Introduction to Causal Inference

Causal Inference using Machine Learning
Master in Economics, UNT

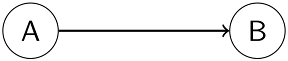
Andres Mena

Spring 2024

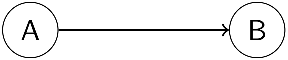
Table of Contents

Content

What is a Causal Question?

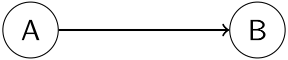


What is a Causal Question?



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

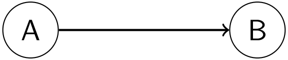
What is a Causal Question?



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

- A could be:

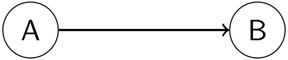
What is a Causal Question?



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))

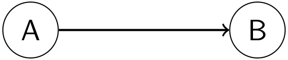
What is a Causal Question?



A 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)

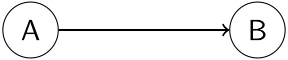
What is a Causal Question?



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)

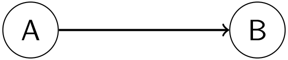
What is a Causal Question?



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)

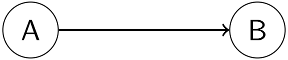
What is a Causal Question?



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:

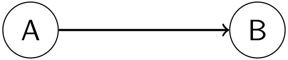
What is a Causal Question?



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)

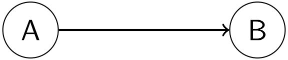
What is a Causal Question?



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)

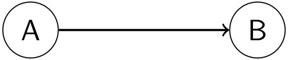
What is a Causal Question?



A 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 $\not\rightarrow$ Causation**

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.

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

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.

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 $\not\rightarrow$ 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.
2. **Causation $\not\rightarrow$ 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.

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

   - 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

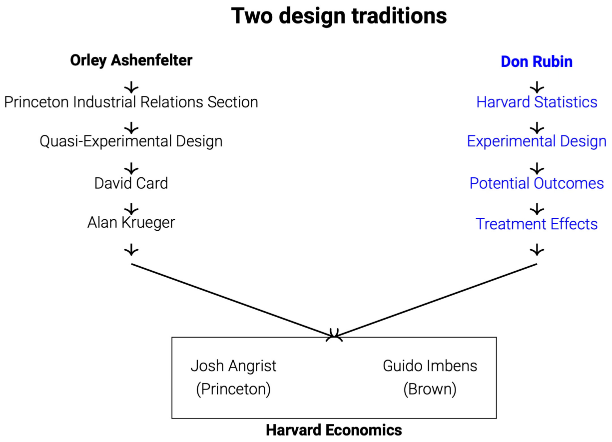


Figure: Source: Scott Cunningham Substack

Experimental Design Tradition

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

- Neyman (1923):
  - Introduced the **potential outcomes** framework for randomized experiments.

Experimental Design Tradition

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

- Neyman (1923):
  - Introduced the **potential outcomes** framework for randomized experiments.
- Fisher (1935):
  - Random assignment removes selection 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 **potential outcomes** framework for randomized experiments.
- Fisher (1935):
  - Random assignment removes selection bias.
- Rubin (1974):
  - Generalized the potential outcomes framework to both observational and experimental studies.

Experimental Design Tradition

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

- Neyman (1923):
  - Introduced the **potential outcomes** framework for randomized experiments.
- Fisher (1935):
  - Random assignment removes selection bias.
- Rubin (1974):
  - Generalized the potential outcomes framework to both observational and experimental studies.
- Banerjee and Duflo (2011):
  - Duflo and Banerjee pioneered the use of RCTs in development economics to evaluate the impact of poverty alleviation policies.

Quasi-experimental Design

Explodes natural variation and perform comparisons under identification assumptions.

Quasi-experimental Design

Explodes natural variation and perform comparisons under identification assumptions.

- **Difference-in-Differences (DiD)**: Snow (1854) used natural variation in water sources to compare cholera outcomes between neighborhoods.
  -*CQ*: Does contaminated water cause cholera?

