

Kausal inferens III: Eksperimentelle designs

Metode 3

Link til slides: kortlink.dk/2acdu

Søren Damsbo-Svendsen
sdas@ifs.ku.dk

Institut for Statskundskab
Københavns Universitet

Uge 15

Sidste holdtime!

Uge	Holdtime	Emne	Øvelsesopgave
6		Kvantitativ indholdsanalyse	1
7		Diskursanalyse I	2
8	25-26 februar 2021	Diskursanalyse II	3
9	04-05 marts 2021	Interaktioner og modelspecifikation	4
10	11-12 marts 2021	Logistisk regression	5
11	18-19 marts 2021	Multilevel analyse	6
12		Kausal inferens I: Kausalitet og instrumentvariable (IV)	7
13		Påskeferie	
14	08-09 april 2021	Kausal inferens II: Paneldata	8
15	15-16 april 2021	Kausal inferens III: Eksperimentelle designs	9
16		Kausal inferens IIII: Regression Discontinuity (RD)	10
17		Social Data Science: Big Data	11
18		Process tracing	12
19		Kriterier for god videnskab	
20		Opsamling/spørgetime	
23		Aflevering af skriftlig hjemmeopgave (07 juni 2021)	

Vejledning

- Husk første vejledning på Zoom 16/4, 19/4 og 23/4
- Tilmeld jer i Google Sheets → <http://kortlink.dk/2b3zy>
 - de fleste har allerede gjort det → skriv til mig, hvis der ikke er flere tider
- Anden vejledning → samme format omkring uge 20

Recap fra sidste gang

- Kausalitet kræver, at X er **eksogent** → fravær af endogenitetsbias, herunder **omitted variable bias (OVB)**
- **Paneldata** er en smart datatype med *gentagne observationer af mange enheder* (fx survey to gange med samme responderter)
- **Fixed effects-modeller** (enheds-, tids- eller two-way) er en udvidelse af OLS, der udnytter paneldatastrukturen
 - kontrollerer "automatisk" for variation (heterogenitet) på enheds- og/eller tidsniveau
 - kan helt **eliminere OVB fra faktorer, der er tids- eller enhedskonstante** - både observerede *og uobserverede* (!)
 - fortolkning ændres en smule til effekten af X på Y **inden for enheder over tid (within unit-estimator)** eller forskelle i udviklingen over tid (*difference-in-differences-estimator*)
- **paneldata og fixed effects er ikke en perfekt løsning** → fortsat mulighed for omvendt kausalitet og OVB fra tids- og enheds-variante faktorer

Dagens program

1. Lidt indledning om eksperimenter
2. Potential outcomes, ATE og randomisering
3. Eksperiment vs. statistisk kontrol
4. Tre centrale antagelser bag eksperimenter
5. Analyse og øvelsesopgaver

Dagens formål

- At forstå *hvorfor* eksperimentet er den bedste løsning, når det kommer til at identificere **kausale effekter** (kausal inferens)
- At forstå fidusen og logikken i eksperimentets hjørnesten: **randomisering**
- At vide mere om nogle af de **faldgruber** (antigelser), man skal være opmærksom på, når man designer eksperimenter
- At kunne analysere og fortolke resultaterne af et eksperiment

Pensum

Gerber, A. S., & Green, D. P. (2012). Field experiments: Design, analysis, and interpretation. WW Norton.

- Kapitel 1: Introduction
- Kapitel 2: Causal Inference and Experimentation

→ Grundsten i moderne eksperimentel politologi og socialvidenskab (eksperiment- og kausal inferens-interesserede kan med fordel læse lidt videre)

Hvad er et eksperiment*?

*aka *randomized controlled trial* (RCT)

- Kaldes sommetider "guldstandarden" inden for kausal inferens, da vi ofte stræber mod, sammenligner med eller efterligner eksperimentet
- Kort fortalt er eksperimentet en metode, hvor man **tilfældigt fordeler nogle subjekter** i (mindst) to grupper
 - en treatmentgruppe og en kontrolgruppe - og tildeler subjekterne i treatmentgruppen et - I gættede det
 - **treatment** (stimulus)
 - Efter tildelingen af treatment måles et outcome (Y) i begge grupper
 - Pga. randomiseringen er grupperne statistisk set ens og treatment er den eneste forskel. **Derfor kan effekten af treatment estimeres som forskellen i de to gruppers outcome**
 - Vi kan mao. helt eliminere endogenitetsbias → stærk evidens for kausalitet

