variabel til å undersøke effekt på gitte avhengige variabler. Her er det viktig at man motstår fristelsen til å si at dette påviser kausalitet, da det er såpass manglende kontroll på konfunderende variabler at man kun kan finne korrelasjon, i en typisk situasjon hvor kvasiekperimentelt design benyttes (det virkelige liv).

Time series design

- Tidsseriedesign er en type within-subjects design der man undersøker effekten av en treatment hos deltakere over tid. Tidsseriedesign er en type kvasiekperimentelt design. Det finnes flere typer, for eksempel brutte tidsseriedesign og ekvivalente tids sample design.

Interrupted time series design

- Tidsseriedesign der det utføres målinger i gitte tidsrom før og etter innføringen av treatment (kvasiavhengige variabler). Eksempel: undersøkelser gjøres i tidsrom 3 år før og 3 år etter en lovendring.

Equivalent time samples design

- Within-subjects design hvor treatment innføres og fjernes gjentakende over tid, slik at man kan undersøke effekter av treatment hos deltagere, og utnytte within-subjects-designets fordeler ifht kontroll osv. Eksempel: René blir testet for ADHD en gang i uka i fire uker uten treatment, som så innføres i fire uker, fjernes i fire uker, og gjeninnføres. Har symptomene økt og sunket tilsvarende har man funnet en effekt.

Nonequivalent control group design

- Et tidsseriedesign hvor effekten av treatment undersøkes, med en kontrollgruppe fra et annet miljø i tillegg.

Pretest-posttest design

- Pretest-posttest-strategien hører til et sant eksperiment, og er i praksis at en pretest introduseres for å undersøke effekten av en treatment gjennom en posttest, sett i lys av allerede eksisterende data hentet gjennom pretest. Eksempel: Fungerer læringsstrategi 2 bedre enn læringsstrategi 1? Tom-René og Erik scorer helt forskjellig i deres kunnskap om dinosaurer, og sett i lys av pretesten kan man ikke hevde at læringsstrategi 2 var bedre enn 1 under posttesting.

Solomon four-group design

- Det benyttes fire scenarioer i pretest-posttest designet:
 1. Pretest-treatment-posttest
 2. Pretest-ingen treatment-posttest
 3. Ingen pretest-treatment-posttest
 4. Ingen pretest-ingen treatment-posttest

Designet benyttes for å minimere eller eliminere carryovereffektene som kan oppstå gjennom pretesting. Her finner man både ut om treatment har effekt (ideelt sett selvfølgelig) uansett om det benyttes en pretest eller ikke.

Cross-sectional design

- I krysseksjonelle design hentes det ut deltagere tilhørende ulike aldersgrupper (aldersmessig krysseksjon av samfunnet) som undersøkes for det man ønsker informasjon om. I en undersøkelse av iq finner man at iq øker fram til voksen alder øker, og holder seg stabil frem til alderdommen, der den synker. Problemet med dette er at det kan oppstå en generasjonseffekt: De som er 70 i dag har gjerne levd under vidt forskjellige kår fra det en 30-åring har, og nedgangen i iq rundt denne alderen kan skyldes andre faktorer, for eksempel dårlig kosthold eller dårligere undervisning i personens oppvekst.

Longitudinal design

- Longitudinelle design baserer seg på at man henter ut representanter av samfunnet (for eksempel fordelt aldersmessig), som man følger over mange år (gjerner flere tiår). Man kan hente en gruppe barn på et tidspunkt, som man følger over 30 år. På denne måten unnslipper man generasjonseffekten nevnt ifbm krysseksjonelle design. Problemet er at andre generasjonseffekter kan finne sted, Shaffer nevnte i 1985 at longitudinelle design kan ha "cross-generational effects", dvs at en longitudinell studie, for eksempel fra 1910-1940, ikke gjelder i dag på grunn av endringer i verden osv. Ulemper er at designet er ekstremt tidskrevende og påvirkes av frafall i fbm død, flytting osv (subject mortality).

Cohort-sequential design

- En slags blanding av krysseksjonelle og longitudinelle design der det hentes ut kohorter av samfunnslag (for eksempel aldersgrupper), som følges over tid. På denne måten kan generasjonseffekter detekteres og tas høyde for ifht resultater osv. Slik kan man for eksempel sammenlikne 10-åringer i 1990 og 10-åringer i 2000, og undersøke kohortgruppe- / generasjonseffekter.

