Estimating Causal Effects with Randomized Experiments

POLS 602

September 23, 2025

Dr. Mike Burnham
Texas A&M Political Science

Causal Inference

Why is this important?

(Data analysis and causal inference)

- Distinguish between correlation and explanation
- Epistemic justification
- Ground explanation in intervention
- Prediction

Today

- Causal Effects
- Functions
- Treatment and outcome variables
- Individual causal effects
- Average causal effects
- Randomized experiments
- Difference-in-means estimator

Causal Effects

- Why do countries go to war?
- Why do insurgencies start?
- What increases voter turnout?
- Which education policies improve student outcomes?

Functions

A rule that relates one quantity to another

Treatment Variable

The variable whose change may produce a change in the outcome. The variable where the change originates

Outcome/Dependent Variable

The variable that may change in response to a change in the treatment variable

Notation For Causal Relationships

X :The treatment variable

Y: The Outcome Variable

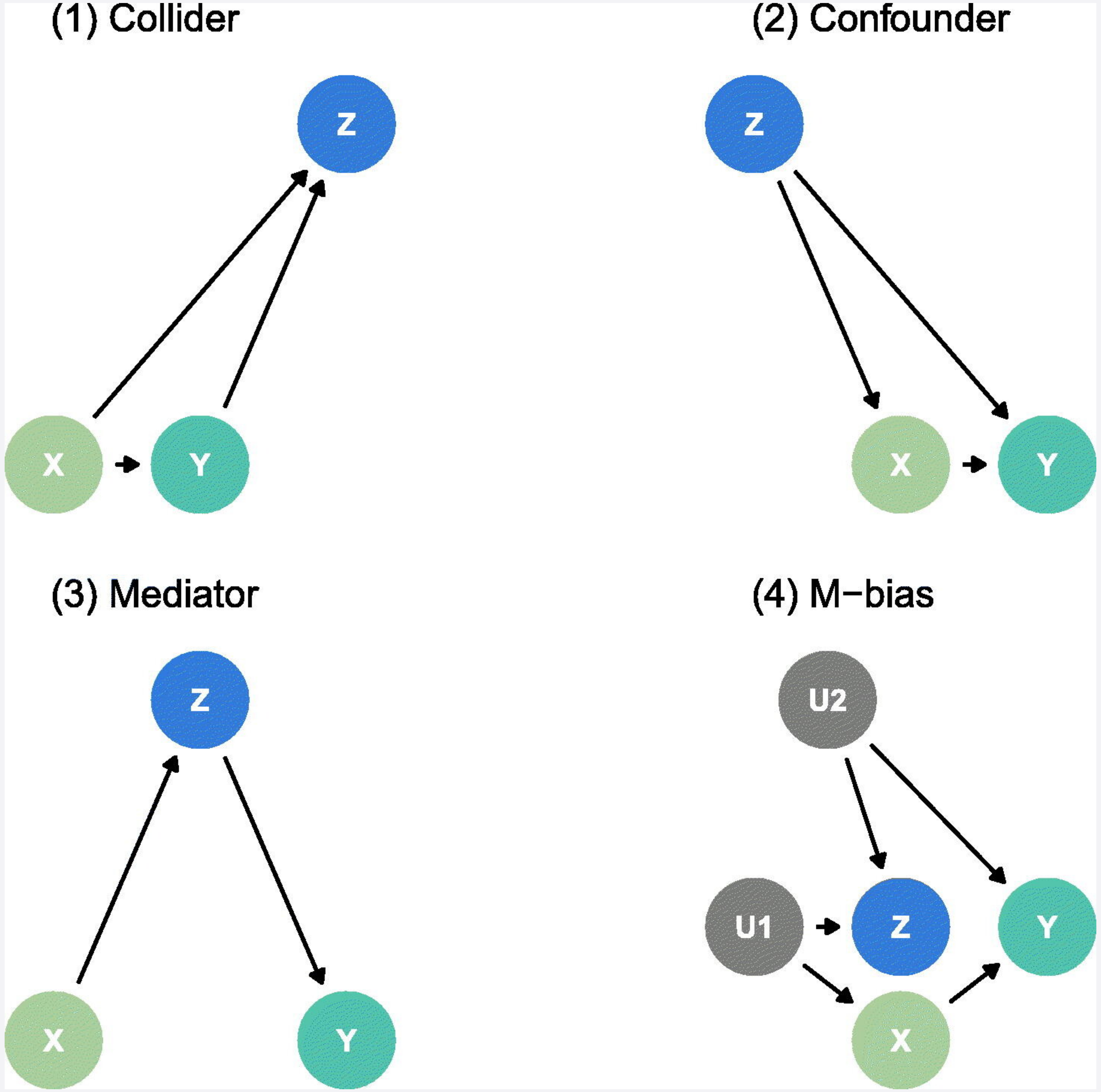
$X - > Y$

Examples:

