Conceptual Causality: IV + Matching

Section 3

Sima Biondi

Spring 2025

Gov 51: Data Analysis and Politics

Overview

1. Brainstorming
2. Recap: Instrumental variables (IV)
3. Matching

Housekeeping

- Problem Set I: out at 5 pm today, due at 11:59 pm the following Thursday
- CA office hours:
  - ↪ Pranav Moudgalya: Walk-in 7-9 on Mondays and 7:45-9pm on Thursdays (Leverett House Dining Hall)
  - ↪ Ben Heilbronn: Tues/Thurs 7:30pm-9:30pm @ Eliot Dining Hall; by appointment

Overview

Where do I start?

Two pathways to a research project

1. Have a question, find some data
2. Find some data, ask a question

Where do I start?

Two pathways to a research project

1. Have a question, find some data
2. Find some data, ask a question

In reality, we do both!

Where do I find data?

- R has data! `https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/00Index.html`
- Other sources:
  - `https://data.gov/`
  - `https://gov50.mattblackwell.org/assignments/final-project.html#finding-a-data-source`

10-minute brainstorm!

- Look for data
- What am I interested in exploring?
- *If you need a group*: post about it in the Slack channel

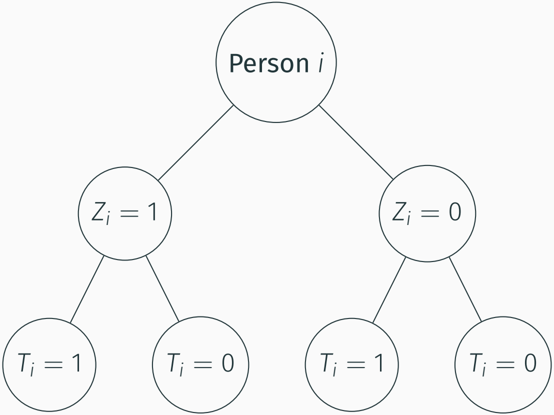
Overview

1 Brainstorming

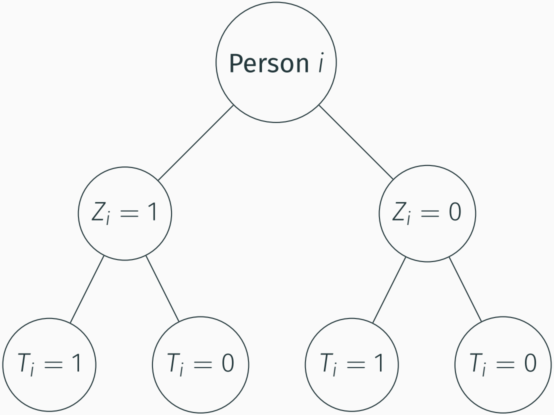
2 Recap: Instrumental variables (IV)

3 Matching

IV set up

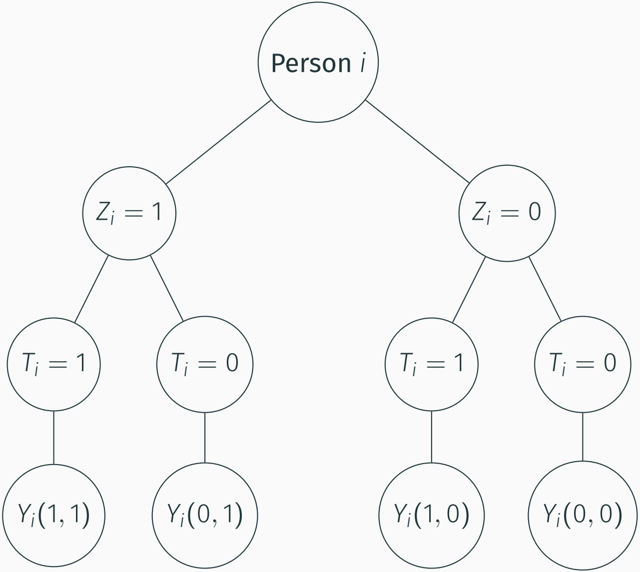


IV set up



Where are the potential outcomes?

