PSY 503: Foundations of Psychological Methods

Lecture 4: Advanced Topics in Causality, Potential Outcomes, and Experimental Design

Robin Gomila

Princeton

September 9 & 14, 2020

Notation

- *Indexing experimental individuals/units:* the subscript i refers to unit 1 to N
- *Defining treatment assignment:* The variable z_i indicates whether the ith individual is assigned to receive the treatment
- *Defining treatment:* The variable d_i indicates whether the ith subject is treated
- $z_i = 1$ means the ith subject was assigned to receive the treatment
- $z_i = 0$ means the ith subject was not assigned to receive the treatment
- $d_i = 1$ means the ith subject receives the treatment
- $d_i = 0$ means the ith subject does not receive the treatment

Potential outcomes

- Regardless of which treatment an individual receives, all individuals have a potential response in the event that treatment is or is not received
- Potential outcomes are written $Y_i(d)$, where the argument d indexes the treatment

Potential outcomes

- $Y_i(1)$ is the potential outcome if the ith individual was treated
- $Y_i(0)$ is the potential outcome if the ith individual was not treated
- Potential outcomes are fixed attributes of each individual and represent the outcome that would be observed hypothetically if that individual were treated or untreated

Conditional potential outcomes

- Potential outcomes for a subset of subjects
- $Y_i(d)|X = x$ denotes potential outcomes when the condition $X = x$ holds for individual i

Conditional potential outcomes

- $Y_i(0)|d_i = 0$: untreated potential outcome for individuals who do not receive the treatment
- $Y_i(0)|d_i = 1$: untreated potential outcome for individuals who do receive the treatment
- $Y_i(1)|d_i = 0$: treated potential outcome for individuals who do not receive the treatment
- $Y_i(1)|d_i = 1$: treated potential outcome for individuals who do receive the treatment

Potential and observed outcomes

- The $Y_i(1)$s are observed for individuals who are treated, and the $Y_i(0)$s are observed for individuals who are not treated.
- For any individual, we observe either $Y_i(1)$ or $Y_i(0)$, never both at the same time
- We can express the connection between the observed outcome Y_i and the underlying potential outcomes through the "switching equation":

$$Y_i = Y_i(1)d_i + Y_i(0)(1 - d_i) \tag{1}$$

ATE

$$\text{ATE} = \mu_{Y(1)} - \mu_{Y(0)} \tag{2}$$

in which $\mu_{Y(1)}$ is the average of $Y_i(1)$ for all individuals and $\mu_{Y_i(0)}$ is the average of $Y_i(0)$ for all individuals.

Estimation of the ATE

In experimental studies, researchers estimate $\mu_{Y(1)}$ using the mean $\widehat{\mu}_{Y(1)}$ of observed $Y(1)$ and $\mu_{Y(0)}$ using the mean $\widehat{\mu}_{Y(0)}$ of observed $Y_i(0)$. We have:

$$\widehat{\mathrm{ATE}} = \widehat{\mu}_{Y_i(1)} - \widehat{\mu}_{Y_i(0)} \tag{3}$$

in which $\widehat{\mathrm{ATE}}$ is the estimated ATE, $\widehat{\mu}_{Y_i(1)}$ is the estimated $\mu_{Y_i(1)}$, and $\widehat{\mu}_{Y_i(0)}$ is the estimated $\mu_{Y_i(0)}$.

Experimental Assumptions and Threats to Causal Identification

Assumptions and statistical methods: generalities

- Statistical methods involve assumptions
- Problem:
  - Assumptions do not always hold
  - Quite easily violated
- Do everything you can to make sure that assumptions hold
- When they are violated (or you suspect they may be violated):
  - Use corrective methods
  - Discuss possible implications of violated assumptions in your articles

Precision of individual experiment

- Do experiments inevitably provide precise estimates of the ATE?
- An estimate from just one experiment is only a best guess about the true value of the ATE
- ATE is often too high or too low
- Our dataset is just one of many possible data sets that could have been created via random assignment. If we would redo the exact same random assignment procedure, different units would be allocated to treatment and control groups!
- So what is the point?

