Class 8:
Cause and effect

API-201
Quantitative Analysis and Empirical Methods I

Profs. Goel and Taylor
Harvard Kennedy School

Agenda

1. Simpson's paradox
2. A model of causality
3. Randomized experiments

A motivating example
Test preparation

Suppose students who elect to take a test preparation course ultimately end up doing better than those who do not.

What can we conclude about the efficacy of the course?

Motivating example

Graduate school admissions rates at Berkeley [1973]

	Applicants	Admitted
Men	8442	44%
Women	4321	35%

Can we conclude that women are admitted at lower rates *because of* their gender? What else might be going on?
[Discuss with your neighbor.]

Simpson's paradox

Trends across groups can disappear or even reverse when groups are disaggregated.

Difficult [impossible] to fully identify causal effects from *observational* data.

Experiments

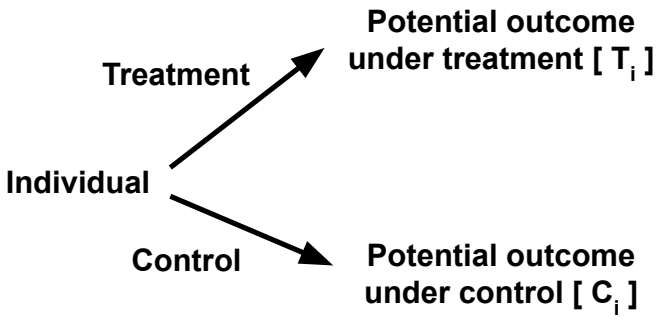
One way to rigorously identify causal effects is through **randomized controlled trials [RCTs]**.

We'll return to this idea shortly. But first we need to develop a theory of causality.

What is causality?

Neyman-Rubin Causal Model

Potential outcomes



Neyman-Rubin Causal Model

Unit-level causal effect

Unit-level causal effect is $T_i - C_i$

- T_i is the potential outcome under treatment. For example, how well a student *would have* done had they taken the test prep.
- C_i is the potential outcome under control. For example, how well a student *would have* done had they not taken the test prep.

Neyman-Rubin Causal Model

Fundamental problem of causal inference

It is *impossible* to observe both potential outcomes for a given individual. Cannot *directly* observe unit-level causal effects.
[Why not?]

But — amazingly — we can estimate the *average* causal effect across individuals.

Neyman-Rubin Causal Model

Average causal effect

$$\bar{T} - \bar{C} = \frac{1}{N} \sum_{i=1}^{N} T_i - C_i$$

Neyman-Rubin Causal Model

Average causal effect

Unit-level treatment effect can't be observed

$$\bar{T} - \bar{C} = \frac{1}{N} \sum_{i=1}^{N} T_i - C_i$$

$$= \frac{1}{N} \sum_{i=1}^{N} T_i - \frac{1}{N} \sum_{i=1}^{N} C_i$$

Average outcome under treatment can be estimated!

Average outcome under control can be estimated!

Neyman-Rubin Causal Model

Estimating the average causal effect

1. Pick a random set of subjects and treat them.
2. Average observed outcome for treatment group is a good estimate of the average treatment potential outcome.
3. Average observed outcome for control group is a good estimate of the average control potential outcome.
4. The difference in these estimates is a good estimate of the average causal effect.

Estimating the effect of test prep

To estimate the effect of test prep, we can run a **randomized controlled trial**:

1. Offer test prep to a **randomly** selected subset of students.
2. Look at the **difference** in average test performance between those students who were offered test prep and those who were not [difference-in-means estimator].

Interventions

Causal effects are inherently tied to *interventions*. We always ask, "what is the effect of this intervention on outcomes?"

The mantra of causal inference is "no causation without manipulation" [Holland, 1986]

Interventions

We can ask about the "**effect of a cause**" but not clear how to make statistical sense of the "**cause of an effect**."

- We can ask, "What is the effect of test prep on academic performance?"
- Harder to make sense of , "Why do some students have better academic performance?"

But policymakers and others are inclined to ask "why" questions, which are inherently hard / impossible to answer!

With your neighbor come up with "why" questions, and transform them into "intervention" questions.

Graduation admissions redux

What would it mean to say gender or race *causes* lower admissions rates? What is the intervention?
[Discuss with your neighbor.]

Implicit assumptions

Non-interference

The Neyman-Rubin causal model is simple and elegant. But it has several limitations.

A key assumption is **non-interference**: potential outcomes for a unit depend only on the treatment assignment for that unit. We assume no one's treatment assignment affects anyone else's potential outcomes.

Imagine conducting a randomized controlled trial to measure the effectiveness of a COVID vaccine. Is non-interference violated? [Discuss with your neighbor.]

Non-compliance

Assignment ≠ treatment

Not everyone who is randomly assigned to the treatment condition will ultimately undergo treatment [and not everyone who's assigned to control condition will comply].

Non-compliance

Assignment ≠ treatment

In our test prep example, *offering* students test prep is not the same as ensuring they *complete* test prep.

The difference in performance between those [randomly] offered test prep and those [randomly] not offered test prep is called the **intention-to-treat** estimate.

Come up with another example where compliance would likely be an issue. [Discuss with your neighbor.]

Key takeaways

The potential outcomes framework is a powerful way to conceptualize causal effects.

Key to this approach is thinking about the **intervention**. *What would have happened had one done something else?*
["No causation without manipulation."]

We can use **randomized controlled trials [RCTs]** to estimate average treatment effects. [Next time we'll talk about estimating causal effects with observational data.]