IV set up



IV setup

	$T_i(Z_i = 0) = 1$	$T_i(Z_i = 0) = 0$
$T_i(Z_i = 1) = 1$	$Y_i(1,1) - Y_i(1,0) = 0$ *always-taker*	$Y_i(1,1) - Y_i(0,0) = *$ *complier*
$T_i(Z_i = 1) = 0$	$Y_i(0,1) - Y_i(1,0)$ *defier*	$Y_i(0,1) - Y_i(0,0) = 0$ *never-taker*

IV assumptions

- For an IV to be identified, it must:

IV assumptions

- For an IV to be identified, it must:
    1. Be assigned as-if random

IV assumptions

- For an IV to be identified, it must:
    1. Be assigned as-if random
    2. Affect treatment assignment

IV assumptions

- For an IV to be identified, it must:
  1. Be assigned as-if random
  2. Affect treatment assignment
  3. Only affect outcome through treatment (exclusion restriction)

IV assumptions

- For an IV to be identified, it must:
    1. Be assigned as-if random
    2. Affect treatment assignment
    3. Only affect outcome through treatment (exclusion restriction)
    4. Monotonicity: no defiers

IV assumptions

- For an IV to be identified, it must:
    1. Be assigned as-if random
    2. Affect treatment assignment
    3. Only affect outcome through treatment (exclusion restriction)
    4. Monotonicity: no defiers
- IF we meet these assumptions $\rightarrow$ consistent estimate of the local Average Treatment Effect (LATE)

What is encouragement?

$$\frac{\sum_{i=1}^{N} T_i Z_i}{\sum_{i=1}^{N} Z_i} - \frac{\sum_{i=1}^{N} T_i (1 - Z_i)}{\sum_{i=1}^{N} (1 - Z_i)}$$

$$\Rightarrow E(T_i \mid Z_i = 1) - E(T_i \mid Z_i = 0)$$

This formula compares the fraction of people who actually took the treatment (T_i = 1) in two groups:

1. Those who were encouraged to take it (Z_i = 1)
2. Those who were not encouraged (Z_i = 0)

Encouragement measures how much more likely people are to take the treatment if they were encouraged.

IV: ITT

What is the ITT?

$$\frac{\sum_{i=1}^{N} Y_i Z_i}{\sum_{i=1}^{N} Z_i} - \frac{\sum_{i=1}^{N} Y_i (1 - Z_i)}{\sum_{i=1}^{N} (1 - Z_i)}$$

$$\Rightarrow E(Y_i \mid Z_i = 1) - E(Y_i \mid Z_i = 0)$$

This formula compares the average outcome (Y_i) between two groups based on an instrumental variable (Z_i):

- People who were assigned Z_i = 1 (e.g., those who received some kind of encouragement or assignment to treatment).
- People who were assigned Z_i = 0 (e.g., those who did not receive encouragement or assignment).

It calculates the difference in average outcomes between these two groups (how much the outcome changes, on average, between the two levels of the instrumental variable)

IV: Wald estimator

$$\frac{\text{ITT}}{\text{Encouragement}}$$

$$\frac{E(Y_i \mid Z_i = 1) - E(Y_i \mid Z_i = 0)}{E(T_i \mid Z_i = 1) - E(T_i \mid Z_i = 0)}$$

By dividing ITT by encouragement, we isolate the causal effect of treatment for the group that actually complies with their assignment.

Overview

1. Brainstorming
2. Recap: Instrumental variables (IV)
3. Matching

Mixing and matching

What if there aren't any valid or strong instruments? What if the parallel trends assumption doesn't hold? How do we make valid causal estimates?

Mixing and matching

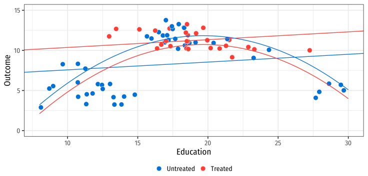
What if there aren't any valid or strong instruments? What if the parallel trends assumption doesn't hold? How do we make valid causal estimates? $\rightarrow$ one potential solution is matching!

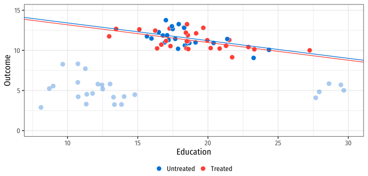
Mixing and matching

What if there aren't any valid or strong instruments? What if the parallel trends assumption doesn't hold? How do we make valid causal estimates? $\rightarrow$ one potential solution is matching!

What is matching?

- A group of methods that pair/group similar observations to calculate an average treatment effect for the treated (ATT)





Basics of matching

- Why match?

Basics of matching

- Why match? Another way to estimate the counterfactual
  - ⋆ Stratification: find strata within the data for which treatment assignment is uncorrelated with potential outcomes
  - ⋆ Weighting: re-weighting observations
  - ⋆ Pre-processing: reduces model dependence
- What estimand does matching produce?

Basics of matching

- Why match? Another way to estimate the counterfactual
  - ⋆ Stratification: find strata within the data for which treatment assignment is uncorrelated with potential outcomes
  - ⋆ Weighting: re-weighting observations
  - ⋆ Pre-processing: reduces model dependence
- What estimand does matching produce? Average Treatment Effect on the Treated (ATT)

Matching example

Building on last week's example: get out the vote

Next week

- Upcoming: problem set released soon
- Questions about IV or matching? Come to office hours!