

Validitetstypologi

Erik Gahner Larsen

Kausalanalyse i offentlig politik

Dagsorden

- ▶ Kausalitet og validitet
- ▶ Typologi
- ▶ Validitetsudfordringer
- ▶ Validitet og potentielle udfald

Seminaropgaven

- ▶ Næste deadline: 29. februar
- ▶ Omkring 5 sider
- ▶ Indhold:
 - ▶ Kort beskrivelse af emne
 - ▶ Gerne én eller to konkrete ideer
 - ▶ En reference eller to til relevant litteratur
- ▶ Send til: egl@sam.sdu.dk

Sidste uge og i dag

- ▶ Sidste uge: Rubins kausalmodel
 - ▶ Fokus på potentielle udfald
 - ▶ Definition af en kausaleffekt
 - ▶ Antagelser (SITA, SUTVA)
- ▶ I dag: Campbells kausalmodel
 - ▶ Validitetstypologi
 - ▶ Forskellige validitetsbegreber
 - ▶ Validitetstrusler

Validitet og potentielle udfald

- ▶ Flere ligheder end forskelle
 - ▶ Eksperimentel logik til observationsstudier
 - ▶ Fremhæver manipulation som afgørende for kausal inferens
 - ▶ Prioriterer simple deskriptive kausaludsagn over komplekse kontekstuelle kausalforklaringer
 - ▶ Fokuserer på en ukendt effekt af et kendt stimuli snarere end en kendt effekt af et ukendt stimuli
- ▶ Fundamentale forskelle
 - ▶ Rubin: Formel definition af en kausal effekt
 - ▶ Campbell: Generel typologi med validitetstrusler

Validitetstypologi

- ▶ Hvad er validitet?
 - ▶ Gyldighed
 - ▶ Den appromikserede sandhed af en inferens
- ▶ Validitet
 - ▶ Vi taler ofte om validiteten af et studie som værende lav eller høj
 - ▶ Men: validitet er ikke endimensionelt
- ▶ Typologi
 - ▶ Forskellige validitetsbegreber
 - ▶ Trusler til validitet (alternative forklaringer, plausible rivaliserende hypoteser)

Fire typer af validitet

- ▶ Statistisk validitet
- ▶ Intern validitet
- ▶ Konstruktionsvaliditet
- ▶ Ekstern validitet

Statistisk validitet

- ▶ På engelsk: *Statistical conclusion validity*
- ▶ Validiteten af inferensen relateret til korrelationen (kovariation) mellem stimuli (*uafhængig variabel*) og udfald (*afhængig variabel*)
- ▶ To aspekter:
 1. Er der en korrelation?
 - ▶ Hypotesetest
 2. Hvor stærkt korrelerer de?

Statistisk validitet: Ni trusler

1. Lav statistisk styrke (*statistical power*)

- ▶ Type I og II fejl
- ▶ Lav stikprøve: svært at finde en statistisk signifikant effekt
- ▶ Tommefingerregel (husk lektion 2): Jo flere observationer, desto bedre
- ▶ Statistisk signifikans \neq substantiel signifikans
- ▶ Forskellige måder at øge den statistiske styrke
 - ▶ Større stikprøve
 - ▶ Bedre mål (flere målinger, undgå *floor effects* etc.)
 - ▶ Forøg styrken i stimuli
 - ▶ Brug et *within-participants* design
 - ▶ osv.

