

POTENTIAL OUTCOMES

Denis Cohen

Mannheim Centre for European Social Research (MZES)

✉ denis.cohen@mzes.uni-mannheim.de

🌐 denis-cohen.github.io

🐦 [@denis_cohen](https://twitter.com/denis_cohen)

Social Science Data Lab

MZES, University of Mannheim


September 10, 2019



Logistics

 Workshop materials:

github.com/socialsciencedatalab/potential-outcomes

 Live stream: zoom.us/j/386654642

General information

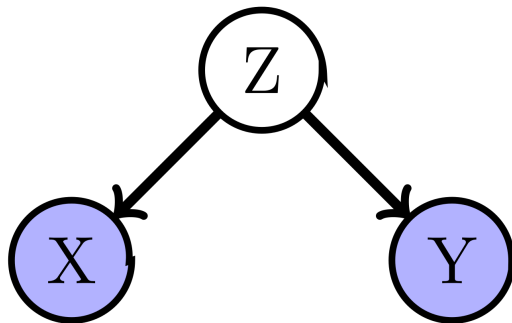
- Introduction to Potential Outcomes (60 min)
- Simulation Exercise (30 min)
- Prep course for Florian Foos “Randomized Experiments and Randomization Inference”, Sept 16, 15:30-17:00

Introduction

USING CORRELATIONS/ASSOCIATIONS FOR CAUSAL INFERENCE

- we nearly always find notable correlations/associations among variables in (observational) data
- correlation/association indicates that there *may* be an effect that *may* have a causal interpretation
- exact identification hinges on the assumption of **no unobserved heterogeneity** (aka exogeneity, no confounding bias)

A pretreatment variable that is associated with both the treatment and the outcome variables is called a **confounder** and is a source of **confounding bias** in the estimation of the treatment effect.



- We can try to make this assumption more credible, for instance by
 - subsetting the sample to rather homogeneous subgroups
 - using covariate adjustment to reduce unit heterogeneity
 - matching
 - randomization
- Ultimately, causal identification requires that all observed and unobserved confounders are accounted for – or ‘eliminated’ by design
- This assumption remains **untestable** – no matter the research design!
- Causal inference thus requires making untestable assumption
- **Causal identification**: Development and application of an **identification strategy** that makes these assumptions as plausible and defensible as possible

Potential Outcomes

POTENTIAL OUTCOMES

- “What if?”: Potential outcomes is all about hypothetical **counterfactuals**
- More specifically: Comparing factual and the counterfactual states
- **Core assumption**: Every unit in the population has a potential outcome Y_i under each possible treatment condition.
- For a binary treatment D (received=1, not received=0), we speak of treatment and control
- Causal effects can only be identified in comparison of different treatment states at the **unit level**: $\tau_i = Y_i(D_i = 1) - Y_i(D_i = 0)$

POTENTIAL OUTCOMES: NOTATION

- $i = 1, \dots, N$: subscript to denote a unit
- $D_i \in \{0, 1\}$: hypothetical treatment status (here: binary)
- $d_i \in \{0, 1\}$: factual treatment status
- $z_i \in \{0, 1\}$: assignment to treatment/control group (may differ from d_i !)
- $(Y_i(0), Y_i(1))$: potential outcomes of unit i in its untreated and treated state (general: $Y_i(D_i)$)
- $Y_i(d_i)$: factual/observed outcome of unit i
- X_i : covariate(s) (potential confounder(s)) for unit i
- $E[Y_i(1)]$: expectation of outcome across all units under treatment
- $E[Y_i(1)|D_i = 0]$: expectation of outcome across all units under treatment conditional on being in the control group – a counterfactual conditional expectation

OUR FICTIONAL EXAMPLE FOR TODAY

Setup

- We have a sample of $N = 1000$ students. Our outcome of interest, Y_i , is knowledge about counterfactual causality measured by test scores in a quiz on causal inference.
- The treatment of interest, D_i , is attending a 90 minute talk titled “Introduction to the Potential Outcomes Framework”.
- Assignment of treatment status equals actual receipt of treatment, $z_i = d_i$.
- Students’ potential outcomes under the control, $Y_i(0)$, range between 20 and 80.
- Individual treatment effects, τ_i , range between 0 and 20.
- Potential outcomes under the treatment, $Y_i(1)$, range between 20 and 100.