Kapittel 12

Baseline design

- I et baseline design blir i motsetning til gruppebaserte design et enkelt individs atferd undersøkt på tvers av treatment conditions, for å etablere en såkalt behavioral baseline. På denne måten kan et individs "baseline" for atferd etableres, dvs hvordan den naturlige

forekomsten av målatferden (avhengig variabel) arter seg. Baseline design ble hovedsakelig utviklet av B. F. Skinner.

Behavioral baseline

- Et behavioral baseline er en kartlegging av resultater fra repeterte baselinetester (pretester ifb med atferd). Behavioral baseline endrer seg som regel i større grad ved innføring av påvirkende treatment, men disse endringene vil som regel stabiliseres, utenom litt variabilitet grunnet tredjevariabler (usystematisk fluktuering). Stabiliseringen av baseline er fullendt når et "stability criterion" oppfylles.

Stability criterion

- Et stability criterion (stabilitetskriterie?) er et gitt punkt hvor baseline har blitt oppnådd, dvs at baseline har stabilisert seg ifb med en gitt treatment. Heretter kan deltager utsettes for ytterligere treatments, som så stabiliserer seg. Så blir denne prosessen repetert.

Baseline phase

- Baselinefasen er fasen hvor baseline (målatferd / avhengig variabel) er etablert, uten at deltager har blitt utsatt for treatments. Den enkleste formen for baselinedesign inneholder to faser, baselinefase og intervensjonsfase.

Intervention phase

- Intervensjonsfasen er fasen hvor deltager har blitt utsatt for treatment, og her kartlegges deltagers atferd ifb med administrering av treatment.

ABAB design

- ABAB-designet er kort fortalt at subjekt utsettes for baselinefasen (A), og deretter intervensjonsfasen (B), to ganger, slik at en A-B-A-B-rekkefølge oppstår.

Intrasubject replication

- Intrasubject replication er at målingen av baseline- og intervensjonsfasene gjentas, for å undersøke reliabiliteten mellom hver fase (er det konsekvent variabilitet mellom fasene?), og om reliabiliteten ser ut til å være tilstede, har man etablert intern validitet.

Reversal strategy

- For at et ABAB-design skal fungere (intrasubject replication), må man vite og ha undersøkt om effektene av B vil avta etter eksponering. Reverseringsstrategien er altså at intervensjonsfasen avsluttes, og den nye baselinefasen (forhåpentligvis) tiltar og derfor kan vurderes. Er effektene av B reversible eller i større grad varige?

Intersubject replication

- Selv om single-subject design heter det det heter (n=1 design), er det vanlig å undersøke ekstern validitet gjennom at flere forskjellige individer utsettes for single-subject designet. Dette kalles intersubject replication, altså at single-subject design repeteres med forskjellige subjekter for å se om funnene kan generaliseres. Tre til seks subjekter er et vanlig antall.

Systematic replications

- Systematisk replikasjon innebærer i n=1-design å replikere eksperimenter med et utvidet design som inneholder flere manipulasjoner av uavhengig variabel, flere variabler, og flere typer subjekter som testes, til forskjell fra direkte replikasjoner hvor den første eksperimentelle settingen repliseres identisk.

Direct replication

- Direkte replikasjon innebærer at man repliserer et tidligere forsøksdesign og forsøk helt likt, dvs at det ikke gjøres noen endringer av variabelers manipulasjon, antallet og typen variabler eller subjekt-typer (duer vs kaniner for eksempel).

Multiple-baseline design

- I eksperimenter der reverseringsstrategien ikke fungerer, altså at reverseringen av effekt fra intervensjons- til ny baselinefase ikke forekommer, benyttes multiple-baseline design. Multiple-baseline design kan benyttes når subjekter har to ulike atferdsmønstre (behaviors), der det er forventet at intervensjon av disse ikke vil reverseres etter endt intervensjon. Her kartlegges først baseline for én atferd, som deretter endres med intervensjon. Deretter kartlegges baseline for atferd 2, som deretter endres med intervensjon. I multiple-baseline design med flere subjekter undersøkes én atferd, der baseline kartlegges likt, mens intervensjon påbegynner i ulike faser for de forskjellige subjektene.

