GOV1368 Section 3

Mohit Karnani

Harvard

Fall 2024

Recap

Last time we covered:

- Statistical Models
- Linear Regression
- Causality

Today we will learn about the “gold standard” of causal inference: **Randomized Controlled Trials (RCTs)**.

Agenda

Estimands and Estimators

Potential Outcomes and Causal Inference

Randomized Controlled Trials

Application

Contents

Some Definitions

Recall that last week we used the linear regression model to *estimate* the relationship between the number of hours studied and the test score.

Our goal was to estimate the slope of a linear function. This slope β_1 was our *estimand*, i.e. it was the target population parameter that we wanted to approximate using data.

Our *estimator* of this slope was $\hat{\beta}_1$, which was the slope of the line that "best fit" the data. This was our best guess of the true slope β_1.

The *estimator* is a function of the data, and it is random. The *estimand* is a fixed, unknown parameter of the population. An *estimate* is a realization of the estimator.

We can study other parameters of interest, not just the slopes of linear regression models.

Properties of Estimators: Bias and Consistency

Two important properties of estimators are *bias* and *consistency*.

The *bias* of an estimator is the difference between the expected value of the estimator and the true value of the parameter being estimated. An estimator is *unbiased* if its expected value is equal to the true value of the parameter.

The *consistency* of an estimator is the property that the estimator converges to the true value of the parameter as the sample size increases. An estimator is *consistent* if it converges to the true value of the parameter as the sample size goes to infinity.

Estimators can be unbiased but not consistent, consistent but biased, both, or neither. For example, when targetting $\mathbb{E}[X_i]$, $\sum_{i=1}^{n} X_i/n$ is unbiased and consistent, X_1 is unbiased and inconsistent, $\sum_{i=1}^{n} X_i/(n-1)$ is biased and consistent, and X_n^2 is biased and inconsistent.

Contents

A Tale of Two Worlds

Suppose we are measuring the effects of a tutoring program on test scores. We have a group of students who receive the tutoring program, and a group of students who do not.

We can think of each student as having two *potential outcomes*: the test score they would have if they received the tutoring program, and the test score they would have if they did not receive the tutoring program. We can only observe one of these outcomes, not both.

Let's denote the test score of student i as Y_i, the treatment received by student i as D_i, and the potential outcomes as $Y_i(1)$ and $Y_i(0)$. The observed outcome is

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0).$$

The *observed outcome* is the test score that we actually observe. The *counterfactual outcome* is the test score that we would observe if the student received the other treatment.

Note: both $Y_i(1)$ and $Y_i(0)$ are random variables, but we can only observe one of them.

Counterfactuals

What we would like to see:

D_i	$Y_i(1)$	$Y_i(0)$	Y_i
1	11.9	6.6	11.9
1	10	8.5	10
1	9.7	9.4	9.7
1	9.5	7	9.5
1	11.4	7.4	11.4
0	9.6	7.6	7.6
0	9.1	7.1	7.1
0	10.4	7.7	7.7
0	10.4	8	8
0	12.4	7.8	7.8

What we actually see:

D_i	$Y_i(1)$	$Y_i(0)$	Y_i
1	11.9		11.9
1	10		10
1	9.7		9.7
1	9.5		9.5
1	11.4		11.4
0		7.6	7.6
0		7.1	7.1
0		7.7	7.7
0		8	8
0		7.8	7.8

Not observing both potential outcomes is the *fundamental problem of causal inference*.

Measuring the Causal Effect of a Treatment

Let's study a *causal* question: what is the effect of the tutoring program on test scores?

Specifically, say the estimand we want to target is the *average treatment effect* (ATE) of the tutoring program. This is the average difference in the potential outcomes:

$$ATE = \mathbb{E}[Y_i(1) - Y_i(0)].$$

(We could target other estimands, such as the individual treatment effects $Y_i(1) - Y_i(0)$)

If we could observe both potential outcomes, we would subtract them for each student and average these differences (sample mean approximates the population expectation)...

...but we can't observe both potential outcomes (causal inference would be too easy!)

Difference in Means and Selection Bias

What if we take the average difference in test scores between students who receive the tutoring program and students who do not?

(No need to follow the math)

$$
\begin{aligned}
\sum_{D_i=1} Y_i/n_T - \sum_{D_i=0} Y_i/n_C \rightarrow & \mathbb{E}[Y_i|D_i=1] - \mathbb{E}[Y_i|D_i=0] \\
= & \mathbb{E}[Y_i(1)|D_i=1] - \mathbb{E}[Y_i(0)|D_i=0] \\
= & \mathbb{E}[Y_i(1)|D_i=1] - \mathbb{E}[Y_i(0)|D_i=0] \pm \mathbb{E}[Y_i(0)|D_i=1] \\
= & \underbrace{\mathbb{E}[Y_i(1)-Y_i(0)|D_i=1]}_{ATET} + \underbrace{\mathbb{E}[Y_i(0)|D_i=1] - \mathbb{E}[Y_i(0)|D_i=0]}_{\text{Selection Bias}}
\end{aligned}
$$

We end up with *selection bias*: students that received tutoring might be fundamentally different from those who didn't (they can have different average potential outcomes).

Contents

Randomized Controlled Trials

One way of overcoming selection bias is to conduct a *randomized controlled trial* (RCT).

In an RCT, participants are *randomly assigned* into either a *treatment* or *control* group.

Thus, the value that D_i takes is determined by chance: there are no other factors that can affect who receives the tutoring intervention and who doesn't, i.e. the treatment assignment is *independent* of *all* other random variables, including potential outcomes.

Intuitively, students who are randomly assigned to the group that receives tutoring are *on average* equivalent to students who were randomly assigned to the group that does not.

Thus, any average differences observed in Y must be due to the intervention!

RCTs are extremely powerful tool, they are the "gold standard" of causal inference.

Contents

Application: Project STAR

Stata