Counterfactuals, Clones, and Cool Regression Stuff

Econ 140, Section 2

Jonathan Old

Roadmap

1. Potential outcomes

2. Selection bias

3. How to get rid of selection bias

Any questions?

… Remember – Every question is useful!

Potential outcomes

Potential outcomes

- "Potential outcomes" is a framework that can help us **think through causal claims**: An alternative to math, drawing errors, or thinking things through
- We like to **write things down** in a rigorous way: Transparent, easy to verify, easy to replicate
- Potential outcomes are hypothetical outcomes
- Example: Your exam score when you go to all sections vs. when you go to no sections
- Think of potential outcomes as "parallel universes"

Our main challenge: We NEVER observe an individual at more than one status at the same time!
$\rightarrow$ "The **fundamental problem of unobservability**"!

Potential outcomes: Notation

We write the potential outcomes as:

$$\left.\begin{aligned} Y_{i0} &= \text{Outcome of individual } i \text{ with "status" } 0 \\ Y_{i1} &= \text{Outcome of individual } i \text{ with "status" } 1 \end{aligned}\right\} \text{Counterfactual Outcomes}$$

Alternative way of writing it: $Y_i(0)$ and $Y_i(1)$

"Status" can be anything

- Treatment assignment: 0 or 1
- Actual treatment: 0 or 1
- Drinking expensive whiskey or not
- Can also be: Multi-valued (number of children) or continuous (hours studied)

Potential outcomes: Notation

- We are often interested in the expected (think: average) potential outcome of a group of individuals with a given status.
- We write the group behind a conditional sign: $\mathrm{E}\left[\text{Score}_{i0} \mid \text{iPad}_i = 0\right]$ gives the potential outcome of a group of people that had no iPad, in the ”parallel universe” where they don’t have an iPad.
- Then, $\mathrm{E}\left[\text{Score}_{i1} \mid \text{iPad}_i = 0\right]$ gives the potental outcome of the same group (that currently have no iPad), in the ”parallel universe” where do have an iPad.

Treatment effects

We mostly use the potential outcomes framework to guide our thinking when we talk about estimating ”treatment effects”

- The individual treatment effect for agent i is: $Y_i(1) - Y_i(0)$
- The average treatment effect (ATE) is the expected value of the individual treatment effects: $ATE = E[Y_i(1) - Y_1(0)]$
- The Averae Treatment Effect on the Treated (ATT) is the same thing but conditioning on being treated: $ATT = E[Y_i(1) - Y_1(0)|T_i = 1]$
- Is the ATE always the same as the ATT?

Treatment effects

We mostly use the potential outcomes framework to guide our thinking when we talk about estimating ”treatment effects”

- The individual treatment effect for agent i is: $Y_i(1) - Y_i(0)$
- The average treatment effect (ATE) is the expected value of the individual treatment effects: $ATE = E[Y_i(1) - Y_1(0)]$
- The Averae Treatment Effect on the Treated (ATT) is the same thing but conditioning on being treated: $ATT = E[Y_i(1) - Y_1(0)|T_i = 1]$
- Is the ATE always the same as the ATT? **No!** It can be different from the ATE because the treated individuals need not be a random sample of the whole population.

Estimating the effect of iPads on grades

Let us start with a difference-in-means comparison:

$$\begin{aligned}
\Delta &= E[\text{Grade}_i|\text{iPad}_i = 1] - E[\text{Grade}_i|\text{iPad}_i = 0] \\
&\quad \text{Add and subtract } E[\text{Grade}_i(0)|\text{iPad}_i = 1]: \\
&= E[\text{Grade}_i(1)|\text{iPad}_i = 1] - E[\text{Grade}_i(0)|\text{iPad}_i = 1] \quad + \\
&\qquad E[\text{Grade}_i(0)|\text{iPad}_i = 1] - E[\text{Grade}_i(0)|\text{iPad}_i = 0] \\
&\quad \text{Use properties of expectations:} \\
&= E[\text{Grade}_i(1) - \text{Grade}_i(0)|\text{iPad}_i = 1] \quad + \\
&\qquad E[\text{Grade}_i(0)|\text{iPad}_i = 1] - E[\text{Grade}_i(0)|\text{iPad}_i = 0] \\
&= \text{ATT} + \text{Selection bias}
\end{aligned}$$

Selection bias: Students with and without iPad have different potential grades: **even if they both *had* iPads, they would be different**.