Statistisk validitet: Ni trusler

2. Forudsætningsbrud

- ▶ Alle statistiske tests bygger på forudsætninger
 - ▶ Nogle tests bygger på flere forudsætninger end andre
- ▶ Har vi specificeret vores model korrekt?
- ▶ Kræver vi linearitet i vores parametre?
- ▶ Kræver vi normalfordelte fejllid?
- ▶ Tænk tilbage på Metode II

Statistisk validitet: Ni trusler

3. *P-value fishing*

- ▶ Hvis vi kun rapporterer signifikante resultater, kan konklusionerne være misvisende
- ▶ Ofte kører forskere statistiske tests til der kommer noget signifikant
 - ▶ “If you torture the data long enough, it will confess.” - Ronald Coase
- ▶ Kan også finde sted kun med én test
 - ▶ The garden of forking paths (Gelman)
- ▶ Flere løsninger
 - ▶ Bonferroni korrektionen
 - ▶ Preregistreret studie
 - ▶ Gør negative fund teoretisk interessante

Statistisk validitet: Ni trusler

4. Målefejl

- ▶ Reliabilitet
- ▶ Intet empirisk mål er perfekt
- ▶ Latent variabel, indekskonstruktion, faktoranalyse
- ▶ Løsning:
 - ▶ Flere mål (flere items, flere kodere)
 - ▶ Bedre mål (bedre items, træning af kodere)
- ▶ Mere i lektion 7

Statistisk validitet: Ni trusler

5. Intervalbegrænsning

- ▶ Ofte undersøger vi kun forskelle inden for et begrænset interval
 - ▶ Både i forhold til stimuli (uafhængig variabel) og udfald (afhængig variabel)
- ▶ Undgå at begrænse antallet af intervaller
- ▶ Gulv og loft effekter
- ▶ Løsning:
 - ▶ Flere doser af stimuli (i bedste fald “fuld dosis” versus “ingen dosis”)
 - ▶ Item response theory analysis
 - ▶ Arbejd på afhængige variable der indfanger al relevant variation

Statistisk validitet: Ni trusler

6. Manglende overholdelse af stimultildeling

- ▶ Implementeringen af stimuli er afgørende
 - ▶ Er der forskelle i sandsynligheden for at få stimuli?
 - ▶ Er der selvseleksion?
- ▶ Får personer samme stimuli?
- ▶ Overholder personerne deres tildeling af stimuli?
 - ▶ Se lektion 3 slides

Statistisk validitet: Ni trusler

7. Kontekststøj

- ▶ Konteksten for et studie er fyldt med faktorer, der kan påvirke relationen mellem stimuli og udfald
- ▶ Der er tilfældigheder, der kan påvirke den statistiske korrelation mellem to variable
- ▶ Løsning:
 - ▶ Jo mere kontrol over konteksten, desto bedre
 - ▶ Mål forhold ved konteksten

Statistisk validitet: Ni trusler

8. Enhedsheterogenitet

- ▶ Jo mere enheder varierer inden for et stimuli på udfaldsvariablen, desto større vil standardafvigelsen være på den variabel
- ▶ Dette vil gøre det vanskeligere at estimere en systematisk korrelation mellem stimuli og udfald
 - ▶ Især hvis folk reagerer forskelligt på et stimuli
- ▶ Løsning:
 - ▶ Reducer enhedsheterogenitet i designstadiet
 - ▶ Mål og brug relevante kovariate
 - ▶ Interaktionsmodeller

Statistisk validitet: Ni trusler

9. Unøjagtig kalkulation af effektstørrelse

- ▶ Når effektstørrelsen er kalkuleret dårligt, får vi upræcise resultater
- ▶ Problemer med ekstreme observationer
- ▶ Karakteren af udfaldsvariablen (odds ratios for binære variable)
- ▶ Løsning: Brug den korrekte statistiske test
 - ▶ Med de rigtige forudsætninger (jvf. punkt 2)

Intern validitet

- ▶ Validiteten af en inferens omkring hvorvidt et stimuli har en kausal effekt på en udfaldsvariabel
- ▶ Fokus på kausalitet, ikke blot en korrelation
 - ▶ Årsag og virkning
- ▶ Jo stærkere intern validitet, desto mere tiltro til at der er tale om en kausal relation
- ▶ Forskellige forhold, der kan problematisere den interne validitet