Changing criterion design

- I situasjoner hvor det ikke er ønskelig å introdusere en "hel intervensjon" i én enkelt fase benyttes changing criterions. Altså endrer kriteriene som må oppnås for at en fase skal gå over, fra fase til fase. For eksempel kan et overordnet mål for et subjekt være å slutte å røyke (null sigaretter per dag). Her kartlegges baseline, og det fjernes én sigarett per dag per fase, slik at en ny baseline etableres (kriteriet nås). Neste intervensjon fjernes nok en sigarett fra budsjettet, og når dette er gjort har det nye kriteriet blitt nådd. Kalles "shaping" innenfor operant betingning.

Dynamic design

- Et dynamisk design innebærer bruk av "continuous" endring uavhengig variabel (uavhengig variabels verdi varierer innenfor et range, med uendelige verdier innenfor denne rangen), sammenliknet med "discrete" endring, dvs endring av uavhengig variabel som er begrenset og målbar). Dynamiske design deler en del karakteristika med baseline design, men på grunn av mangelen av "discrete variables", og heller bruk av kontinuerlige, kaller man det dynamisk. Dynamiske design kan være like informative som baseline design om endringen av uavhengig variabel ikke er raskere enn atferden man ønsker å undersøke, og at endringene er reversible til en viss grad (s.419). Designet benyttes for å undersøke atferdsdynamikk, altså overgangen mellom ulike grader av treatment (s.109 Fostervold).

Discrete trials design

- Discrete trials design benyttes der det ikke er produktivt å dele faser inn i baseline og intervensjon, da disse fasene heller må "blandes" fordi subjektet raskt vil forstå at (det repeterte) forholdet i hver fase. Designet brukes ofte innenfor psykofysikk for å undersøke menneskers evne til å skilde stimulus (signal) fra støy, og bygger på at subjekter utsettes for et gitt ("høyt") antall conditions av hver type (randomisert eller motbalansert). Se. S.421 B & A, og s. 109 Fostervold).

Kapittel 13

Descriptive statistics

- Deskriptiv statistikk brukes for å få overblikk over datamaterialets utseende, for eksempel hvordan ulike grupper fordeler seg, hvordan den uavhengige variabelen er manipulert og hvordan spredningen i materialet er. Slik kan man avsløre feilkilder i datamaterialet, detektere mønstre i dataen, og sjekke hypoteser man hadde før datainnsamlingen begynte.

Exploratory data analysis (EDA)

- EDA benyttes for å gjøre deskriptiv statistikk, dvs - EDA er å bedrive deskriptiv statistikk ift datasettet man har samlet, for å sjekke etter mønstre, feilkilder og hypoteser man hadde før innsamlingen skjedde.

Dummy code

- Dummy code benyttes under organisering av data, ved at forskjellige kategorier tildeles tilsvarende tall, for eksempel å skrive "1" i stedet for "mann". Dette forenkler organiseringsprosessen, da det for eksempel er lettere å skrive "1", enn "mann". Dette er tidsbesparende da tabeller inneholdende data kan bli svært store og lange.

Bar graph

- Søylediagram består av søyler fordelt langs en x-akse (akse for uavhengig / prediktorvariabel) hvor lengden avgjøres av graden av avhengig variabel på y-aksen. Søylediagram er nyttige hvor forskjellige grader av uavhengig variabel er kategoriske, for eksempel når uavhengig variabel tilsvarer forskjellige typer treatment (f. Eks. Medikamenter), sammenliknet med for eksempel dose av ett gitt medikament.

Line graph

- Linjediagrammer er nyttige hvor det er et kontinuerlig forhold eller en trend mellom y- og x-aksen, og benyttes ofte i eksperimentelle forsøk. Altså er linjediagrammer nyttige om man ønsker å illustrere funnet kausalitet, og funksjonen som står bak endringene. For eksempel kan hvert punkt langs x-aksen være en grad av én treatment (dose), mens y-aksen reflekterer den gitte graden av avhengig variabel. Slik kan man lett se forskjellene som ulike doser gir. Linjediagrammer kan også benyttes i flerfaktoreksperimenter, ved å benytte to ulikt utseende linjer langs hverandre.