Quasi-experimental Design

Explodes natural variation and perform comparisons under identification assumptions.

- **Difference-in-Differences (DiD)**: Snow (1854) used natural variation in water sources to compare cholera outcomes between neighborhoods.
  -*CQ*: Does contaminated water cause cholera?
- **Instrumental Variables (IV)**: Wright (1928) used exogenous supply-side instruments (rain) to estimate demand for agricultural products.
  -*CQ*: What is the effect of price on the quantity demanded?

Quasi-experimental Design

Explodes natural variation and perform comparisons under identification assumptions.

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

Quasi-experimental Design

Explodes natural variation and perform comparisons under identification assumptions.

- **Difference-in-Differences (DiD)**: Snow (1854) used natural variation in water sources to compare cholera outcomes between neighborhoods.
  -*CQ*: Does contaminated water cause cholera?
- **Instrumental Variables (IV)**: Wright (1928) used exogenous supply-side instruments (rain) to estimate demand for agricultural products.
  -*CQ*: What is the effect of price on the quantity demanded?
- **Regression Discontinuity (RD)**: Thistlethwaite and Campbell (1960) studied the effect of merit awards on student performance.
  -*CQ*: What is the effect of winning a merit award on future academic success?
- **Synthetic Control (SCM)**: Abadie and Gardeazabal (2003): developed SCM to run comparative case studies.
  -*CQ*: What is the impact of Terrorism on Economic Performance?

Machine Learning and Causal Inference

What is Machine Learning?

- Machine Learning (ML) refers to algorithms and statistical models that allow computers to learn from and make predictions or decisions based on data.
- ML excels at identifying complex patterns and making predictions in high-dimensional settings.

Machine Learning and Causal Inference

What is Machine Learning?

- Machine Learning (ML) refers to algorithms and statistical models that allow computers to learn from and make predictions or decisions based on data.
- ML excels at identifying complex patterns and making predictions in high-dimensional settings.

How Machine Learning is Used in Causal Inference

- Machine Learning can be used to estimate nuisance parameters (e.g., propensity scores, regression functions) in causal inference models.
- Chernozhukov et al. (2018) introduced **Double/Debiased Machine Learning (DML)**, which combines machine learning for estimating high-dimensional nuisance functions with traditional econometric techniques to ensure valid causal inference.
- ML algorithms are employed in tasks like instrumental variable estimation, heterogeneous treatment effects, and controlling for high-dimensional confounders.

Machine Learning and Causal Inference

What is Machine Learning?

- Machine Learning (ML) refers to algorithms and statistical models that allow computers to learn from and make predictions or decisions based on data.
- ML excels at identifying complex patterns and making predictions in high-dimensional settings.

How Machine Learning is Used in Causal Inference

- Machine Learning can be used to estimate nuisance parameters (e.g., propensity scores, regression functions) in causal inference models.
- Chernozhukov et al. (2018) introduced **Double/Debiased Machine Learning (DML)**, which combines machine learning for estimating high-dimensional nuisance functions with traditional econometric techniques to ensure valid causal inference.
- ML algorithms are employed in tasks like instrumental variable estimation, heterogeneous treatment effects, and controlling for high-dimensional confounders.

Content

1. What's the Causal Relationship of Interest?

Causal inference begins by identifying the causal relationship of interest. Some classic papers are examples of the main methodologies:

- Difference-in-Differences (DiD): How do historical institutions impact modern economic development? Acemoglu et al. (2001)

1. What's the Causal Relationship of Interest?

Causal inference begins by identifying the causal relationship of interest. Some classic papers are examples of the main methodologies:

- Difference-in-Differences (DiD): How do historical institutions impact modern economic development? Acemoglu et al. (2001)
- Instrumental Variables (IV): What is the causal effect of compulsory schooling on earnings? Angrist and Krueger (1991)

1. What's the Causal Relationship of Interest?

Causal inference begins by identifying the causal relationship of interest. Some classic papers are examples of the main methodologies:

- Difference-in-Differences (DiD): How do historical institutions impact modern economic development? Acemoglu et al. (2001)
- Instrumental Variables (IV): What is the causal effect of compulsory schooling on earnings? Angrist and Krueger (1991)
- Comparative Case Study: How do immigration shocks affect local labor markets? Card (1990)

These questions seek to determine the effect of specific treatments (institutions, education, immigration) on outcomes of interest (development, earnings, labor markets).