Intern validitet: Ni trusler

1. Uklar tidsmæssig rækkefølge

- ▶ Hvilken variabel ændrede sig først?
- ▶ Hvad er årsag og hvad er effekt?
- ▶ Der kan være tale om reciprokke effekter
- ▶ Det er ikke altid klart, om en ændring i stimuli kommer før ændringen i udfaldsvariablen

Intern validitet: Ni trusler

2. Selektionseffekt

- ▶ Systematiske forskelle mellem stimuligrupper
- ▶ Forskellige sandsynligheder for at modtage et stimuli
- ▶ Ikke muligt at afgøre om det er stimuli eller de systematiske forskelle, der fører til ændringer i udfaldsvariablen

Intern validitet: Ni trusler

3. Historieeffekt

- ▶ Alle begivenheder der foregår mellem et stimuli og målingen af et udfald kan have bestemte effekter
- ▶ Begivenheder, der foregår samtidig med stimultildelingen, kan drive en effekt
- ▶ Mindre problematisk i laboratorieeksperimenter, hvor man kan udelukke ændringer i bestemte forhold

Intern validitet: Ni trusler

4. Modningseffekter

- ▶ Naturlige hændelser over tid
- ▶ Mange ting finder sted
 - ▶ Ældre, sulten, klogere, stærkere, erfaren
- ▶ Ikke et problem hvis de ikke er systematisk relateret til stimuli eller udfaldsvariablen
- ▶ Undersøg om bestemte forhold betinger effekterne

Intern validitet: Ni trusler

5. Regressionsartefakter

- ▶ Ofte vælges forsøgspersoner på baggrund af lave eller høje værdier på en variabel
- ▶ Jo mere ekstrem værdi på én måling, desto større er sandsynligheden for at der vil være en meget lavere værdi på næste måling
- ▶ *Regression toward the mean*
- ▶ Tommelfingerregel: Vær opmærksom på dette når observationer er valgt på baggrund af bestemte værdier (eller kan selektere sig ind på baggrund af bestemte værdier), der afviger substantielt fra gennemsnittet

Intern validitet: Ni trusler

6. Nedslidningseffekt

- ▶ Ikke alle personer i et studie vil være med hele vejen
- ▶ Nogle personer er mere tilbøjelige til at være med
- ▶ Kan være relateret til stimuli
 - ▶ Selektionsbias *efter* stimultildeling
- ▶ Tab af respondenter kan være systematisk korreleret med stimuli og udfaldsvariablen
 - ▶ Især hvis stimuli kan motivere deltagerne til at deltage i studiet

Intern validitet: Ni trusler

7. Testeffekt

- ▶ Det at blive udsat for en test kan påvirke senere målinger
- ▶ En test vil påvirke resultaterne på en senere test
- ▶ Eksempel: At veje nogen kan få dem til at tabe sig
- ▶ Løsning: Item response theory

Intern validitet: Ni trusler

8. Instrumentering

- ▶ Den måde hvorpå vi måler noget, eller virkningen heraf, kan ændre sig over tid
- ▶ Kodere kan blive mere erfarne over tid
 - ▶ Mere præcise kodninger
 - ▶ Kan være systematisk korreleret med stimuli

Intern validitet: Ni trusler

9. Additive effekter og interaktionseffekter

- ▶ Trusler til den interne validitet opererer ikke uafhængigt af hinanden
 - ▶ Det at der er én trussel, gør ikke, at vi *kun* skal adressere denne
- ▶ En trussel til den interne validitet kan tilføjes til andre trusler til den interne validitet
- ▶ Effekten af én trussel kan være betinget af omfanget af andre trusler til den interne validitet

Relationen mellem intern og statistisk validitet

- ▶ Intern og statistisk validitet er tæt relateret
- ▶ Statistisk validitet er interesseret i fejl i den statistiske korrelation
- ▶ Intern validitet er interesseret i fejl i vores kausale inferens
- ▶ Selv ved høj statistisk validitet, ikke garanti for kausalitet
- ▶ Med andre ord: korrelation (*statistisk validitet*) er ikke det samme som kausalitet (*intern validitet*)

