Introduction

Professor Ji-Woong Chung
Korea University

Outline

Outline

Let's go over the syllabus

Outline

Goals of Econometrics/Statistics/Data analysis

Three types of questions we want to answer:

1. Description (how things are/were: statistical properties and relationships)
   - Has industry concentration increased over time?
   - How much is the cost of financial distress?
2. Prediction (guessing an unknown value, without interfering)
   - What will the GDP be next quarter?
   - What is the income of a person visiting your company's website?
3. Causality (how changing one variable would affect another, **all else equal**)
   - How do increases in disclosure requirement affect firms?
   - How do increases in the minimum wage affect employment?

What is Causality?

- Read the Wikepidia on Causality:
  `https://en.wikipedia.org/wiki/Causality`
- Hope you don't get lost. :-)
- Our narrow(?) definition of causality
  - **Ceteris paribus**, how does X affect Y?
  - I.e., if we know what would have happend to Y without X, then we could compare it with Y with X and find a causal effect.
- "Everything else equal"
  - Need to know "counterfactual" outcomes
  - Y without X vs. Y with X

The Fundamental Problem of Causal Inference (Holland, 1986)

Prediction vs. Causality

- Causal inference **predicts** a counterfactual outcome based on a specific choice.
- Causal inference uses the predicted counterfactual to estimate the causal effect of similar future choices.

- Causal inference is a missing data problem (ML comes to play)
  - In the counterfactual outcomes framework, causal inference and missing data are tightly linked.
  - Any causal answer uses assumptions to infer the missing counterfactual

This course will discuss several ways to solve these types of problems

On Prediction Accurate prediction may be possible without knowing anything about the underlying model/causal relationships in the data

- Indeed, this is the promise of many new **machine learning** methods
- From this perspective, prediction is a more descriptive task: estimate stable relationships between past and future data, as observed

But "theory-free" prediction can fall short when we need it most

- An important research topic: when predictions fail?

Causal inference tools are useful for avoiding such blind spots. Interested?

> *Applied Causal Inference Powered by ML and AI*, by Victor Chernozhukov, Christian Hansen, Nathan Kallus, Martin Spindler, Vasilis Syrgkanis: `https://causalml-book.org`

Correlation vs. Causality

Correlation and causality are different concepts

- Carrying umbrellas causes rain?
- Correlation may imply another causality than this one.

Coming first may not mean causality!

- Do roosters cause the sun to rise?

Causality may mask correlations!

- Perfect doctor: treatment and life expectancy

Potential Outcomes Framework

Neyman (1923)-Rubin (1974) causal model

Hypothesis: a positive causal effect of going to college on later-life earnings (Becker, 1957))

Let D_i be an indicator ("dummy" variable) for the college "treatment":

- $D_i = 1$ if individual i went to college and $D_i = 0$ if she did not

For each individual i, we imagine two **potential** earnings outcomes:

- $Y_i(1) =$ outcome if "treated" (earnings if i went to college)
- $Y_i(0) =$ outcome if "untreated" (earnings if i didn't go to college)

We never observe both!

Common Causal Parameters

- Individual Treatment Effect: $\tau_i = Y_i(1) - Y_i(0)$
- Average Treatment Effect: $\tau_{ATE} = E[Y_i(1) - Y_i(0)]$
  - Defined over the full population, and includes individuals who may never recieved the treatment.
- Average Treatment Effect on the Treated: $\tau_{ATT} = E[Y_i(1) - Y_i(0)|D_i = 1]$
  - For individuals who *received* the treatment
  - Note that one piece of this measure is purely observed data: $E[Y_i(1)|D_i = 1]$
- Average Treatment Effect on the Unreated: $\tau_{ATU} = E[Y_i(1) - Y_i(0)|D_i = 0]$
- The Conditional Average Treatment Effect: $\tau_{CATE} = E[Y_i(1) - Y_i(0)|X_i = x]$, where X_i is some characteristic.

Target Parameters

These are called "target parameters": what we actually want to know about the population

- Average earnings effect of going to college
- A fixed number, because the population is fixed

Suppose we are interested in ATE: $E[Y_i(1) - Y_i(0)]$

$$\tau^* = \underbrace{E[Y_i(1)]}_{\text{Avg. } \textit{potential college earnings}} - \underbrace{E[Y_i(0)]}_{\text{Avg. } \textit{potential no-college earnings}}$$

How do we find this effect?

Choice of Parameter Whether we are interested in the ATE, ATT, or ATU (or a different parameter all together) depends on the causal question.

- ATE gives the expected returns to education for a randomly selected individual;
- ATT gives the expected returns to education for college graduates;
- ATU gives the expected returns to education for non-graduates.