POTENTIAL OUTCOMES, EXAMPLE

i	$Y_i(0)$	$Y_i(1)$	τ_i
1	42.1	49.1	7
2	33.3	46	12.7
3	60.7	71.4	10.8
\vdots	\vdots	\vdots	\vdots
998	61.1	73.6	12.5
999	54.8	61.8	7
1000	43.6	52.9	9.3
(mean)	50.5	60.6	10.1

Table 1: What if students took the class? What if they didn't? Potential Outcomes under treatment and control plus idiosyncratic treatment effects

THE FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE (FPCI)

- we never observe more than one potential outcome for a given unit at a time
- if $D_i \in \{0, 1\}$, we observe $Y_i = d_i Y_i(1) + (1 - d_i) Y_i(0)$
 - we observe $Y_i(1)$ if $d_i = 1$
 - we observe $Y_i(0)$ if $d_i = 0$
- if $D_i \in \{0, 1, \dots, K\}$, we observe $Y_i = \sum_{k=1}^K 1\{d_i = k\} Y_i(k)$
i.e., we observe $Y_i(k)$ if and only if $d_i = k$

i	$Y_i(0)$	$Y_i(1)$	τ_i
1	42.1	?	?
2	?	46	?
3	?	71.4	?
\vdots	\vdots	\vdots	\vdots
998	61.1	?	?
999	?	61.8	?
1000	?	52.9	?
(mean*)	51.2	59.7	8.5

Table 2: The fundamental problem of causal inference illustrated.

* sample means based on observed values.

Remember: Idiosyncratic causal effects can only be identified in comparison of different treatment states at the **unit level**: $\tau_i = Y_i(1) - Y_i(0)$.

① Assume temporal stability and causal transience

- expose each unit i to control, measure $Y_i(0)$; then expose to treatment, measure $Y_i(1)$
- calculate $\tau_i = Y_i(1) - Y_i(0)$

② Assume unit homogeneity

- expose one unit to control, measure $Y_i(0)$; expose other unit to treatment, measure $Y_j(1)$
- calculate $\tau_{ij} = Y_j(1) - Y_i(0)$

③ Statistical Method

- shift the focus to **expected** causal effects
- observe average outcomes of units in treatment and control groups
- estimate $E[\tau] = E[Y_i(1)] - E[Y_i(0)]$ per $\overline{Y(1)} - \overline{Y(0)}$

AVERAGE TREATMENT EFFECT (MORGAN AND WINSHIP 2015, CH. 2)

- **Remember:** $Y_i(1)$ and $Y_i(0)$ are hypothetical random variables defined over all units for different treatment conditions
- ATE is the average in potential outcomes:

$$\begin{aligned}E[\tau_i] &= E[Y_i(1) - Y_i(0)] \\&= E[Y_i(1)] - E[Y_i(0)] \\&= \{\pi E[Y_i(1)|D_i = 1] + (1 - \pi)E[Y_i(1)|D_i = 0]\} \\&\quad - \{\pi E[Y_i(0)|D_i = 1] + (1 - \pi)E[Y_i(0)|D_i = 0]\}\end{aligned}$$

- Which of these can we estimate from observed data? Which are unobservable counterfactual quantities?