Scatter plot

- Scatter plots består av to akser som korresponderer til to gitte variabler, der det markeres punkter for hver korrelasjon gjennom koordinater. Om variabel A og B har verdiene 1 og 2 for subjekt 1, kan dette markeres som et punkt i grafen der $x = 1$ og $y = 2$. Scatter plots kan legges til rette for regresjonsanalyse senere.

Frequency distribution

- Et av de første stegene innen dataanalyse har med undersøkelse av frekvensdistribusjon å gjøre. Her deles de ulike variablene i en korrelasjonsstudie eller den avhengige variabelen i et eksperiment inn i ulike kategorier for hver kvantitet. Om du undersøker IQ, er det nyttig å etablere ulike "ranges" av funnet IQ, for eksempel 65-74, 75-84 osv, og så plottes det antallet deltagere som falt inn i hver range. Kompakte frekvensdistribusjonstabeller er dessverre ikke veldig praktiske for henting av informasjon om center, spread eller shape, og derfor kan man lage histogram og stemplots fra tabellen.

Histogram

- Histogram likner på søylediagrammer, men det mangler mellomrom mellom de ulike søylene for å illustrere at det ikke er noe mellomrom mellom de ulike kategoriene (ranges) som vises. I et histogram representerer y-aksen frekvensen som hver klasse på x-aksen representerer, og man kan enkelt se hvordan fordelingen av de ulike klassene ser ut.

Stemplot

- Stemplots bygger på at innsamlet data deles inn i to sider av en vertikal linje, der sifferet på venstre side er det tallet i hvert stykke data som ligger lengst til venstre, mens dataen til høyre for linjen viser de ulike tallene som ville ligget bak for hvert stykke data (eks. 6 | 022, her har man tre målinger på 60, 62 og 62.) Her bevares alle stykker med data, som er en

fordel om det ikke er så veldig mange målinger som ligger bak grafen. Histogrammer er altså å foretrekke med svært store datasett (eks. Stemplot med mange målinger: 10 | 000011111122223333334444555555666677777788888888889999999999, dette blir raskt uoversiktlig...).

Skewed distribution

- En "skewed distribution" (skjev fordeling) er en distribusjon av skårer der flest klasser forekommer skjevt, dvs at de ikke ligger sentrert blant klassene. En positivt skjev distribusjon innehar flere skårer som ligger høyt, mens en negativt skjev distribusjon har flere lave skårer (relativt til måleklassenes verdi). En skjev distribusjon er en "shape" (form) på en fordeling. På et histogram kan dette sees ved at flest skårer forekommer høyt eller lavt, og ser ut som en opp- eller nedoverbakke (fra venstre).

Normal distribution

- En normalfordeling forekommer når en fordeling av skårer ser noenlunde symmetrisk ut, dvs at de fleste skårer havner rundt (det forventede) gjennomsnittet, med jevn avtagelse av skårer på hver side av snittet. En del psykologiske målinger oppstår en normalfordeling (f. Eks. IQ), men ikke alltid. En del vanlige slutningsstatistikker kan gjøres ifb med normalfordeling, og det er viktig å undersøke om dette gjelder en gitt fordeling man ønsker å undersøke.

Outliers

- Outliers er skårer som ser ut til å ikke "passe" med resten av en gitt fordeling, for eksempel om man finner svært mange med IQ på 65-74 relativt til en vanlig fordeling. I et histogram kan dette sees med at det finnes en tilsynelatende "for høy" søyle. Det er viktig å undersøke outliers og hva de betyr, da de kan være indikative for feil under måling, men ikke alltid.

Resistant measures

- Resistant measures er mål som i mindre grad lar seg påvirke av skjeve fordelinger, eller fordelinger preget av outliers. Eksempler er "measures of center" og "spread" (sentraltendenser og varians).

Measures of center

- "Measures of center", eller "measure of central tendency" (mål på sentraltendenser) er mål som kan gi et inntrykk av den generelle verdien av skårer i en fordeling, for eksempel gjennomsnitt (mean), typetall / modus (mode) og median. Hver av disse har styrker, svakheter og situasjoner hvor de ikke kan brukes.