There are lots of prospective students touring campus these days. Suppose we are considering to encourage them from pursuing college. Which parameter would be most relevant?

Instead, say we are considering a policy that encourages students from high schools with traditionally low college-enrollment rates to pursue higher education. Which parameter would be most relevant?

Estimand

Again we cannot observe both potential outcomes.

Consider, instead, the corresponding difference in population means:

$$\tau = \underbrace{E[Y_i|D_i = 1]}_{\text{Avg. college earnings in population}} - \underbrace{E[Y_i|D_i = 0]}_{\text{Avg. no-college earnings in population}}$$

This is called an **estimand**: a function of distribution of observable data in the **population**

- Difference in earnings between **all** individuals with college and without college degree
- A fixed number, again because the population is fixed

Target Parameter = Estimand?

Population

i	$Y_i(1)$	$Y_i(0)$	D_i
1	8	3	1
2	6	2	1
3	5	3	1
4	8	2	1
5	7	3	1
6	4	4	0
7	8	6	0
8	6	2	0
9	8	2	0
10	9	3	0
Mean	6.9	3	

$E[Y_i(1)] - E[Y_i(0)] = 6.9 - 3 = 3.9$
$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = 6.8 - 3.4 = 3.4$

Identification

For $E[Y_i(1)] - E[Y_i(0)] = E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$, we need assumptions.

$\Rightarrow$ Identifying Assumptions!

Stable Unit Treatment Value Assumption (SUTVA)

Observed outcome Y_i is $Y_i(1)$ if $D_i = 1$ and $Y_i(0)$ if $D_i = 0$.

- We can write this as $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$

Underlying assumptions...

SUTVA (Rubin, 1978)

1. The potential outcomes for any unit do not vary with the treatments assigned to other units.
2. For each unit, there are no different forms or versions of each treatment level, which lead to different potential outcomes.

Ignorability (Exchangeability)

Strong Ignorability (Rubin 1978)

D_i is *strongly ignorable* conditional on a vector $\mathbf{X}_i$ if

1. $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp D_i \mid \mathbf{X}_i$
2. $0 < \Pr(D_i = 1|X_i) < 1$ (Positivity)

- The first condition asserts independence of the treatment from the "potential" outcomes
- The second condition asserts that there are both treated and untreated individuals
- Also called,
  - unconfoundedness
  - conditional independence
  - exogeneity
- Drop "$|X$", then it's called "ignorability."

SUTVA + Ignorability

$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$ and $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp D_i$

$\Rightarrow E(Y_i|D_i = d) = E(Y_i(d))$ for $d = 0, 1$

$$
\begin{aligned}
E(Y_i|D_i = d) &= E(D_i Y_i(1) + (1 - D_i) Y_i(0)|D_i = d) \\
&= \begin{cases} E(Y_i(1)|D_i = 1) = E(Y_i(1)) \text{ if } d = 1 \\ E(Y_i(0)|D_i = 0) = E(Y_i(0)) \text{ if } d = 0 \end{cases} \\
&= E(Y_i(d))
\end{aligned}
$$

The mean of the observed outcomes equals the mean of the potential outcomes. $\Rightarrow E(Y_i(d))$ is identified!

Under These Identifying Assumptions

$$\begin{aligned}\tau_{ATE} &= E[Y_i(1) - Y_i(0)] \\ &= E[Y_i(1)] - E[Y_i(0)] \\ &= E[Y_i|D_i = 1] - E[Y_i|D_i = 0]\end{aligned}$$

OK. Now how do we estimate this estimand from a sample?

When the target parameter is ATT

$$
\begin{aligned}
\tau_{ATT} &= E[Y_i(1) - Y_i(0)|D_i = 1] \\
&= E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] \\
&= E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0] \\
&\qquad + \underbrace{E[Y_i(0)|D_i = 0] - E[Y_i(0)|D_i = 1]}_{\text{Selection bias}} \\
&= E[Y_i|D_i = 1] - E[Y_i|D_i = 0]
\end{aligned}
$$

So, the estimands for the ATE and ATT are the same.