The Experimental Ideal

What is the ideal experiment to capture the treatment/causal effect?

The Experimental Ideal

What is the ideal experiment to capture the treatment/causal effect?

- The experimental ideal serves as a benchmark for understanding causal effects, helping to evaluate the identification strategy.

The Experimental Ideal

What is the ideal experiment to capture the treatment/causal effect?

- The experimental ideal serves as a benchmark for understanding causal effects, helping to evaluate the identification strategy.
-
- It is often an abstract or hypothetical experiment with no practical, ethical, or financial constraints.

The Experimental Ideal

What is the ideal experiment to capture the treatment/causal effect?

- The experimental ideal serves as a benchmark for understanding causal effects, helping to evaluate the identification strategy.
-
- It is often an abstract or hypothetical experiment with no practical, ethical, or financial constraints.
-
- Example: Bauer (2015)

The Experimental Ideal

What is the ideal experiment to capture the treatment/causal effect?

- The experimental ideal serves as a benchmark for understanding causal effects, helping to evaluate the identification strategy.
-
- It is often an abstract or hypothetical experiment with no practical, ethical, or financial constraints.
-
- Example: Bauer (2015)
  - Research Question: What is the causal effect of victimization (being robbed) on social trust?

The Experimental Ideal

What is the ideal experiment to capture the treatment/causal effect?

- The experimental ideal serves as a benchmark for understanding causal effects, helping to evaluate the identification strategy.
-
- It is often an abstract or hypothetical experiment with no practical, ethical, or financial constraints.
-
- Example: Bauer (2015)
  - Research Question: What is the causal effect of victimization (being robbed) on social trust?
  -
  - Ideal Experiment:

The Experimental Ideal

What is the ideal experiment to capture the treatment/causal effect?

- The experimental ideal serves as a benchmark for understanding causal effects, helping to evaluate the identification strategy.
-
- It is often an abstract or hypothetical experiment with no practical, ethical, or financial constraints.
-
- Example: Bauer (2015)
  - Research Question: What is the causal effect of victimization (being robbed) on social trust?
  -
  - Ideal Experiment:
    - Randomly sample individuals from the target population (e.g., Swiss population).

The Experimental Ideal

What is the ideal experiment to capture the treatment/causal effect?

- The experimental ideal serves as a benchmark for understanding causal effects, helping to evaluate the identification strategy.
-
- It is often an abstract or hypothetical experiment with no practical, ethical, or financial constraints.
-
- Example: Bauer (2015)
  - Research Question: What is the causal effect of victimization (being robbed) on social trust?
  -
  - Ideal Experiment:
    - Randomly sample individuals from the target population (e.g., Swiss population).
    -
    - Homogeneous robbers administer the same factual negative experience to the treatment group, while others are left undisturbed (control

The Experimental Ideal

What is the ideal experiment to capture the treatment/causal effect?

- The experimental ideal serves as a benchmark for understanding causal effects, helping to evaluate the identification strategy.
-
- It is often an abstract or hypothetical experiment with no practical, ethical, or financial constraints.
-
- Example: Bauer (2015)
  - Research Question: What is the causal effect of victimization (being robbed) on social trust?
  -
  - Ideal Experiment:
    - Randomly sample individuals from the target population (e.g., Swiss population).
    -
    - Homogeneous robbers administer the same factual negative experience to the treatment group, while others are left undisturbed (control

3. What's the Identification Strategy?

Causal inference requires a strategy to identify the causal effect. The strategy relies on exogenous variability and assumptions:

3. What's the Identification Strategy?

Causal inference requires a strategy to identify the causal effect. The strategy relies on exogenous variability and assumptions:

- **Instrumental Variables (IV)**:
  - Strategy: Use of instruments that affect the treatment but not the outcome directly (e.g., proximity to a college Angrist and Krueger (1991)).
  - Assumption: **Exclusion Restriction**, **Relevance**.