Mode

- Modus er den oftest forekommende verdien (skåren) i en distribusjon, og kan ikke benyttes i en fordeling der hver skåre er unik. I tillegg kan det finnes flere ulike modus, som i bimodale fordelinger (fordelinger med to modus). Modus er ikke det beste målet av sentraltendenser da man ofte ikke kan bruke det til sammenlikning av to fordelinger fordi fordelinger kan være svært forskjellige men likevel ha liknende modus. Modus må brukes på nominelle skaler, kan brukes sammen med median på ordinale skalaer. Modus er eneste nyttige mål på bimodale distribusjoner.

Median

- Median er den midterste skåren rangert etter verdi i en fordeling der antallet skårer er et oddetall, og snittet av de to midterste skårene i en fordeling der antallet skårer er et partall. Median gir begrenset (ingen) informasjon om hvordan en fordeling ser ut utenom "midten", da det kan være svært forskjellig utseende fordeling ellers. Eksempel: Median kan være 5,

men det finnes likevel svært mange skårer på 2 og 20, som median ikke indikerer på noen måte. Derfor brukes median stort sett der gjennomsnitt ikke kan benyttes. Median kan brukes på ordinale skalaer, kan ikke brukes på bimodale distribusjoner, og er beste mål på svært skjevfordelte distribusjoner.

Mean

- Mean (M), eller gjennomsnitt, er et mål på sentraltendens der alle skårer oppsummeres og deretter deles på antallet skårer. Gjennomsnitt er det vanligst benyttede målet på sentraltendens. Fordelen er at verdien på snittet er direkte påvirket av verdien på hver enkelt skåre, men dette gjør også at verdien av snittet er svært sensitiv for påvirkning fra outliers. Derfor kan outliers gjøre gjennomsnittet kunstig høyt eller kunstig lavt, sammenliknet med hvordan en fordeling faktisk er ut. Det beste målet på sentraltendens er det som best reflekterer hvordan en sentrum av fordeling faktisk er. Gjennomsnitt kan ikke brukes på nominelle skalaer, bør ikke brukes på ordinale skalaer, eller bimodale fordelinger, eller på skjeve fordelinger.

Measures of spread

- Spread er variabiliteten observert ift verdier på en avhengig variabel, eller skårer på en fordeling. Spread er altså hvor mye skårer avviker fra hverandre, eller hvor "consistent" skårer er. Fire mål på spredning er "range", "interquartile range", "variance" (varians) og "standard deviation" (standardavvik). Man bør både undersøke spredning og sentrum i et sett med skårer i en fordeling.

Range

- Range er den enkelste målingen av spread, og innebærer at den høyest skåren i en fordeling subtraheres med den laveste. Range er sensitiv for outliers, og den tar heller ikke hensyn til størrelsen på skårene mellom høyeste og laveste skåre.

Interquartile range

- Interkvartil range bygger på at en rangert fordeling av skårer deles inn i fire like store andeler, og så finner man rangen mellom topp og bunn 25% (Q3 minus Q1). Interkvartil range er mindre sensitiv for outliers, og er nyttig der man ønsker å finne en omtrentlig måling av spredning som ikke er sensitiv for ekstreme skårer eller skjevheter i fordelingen.

Variance (s^2)

- Variansen er den gjennomsnittlige kvadrerte avstanden fra gjennomsnittet hver skåre har.

$$s^2 = \frac{\text{Summen av } (X-M) \cdot (X-M)}{n-1}$$

X er hver individuelle skåre fordelingen består av, M er gjennomsnittet for hver skåre, og n er antallet skårer i distribusjonen.

Standard deviation

- SD er roten av s, og er et nyttig mål da det uttrykker variansen i samme "form" som resten av uttrykkene i en analyse (ikke kvadrert verdi). Eks. $S^2 = \text{roten av } S$, roten av 6,8 = 2,61. SD / varians er ikke nyttig å bruke i en svært skjev fordeling (bygger på gjennomsnitt).