Dissecting Bad Causal Claims

Discuss in groups of 2: Why is this statement problematic?

Over the past 60 years, more spending on police hasn't meant less crime

The Washington Post

Intuitively, one might worry that reducing police spending would lead to a spike in crime. A review of spending on state and local police over the past 60 years, though, shows no correlation nationally between spending and crime rates.

In 1960, about $2 billion was spent by state and local governments on police. There were about 1,887 crimes per 100,000 Americans, including 161 violent crimes. By 1980, spending had increased to $14.6 billion — and crime rates had soared to 5,950 crimes per 100,000 Americans and 597 violent crimes. Over the next two decades, those rates thankfully fell, down to about 4,120 crimes per 100,000 people and 507 violent crimes. Spending spiked to more than $67 billion. Eighteen years later — by 2018, the most recent year for which full data are available — crime rates had fallen further to 2,580 crimes per 100,000, including 381 violent crimes.

Figure 1: Police spending and Crime (Source)

Dissecting Bad Causal Claims II

Discuss in groups of 2: Why is this statement problematic?

Three's a crowd: Having more than 2 kids linked to weaker brain function

NEW YORK – Everything in moderation – even kids? Researchers from Columbia University and Université Paris-Dauphine report having more than two kids may have a negative impact on late-life cognition. The study shows that older parents with just two children appeared sharper cognitively than those with three.

Figure 2: Number of Children and Cognitive Function(Source)

Dissecting Bad Causal Claims III

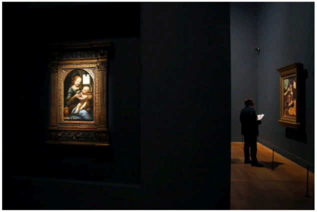
Discuss in groups of 2: Why is this statement problematic?

The New York Times

Another Benefit to Going to Museums? You May Live Longer

Researchers in Britain found that people who go to museums, the theater and the opera were less likely to die in the study period than those who didn't.

Give this article



Leonardo da Vinci's "The Virgin and Child" at the Louvre in Paris. A British study found that people who engaged with the arts had a lower likelihood of dying during the study period than those who did not. Thibault Camus/Associated Press

Figure 3: Museums and longevity (Source)

Selection bias

How to think about Selection Bias

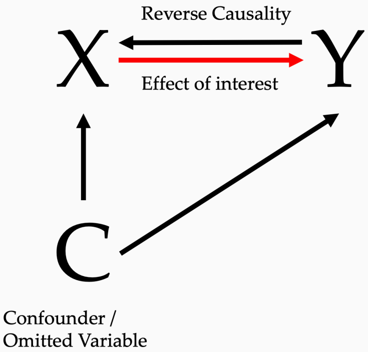


Figure 4: Selection bias

How to get rid of selection bias

RCTs solve selection bias

We had:

$$\Delta = E[\text{Grade}_i(1) - \text{Grade}_i(0)|\text{iPad}_i = 1] \quad + \\ E[\text{Grade}_i(0)|\text{iPad}_i = 1] - E[\text{Grade}_i(0)|\text{iPad}_i = 0]$$

- The second line was selection bias: The potential grade of individuals with and without iPad is different
- If the treatment (iPad) is independent of the potential outcomes, then:

$$\text{iPad}_i \perp (\text{Grade}_i(1), \text{Grade}_i(0))$$
$$\Rightarrow E[\text{Grade}_i(0)|\text{iPad}_i = 1] = E[\text{Grade}_i(0)|\text{iPad}_i = 0]$$

  and selection bias will be zero.
- The the difference is equal to the ATT and **also the ATE**:

$$\Delta = E[\text{Grade}_i(1) - \text{Grade}_i(0)|\text{iPad}_i = 1] \\ = E[\text{Grade}_i(1) - \text{Grade}_i(0)|\text{iPad}_i = 0]$$