3. What's the Identification Strategy?

Causal inference requires a strategy to identify the causal effect. The strategy relies on exogenous variability and assumptions:

- **Instrumental Variables (IV)**:
  - Strategy: Use of instruments that affect the treatment but not the outcome directly (e.g., proximity to a college Angrist and Krueger (1991)).
  - Assumption: **Exclusion Restriction**, **Relevance**.
- **Difference-in-Differences (DiD)**:
  - Strategy: Compare treated and control groups over time.
  - Assumption: **Parallel Trends**.

3. What's the Identification Strategy?

Causal inference requires a strategy to identify the causal effect. The strategy relies on exogenous variability and assumptions:

- **Instrumental Variables (IV)**:
  - Strategy: Use of instruments that affect the treatment but not the outcome directly (e.g., proximity to a college Angrist and Krueger (1991)).
  - Assumption: **Exclusion Restriction**, **Relevance**.
- **Difference-in-Differences (DiD)**:
  - Strategy: Compare treated and control groups over time.
  - Assumption: **Parallel Trends**.
- **Regression Discontinuity (RD)**:
  - Strategy: Use a cutoff or threshold to identify the causal effect.
  - Assumption: **Local Randomization**, **Continuity of Potential Outcomes**.

4. What's the Inference Strategy?

After identifying the causal effect, we need to make valid statistical inferences. Common inference strategies include:

- **Delta Method**:
  - pproximate the variance (or standard error) of a function of an estimator, by using a first-order Taylor expansion.

4. What's the Inference Strategy?

After identifying the causal effect, we need to make valid statistical inferences. Common inference strategies include:

- **Delta Method**:
  - pproximate the variance (or standard error) of a function of an estimator, by using a first-order Taylor expansion.
- **Bootstrap**:
  - Re-sample the data multiple times to approximate the sampling distribution of an estimator.

4. What's the Inference Strategy?

After identifying the causal effect, we need to make valid statistical inferences. Common inference strategies include:

- **Delta Method**:
  - pproximate the variance (or standard error) of a function of an estimator, by using a first-order Taylor expansion.
- **Bootstrap**:
  - Re-sample the data multiple times to approximate the sampling distribution of an estimator.
- **Influence Functions**:
  - Analyze the sensitivity of the estimator to small changes in the sample. Common in non-parametric estimation.

Each strategy allows for valid inference under different circumstances, depending on the complexity of the model and data.

Content

Random Variable and Probability Measure

Random Variable:

- A random variable X is a measurable function from the sample space Ω to the real numbers:

$$X : \Omega \to \mathbb{R}$$

Random Variable:

- A random variable X is a measurable function from the sample space Ω to the real numbers:

$$X : \Omega \to \mathbb{R}$$

Probability Measure:

- A probability measure P is a function that assigns a probability to each event $A \in \mathcal{F}$, where $\mathcal{F}$ is a sigma-algebra over the sample space Ω, such that:

Random Variable and Probability Measure

Random Variable:

- A random variable X is a measurable function from the sample space Ω to the real numbers:

$$X : \Omega \to \mathbb{R}$$

Probability Measure:

- A probability measure P is a function that assigns a probability to each event $A \in \mathcal{F}$, where $\mathcal{F}$ is a sigma-algebra over the sample space Ω, such that:

1. **Non-negativity**: $P(A) \geq 0$ for any event $A \in \mathcal{F}$.

Random Variable and Probability Measure

Random Variable:

- A random variable X is a measurable function from the sample space Ω to the real numbers:

$$X : \Omega \to \mathbb{R}$$

Probability Measure:

- A probability measure P is a function that assigns a probability to each event $A \in \mathcal{F}$, where $\mathcal{F}$ is a sigma-algebra over the sample space Ω, such that:

1. **Non-negativity**: $P(A) \geq 0$ for any event $A \in \mathcal{F}$.
2. **Normalization**: $P(\Omega) = 1$.