Five-number summary

- Five-number summary består av mål på maksimum, det første kvartilet, medianen (andre kvartil), tredje kvartil og minimum. Denne oppsummeringen inneholder tre mål som er resistente mot outliers og ekstreme verdier, første, andre og tredje kvartil, og to som ikke er resistente, minimum og maksimum. Med minimum og maksimum kan range enkelt regnes ut (maksimum - minimum).

Box plot

- Box plots kan illustrere dataen vist i en five-number summary, gjerne sammenliknet med andre fordelinger / utvalg med egne verdier for sin five-number summary. Box plots består av én firkant med median (som en horisontal strek), og to "whiskers" som stikker ut (maksimum og minimum). Firkanten i midten viser median, første kvartil og tredje kvartil. I en skjevfordelt fordeling vil medianstreken være høyere eller lavere relativt til firkanten som viser første og tredje kvartil (avhengig av positiv eller negativ skjevhet).

Measure of association

- Measure of association (mål på korrelasjon), brukes for å undersøke i hvilken grad to fordelinger korrelerer. Vanligvis benyttes Pearson r (Pearson product-moment correlation coefficient).

Pearson r

- Pearson r brukes som et mål på om hvorvidt to utvalg korrelerer, på en intervall- eller ratioskala. Verdien på Pearson r kan variere mellom -1 og 1, hvor -1 viser negativ korrelasjon, +1 viser positiv korrelasjon, og 0 viser ingen korrelasjon i det hele tatt. Altså illustreres det i hvilken grad to fordelinger ser ut til å danne en lineær linje i et boxplot (lineær korrelasjon). Outliers, skjevfordeling og kurvelineære forhold vil påvirke Pearson r, og det bør derfor sikres at en fordeling er noenlunde normalfordelt før Pearson r brukes som mål på assosiasjon.

Point-biserial correlation

- Point-biserial correlation er et mål på assosiasjon som kan benyttes om to variabler er målt på to ulike skalaer, for eksempel en intervallskala og en nominell skala. Om man bruker dummy code, og gir den nominelle skalaens mål tallverdier, kan man regne ut korrelasjonen likt som med Pearson r om den nominelle skalaen er dikotom. Om det er ulikt antall deltagere i hver nominelle kategori, vil man ikke kunne få et svar likt 1,0, selv med perfekt korrelasjon. Derfor kan man lett undervurdere en korrelasjon om det er slik (dette må man være obs på). I tillegg kan korrelasjonens sanne styrke bli skjult om en kontinuerlig variabel gjøres dikotom (angst = høy eller lav).

Spearman rank-order correlation, rho

- Rho benyttes som mål på assosiasjon om dataen er målt på ordinale eller større skalaer, eller når man ønsker å finne ut om forholdet mellom variabler er monotonisk.

Phi coefficient

- Phi-koeffisienten benyttes når begge variabler er målt på dikotome skalaer. Her gjelder de samme begrensningene og fordelene som med point-biserial korrelasjon, bare i dobbelt så høy grad. Altså kan verdiene dummy kodes og regnes ut med Pearson r, men begrensningene, at kontinuerlige variabler som gjøres dikotome kan skjule forholdets sanne styrke. I tillegg spiller forskjeller i antall som skårer forskjellig inn, som med point-biserial correlation.

Linear regression

- Lineær regresjon kan brukes til å estimere verdiene til variabler sammenliknet med andre variabler, på bakgrunn av kunnskap om assosiasjon mellom disse. Om du vet at $y=1$ når $x=2$, og at $x=6$ når $y=3$, kan du predikere at $x=4$ når $y=2$ (med lineært forhold).

Bivariate linear regression

- Bivariate linear regression er i lineær regresjon ifb med fordelinger basert på to variabler. Her er målet å finne en rett linje som best passer sammen med data vist i et scatter plot, dvs

illustrerer forholdet best. Den beste linjen her er den som minimerer summen av avstandene mellom linjen og punktene på grafen på y-aksen som målt av mengden kvadrater "mellom" punktene og linjen.

Least-squares regression line

- For å finne den beste og mest informative linjen for bruk i lineær regresjon, ønsker man at denne skal i snitt ligge nærmest mulig alle punkter langs y-aksen i kvadrert avstand.

