POLI 30 D: Political Inquiry

Professor Umberto Mignozzetti

(Based on DSS Materials)

Lecture 06 | Estimating Causal Effects with Randomized Experiments

Before we start

Announcements:

- Quizzes and Participation:
  - On Canvas. Check the board for the attendance check password!
- How was your Lab last week?
  - I hope R is treating your well
  - And don't forget: we are here to help!
- Github page: https://github.com/umbertomig/POLI30Dpublic
- Piazza forum: https://piazza.com/ucsd/winter2023/17221

Before we start

Recap:

- We learned the definitions of Theory, Scientific Theory, and Hypotheses.
- Data, datasets, variables, and how to compute means.

Great job!

- Do you have any questions about these contents?

Plan for Today

- Causal Effects
- Treatment and Outcome Variables
- Individual Causal Effects
- Average Causal Effects
- Randomized Experiments
- Difference-in-Means Estimator

Causal Inference

Why Do We Analyze Data?

1. MEASURE: **To infer population characteristics via survey research**
   - what proportion of constituents support a particular policy?

2. PREDICT: **To make predictions**
   - who is the most likely candidate to win an upcoming election?

3. EXPLAIN: **To estimate the causal effect of a treatment on an outcome**
   - what is the effect of small classrooms on student performance?

- We will progress from simple to more complex methods
- We begin with EXPLAIN by learning how to estimate causal effects with randomized experiments
  - involves relatively simple math
- Then, we will learn how to MEASURE the characteristics of an entire population from a sample of survey respondents
  - visualizations, descriptive statistics, correlation
- Then, we will learn how to PREDICT outcome variables
  - simple linear regression
- Then, we will return to EXPLAIN and estimate causal effects with observational data
  - multiple linear regression

Why Do We Analyze Data?

1. MEASURE: **To infer population characteristics via survey research**
   - what proportion of constituents support a particular policy?

2. PREDICT: **To make predictions**
   - who is the most likely candidate to win an upcoming election?

3. EXPLAIN: **To estimate the causal effect of a treatment on an outcome**
   - what is the effect of small classrooms on student performance?

Does Social Pressure Affect Turnout?

(Based on Alan S. Gerber, Donald P. Green, and Christopher W. Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *American Political Science Review*, 102 (1): 33-48.)

- To answer, we will analyze data from a randomized experiment where registered voters in Michigan were randomly assigned to either
  - (a) receive a message designed to induce social pressure to vote, or
  - (b) receive nothing
- The message told registered voters that after the election their neighbors would be informed about whether they voted in the election or not

Dear Registered Voter: **WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?**... We're sending this mailing to you and your neighbors to publicize who does and does not vote. The chart shows the names of some of your neighbors, showing which have voted in the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not. **DO YOUR CIVIC DUTY–VOTE!**

MAPLE DR	Aug 2004	Nov 2004	Aug 2006
9995 JOSEPH JAMES SMITH	Voted	Voted	???
995 JENNIFER KAY SMITH	Didn't vote	Voted	???
9997 RICHARD B JACKSON	Didn't vote	Voted	???
9999 KATHY MARIE JACKSON	Didn't vote	Voted	???

The *voting* dataset

Unit of observation: registered voters

Description of variables:

variable	description
birth	year of birth of registered voter
message	whether registered voter received message: "yes", "no"
voted	whether registered voter voted: 1=voted, 0=didn't vote

Causal Effects

- Many of the most important research questions in politics involve estimating a causal effect:
  - Does foreign aid promote democratic government?
  - Do women promote different policies than men?
  - Do small classes improve student performance?
  - Does social pressure increase the probability of turning out to vote?

- **Causal effects** refer to the cause-and-effect connection between two variables
  - the **treatment variable (X)**: variable whose change may produce a change in the outcome variable
  - the **outcome variable (Y)**: variable that may change as a result of a change in the treatment variable
- The causal relationship we are interested in is:

$$X \rightarrow Y$$

- In the voting dataset we have three variables, *birth*, *message*, and *voted*, and we aim to answer the research question: "Does social pressure increase the probability of turning out to vote?"
- What is the treatment variable?
  - *message*: indicates whether register voter received the message inducing social pressure
- What is the outcome variable?
  - *voted*: indicates whether register voter voted
- The causal relationship we are interested in is:

$$\textit{message} \rightarrow \textit{voted}$$

Treatment Variables

- In this class, treatment variables will always be binary

$$X_i = \begin{cases} 1 \text{ if individual } i \text{ takes the treatment} \\ 0 \text{ if inidividual } i \text{ does not take the treatment} \end{cases}$$

- In the voting experiment, the treatment variable is:

$$message_i = \begin{cases} 1 \text{ if registered voter } i \\ \quad \text{received the message} \\ 0 \text{ if registered voter } i \\ \quad \text{did not receive the message} \end{cases}$$

- Based on whether the individual takes the treatment, we speak of two different conditions:
  - **treatment** is the condition with the treatment
    - $X_i = 1$
  - **control** is the condition without the treatment
    - $X_i = 0$

Outcome Variables

- We will see different types of outcome variables
  - binary
  - non-binary
- In the voting experiment, the outcome variable is:

$$voted_i = \begin{cases} 1 \text{ if registered voter } i \text{ voted} \\ 0 \text{ if registered voter } i \text{ didn't vote} \end{cases}$$

  - what type of variable do you think this is?