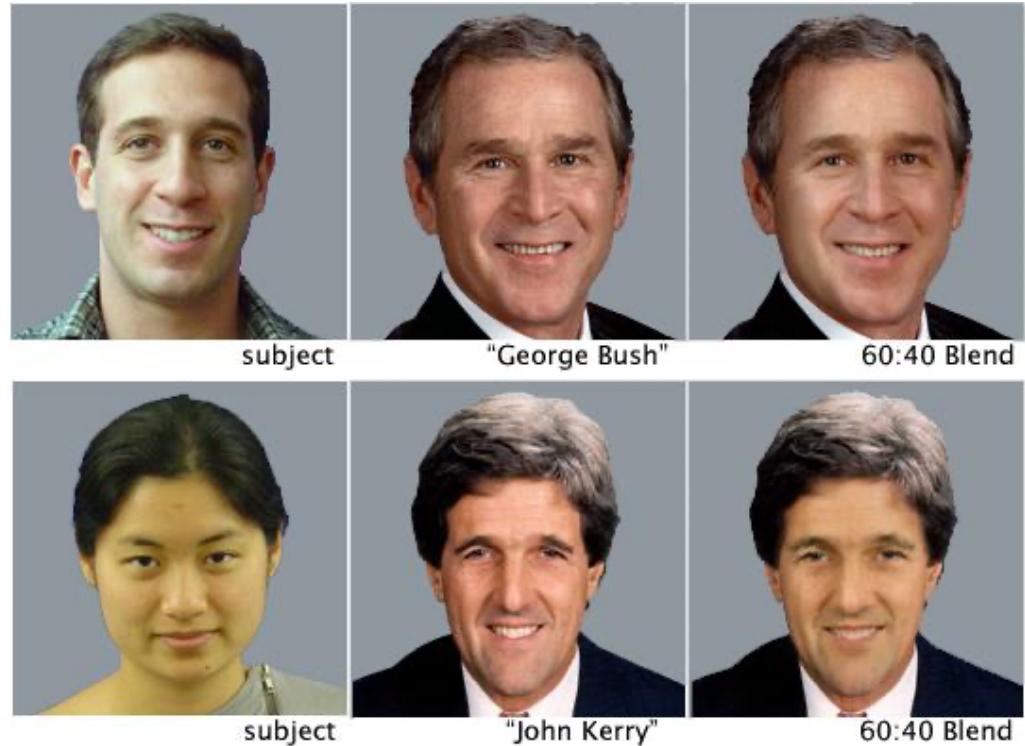
"*Random allocation is the dividing line that separates experimental from non-experimental research in the social sciences*" (Gerber & Green, 2012, p. 8)

Typer af eksperimenter

- Eksperimenter kan designes og "findes" i mange former, herunder
 1. Det klassiske laboratorieeksperiment
 2. Surveyeksperimenter
 3. Feltekspimenter (eksperimenter udført i "den virkelige verden")
 4. Naturlige eksperimenter (IV-regression, RDD mv.)
 5. Kvasiekspimenter (observationelle studier, hvor vi kan argumentere for, at X er eksogen eller så godt-som-tilfældigt fordelt)
- Der er fordele og ulemper ved alle → fx har laboratorieeksperimenter ofte størst **intern validitet**, fordi forskeren har fuldstændig kontrol over situationen
- Til gengæld kan flere af de andre have en højere grad af eksperimentel realisme eller **økologisk validitet**, fordi de undersøger den virkelig verden, som den faktisk udfolder sig
 - det gør at vores resultater lettere kan overføres/generaliseres til "virkeligheden"
 - afhænger af autenciteten af (1) treatment, (2) subjekterne, (3) konteksten og (4) outcome (se Gerber & Green, 2012, pp. 9-13)

Et eksempel på et politologisk eksperiment

1. "researchers obtained digital photographs of 172 registered voters"
2. "these same participants were asked to participate in an online survey [...] about the presidential candidates (President George W. Bush and Senator John Kerry)"
3. "The screens for these candidate questions included photographs of the two candidates displayed side by side. **Within this split-panel presentation, participants had their own face either morphed with Bush or Kerry** at a ratio of 60% of the candidate and 40% of themselves"
4. "weak partisans and independents [...] **moved in the direction of the more similar candidate**"



Kilde: Iyengar (2009) og Bailenson, Iyengar, and Yee (2009)

Potential outcomes

Hvad er potential outcomes og hvordan hjælper de os med at forstå kausale effekter?