Konstruktionsvaliditet

- ▶ Kausal generalisering som representation
 - ▶ Operationalisering af det koncept, der studeres
- ▶ Validiteten i den inferens vi drager mellem vores operationalisering og de teoretiske koncepter, de repræsenterer
- ▶ Processen fra konkrete enheder, stimuli og observationer til de koncepter, de repræsenterer

Konstruktionsvaliditet: 14 trusler

1. Mangelfuld beskrivelse af koncept

- ▶ Hvad er vores koncept, der skal operationaliseres?
- ▶ Mangelfuld beskrivelse gør det vanskeligt at generalisere fra operationalisering til koncept
- ▶ Problemer:
 - ▶ For generelle koncepter
 - ▶ For specifikke koncepter
 - ▶ Forkerte operationaliseringer
 - ▶ Forkerte koncepter

Konstruktionsvaliditet: 14 trusler

2. Operationaliseringsuafhængighed

- ▶ Operationaliseringer kan være relateret til andre fænomener
 - ▶ En operationalisering kan korreilere med en anden
- ▶ Eksempel: Operationaliseringen af en arbejdsløs
- ▶ En operationalisering kan være relateret til andre operationaliseringer, der ikke er meningsfulde indikatorer for ens koncept

Konstruktionsvaliditet: 14 trusler

3. Operationaliseringshomogenitet

- ▶ Én operationalisering underrepræsenterer konceptet
 - ▶ Én operationalisering = mindre konstruktionsvaliditet
- ▶ Operationalisering kan også måle andre/irelevante koncepter
- ▶ Jo flere forskellige indikatorer, desto bedre
 - ▶ Skal indfange relevant variation, repræsentere koncept

Konstruktionsvaliditet: 14 trusler

4. Metodehomogenitet

- ▶ Vi har sjældent variation på den anvendte metode
- ▶ Når alle operationaliseringer anvender samme metode, er metoden en del af det koncept, man studerer
- ▶ Et stimuli præsenteres ofte på én bestemt måde
- ▶ Eksempel: selvrapporterede mål
 - ▶ Løsning: adfærdsmål

Konstruktionsvaliditet: 14 trusler

5. Intervalbegrænsning

- ▶ Ofte indfanger vi kun variation i vores operationalisering inden for et bestemt interval
 - ▶ Uklart hvor repræsentativt dette interval er for andre intervaller
- ▶ De indikatorer der bedst beskriver vores koncept, kan indfange et begrænset antal af niveauer
- ▶ Løsning: Forskellige niveauer ved forskellige stimuli

Konstruktionsvaliditet: 14 trusler

6. Stimulisensitiv faktorstruktur

- ▶ Strukturen af et mål kan ændres som et resultat af et stimuli
- ▶ Dem der har fået et stimuli, kan se mere kompliceret på et spørgsmål
- ▶ Forskellige indikatorers relation kan få en mere kompleks struktur

7. Reaktiv selvrapporterede ændringer

- ▶ Selvrapporteringer kan blive påvirket af stimultildelingen
 - ▶ Selv før stimuli tildeles (dem der gerne vil have stimuli, rapporterer større behov for at få stimuli)
- ▶ Motivationen *efter* stimuli kan stoppe for dem der har fået stimuli, men forsætte for dem, der ikke har fået det
- ▶ Løsninger:
 - ▶ Adfærdsmål
 - ▶ Incitamerter til at svare rigtigt

8. Opfattelse af studiesituation

- ▶ Forsøgsdeltagere påvirkes ikke alene af stimuli og mål, men også deres opfattelse af studiesituationen
 - ▶ Også (og måske især) hvis de ved, at de er en del af et eksperiment
 - ▶ Forsøgsdeltagere vil forsøge at "gætte eksperimentet"
- ▶ Løsninger:
 - ▶ Udfaldsvariablen skal ikke være åbenlys (mål udfaldsvariablen flere gange - og gerne senere i tid)
 - ▶ Undgå premålinger
 - ▶ Test hvor stor en effekt viden om eksperimentet har
 - ▶ Lyv om forsøgets formål
 - ▶ osv.