Regression weight

- Den matematiske formelen som beskriver regresjonslinjen matematisk er $\hat{y}=a+bX$. Her er b linjens helning, og kalles "regresjonsvekten" (regression weight). Se s.462 B&E for informasjon om videre utregninger.

Standard error of estimate

- \hat{Y} er den forventede Y -verdien predikert med regresjonslinjen. Denne stemmer ikke alltid, og forskjellen mellom \hat{y} og faktisk y kalles en residual. Disse oppstår når korrelasjonen ikke er perfekt (stort sett aldri), og er i praksis avstanden mellom \hat{y} og y . En "standard error of estimate" er altså målet på avstandene mellom \hat{y} og y , for en gitt regresjonslinje i et gitt scatter plot, og formelen finnes på s.462 i B&A.

Coefficient of determination

- Om en gitt korrelasjonskoeffisient kvadreres, får man koeffisienten for bestemmelse (coefficient of determination). Denne måler proporsjonen av varians som de to testede variablene deler, og kan indikere om variabilitet av X forklarer variabilitet av Y og i hvilken grad. Om $r=0,60$, er coefficient of determination lik $0,36$. Dette betyr at ca 36% av variasjonen av y er skyldes variasjon av x . Det kan være fristende å trekke slutninger om kausalitet her, la være, eller vennligst vær forsiktig. Kalles også delt varians.

Coefficient of nondetermination

- Coefficient of nondetermination er lik $1 - \text{coefficient of determination}$, her for eksempel lik $1 - 0,36 = 0,64$. Her har vi altså kalkulert hvor mye av variasjonen av en variabel som ikke skyldes variasjon i en annen, og dette er i praksis "uforklart varians", forårsaket av umålte faktorer. Uforklart varians kan være lav fordi korrelasjonen ikke er lineær... Gjerne sjekk dette med et scatter plot.

Correlation matrix

- En korrelasjonsmatrise er en tabell hvor korrelasjonen mellom ulike variabler kan vises. Her lar man være å vise verdiene langs de "diagonale", da det holder å vise korrelasjonen mellom for eksempel variabel 3 og 5 én gang (se. Table 13-5, s. 462 B&A).

Kapittel 14 (ikke hele, det som er relevant (så vidt jeg har forstått))

Inferential statistics

- Slutningsstatistikk er statistikk som gjøres for å undersøke reliabiliteten av funn i forskning. Slutningsstatistikk henger tett opp imot NHST (Null hypotheses significance testing), det vil si vurderinger som gjøres ift om hvorvidt nullhypotesen er støttet eller ikke. For at nullhypotesen ikke skal være støttet må en alternativ hypotese være støttet i stedet.

Standard error of the mean

- Standardfeil er standardavviket i et utvalgs fordeling, og standard error of the mean (s_M / SE) er det vanligste målet. Her ønsker man å finne ut hvor mye gjennomsnittet i en fordeling kan avvike fra gjennomsnittet fra populasjonsgjennomsnittet. Formelen er slik:

$$SE = \frac{s}{(\text{roten av } n)}$$

S= standardavviket i utvalget, n= antall observasjoner.

Degrees of freedom

- Frihetsgrader (df) er antallet skårer i en distribusjon som kan variere rundt gjennomsnittet. I B&A gis et eksempel med 10 observasjoner, med gjennomsnittsskåre på 6,5. Her er kun 9 av observasjonene frie til å variere utenom gjennomsnittet, altså er df gitt av n-1. Om du har tre grupper, med et kjent gjennomsnitt (for alle tre totalt), er df lik k-1.

Estimation

- Estimering er å bruke data fra hentet fra utvalg til å estimere populasjonsparametre. Usikkerheten i en estimering oppgis med en gitt feilmargin (f. Eks. "+/- 2%").

Confidence interval (CI)

- Et konfidensintervall er en gitt "sikkerhet" man ønsker å operere med, for eksempel 95%, og det er dette tallet som gir opphav til feilmarginen oppgitt. Høyere CI vil gå på bekostning av presisjon (se eksempel med IQ s. 472 B&A). Dette er fordi CI bygger på at man for eksempel er sikker på at 95% av populasjonen vil falle innenfor konfidensintervallet, og om man skal være like sikker på at 99% av befolkningen faller innenfor, må intervallet forlenges.