- Potential outcomes er en måde at formalisere den kausale logik
- Forestil jer, at vi slår **plat eller krone**, om hvorvidt et subjekt, i , skal tildeles et treatment (D) eller ej:
 - $D_i = 1$, treatment tildeles
 - $D_i = 0$, treatment tildeles ikke
- Hvert subjekt i har **to potentielle udfald af outcome** Y_i :
 - subjekt i viser sit treatede potential outcome $Y_i(1)$, hvis vi tildeler det et treatment, $D_i = 1$
 - omvendt viser subjekt i sit untreatede potential outcome $Y_i(0)$, hvis vi *ikke* tildeler det et treatment, $D_i = 0$
- Kun ét af scenarierne kommer faktisk til at udspille sig. Vi vil aldrig kunne observere begge udfald for samme enhed

Potential outcomes

Hvad forsøger Gerber & Green at illustrere med disse to næsten ens tabeller? Hvad er forskellen?

TABLE 2.1

Illustration of potential outcomes for local budgets when village council heads are women or men. (Entries are shares of local budgets allocated to water sanitation.)

Village i	$Y(0)$ Budget share if village head is male	$Y(1)$ Budget share if village head is female	τ_i Treatment effect
Village 1	10	15	5
Village 2	15	15	0
Village 3	20	30	10
Village 4	20	15	-5
Village 5	10	20	10
Village 6	15	15	0
Village 7	15	30	15
Average	15	20	5

Syv landsbyer i to parallelle verdener
(Gerber & Green, 2012, p. 23)

TABLE 2.2

Illustration of observed outcomes for local budgets when two village councils are headed by women.

Village i	$Y(0)$ Budget share if village head is male	$Y(1)$ Budget share if village head is female	τ_i Treatment effect
Village 1	?	15	?
Village 2	15	?	?
Village 3	20	?	?
Village 4	20	?	?
Village 5	10	?	?
Village 6	15	?	?
Village 7	?	30	?
Estimated average based on observed data	16	22.5	6.5

Syv landsbyer i vores realiserede verden
(Gerber & Green, 2012, p. 31)

Average Treatment Effect (ATE)

Hvert subjekt har i teorien sin egen **treatmenteffekt** defineret ved:

$$\tau_i = Y_i(1) - Y_i(0) \rightarrow \text{dvs. forskellen på } Y_i \text{ med og uden treatment } D_i \text{ (to parallelle verdener)}$$

- Men er τ_i observerbar?
- Nej. Vi kan kun observere det realiserede (treatede eller untreatede) potential outcome for hvert subjekt - aldrig begge
- Ærgerligt?! For det vi er interesserede i at finde er gennemsnittet af alle subjekters τ_i også kaldet **den gennemsnitlige treatment effekt (ATE)**
- Men der findes en genial løsning på problemet ... nogen gæt?
- **Tilfældig tildeling af treatment (randomisering)!**

Tilfældig tildeling af treatment (I)

Hvordan gør *randomisering* os i stand til at estimere ATE, der, som sagt, aldrig kan observeres direkte?

- Svar: Randomiseringen skaber grupper, der i udgangspunktet (pre-treatment) er identiske
 - gælder alle observerbare såvel som uobserverbare karakteristika
 - betyder at grupperne i forventningen er ens → *statistisk set* ens
- Det betyder, at der ikke er nogen (systematisk) sammenhæng mellem **gruppernes karakteristika** (baggrundsvariable) og deres **treatmentstatus**
 - i kausalmodellen er der altså ingen pile, der peger ind på X
- I økonometri-lingo er den **uafhængige variabel (X)** uafhængig af fejlledet, ϵ_i . Man varierer ϵ_i kun rent tilfældigt ift. X_i i modellen $Y_i = \alpha + \beta_1 X_i + \epsilon_i$
 - det betyder, at den uafhængige variabel per design er **eksogen** → vi har elimineret **endogenitetsbias**