9. Forsøgsforventninger

- ▶ Forskeren (eller anden der driver et studie), kan påvirke deltagerens svar ved at give dem forventninger til ønskede/hensigtsmæssige svar
- ▶ Placeboeffekter, social desirability
- ▶ Løsninger for forskere der gennemfører forsøget:
 - ▶ Brug flere forskere
 - ▶ Observer forskeradfærd
 - ▶ Undgå at forskerne kender hypotesen
 - ▶ Reducer interaktion med forsøgsparticipanter
 - ▶ Brug kontrolgrupper til at undersøge om det er et problem

10. Forstyrrelseseffekt

- ▶ De svar der gives, påvirkes af rutiner og brud på disse
- ▶ Nyt stimuli (en innovation) kan medføre entusiasme, beundring m.v.
 - ▶ Tænk nudging
- ▶ Nyt stimuli kan også medføre irritation
- ▶ Hvis der aldrig har været foretaget ændringer, vil en ændring kunne forstyrre mere

11. Kompenserende udligning

- ▶ Når ikke alle får stimuli, kompenseres der nogle gange til dem, der ikke får det
- ▶ Vi måler ikke effekten af et stimuli, men effekten af et stimuli og en kompensation
- ▶ Dette kan udligne forskelle mellem grupperne, når begge får en service
- ▶ Udligning kan også finde sted ved at reducere et stimuli
- ▶ Især relevant i forhold til randomiserede studier, hvor markarbejdere skal stå for tildelingen af stimuli

12. Kompenserende rivalisering

- ▶ Social konkurrence
- ▶ Dem der ikke modtager et stimuli kan være motiveret til at vise, at de kan klare sig lige så godt, som dem der har modtaget stimuli
- ▶ Potentielle løsninger:
 - ▶ Kvalitative metoder (interviews)
 - ▶ Difference-in-difference

Konstruktionsvaliditet: 14 trusler

13. Uønsket stimultildeling

- ▶ Ikke alle forsøgspersoner får det ønskede stimuli
 - ▶ Mere negative svar
- ▶ Løsning:
 - ▶ Undgå at informere om stimultildelingsmekanismen
 - ▶ Undersøg frafaldsrater

Konstruktionsvaliditet: 14 trusler

14. Stimulidiffusion

- ▶ Forsøgsdeltagere kan modtage stimuli som de ikke er tildelt
- ▶ Især aktuelt når der er lille geografisk distance mellem enheder
- ▶ Tænk SUTVA
- ▶ Løsninger:
 - ▶ Undgå at dem der står for at implementere stimuli kender til kontrol
 - ▶ Undgå kontakt mellem enheder
 - ▶ Test for ligevægtseffekter

Ekstern validitet

- ▶ De fleste studier er foretaget i en specifik kontekst
- ▶ Hvor meget kan resultaterne fra ét studie generaliseres?
- ▶ Kausal generalisering som ekstrapolation
- ▶ Aspekter af ekstern validitet
 - ▶ Enheder
 - ▶ Stimuli
 - ▶ Udfald
 - ▶ Kontekst
- ▶ Ekstern validitet kan omhandle variationer i personer, kontekst, stimuli og udfald der er og ikke er en del af studiet

Ekstern validitet og forskellige typer af generaliseringer

- ▶ Fra smal til smal generalisering
 - ▶ Eksempel: Er resultaterne fra én kommune generaliserbare til nabokommunen?
- ▶ Smal til bred generalisering
 - ▶ Eksempel: Er resultaterne fra én kommune generaliserbare til alle kommuner?
- ▶ Fra bred til smal generalisering
 - ▶ Eksempel: Er resultaterne for alle kommuner generaliserbare til én bestemt kommune?
- ▶ Fra lignende/forskellige til lignende/forskellige
 - ▶ Eksempel: Er resultaterne fra en kommune generaliserbare til en region?
- ▶ Tilfældig udvælgelse
 - ▶ Eksempel: Er 1000 vælgere generaliserbare til alle vælgere?

