Introduction to Causal Inference

Causal Inference using Machine Learning
Master in Economics, UNT

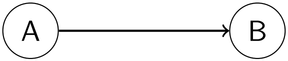
Andres Mena

Spring 2024

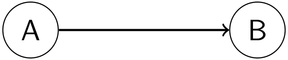
Table of Contents

Content

What is a Causal Question?

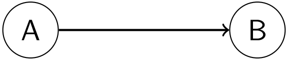


What is a Causal Question?



A causal question asks about the effect of a cause A on an outcome B.

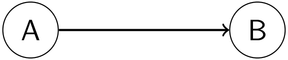
What is a Causal Question?



A causal question asks about the effect of a cause A on an outcome B.

- A could be:

What is a Causal Question?



A causal question asks about the effect of a cause A on an outcome B.

- A could be:
  - Policy (e.g., a new transfer policy impacting household income (AUH))

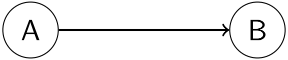
What is a Causal Question?

$$A \longrightarrow B$$

A causal question asks about the effect of a cause A on an outcome B.

- A could be:
  - Policy (e.g., a new transfer policy impacting household income (AUH))
  - Treatment (e.g., a drug trial reducing heart attack risk)

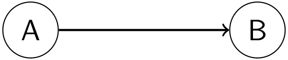
What is a Causal Question?



A causal question asks about the effect of a cause A on an outcome B.

- A could be:
  - Policy (e.g., a new transfer policy impacting household income (AUH))
  - Treatment (e.g., a drug trial reducing heart attack risk)
  - Business strategy (e.g., a marketing campaign affecting sales)

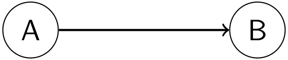
What is a Causal Question?



A causal question asks about the effect of a cause A on an outcome B.

- A could be:
  - Policy (e.g., a new transfer policy impacting household income (AUH))
  - Treatment (e.g., a drug trial reducing heart attack risk)
  - Business strategy (e.g., a marketing campaign affecting sales)
  - Technological innovation (e.g., the introduction of innovation of supply chain)

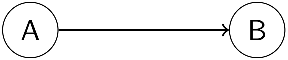
What is a Causal Question?



A causal question asks about the effect of a cause A on an outcome B.

- A could be:
  - Policy (e.g., a new transfer policy impacting household income (AUH))
  - Treatment (e.g., a drug trial reducing heart attack risk)
  - Business strategy (e.g., a marketing campaign affecting sales)
  - Technological innovation (e.g., the introduction of innovation of supply chain)
- B could be:

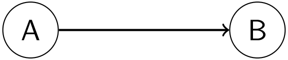
What is a Causal Question?



A causal question asks about the effect of a cause A on an outcome B.

- A could be:
  - Policy (e.g., a new transfer policy impacting household income (AUH))
  - Treatment (e.g., a drug trial reducing heart attack risk)
  - Business strategy (e.g., a marketing campaign affecting sales)
  - Technological innovation (e.g., the introduction of innovation of supply chain)
- B could be:
  - Welfare (e.g., income, poverty, crime, happiness, inequality)

What is a Causal Question?



A causal question asks about the effect of a cause A on an outcome B.

- A could be:
  - Policy (e.g., a new transfer policy impacting household income (AUH))
  - Treatment (e.g., a drug trial reducing heart attack risk)
  - Business strategy (e.g., a marketing campaign affecting sales)
  - Technological innovation (e.g., the introduction of innovation of supply chain)
- B could be:
  - Welfare (e.g., income, poverty, crime, happiness, inequality)
  - Health (e.g., life expectancy, disease risk, heart attacks)

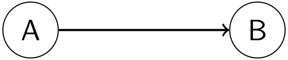
What is a Causal Question?

$$A \longrightarrow B$$

A causal question asks about the effect of a cause A on an outcome B.

- A could be:
  - Policy (e.g., a new transfer policy impacting household income (AUH))
  - Treatment (e.g., a drug trial reducing heart attack risk)
  - Business strategy (e.g., a marketing campaign affecting sales)
  - Technological innovation (e.g., the introduction of innovation of supply chain)
- B could be:
  - Welfare (e.g., income, poverty, crime, happiness, inequality)
  - Health (e.g., life expectancy, disease risk, heart attacks)
  - Profits (e.g., company earnings, market share, costs)

What is a Causal Question?