Tilfældig tildeling af treatment (II)

- Da grupperne er statistisk set ens, er deres potentielle outcomes også identiske
- Derfor udgør kontrolgruppen *per design* et validt kontrafaktum til treatmentgruppen → vi kan sammenligne dem
- Forskellen i gruppernes gennemsnit vil udtrykke den gennemsnitlige τ_i (ATE) - netop hvad vi søger!
- Denne *difference-in-means*-estimator er en **unbiased estimator af ATE** (som vi søger), når vores uafhængige variabel er tilfældigt tildelt
 - unbiased: vi kan godt ramme forbi den sande ATE med vores konkrete undersøgelse, men hvis vi gentog den 100 gange, ville vi ramme rigtigt
- **Tilfældig tildeling af X er fundamentet for designbaseret inferens, og det er den første og vigtigste antagelse for eksperimentet** (de to andre kommer lige straks)

Tilfældig tildeling af treatment (III)

Randomisering i praksis

- Sandsynligheden for at blive udtrukket til treatment- og kontrolgruppen behøver ikke at være 50/50. Det kunne også være f.eks. 20/80

Simpel vs. komplet randomisering

Vi antager en 20/80-fordeling af treatment- og kontrolgruppen:

- **Komplet randomisering** sørger for, at der nøjagtigt er 20 i treatmentgruppen og 80 i kontrolgruppen
- **Simpel randomisering** betyder blot, at der er 20 % sandsynlighed for at hvert subjekt placeres i treatmentgruppen, men reelt kan det (tilfældigvis) ende med at være 12, 17 eller 25
- Se **konkret procedure** (komplet randomisering) i Frederik Hjorths slides eller Gerber & Green (2012)
- Sommetider tager vi en mere praktisk tilgang, fx ved at tildele treatment pba. lige/ulige fødselsdatoer i surveyeksperimenter

Modelbaseret inferens

Kunne vi have opnået samme valide sammenligningsgrundlag - dvs. to statistisk set ens grupper - vha. statistisk kontrol?

- Hvad ville det kræve?
- At vi kontrollerer for samtliga potentielt relevante variable
 - at alle relevante variable er observerbare
 - at vi faktisk *har* observeret (målt) dem → vi har data
 - at der ikke findes yderligere relevante variable, som vi pt. ikke engang har fantasi til at forestille os
- Meget **stærk antagelse** → og vi kan aldrig være sikre!
- Netop **derfor er eksperimentet genialt**: Randomiseringen eliminerer alle systematiske observerbare **og uobserverbare** forskelle mellem treatmentgruppen og kontrolgruppen. Det eliminerer *alle* confounders af X → Y
- **Tilfældig tildeling af X giver det stærkest mulige grundlag for kausal inferens** og er derfor altid at foretrække frem for *modelbaseret inferens*, når det kommer til kausale spørgsmål

De tre centrale antagelser* bag eksperimenter

**nødvendige antagelser for at et eksperiment estimerer gennemsnitseffekten (ATE) på en unbiased måde*

1. Tilfældig tildeling af treatment (randomisering)

Antagelse 2: Ekskluderbarhed

Treatmenttildelingen påvirker kun Y gennem selve treatment

- Truet når der opstår forskelle mellem treatment- og kontrolgruppen, **udo over selve treatment**, som kan påvirke outcome, fx hvis:
 - **treatment er "bundled"**, dvs. indeholder flere X'er. F.eks. at få et brev, som opfordrer en til at stemme.
Her modtager treatmentgruppen både et brev og informationen i brevet
 - **subjekter er bevidste** om deres treatmentstatus og kan handle på den
 - **outcome måles forskelligt** i de to grupper
- Antagelsen om ekskluderbarhed er grundten til, at vi eksempelvis:
 - sørger for at subjekter og forskere ikke kender treatmentstatus → (double) **blinding**
 - tildeler kontrolgruppen et **placebo** (fx kalktabletter eller en "neutral" besked)
 - generelt undersøger alle subjekter under nogenlunde **samme betingelser** (fx tidspunkt, vejrs)
- Vi kan **visualisere** problemet i en kausalmmodel...

