Thinking about Causality

ECON 490 (Spring 2024)

Email: tmackay@fullerton.edu || **Office Hours:** Mondays 5:45 to 6:45 PM in SMGH 3340

Overview

In these slides, we'll talk about:

- An intuitive definition of causal analysis
- Defining the key challenge of causal analysis

Setting the Stage

Article we read last week looked at callback rates for job applications

Intuitively, all kinds of things are likely correlated with getting job offers

- E.g., education, experience, race, etc.
- How do callbacks vary across these characteristics?

Use ***descriptive analysis*** to document patterns in data

- Calculate correlations between getting a callback and characteristics above

What Descriptive Analysis Can and Can't Tell Us

Descriptive analysis gives us correlations between characteristics and callbacks

- But lots of characteristics are likely correlated *with each other*
- E.g., job experience is naturally limited for previously incarcerated applicants

We can calculate the correlation between criminal records and callbacks…

- But how much of that correlation is really driven by limited experience?
- What about folks with limited experience but no incarceration history?

The goal of causal analysis is to ***isolate*** the effect of particular variables or factors

Establishing Causality

In causal analyses, we're typically interested in one ***treatment*** variable

- In the paper we read, the treatment was prior criminal record
- What's the effect of treatment on callback rates?

What if we just took a sample of real people's resumes and sent them out?

- Prior criminal records are likely correlated with other characteristics
- What would the average difference in callback rates tell us then?

Key feature of paper – treatment was ***randomized***

- Implies avg. differences can ***only*** be driven by differences in criminal records

Causality and Identification

Identification is a key concept in causal analysis (and economics more generally)

- "Does this analysis ***identify*** the effect of X on Y?"
- Intuitively, "Does this estimate tell us what we really want to know?"

In the callbacks example, we want to know how criminal records impact callbacks

- Given experiment setup, calculate average differences in callbacks
- Did we ***identify*** a causal effect?

The idea is that we want to rule out all other possible explanations

- If something like typos also differ systematically across applications, how do we know what's driving the differences we observe?

Causal Analysis – *An Example*

Suppose you did poorly on an exam and want to do better on the next test

- You decide to change your study routine
- 3 new strategies – flashcards, YouTube videos, and more practice problems

Good news – you take the next test, and your score goes up!

Think causally – which of the changes you made ***caused*** that improvement?

- In causal inference terms, can you ***identify*** the key factor?

Causal Analysis – *An Example*

Suppose the YouTube videos you watched weren't very informative

- This might let you rule out one potential explanation…
- But this still leaves both flashcards and practice problems

We can think about causal inference as an ***attribution*** problem

- We want to attribute the change in test score to particular factors
- Because both factors changed at the same time, this is hard to do!

In general, outcomes we care about are generally driven by all kinds of factors

Key Challenge of Causal Analysis

When doing causal analysis, we want to explain a portion of the variation in the outcome we're interested in as being **specifically** caused by some treatment

The key challenge – knowing if that variation is **only** driven by treatment

- The key concern of causal analysis – *what if something else is going on?*
- In regression terms, we'll refer to this as omitted variables bias (OVB)

In intuitive terms, good causal analysis requires ruling out all other explanations

Approaches to Causal Analysis

We've already seen one way of ruling out other explanations – ***randomization***

- The "gold standard" of causal inference
- Used in pharmaceutical trials, field experiments, etc.

What happens if we can't run experiments and assign treatment ourselves?

- Causal inference with observational data – finding (or creating) randomization
- One of the key contributions of economics to social science

A Final Note

We'll talk about causal analysis throughout this class

- That doesn't mean descriptive analysis isn't important!
- It's what you'll do for the capstone paper (and most data work in industry)

When we switch over to talking about regression and using R, we're generally be thinking about exploring ***descriptive*** questions