A causal question asks about the effect of a cause A on an outcome B.

- A could be:
  - Policy (e.g., a new transfer policy impacting household income (AUH))
  - Treatment (e.g., a drug trial reducing heart attack risk)
  - Business strategy (e.g., a marketing campaign affecting sales)
  - Technological innovation (e.g., the introduction of innovation of supply chain)
- B could be:
  - Welfare (e.g., income, poverty, crime, happiness, inequality)
  - Health (e.g., life expectancy, disease risk, heart attacks)
  - Profits (e.g., company earnings, market share, costs)
  - Efficiency (e.g., speed of production or resource utilization)

Correlation is not Causation (and vice versa)

1. **Correlation $\not\to$ Causation**

Correlation is not Causation (and vice versa)

1. **Correlation $\not\to$ Causation**
   - **Correlation**: A statistical relationship between two variables, where changes in one variable are associated with changes in another.

$$\mathrm{Corr}(A, B) = \frac{\mathrm{Cov}(A, B)}{\sigma_A \sigma_B}$$

Correlation is not Causation (and vice versa)

1. **Correlation $\not\rightarrow$ Causation**
   - **Correlation**: A statistical relationship between two variables, where changes in one variable are associated with changes in another.

$$\mathrm{Corr}(A, B) = \frac{\mathrm{Cov}(A, B)}{\sigma_A \sigma_B}$$

   - Example: Rain and umbrellas.

Correlation is not Causation (and vice versa)

1. **Correlation $\not\rightarrow$ Causation**
   - **Correlation**: A statistical relationship between two variables, where changes in one variable are associated with changes in another.

$$\text{Corr}(A, B) = \frac{\text{Cov}(A, B)}{\sigma_A \sigma_B}$$

   - Example: Rain and umbrellas.
2. **Causation $\not\rightarrow$ Correlation**

Correlation is not Causation (and vice versa)

1. **Correlation $\not\to$ Causation**
   - **Correlation**: A statistical relationship between two variables, where changes in one variable are associated with changes in another.

   $$\text{Corr}(A, B) = \frac{\text{Cov}(A, B)}{\sigma_A \sigma_B}$$

   - Example: Rain and umbrellas.
2. **Causation $\not\to$ Correlation**
   - **Causation**: A direct effect of one variable on another, where A causes B.

   $$B = f(A)$$

Correlation is not Causation (and vice versa)

1. **Correlation $\not\rightarrow$ Causation**
   - **Correlation**: A statistical relationship between two variables, where changes in one variable are associated with changes in another.

     $$\text{Corr}(A, B) = \frac{\text{Cov}(A, B)}{\sigma_A \sigma_B}$$

   - Example: Rain and umbrellas.
2. **Causation $\not\rightarrow$ Correlation**
   - **Causation**: A direct effect of one variable on another, where A causes B.

     $$B = f(A)$$

   - Example: Eating more food can cause weight gain, but if food intake and exercise both increase proportionally, we may observe no correlation between food and weight in the data, even though causation exists.

Causal Inference Tree

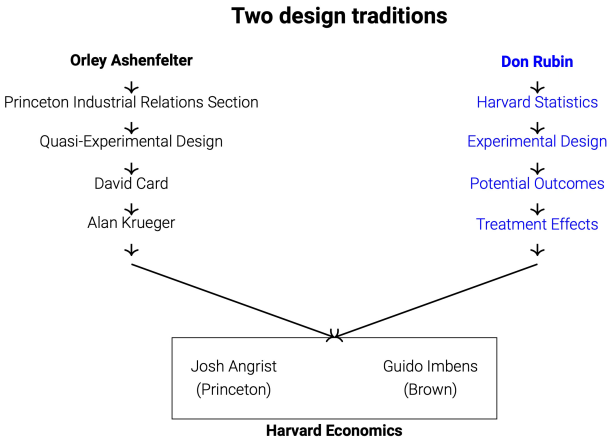


Figure: Source: Scott Cunningham Substack

Experimental Design Tradition

Experimental design relies on randomized controlled trials (RCTs) to establish causality through direct manipulation of the treatment.

- **Neyman (1923)**:
  - Introduced the concept of randomization in experiments and formalized the **potential outcomes** framework.
  - *Contribution*: Randomization as a tool to eliminate bias.