Antagelse 3: Non-interferens

Ingen subjekter påvirkes af andre subjekters treatmentstatus (ingen spill-over mellem subjekter)

- Truet når der er spill-over mellem subjekter, fx hvis:
 - Subjekter kender andre subjekters **treatmentstatus**
 - Treatment kan "smitte" fra treatede til untreatede (og subjekterne er tidsligt og rumligt i nærheden af hinanden)
- Endnu en god grund til **blinding** (og evt. at adskille subjekter i rum/tid)
- Også dette kan vi **visualisere** i en kausalmodel...

Jamen, hvordan analyserer man så et eksperiment?

- Med designbaseret (kausal) inferens ligger alt det hårde arbejde i *designet*, så analysen er nem!
- **Effekten** af et treatment estimeres simpelthen som **forskellen i gennemsnittet af outcome mellem treatment- og kontrolgruppen** (*difference-in-means*)
- Men vi vil også gerne vide noget om **præcisionen eller usikkerheden** i estimatet
 - Vi kunne for eksempel lave en **t-test** ("ttest outcome, by(treatmentgruppe)"), men et langt mere fleksibelt værktøj er ... 🎉

OLS!



reg outcome i.treatmentgruppe

- β_1 er den estimerede effekt (ATE) og p-værdien viser, om den er statistisk signifikant
- Vi ser det om lidt i øvelsesopgaven

Øvelsesopgave 9

Øvelsesopgave 9: Baggrund

Dansk surveyeksperiment af holdning til invandring/multikulturalisme som funktion af priming af (anti-)racisme og politisk korrekthed

Stimuli (*splitku5*) er tre tilfældigt tildelte vignetter:

Vignette 1 [*kontrolgruppen*]: "Danmark bliver i disse år et mere etnisk og religiøst mangfoldigt samfund. Der er delte meninger om, hvorvidt det udgør et problem eller en mulighed"

Vignette 2 [*treatment 1*]: "Danmark bliver i disse år et mere etnisk og religiøst mangfoldigt samfund. Der er delte meninger om, hvorvidt det udgør et problem eller en mulighed. **Nogle siger, at der er "politisk korrekt" at kalde mangfoldighed en mulighed**"

Vignette 3 [*treatment 2*]: "Danmark bliver i disse år et mere etnisk og religiøst mangfoldigt samfund. Der er delte meninger om, hvorvidt det udgør et problem eller en mulighed. **Nogle siger, at det er racistisk at kalde mangfoldighed et problem**"

Outcome (*ku5*) er holdning til mangfoldighed:

- "Hvad med dig? Betragter du mangfoldighed mest som **et problem eller mest som en mulighed?**" [1-5]

Øvelsesopgave 9: Break-out

Jeg har tilføjet lidt flere spørgsmål til **øvelsesopgave 9.1 og 9.2** (tre spørgsmål til hver, de står på de næste slides, man kan godt svare uden at have kigget på øvelsesopgaven hjemmefra)

1. Vi gennemgår spørgsmålene
2. Diskutér i grupper → **break-out** i ca. 10 min.
3. Bagefter samler vi op og tager opgave 9.3 (analyse) sammen
4. Der kommer en udførlig **do-fil** på Absalon med svar og Stata-kode til det hele

Øvelsesopgave 9.1: Antagelser

1. Hvad er den konkrete og den konceptuelle forskel på de tre vignetter (se forrige slide)?
2. Hvad er **treatment**?
3. Hvad er idéen med at give kontrolgruppen "neutral" information frem for blot at undlade at give dem en vignette?
 - tænk på antagelsen om ekskluderbarhed (at randomiseringen *kun* virker gennem treatment)

Øvelsesopgave 9.2: Randomisering og balancetest

1. Det er ikke muligt at undersøge statistisk, om randomiseringen - grundstenen i et eksperiment - er *lykkedes*. Hvorfor?
2. Men vi kan lave en **balancetest**, dvs. en statistisk test af, om grupperne empirisk set er nogenlunde ens (balancede) på en række baggrundsvariable, fx køn (*gender*), alder (*age*), uddannelse (*profile_education*) og pre-treatment holdning (*ku1*)
3. Hvis vi lavede en **regression af treatment på de fire baggrundsvariable** (obs.: her er treatment den *afhængige* variabel), og vi fandt en **statistisk signifikant sammenhæng mellem alder og treatment**,
 - a. hvad ville dette så fortælle os om sammensætningen af vores treatment- vs. kontrolgrupper?
 - b. ville vi kunne vurdere, om randomiseringen var succesfuld?