What about ATU? Under the same set of assumptions,
ATE = ATT = ATU

Estimator

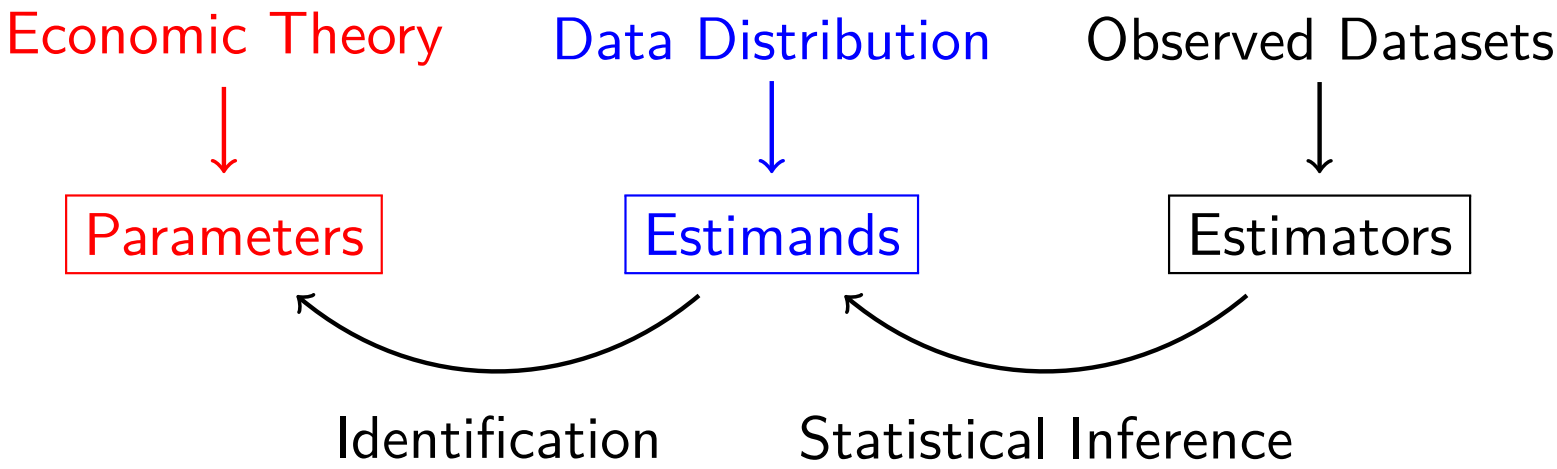
Estimator: a function of data for the sample: what we actually see

- Difference in earnings between **surveyed** college vs no-college
- A random object, because the sampled data are random

Take as our **estimator** the difference in sample average earnings for the N_1 people who did go to college vs. the N_0 people who didn't go to college:

$$\hat{\tau}_N = \underbrace{\frac{1}{N_1} \sum_{i:D_i=1} Y_i}_{\text{Avg. college earnings in sample}} - \underbrace{\frac{1}{N_0} \sum_{i:D_i=0} Y_i}_{\text{Avg. no-college earnings in sample}}$$

Parameters, Estimands, and Estimators



Statistical inference: the process of learning about the estimand from the estimator (statistical task)

- How does the sample we observe relate to the population of interest?

Identification: the process of learning about the parameter from the estimand (modeling task)

- How do observable features of the population relate to (causal) parameters we care about?

From Estimator to Estimand

The process of learning about unobserved population means $E[Y_i|D_i = d]$ from observed sample means $\frac{1}{N_d}\sum_{i:D_i=d} Y_i$ (**statistical inference**).

We ask, in large samples:

- If sample means are "close to" population means (by the LLN): $\hat{\tau}_N \xrightarrow{p} \tau$
- Deviations between sample and population means tend to follow a known distribution (by the CLT): $\delta_N(\hat{\tau}_N - \tau) \xrightarrow{d} P(\theta)$
- We can use these facts to make inferences about population means from observed sample means (e.g. form 95% CIs).

From Estimand to Parameter

The process of identifying assumptions required to make sure our estimand is equal to our target parameter.

$$\text{ATE} = E[Y_i(1) - Y_i(0)] \overset{?}{=} E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$$

Selection bias: people who do/don't go to college (D_i) have different potential earnings with/without college ($Y_i(d)$)

- Differences in academic ability, family background, career goals, etc.
- Ideally, want to control for all of these.
- Sometimes referred to as omitted variables or confounding factors

Identification Strategies

An identification strategy is a research design intended to solve the causal inference identification problem (Angrist and Pischke 2010).

1. What are identifying assumptions (what we discussed so far)

   Another part is:

2. Which empirical design to employ to satisfy the assumptions?
   - E.g., Randomized experiments

Identification through Randomization

The gold standard for learning about causal effects is a randomized controlled trial (RCT), aka an **experiment**

Suppose that we were somehow able to randomize who went to college

- This makes D_i statistically independent of $(Y_i(0), Y_i(1))$: same potential outcomes among those randomized to go to college and not.
- Suppose SUTVA holds.

Randomization eliminates selection bias

But randomization is often hard/impossible/unethical in economics

- We will develop tools to "mimic" the RCT gold standard, in non-experimental ("**observational**") data
  - natural experiments, instrumental variables, regression discontinuity **...**