Experimental Design Tradition

Experimental design relies on randomized controlled trials (RCTs) to establish causality through direct manipulation of the treatment.

- **Neyman (1923)**:
  - Introduced the concept of randomization in experiments and formalized the **potential outcomes** framework.
  - *Contribution*: Randomization as a tool to eliminate bias.
- **Rubin (1974)**:
  - Developed the potential outcomes model in the context of randomized experiments.
  - *Contribution*: Introduced the Rubin Causal Model (RCM) as a framework for understanding causal effects.

Experimental Design Tradition

Experimental design relies on randomized controlled trials (RCTs) to establish causality through direct manipulation of the treatment.

- **Neyman (1923)**:
  - Introduced the concept of randomization in experiments and formalized the **potential outcomes** framework.
  - *Contribution*: Randomization as a tool to eliminate bias.
- **Rubin (1974)**:
  - Developed the potential outcomes model in the context of randomized experiments.
  - *Contribution*: Introduced the Rubin Causal Model (RCM) as a framework for understanding causal effects.
- **Esther Duflo (Nobel Prize 2019)**:
  - Pioneered the use of randomized controlled trials in development economics to study policy interventions.
  - *Contribution*: Applied RCTs to measure the effectiveness of educational and poverty interventions in developing countries.

Quasi-experimental Design

Quasi-experimental designs use natural variation and perform comparisons under assumptions.

Quasi-experimental Design

Quasi-experimental designs use natural variation and perform comparisons under assumptions.

- **Difference-in-Differences (DiD)**:
  - **John Snow's Cholera Study (1854)**: Snow used natural variation in water sources to compare cholera outcomes between neighborhoods (Snow, 1854).
  - *Causal Question*: Does contaminated water cause cholera?

Quasi-experimental Design

Quasi-experimental designs use natural variation and perform comparisons under assumptions.

- **Difference-in-Differences (DiD)**:
  - **John Snow's Cholera Study (1854)**: Snow used natural variation in water sources to compare cholera outcomes between neighborhoods (Snow, 1854).
  - *Causal Question*: Does contaminated water cause cholera?
- **Instrumental Variables (IV)**:
  - **Wright (1928)**: Used supply-side instruments to estimate demand for agricultural products (Wright, 1928).
  - *Causal Question*: What is the causal effect of price on the quantity demanded?

Quasi-experimental Design

Quasi-experimental designs use natural variation and perform comparisons under assumptions.

- **Difference-in-Differences (DiD)**:
  - **John Snow's Cholera Study (1854)**: Snow used natural variation in water sources to compare cholera outcomes between neighborhoods (Snow, 1854).
  - *Causal Question*: Does contaminated water cause cholera?
- **Instrumental Variables (IV)**:
  - **Wright (1928)**: Used supply-side instruments to estimate demand for agricultural products (Wright, 1928).
  - *Causal Question*: What is the causal effect of price on the quantity demanded?
- **Regression Discontinuity (RD)**:
  - **Thistlethwaite and Campbell (1960)**: Studied the effect of merit awards on student performance (Thistlethwaite and Campbell, 1960).
  - *Causal Question*: What is the effect of winning a merit award on future academic success?

Quasi-experimental Design

Quasi-experimental designs use natural variation and perform comparisons under assumptions.

- **Difference-in-Differences (DiD)**:
  - **John Snow's Cholera Study (1854)**: Snow used natural variation in water sources to compare cholera outcomes between neighborhoods (Snow, 1854).
  - *Causal Question*: Does contaminated water cause cholera?
- **Instrumental Variables (IV)**:
  - **Wright (1928)**: Used supply-side instruments to estimate demand for agricultural products (Wright, 1928).
  - *Causal Question*: What is the causal effect of price on the quantity demanded?
- **Regression Discontinuity (RD)**:
  - **Thistlethwaite and Campbell (1960)**: Studied the effect of merit awards on student performance (Thistlethwaite and Campbell, 1960).
  - *Causal Question*: What is the effect of winning a merit award on future academic success?

Content

Content

Content

References I

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *American Economic Review*, 105(3):391–425.

Snow, J. (1854). *On the Mode of Communication of Cholera*. London: John Churchill.

Thistlethwaite, D. L. and Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51(6):309–317.

Wright, P. G. (1928). The tariff on animal and vegetable oils. *Quarterly Journal of Economics*, 43(4):599–607.