AVERAGE TREATMENT EFFECT (ATE)

What we want to know: $E[\tau_i] = E[Y_i(1)] - E[Y_i(0)]$

$$\begin{aligned} & \{ \underbrace{\pi E[Y_i(1)|D_i = 1]}_{\text{factual}} + (1 - \pi) \underbrace{E[Y_i(1)|D_i = 0]}_{\text{counterfactual}} \} \\ & - \{ \pi \underbrace{E[Y_i(0)|D_i = 1]}_{\text{counterfactual}} + (1 - \pi) \underbrace{E[Y_i(0)|D_i = 0]}_{\text{factual}} \} \end{aligned}$$

What we do know:

$$\hat{\tau} = \underbrace{E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0]}_{\text{naive estimator}}$$

ESTIMATING THE ATE

- The naive estimator: **difference-in-means**, $\hat{\tau} = \overline{Y(1)} - \overline{Y(0)}$
- When does this estimator yield an unbiased estimate? What assumption do we make?

$$\begin{aligned} & \underbrace{\{ \pi E[Y_i(1)|D_i = 1] + (1 - \pi) E[Y_i(1)|D_i = 0] \}}_{\text{as observed}} \\ & \quad \underbrace{\hspace{10em}}_{\text{assume} = E[Y_i(1)|D_i=1]} \\ & - \{ \pi \underbrace{E[Y_i(0)|D_i = 1]}_{\text{assume} = E[Y_i(0)|D_i=0]} + (1 - \pi) \underbrace{E[Y_i(0)|D_i = 0]}_{\text{as observed}} \} \end{aligned}$$

- If these assumptions hold:

$$\begin{aligned} & \underbrace{\{ \pi E[Y_i(1)|D_i = 1] + (1 - \pi) E[Y_i(1)|D_i = 0] \}}_{=E[Y_i(1)]} \\ & - \underbrace{\{ \pi E[Y_i(0)|D_i = 1] + (1 - \pi) E[Y_i(0)|D_i = 0] \}}_{=E[Y_i(0)]} \end{aligned}$$

- But when are these defensible assumptions to make?

Imagine you were randomly split into two groups. What would be your expectation regarding the distribution of...

- observable characteristics?
- unobservable characteristics?
- idiosyncracies that result in random or systematic measurement error?
- possible responses to an experimental treatment?

Thus, under randomization:

- $Y_i(0), Y_i(1) \perp D_i$ for all $i = 1, \dots, N$
- $X_i \perp D_i$ for all $i = 1, \dots, N$

$$\begin{aligned} E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0] &= \underbrace{E[\tau_i]}_{\text{True ATE}} \\ &+ \underbrace{\{E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]\}}_{\text{selection bias}} \\ &+ \underbrace{(1 - \pi)\{E[\tau_i|D_i = 1] - E[\tau_i|D_i = 0]\}}_{\text{differential treatment effect bias}} \end{aligned}$$

...so?

$$E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0] = E[\tau_i]$$

only holds if:

- 1 no selection bias:

$$E[Y_i(0)|D_i = 1] = E[Y_i(0)|D_i = 0]$$

- 2 no differential treatment effects

$$E[\tau_i|D_i = 1] = E[\tau_i|D_i = 0]$$

$$E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] = E[Y_i(1)|D_i = 0] - E[Y_i(0)|D_i = 0]$$

$$E[Y_i(1)|D_i = 1] - E[Y_i(1)|D_i = 0] = E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]$$

TWO CORE ASSUMPTIONS ABOUT POTENTIAL OUTCOMES

1 Excludability (Exclusion Restriction)

- Potential outcomes only respond to d , the actual receipt of the treatment
- Over and beyond d , they are unaffected by assignment to the treatment group, z

2 Non-Interference (SUTVA)

- Potential outcomes are unaffected by the pattern of actual or assigned treatments of other units

Summary

- The Fundamental Problem of Causal Inference does not make causal inference impossible. It makes causal inference without making untestable assumptions impossible.
- An identification strategy encompasses a clear statement of these assumptions and a research design that makes them as plausible and defensible as possible.
- The most defensible design is the experimental ideal: Randomized control trials.