Individual Causal Effects

- The **causal effect of X on Y** is *the change in the outcome variable* caused by a change in the treatment variable

- Ideally, we would like to compare two potential outcomes:
  - outcome when the treatment is present: $Y_i(X_i = 1)$
  - outcome when the treatment is absent: $Y_i(X_i = 0)$

- If we could observe *both* potential outcomes for each individual i, the individual causal effect would be:

$$\triangle Y_i = Y_i(X_i = 1) - Y_i(X_i = 0)$$

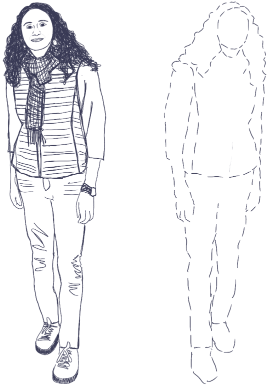
  - $\triangle Y_i$ represents the change in Y for individual i

- In the voting experiment, we aim to measure the extent to which the probability of voting changes as a result of receiving the social pressure message
- Ideally, for each registered voter we would like to observe:
    - whether they voted after receiving the social pressure message: $voted_i(message_i{=}1)$
    - whether they voter after NOT receiving the social pressure message: $voted_i(message_i{=}0)$
- If this were possible, we could measure the effect of receiving the social pressure message on the probability of voting as:

  $$\triangle voted_i = voted_i(message_i = 1) - voted_i(message_i = 0)$$

    - should be interpreted as an increase if positive, a decrease if negative, and as no effect if zero

- Do we ever observe both potential outcomes for the same individual at the exact same time under the same circumstances?
  - No, we only observe what happens in reality: **the factual outcome**
  - We can never observe what would have happened had we made different decisions: **the counterfactual outcome**



- Fundamental problem of causal inference: We can never observe the counterfactual outcome
- As a result, we cannot compute causal effects at the individual level

Average Causal Effects

- To get around the fundamental problem of causal inference, we must find good approximations for the counterfactual outcomes

- To accomplish this, we move away from individual-level effects and focus on the average causal effects across a group of individuals

- The **average causal effect** of the treatment X on the outcome Y (also known as the average treatment effect) is the average of all the individual causal effects of X on Y within a group
  - It is the average change in Y caused by a change in X for a group of individuals

- How can we obtain good approximations for the counterfactual outcomes?
    - We must find or create a situation in which the observations treated and the observations untreated are, at the aggregate level, similar with respect to all the variables that might affect the outcome other than the treatment variable itself
        - Then, we can use the factual outcome of one group as a proxy for the counterfactual outcome of the other
- The best way to accomplish this is by conducting a randomized experiment

Randomized Experiments

- A **randomized experiment** is a type of study design in which treatment assignment is randomized
  - researchers decide who takes the treatment based on a *random* process such as the flip of a coin
- Once treatment is administered, we can differentiate between:
  - the **treatment group**: individuals who received the treatment
  - the **control group**: individuals who did not receive the treatment
- In the voting experiment, what are the treatment and control groups?

Random treatment assignment makes the treatment and control groups *on average* identical to each other in all observed and unobserved pre-treatment characteristics

- When treatment assignment is randomized, the only thing that distinguishes the treatment group from the control group, besides the treatment itself, is chance
  - although the treatment and control groups consist of different individuals, the two groups are, as a whole, comparable to each other in terms of their pre-treatment characteristics (characteristics before treatment was administered)

- If the treatment and control groups are comparable before the treatment is administered
  - we can use the factual outcome of one group as a proxy for the counterfactual outcome of the other
  - we can estimate the average treatment effect by calculating the difference-in-means estimator

Difference-in-Means Estimator

$$\overline{Y}_{\text{treatment group}} - \overline{Y}_{\text{control group}}$$

$\overline{Y}_{\text{treatment group}}$: average outcome for the treatment group
$\overline{Y}_{\text{control group}}$: average outcome for the control group

- Only when the treatment and control groups are comparable does the diffs-in-means estimator produce a *valid* estimate of the average treatment effect
  - $\widehat{\text{average\_effect}} = \overline{Y_{\text{treatment group}}} - \overline{Y_{\text{control group}}}$
  - the "hat" on top of the name denotes that this is an estimate

- In the voting experiment, since treatment was randomly assigned, we can assume that the treatment and control groups are comparable and, thus, can estimate the average causal effect of receiving the message on the probability of voting by using the diffs-in-means estimator:

$$\overline{voted}_{\text{treatment group}} - \overline{voted}_{\text{control group}}$$

- $\overline{voted}_{\text{treatment group}}$: proportion of registered voters who voted among those who received the message
- $\overline{voted}_{\text{control group}}$: proportion of registered voters who voted among those who did not receive the message
    - why proportions and not averages?
        - because *voted* is binary so the average of *voted* should be interpreted as a proportion, not an average

How to Run an Experiment

(link to video)

Summary

- **Today's Class:**
  - Causal Effects
  - Treatment and Outcome Variables
  - Individual vs. Average Causal Effects
  - Randomized Experiments
  - Difference-in-Means Estimator
- Next class:
  - Hands on! We are going to analyze the voting experiment dataset!

Questions?

See you in the next class!