

# Machine Learning for econometrics

Causal perspective

---

Matthieu Doutreligne

January 10, 2025

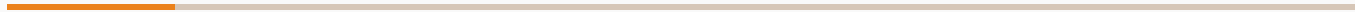
# Table of contents

1. Introduction
2. How to ask a sound causal question
3. How to ask a sound causal question: The PICO framework
4. Causal graphs
5. The four steps of causal inferenceidentification, statistical estimand, statistical inference
6. Potential outcomes
7. Causal estimands

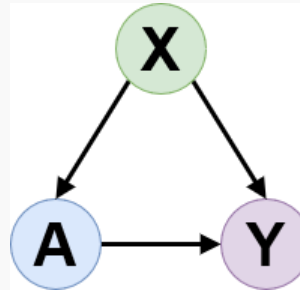
# Table of contents

- 8. Causal graphs
- 9. Statistical estimand
- 10. Statistical inference ie. estimation
- 11. Related concepts

# Introduction

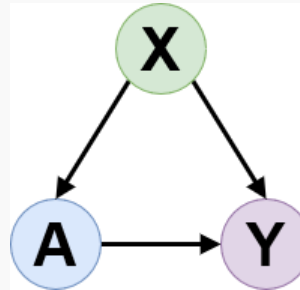


Causal inference: subfield of statistics dealing with “why questions”.



At the center of epidemiology, econometrics, social sciences...

Causal inference: subfield of statistics dealing with “why questions”.



At the center of epidemiology, econometrics, social sciences...

Now, bridging with Machine Learning (Kaddour, Lynch, Liu, Kusner, & Silva, 2022)

# How to ask a sound causal question

---

## What is a **why question** ?

- Economics: How does supply and demand (causally) depend on price?
- Policy: Are job training programmes actually effective?
- Epidemiology: How does this treatment affect the patient's health?
- Public health : Is this prevention campaign effective?
- Psychology: What is the effect of family structure on children's outcome?
- Sociology: What is the effect of social media on political opinions?



## This is different from a **predictive question**

- What will be the weather tomorrow?
- What will be the outcome of the next election?
- How many people will get infected by flue next season?
- What is the cardio-vascular risk of this patient?
- How much will the price of a stock be tomorrow?

## Why is **prediction different from causation**?

- Prediction (most part of Machine Learning) focus on understanding what usually happens in a given situation.

## Why is **prediction different from causation**?

- Prediction (most part of Machine Learning) focus on understanding what usually happens in a given situation.

It assumes iid between train and test data.

## Why is **prediction different from causation**?

- Prediction (most part of Machine Learning) focus on understanding what usually happens in a given situation.
- Causal inference (most part of economists) focus on what would happen if we changed the system ie. under intervention.

It models the covariate shift between treated and control units.

# How to ask a sound causal question: The PICO framework

---

## Identify the target trial

What would be the ideal **randomized experiment** to answer the question?  
(Hernán & Robins, 2016)

## **PICO framework**

- Population : Who are we interested in?
- Intervention : What treatment/intervention do we study?
- Comparison : What are we comparing it to?
- Outcome : What are we interested in?

## PICO framework (illustration)

- P
- I
- C
- O



# Causal graphs

---

The four steps of causal  
inferenceidentification, statistical  
estimand, statistical inference

---

# Causal estimand

What can we learn from the data?

# Identification

What can we learn from the data?

Knowledge based

Cannot be validated with data

# Potential outcomes

---

# Causal estimands

---

# Causal graphs

---

# Statistical estimand

---



Statistical inference ie. estimation

---

# Related concepts

---

- Structural equations:

## Resources

- <https://web.stanford.edu/~swager/stats361.pdf>
- <https://www.mixtapesessions.io/>
- <https://alejandroschuler.github.io/mci/>

## ***Bibliography***

*Hernán, M. A., & Robins, J. M. (2016). Using big data to emulate a target trial when a randomized trial is not available. American Journal of Epidemiology, 183(8), 758–764.*

*Kaddour, J., Lynch, A., Liu, Q., Kusner, M. J., & Silva, R. (2022). Causal machine learning: A survey and open problems. Arxiv Preprint Arxiv:2206.15475.*