Simulation Exercise

OUR FICTIONAL EXAMPLE FOR TODAY

Setup

- We have a sample of $N = 1000$ students. Our outcome of interest, Y_i , is knowledge about counterfactual causality measured by test scores in a quiz on causal inference.
- The treatment of interest, D_i , is attending a 90 minute talk titled “Introduction to the Potential Outcomes Framework”.
- Assignment of treatment status equals actual receipt of treatment, $z_i = d_i$.
- Students’ potential outcomes under the control, $Y_i(0)$, range between 20 and 80.
- Individual treatment effects, τ_i , range between 0 and 20.
- Potential outcomes under the treatment, $Y_i(1)$, range between 20 and 100.
- Students’ probability of selecting into the class is a direct function of their prior ability, i.e., of their potential outcomes under the control, $Y_i(0)$.
- Similarly, prior ability affects how much they learn (i.e., the size of their idiosyncratic treatment effects, τ_i) – students with a higher prior ability tend get more out of the class.

Two Scenarios

- 1 **Randomization:** Pure chance determines whether students attend the talk or not
- 2 **Self-Selection:** Students self-select into the class, based on prior ability

We want to quantify the magnitude of selection bias and differential treatment effect bias under randomization and under self-selection.

- 1 Before you start, set a seed for reproducibility of your random variable generation (`set.seed()`).
- 2 Generate an integer, `N <- 1000L`, to be used in determining the length of the variables we create below.
- 3 Generate a variable of potential outcomes under the control, `Y_0`, containing random draws from a uniform distribution with support `[20, 80]` (`runif()`).
- 4 Generate a variable of individual treatment effects, `tau`. As `tau` is a function of the potential outcomes under the control, we have differential treatment effects. We draw these from a normal distribution with some random error (`tau <- 0.2 * rnorm(Y_0, 2.5)`) and then truncate the distribution at 0 and 20.
- 5 Show a scatter plot with `Y_0` on the x-axis and `tau` on the y-axis. Briefly interpret the pattern you observe.

- 6 Generate a variable of potential outcomes under the treatment, Y_1 , using the two variables you previously generated.
- 7 Generate a randomly assigned binary treatment indicator, D , which takes on values of either 0 or 1 from a Bernoulli distribution, which is a special case of the binomial distribution with $n = 1$, meaning we take $n = 1$ draw for each observation i (`rbinom`). Suppose everyone has equal probability of being assigned to either treatment or control group.
- 8 Show that the potential outcomes are independent of treatment status. Choose an appropriate test or statistic and give a brief interpretation.
- 9 Generate a variable Y_{obs} that takes the values of Y_0 if $D=0$ and the values of Y_1 if $D=1$.
- 10 Run a regression of Y_{obs} on D to gauge the average treatment effect. Give a brief interpretation.

- 11 Next, generate a variable `D_sel` that indicates receipt of the treatment in a scenario of self selection. For this, generate a variable `prob_sel` equal to the potential outcomes under control divided by 100. This gives you each student's probability of selecting into the treatment. Use these probabilities to draw `D_sel` from a Bernoulli distribution.
- 12 Using the same test you chose above, test if the potential outcomes are independent of `D_sel` and give a brief interpretation.
- 13 Generate a variable `Y_obs_sel` that takes the values of `Y_0` if `D_sel==0` and the values of `Y_1` if `D_sel==1`.
- 14 Run a regression of `Y_obs_sel` on `D_sel` to gauge the average treatment effect under selection bias.
- 15 Specify the magnitude of both selection bias and differential treatment effect bias. Explain how the two differ conceptually and interpret the estimates.

References

- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105 (490): 493–505.
- Angrist, Joshua D, and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- Gerber, Alan S, and Donald P Green. 2012. *Field Experiments. Design, Analysis, and Interpretation*. New York, London: W.W. Norton & Company.
- Imai, Kosuke. 2017. *Quantitative Social Science. An Introduction*. Princeton, NJ: Princeton University Press.
- Morgan, Stephen L, and Christopher Winship. 2015. *Counterfactuals and causal inference*. 2nd. Cambridge: Cambridge University Press.