$small\_class - > graduated$

$graduated - > future\_wages$

Directed Acyclical Graphs (DAGs)



Experiments

What is an experiment?

- Manipulating a treatment variable in a controlled condition
- An intervention intended to establish a causal relationship

Does social pressure affect voter turnout?

- Social Pressure and Voter Turnout: Evidence form a Large-Scale Field Experiment, Gerber et al. 2008, APSR

- Michigan voters were randomly assigned to:

  - (a) receive a message that induces social pressure to vote

  - (b) receive nothing

Does social pressure affect voter turnout?

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!

Dependent Variable

Y_i = 1 if voted, 0 if didn’t vote

Treatment Variables

$X_i = 1$ if treated, 0 if not treated

$X_i = 1$ if received message, 0 if received nothing

Individual Causal Effects

$$individual\_effects_i = \Delta Y_i = Y_i(X_i = 1) - Y_i(X_i = 0)$$

$Y_i(X_i = 1)$: The potential outcome for individual *i* under the treatment condition

$Y_i(X_i = 0)$: The potential outcome for individual *i* under the control condition

Δ: change

Individual Causal Effects

$$\Delta voted_i = voted_i(pressure_i = 1) - voted_i(pressure_i = 0)$$

$voted_i(pressure_i = 1)$: Whether individual *i* would vote after receiving the message

$voted_i(pressure_i = 0)$: Whether individual *i* would vote after *not* receiving the message

The Fundamental Problem of Causal Inference

- We never observe both the treatment and control outcomes for an individual
- We only observe the *factual* outcome, but never observe the *counterfactual* outcome

Average Causal Effects

- The *average causal effect* of X on Y is the average change in Y caused by a one-unit increase in X for a group of individuals
- Average of all the individual causal effects
- How do we calculate the average causal effect if we don't observe individual effects?

Estimating Average Causal Effects

- Divide our experiment into two groups
  - Treatment group
  - Control group
- If the treatment and control groups are comparable, we can approximate the average causal effect by comparing the average outcome in the treatment group to the average outcome in the control group.

Randomized Experiments

Randomized Experiments

- A randomized experiment is a research design in which the treatment assignment is determined through a random process.
- What do we call a study where the treatment is out of control of the researcher?
  - Observational studies

Random Treatment Assignment

- One of the most important concepts in science!
- Why?
  - Random treatment assignment makes the treatment and control groups identical, on average, in all observed and unobserved pre-treatment characteristics.
  - n must be large enough

Difference-In-Means Estimator

When the treatment and control groups are comparable, we can estimate the average causal effects with the difference-in-means estimator:

$$average\widehat{\_causa}l\_effect = \overline{Y}_{treatment\_group} - \overline{Y}_{control\_group}$$

Summary

$$individual\_effects_i = \Delta Y_i = Y_i(X_i = 1) - Y_i(X_i = 0)$$

$Y_i(X_i = 1)$: The potential outcome for individual *i* under the treatment condition

$Y_i(X_i = 0)$: The potential outcome for individual *i* under the control condition

Problem: We can never observe both

$$average\_effects = \overline{\Delta Y} = \overline{Y(X = 1)} - \overline{Y(X = 0)}$$

$\overline{Y(X = 1)}$: The potential outcome for individual *i* under the treatment condition

$\overline{Y(X = 0)}$: The potential outcome for individual *i* under the control condition

Problem: We still never observe the counterfactual

Summary

Through random assignment, we can make two groups comparable. When two groups are comparable, the factual outcome of one group is an unbiased estimate of the counterfactual outcome of the other. Thus:

$$\widehat{average\_causal\_effect} = \overline{Y}_{treatment\_group} - \overline{Y}_{control\_group}$$

The “hat” indicates that the average_causal_effect is a sample statistic, or an estimate.