Random Variable and Probability Measure

Random Variable:

- A random variable X is a measurable function from the sample space Ω to the real numbers:

$$X : \Omega \to \mathbb{R}$$

Probability Measure:

- A probability measure P is a function that assigns a probability to each event $A \in \mathcal{F}$, where $\mathcal{F}$ is a sigma-algebra over the sample space Ω, such that:

1. **Non-negativity**: $P(A) \geq 0$ for any event $A \in \mathcal{F}$.
2. **Normalization**: $P(\Omega) = 1$.
3. **Additivity**: For disjoint events $A_1, A_2, \ldots$,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Binary Outcome and Sigma Algebra

Binary Outcome and Probability Measure:

- Define the sample space $\Omega = \{High, Low\}$.
- The sigma-algebra $\mathcal{F}$ is the power set of Ω, i.e.,

$$\mathcal{F} = \{\emptyset, \{Low\}, \{High\}, \{Low, High\}\}$$

- Define $Y \in \{High, Low\}$, with the following values and probabilities:

i	Y
1	Low
2	Low
3	High
4	High
5	High
6	High

- $P(Y = Low) = 0.33$, $P(Y = High) = 0.66$

Verification of Probability Axioms

Verification of Axioms:

1. **Non-negativity**: For all elements in $\mathcal{F}$,

   $$P(\emptyset) = 0, \quad P(\{Low\}) = 0.33, \quad P(\{High\}) = 0.66, \quad P(\{Low, High\})$$

   All probabilities are non-negative.

Verification of Probability Axioms

Verification of Axioms:

1. **Non-negativity**: For all elements in $\mathcal{F}$,

   $$P(\emptyset) = 0, \quad P(\{Low\}) = 0.33, \quad P(\{High\}) = 0.66, \quad P(\{Low, High\})$$

   All probabilities are non-negative.

2. **Normalization**: The total probability of the sample space:

   $$P(\{Low, High\}) = P(Y = Low) + P(Y = High) = 0.33 + 0.66 = 1$$

Verification of Probability Axioms

Verification of Axioms:

1. **Non-negativity**: For all elements in $\mathcal{F}$,

   $$P(\emptyset) = 0, \quad P(\{Low\}) = 0.33, \quad P(\{High\}) = 0.66, \quad P(\{Low, High\})$$

   All probabilities are non-negative.

2. **Normalization**: The total probability of the sample space:

   $$P(\{Low, High\}) = P(Y = Low) + P(Y = High) = 0.33 + 0.66 = 1$$

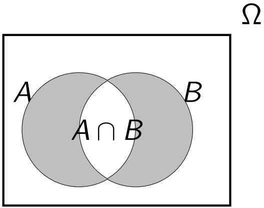
3. **Additivity**: For disjoint events in $\mathcal{F}$,

   $$P(\{Low\} \cup \{High\}) = P(\{Low\}) + P(\{High\}) = 0.33 + 0.66 = 1$$

   $$P(\{Low\} \cup \emptyset) = P(\{Low\}) + P(\emptyset) = 0.33 + 0 = 0.33$$

   $$P(\{High\} \cup \emptyset) = P(\{High\}) + P(\emptyset) = 0.66 + 0 = 0.66$$

Venn Diagram, Conditional Probability, and Independence



Conditional Probability:

- The conditional probability of A given B is defined as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{for } P(B) > 0$$

Independence:

- Two events A and B are independent if:

$$P(A \cap B) = P(A)P(B)$$

Exogenous D

Binary Outcome with New Variable D: $D \sim$ **Bernoulli**(0.5)

i	Y	D
1	Low	0
2	Low	1
3	Low	0
4	High	1
5	High	0
6	Low	1

- $P(D = 1) = 0.5$, $P(Y = High) = 0.33$
- Joint probability: $P(Y = High \cap D = 1) = \frac{1}{6} = 0.167$
- Independence check:

$$P(Y = High)P(D = 1) = 0.33 \times 0.5 = 0.165$$

- Independence check:

$$P(Y = High|P(D = 1) = P(Y = High) = 0.33$$

Endogenous Assignment of D

Binary Outcome with Endogenous D: $D \sim$ Bernoulli$(p(Y))$, where $p(Y^{High}) > p(Y^{Low})$

i	Y	D
1	Low	0
2	High	1
3	Low	0
4	High	1
5	Low	1
6	High	0

- $P(Y = High) = 0.5, \; P(D = 1) = 0.5$
-
- $P(Y = High \cap D = 1) = 0.33 \neq 0.5 * 0.5$
- $P(Y = High | D = 1) = \frac{0.33}{0.5} = 0.67 = \frac{2}{3}$