Null hypothesis significance testing (NHST)

- NHST bygger på at man opererer med en nullhypotese, dvs at man mener at treatment ikke har noen effekt i det hele tatt (H_0), altså at all observert effekt skyldes samplingfeil og ikke faktisk effekt. Man ønsker her å beregne probabiliteten, p, for at effekten vist skyldes treatment og ikke samplingfeil, og om $P=0,05$ (vanligvis) mener man at observert effekt er statistisk signifikant (nullhypotesen avvises), mens alternativhypotesen (H_1 / H_a) styrkes.

Type I error

- En type I-feil har blitt begått når nullhypotesen avvises, til tross for at alternativ hypotesen ikke er støttet, mens nullhypotesen altså er riktig. I et signaldeteksjonsekspesiment kalles dette falsk alarm (intet signal, men likevel detektert).

Type II error

- En Type II-feil (i signaldeteksjon, et "miss" (bom)), er at nullhypotesen ikke forkastes, til tross for at den uavhengige variabelen i eksperimentet hadde effekt. Man ønsker å minimere sannsynligheten for type- I og II-feil, men dessverre vil tiltak som gjøres for å minske sannsynligheten for én type feil, øke sannsynligheten for den andre.

P-value

- P-verdien (p) er den estimerte sannsynligheten for at en observert treatmenteffekt er like stor eller minst like stor (på tvers av utvalg), som den man fant, gitt at nullhypotesen er sann. Altså er p-verdien det samme som kriterieverdien for at en nullhypotese skal forkastes, og bestemmer sannsynligheten for at Type I feil vil bli begått. Lavere kriterie for å forkaste nullhypotese = høyere sannsynlighet for type I feil.

Alpha level (alfa)

- Alpha level er sannsynligheten for at type I-feil vil bli begått. Alfa nivået som velges kalles signifikansnivået (level of significance), dvs sannsynligheten for at observert effekt skyldes

uavhengig variabel og ikke samplingfeil. Typisk settes alfanivået likt 0,05 (1/20 sjanse for at observert effekt skyldes samplingfeil). Lavere alfanivå gir lavere sannsynlighet for type I-feil, men høyere sannsynlighet for type II-feil.

Critical region

- I en fordeling må en viss mengde av de observerte skårene ligge over en viss z-skåre (antall standardavvik), for at funn skal kunne vurderes som statistisk signifikante. Med et alfanivå likt 0,05 tilsvarer den kritiske regionen 1,96 standardavvik i en tohalet test og 1,65 i en enhalet test. Altså må (i en enhalet test) minst 5% av skårer ligge i topp 5% av en fordeling, og i en tohalet test må minst 2,5% av skårer ligge i hver sin side av fordelingen (topp 2,5% på hver side).

Power

- Statistisk power er i hvilken grad statistikk vil reflektere reelle forskjeller som de er ment å vise, det vil si i hvilken grad en slutningsstatistikk vil kunne brukes til å svekke nullhypotesen om denne er feil, og tilsvarer $1 - \beta$. Beta er sannsynligheten for at type II-feil vil forekomme. Altså er en statistikk power i hvilken grad den kan detektere effekt. En rekke faktorer påvirker statistisk power, alfanivå, samplestørrelse og om testen er én- eller tohalet. Om man reduserer alfanivået, vil man redusere sannsynligheten for type I-feil, og derfor også power. Et større sample vil gi mer power enn et mindre et, og tohalet test er mindre "powerful" enn en enhalet test (behøves lavere z-skåre for å nå kritisk verdi).

Effect size

- Effektstørrelse er i hvilken grad manipulasjon av uavhengig variabel påvirker avhengig variabel, og benevnes med Cohens d ($(M_2 - M_1)/s$). Cohens d viser overlappet mellom to fordelinger gitt av to ulike "treatment means" (gjennomsnitt for hver av to treatments). Større effektstørrelse indikerer mindre overlapp og motsatt, og kan altså vise at to treatments er ulike i effekt. Det er i dag større og større press for at mål på effektstørrelse skal indikeres i forskningsartikler, da det er en veldig grei måte å vise om eller om ikke effekt skyldtes manipulasjon av uavhengig variabel (eksperimentelt).