Øvelsesopgave 9.3: Analyse

- Effekten af hvert treatment er simpelthen **forskellen i gennemsnittene** mellem hver treatmentgruppe og kontrolgruppen (*difference-in-means*)
- Hvordan finder vi frem til effekten af at blive treated med vignette 2?
- Effektestimatet for **vignette 2** er $3.050 - 2.920 \approx 0.13$
- Effektestimatet for **vignette 3** er $2.984 - 2.920 \approx 0.06$
- Så nemt var det at finde vores estimer af treatmenteffekterne. **Men hvad mangler der?**
 - Vi vil meget gerne også have et estimat af **usikkerheden**, så vi kan sige, om effekterne er **statistisk signifikante** → løsning: **OLS**

```
. mean ku5, over(splitku5)
```

		Mean estimation		Number of obs = 991	
		Mean	Std. Err.	[95% Conf. Interval]	
c.ku5@splitku5					
	1	2.920489	.0750185	2.773276	3.067703
	2	3.049853	.072306	2.907963	3.191744
	3	2.98452	.0734382	2.840408	3.128633

Øvelsesopgave 9.3: Analyse

OLS

reg ku5 i.splitku5

- Hvordan ser **estimaterne** ud ift. vores egne beregninger?
- Er effekterne **statistisk signifikante**?

. reg ku5 i.splitku5

Source	SS	df	MS	Number of obs	=	991
Model	2.79440498	2	1.39720249	F(2, 988)	=	0.78
Residual	1767.00782	988	1.78846945	Prob > F	=	0.4581
Total	1769.80222	990	1.78767901	R-squared	=	0.0016

ku5	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
splitku5					
2	.1293641	.103509	1.25	0.212	-.0737586 .3324867
3	.0640308	.1049113	0.61	0.542	-.1418438 .2699054
_cons	2.920489	.0739549	39.49	0.000	2.775363 3.065616

En sidste pointe om **kontrolvariable i eksperimenter**. Man kan godt medtage en eller flere **kontrolvariable** i sin regression, selvom det er et eksperiment. Dette kan øge **præcisionen** af estimaterne (mindre standardfejl) uden at tilføre (eller mindske) bias - især **hvis grupperne er ubalancerede** (uens trods den tilfældige fordeling). Man skal dog for det første **undgå at kontrollere for faktorer, der selv kan være påvirket af treatment** ($\rightarrow post-treatment bias$). For det andet skal man ikke bruge det til **p-hacking**. Ideelt set planlægger man på forhånd, hvilke kontrolvariable man evt. vil inddrage, og uanset hvad **fremlægger man også den "rene" analyse** uden kontrolvariable (se eksempel i do-fil)

Spørgsmål?

- Øvelsesopgaven
- Eksperimenter
- Kausal inferens
- De andre metoder
- Eksamens
- Andet

Dagens pointer

- Eksperimenter er med god grund guldstandarden inden for kausal inferens
- Når vi bestemmer, hvem der får treatment, ud fra et tilfældighedsprincip, afskærer vi alle confounders fra at påvirke X og eliminerer dermed både **omitted variable bias** og **omvendt kausalitet**
- Vi er derfor så sikre, man kan være, på, at **vi har isoleret kausaleffekten $X \rightarrow Y$** eller den gennemsnitlige τ_i (Average Treatment Effect, ATE)
- **Eksperimenter er designbaseret inferens** - en tilgang til kausal inferens, der er langt overlegen ift. modelbaseret inferens (kontrolvariabel-tankegangen)
 - der findes mange **typer af eksperimenter** med fordele og ulemper, men de deler alle den samme logik
- **Tre vigtige antagelser** for eksperimentet: (1) Tilfældig tildeling af treatment, (2) ekskluderbarhed, (3) non-interferens
- **Analysen er nem** → typisk laver vi blot **OLS** med outcome som Y og treatmentgruppe som (kategorisk) X

Næste gang

- Nej desværre 
- Jeg skriver til jer på **Absalon** omkring **anden vejledning og spørgetime**
- God fornøjelse med resten af kurset og god arbejdslyst!

Tak for denne gang (ses til vejledning) ❤