Experimental estimates of ATEs are unbiased

- What is bias?
- Estimates are **unbiased** if they yield the correct estimate of the ATE **in expectation** (i.e., on average)
  - The average estimated ATE across all possible random assignments is equal to the true ATE
- Assumptions: necessary conditions for experimental estimates of the ATE to be unbiased

Independence: A necessary condition

Treatment status is statistically independent of potential outcomes and background attributes (X):

$$D_i \perp\!\!\!\perp Y_i(0),\, Y_i(1),\, X$$

This means that knowing whether an individual is treated provides no information about the individual's potential outcomes or background attributes.

Random Assignment

Random assignment: intuition

- **In expectation:** proper randomization of participants into experimental conditions creates groups that are similar on every single dimension except for the treatment
- **In expectation:** Random assignment of individuals to different environments E_0 and E_1 creates subpopulations that have the exact same characteristics at the moment they enter these environments.
  - Same heart rate, amount of sleep, age, income, or level of stress, etc.

Random assignment: Implication

In expectation: any difference in means between the treatment and control groups is caused by the presence of the treatment

Illustration in R

- Randomly assigns individuals from a population of size $N = 500$ to one of two groups
- Repeat this many many times: 100,000 times
- Compare the average characteristics of the individuals that are assigned to each group

Let's open R Studio

...

Lessons from R simulations

- **In expectation: Groups are comparable!**
  - This demonstration is true for every possible characteristic of participants
  - The only difference between treatment and control conditions is the presence vs. absence of treatment (in expectation)

Assumption 1: Excludability (a.k.a. exclusion restriction)

- The *only* relevant causal agent is receipt of the treatment
- The exclusion restriction breaks down if:
  - Treatment assignment z_i sets in motion causes of Y_i other than the treatment d_i
  - Asymmetries in measurement between conditions
  - Noncompliance to the treatment

Treatment assignment brings in other causes

- Study causal effect of writing fiction on students' creativity.
- Treatment group: invitation to "enroll in a writing program that will increase their creativity"

Asymmetries in measurement

- Experimenter in charge of measuring the outcome of interest knows treatment status
- Participants know their treatment status and hypotheses

Noncompliance to the treatment

- Assumption that participants *comply* (or *adhere*) to their randomly assigned experimental condition
- Why would participants not comply?
- How can noncompliance introduce bias in estimates of the population ATE?
  - Invalidates treatment assignment
  - Participants self-select into or select out from their assigned condition
  - Participants who do not comply often have different potential outcomes schedules

Illustration?

Nice guards only comply to treatment...

What to do when noncompliance?

- Instrumental Variable (IV) analysis
- Second part of the semester

Assumption 2: Non-interference

- Often called SUTVA
  - Stable Unit Treatment Value Assumption
- Participants' potential outcomes not affected by the treatment assignment of other participants

Realistic?

Is the non-interference assumption realistic in the prison guards hypothetical study?

What to do when interference between units?

- Emerging literature in political science on estimating average treatment effects under general interference (Aronow & Samii, 2017)

Attrition: A common threat to causal inference:

- Attrition occurs anytime outcome data are missing for some participants
- Why would outcome data be missing?
- How can attrition introduce bias in estimates of the population ATE?
    - Attrition often invalidates treatment assignment procedures
    - Participants who have missing data often have different potential outcomes schedules

Illustration?

Harsh guards in the control condition have missing outcomes...

What to do when attrition?

- Several corrective methods exists
    - Inverse Probability Weighting (IPW)
    - Double Sampling and Bounds
- Next semester

Threats to generalization (or particularization)

- Individual vs. Average causal effects
- External validity

Individual vs. Average causal effects

- Average causal effects cannot be particularized to any single individual
  - Unless additional assumption: constant treatment effect
- Important consideration in clinical and health contexts

External validity

- A single study = one piece of evidence for the existence of an effect within constrains of:
    - Population
    - Treatments
    - Outcome measures
    - Settings
    - Period
- Be cautious

Illustration?

Policies imposing body-worn cameras for prison guards in cultures where violence against prisoners is socially desirable