SciencesPo
DEPARTMENT OF ECONOMICS

ScPoEconometrics Advanced

Causality

Nikiforos Zampetakis based on Florian Oswald's slides
SciencesPo Paris
2024-01-30

Causality

What makes Econometrics different from Statistics?

Causality

Statistics $\longrightarrow$ statistical inference about large populations via sampling.

Econometrics $\longrightarrow$ causal inference via **counterfactuals**.

But

- What is causality?
- What do counterfactuals mean?

What does causality means?

Why did something happen?

More precisely: **Does x causes y**?

Does x has an **effect** on y?

- e.g. Does smoking **causes** lung cancer?

To answer the question what is a **counterfactual** we'll get a bit more rigorous.

Potential Outcomes Model (Rubin's Causal Model)

Two states of the world (Potential Outcomes).

1. Individual i gets the treatment $\rightarrow Y_i^1$.
2. Individual i does **not** get the treatment $\rightarrow Y_i^0$.

The observed outcome is

$$Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0 \quad D_i \in 0, 1$$

That's a "mathy" way to describe a deep question.

What would have happen to i had they not received the treatment D_i?

That's a deep question because we need to **imagine** a parallel universe that the actions taken were different.

Potential outcome Y_i^0 is defined as a **counterfactual**.

Treatment Effect

The treatment effect for i is

$$\delta_i = Y_i^1 - Y_i^0$$

We observe only on of the Y_i^1, Y_i^0 for every $i \rightarrow$ We can *not* compute δ_i.

This is called

The Fundamental Identification Problem of Program Evaluation

But are we stuck??

Question: What about an average effect?

Average Treatment Effect

$$\delta^{ATE} = E[\delta_i] = E\left[Y_i^1\right] - E\left[Y_i^0\right]$$

Fundamentally a **missing data problem**. We observe either Y_i^1 or Y_i^0.

Set up a naive estimator. Difference of the mean outcome of the Treatment and Control group.

$$\hat{\delta} = E\left[Y_i^1 | D_i = 1\right] - E\left[Y_i^0 | D_i = 0\right] = \frac{1}{N_T} \sum_{i \in T}^{N_T} T_i - \frac{1}{N_C} \sum_{j \in C}^{N_C} Y_j$$

Key consideration: δ_i is different for different people. To learn about true δ^{ATE} from $\hat{\delta}$ it matters who gets treated.

If individuals know the gains from treatment $\implies$ is with high gains will select to treatment $\implies$ $\hat{\delta}$ will be upwardly biased from the true δ^{ATE}.

$$E\left[Y_i^1 | D_i = 1\right] \neq \frac{1}{N_T} \sum_{i \in T}^{N_T} T_i$$

Randomization to the rescue

If randomly assign who gets treated or not (e.g. by flipping a coin) people with high expected gains can not select to get treated.

NB: The distrib. of δ_i is still the same in the study pop. $\implies$ there are still people with high and low effects.

Missing data problem solved!!

Whether Y_i^1 or Y_i^0 is observed for each i is random and not a function of any factor that i can act upon.

$$\hat{\delta} = \delta^{ATE}$$

SciencesPo
DEPARTMENT OF ECONOMICS

END

	nikiforos.zampetakis@sciencespo.fr
	Slides
	Book
	@ScPoEcon
	@ScPoEcon