Ekstern validitet: Fem trusler

1. Interaktion med enheder

- ▶ En effekt kan være betinget af, hvem man studerer
- ▶ Måske ikke samme effekt med andre enheder
- ▶ Talrige typer af forskelle, der kan betinge effekter
 - ▶ Alder
 - ▶ Land
 - ▶ Indkomst
 - ▶ Studerende

Ekstern validitet: Fem trusler, WEIRD

- ▶ “Most people are not WEIRD” (Henrich et al. 2010)
 - ▶ Western
 - ▶ Educated
 - ▶ Industrialized
 - ▶ Rich
 - ▶ Democratic
- ▶ “The sample of contemporary Western undergraduates that so overwhelms our database is not just an extraordinarily restricted sample of humanity; it is frequently a distinct outlier vis-a-vis other global samples. It may represent the worst population on which to base our understanding of Homo sapiens.” (Henrich et al. 2010, 82)

Dagens øvelse #1

- ▶ Diskuter:
 1. Hvilke spørgsmål kan vi undersøge med universitetsstuderende, hvor vi med rimelighed kan sige, at resultaterne er generaliserbare?
- ▶ To minutter
- ▶ Alene eller med sidemanden

Ekstern validitet: Fem trusler

2. Interaktion med stimulivariation

- ▶ En effekt kan være betinget af, hvilken udgave af et stimuli, man anvender
- ▶ Interaktion med andre stimuli
 - ▶ Eksempel: Mindre skoleklasser
 - ▶ Eksempel: Negative kampagner

Ekstern validitet: Fem trusler

3. Interaktion med udfald

- ▶ En effekt kan være betinget af, hvilken variabel man måler det på
- ▶ Måske forskellige effekter
 - ▶ Positiv effekt på en udfaldsvariabel
 - ▶ Ingen effekt på en anden
 - ▶ Negativ effekt på en tredje

Ekstern validitet: Fem trusler

4. Interaktion med kontekst

- ▶ En effekt kan være fundet i en kontekst

5. Kontekstafhængig mediator

- ▶ En mediator i en kontekst er ikke nødvendigvis en mediator i en anden kontekst (det samme gælder for moderatorer)

Ekstern validitet og konstruktionsvaliditet

- ▶ Ligheder
 - ▶ Begge vedrører generalisationer
 - ▶ Teoretisk viden omkring koncepter og indikatorer besvarer spørgsmål relateret til den eksterne validitet
- ▶ Forskelle
 - ▶ Det ene garanterer ikke det andet!
 - ▶ Inferens i forhold til forskellige spørgsmål
 - ▶ Konstruktionsvaliditet er i forhold til det koncept, man undersøger
 - ▶ Ekstern validitet er i forhold til personer, stimuli, udfald og kontekst
 - ▶ Ekstern validitet er direkte relateret til den interne validitet
 - ▶ Forskellige metoder til at styrke validiteten

Intern validitet og ekstern validitet

- ▶ Intern validitet er *sine qua non*
- ▶ “It makes no sense to say that some empirical research is low on internal validity but high on external validity.” (Morton and Williams 2010, 275)
- ▶ Intern og ekstern validitet vedrører begge teoretiske spørgsmål

Andre validitetsbegreber

- ▶ Økologisk validitet (ecological validity)
- ▶ Overfladevaliditet (face validity)
- ▶ Indholdsvaliditet (content validity)
- ▶ Kriterievaliditet (criterion-related validity)
- ▶ osv.

Dagens øvelse #2

► To opgaver:

1. Find en hypotese (gerne fra seminaropgaven)
2. Diskuter den eksterne validitet

Næste gang

- ▶ Sidste gang om kausale modeller
 - ▶ *Directed acyclic graphs* (DAGs)
- ▶ Mandag, samme tid og sted