From Counterfactuals to Linear Regression

Econ 140 Spring 2025, Section 2

Jonathan Old

Syllabus/OH bCourses Website Feedback form (**Always open**) RStudio

Roadmap

1. Time for your questions
2. Selection bias
3. Potential outcomes
4. RCTs
5. A non-mathematical intro to OLS – OR: standard errors, t-statistics, and hypothesis tests: What is that all about? Your choice

Your questions

Any questions?

...Remember – this is a safe space! Every question is useful!

Selection bias

How to think about Selection Bias

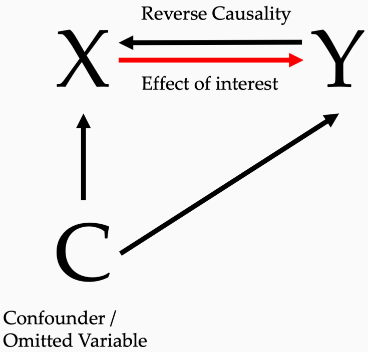
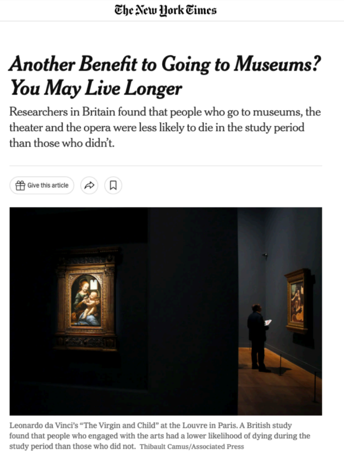


Figure 1: Selection bias

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 2: Museums and longevity (Source)

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!

Potential outcomes: Notation

We write the potential outcomes as:

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

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 they do have an iPad.

Estimating the effect of iPads on grades

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

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

Add and subtract $E[\text{Grade}_i(0)|\text{iPad}_i = 1]$:

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

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

Use properties of expectations:

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

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

$$=$$

Estimating the effect of iPads on grades

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

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

Add and subtract $E[\text{Grade}_i(0) | \text{iPad}_i = 1]$:

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

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

Use properties of expectations:

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

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

$$= \text{Treatment effect for group with iPad} +$$

Estimating the effect of iPads on grades

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

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

Add and subtract $E[\text{Grade}_i(0)|\text{iPad}_i = 1]$:

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

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

Use properties of expectations:

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

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

$$= \text{Treatment effect for group with iPad} + \text{Selection bias}$$

Estimating the effect of iPads on grades

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

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

Add and subtract $E[\text{Grade}_i(0)|\text{iPad}_i = 1]$:

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

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

Use properties of expectations:

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

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

$$= \text{Treatment effect for group with iPad} + \text{Selection bias}$$

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

Selection Bias revisited

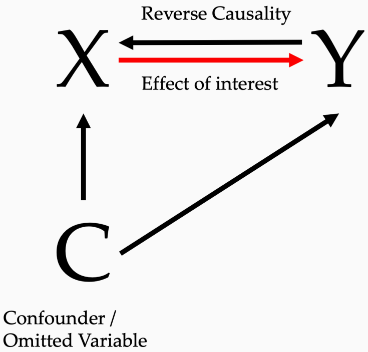


Figure 3: Selection bias

RCTs

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.

RCTs solve the selection bias problem by **randomly assigning treatment** to different groups.

Discuss in groups:

1. Is this the same thing as saying ”We drew a random sample from the population?”
2. What are some problems with RCTs? (ethical/practical/econometric)

RCTs have revolutionized economics

PLANET MONEY

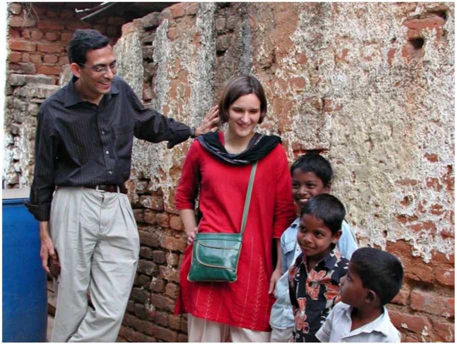
SUBSCRIBE

The Tuesday Podcast: Poor Economics

June 7, 2011 · 7:59 PM ET

26-Minute Listen

+ PLAYLIST



Abhijit Banerjee and Esther Duflo in Hyderabad, India.

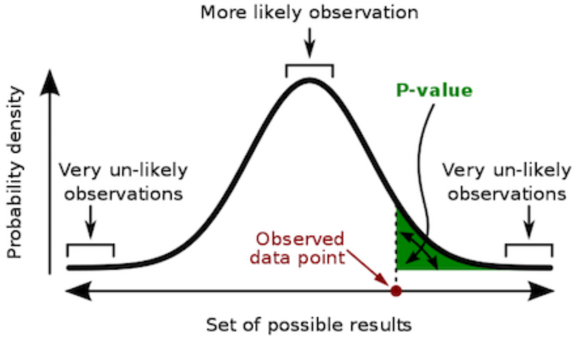
Figure 4: Abhijeet Banerjee and Ester Duflo. (Source)

OLS or hypothesis tests: Your choice

Hypothesis testing

- Random variables: Our estimator is a random variable (randomly drawn from population)
- Standard error: Random variables have a standard deviation, estimators have standard errors. This quantifies their uncertainty
- Statistics: The two keywords are the law of large numbers and the central limit theorem: The sum/mean over many draws from a random variable will be normally distributed
- For hypothesis test: null and alternative hypothesis.
- Assume the null hypothesis is true, then see how plausible results are if it is actually true.
- If they are implausible – we reject the null hypothesis! Otherwise: Fail to reject.

It's all connected



A **p-value** (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true.