Law of Total Probability:

$$P(Y = H) = P(Y = H|D = 1)P(D = 1) + P(Y = H|D = 0)P(D = 0)$$

Using the values from our example:

- $P(Y = High|D = 1) = \frac{2}{3} = 0.67$
- $P(D = 1) = 0.5$
- $P(Y = High|D = 0) = \frac{1}{3} = 0.33$
- $P(D = 0) = 0.5$

Therefore, applying the law of total probability:

$$P(Y = High) = 0.67 \times 0.5 + 0.33 \times 0.5 = 0.335 + 0.165 = 0.5$$

This matches the marginal probability $P(Y = High) = 0.5$ calculated earlier.

Expectation

Expectation (Discrete Variable):

- For a discrete random variable X with probability mass function $p(x)$:

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x \cdot p(x)$$

where $\mathcal{X}$ is the set of possible values of X.

Expectation

Expectation (Discrete Variable):

- For a discrete random variable X with probability mass function $p(x)$:

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x \cdot p(x)$$

where $\mathcal{X}$ is the set of possible values of X.

Example: Bernoulli(0,1)

- $X \in \{0, 1\} \quad P(X = 1) = 0.5$
- $\mathbb{E}[X] = 1 \cdot 0.5 + 0 \cdot 0.5 = 0.5$

Expectation

Expectation (Discrete Variable):

- For a discrete random variable X with probability mass function $p(x)$:

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x \cdot p(x)$$

where $\mathcal{X}$ is the set of possible values of X.

Example: Bernoulli(0,1)

- $X \in \{0, 1\} \quad P(X = 1) = 0.5$
- $\mathbb{E}[X] = 1 \cdot 0.5 + 0 \cdot 0.5 = 0.5$

Expectation (Continuous Variable):

- For a continuous random variable X with probability density function $f(x)$:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

where $f(x)$ is the PDF of X.

Expectation

Expectation (Discrete Variable):

- For a discrete random variable X with probability mass function $p(x)$:

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x \cdot p(x)$$

where $\mathcal{X}$ is the set of possible values of X.

Example: Bernoulli(0,1)

- $X \in \{0, 1\} \quad P(X = 1) = 0.5$
- $\mathbb{E}[X] = 1 \cdot 0.5 + 0 \cdot 0.5 = 0.5$

Expectation (Continuous Variable):

- For a continuous random variable X with probability density function $f(x)$:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

where $f(x)$ is the PDF of X.

Example: Normal(0,1)

- $X \sim N(0, 1)$, where the PDF is $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$
- $\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 0$

Conditional Expectation

Conditional Expectation (Discrete Variable):

- For a discrete random variable X with conditional probability $P(X = x|Y = y)$:

$$\mathbb{E}[X|Y = y] = \sum_{x \in \mathcal{X}} x \cdot P(X = x|Y = y)$$

Conditional Expectation

Conditional Expectation (Discrete Variable):

- For a discrete random variable X with conditional probability $P(X = x|Y = y)$:

$$\mathbb{E}[X|Y = y] = \sum_{x \in \mathcal{X}} x \cdot P(X = x|Y = y)$$

Example: Bernoulli(0,1)

- $X \in \{0, 1\}$ with $P(X = 1|Y = y) = p(y)$
- $\mathbb{E}[X|Y = y] = 1 \cdot p(y) + 0 \cdot (1 - p(y)) = p(y)$

Conditional Expectation

Conditional Expectation (Discrete Variable):

- For a discrete random variable X with conditional probability $P(X = x|Y = y)$:

$$\mathbb{E}[X|Y = y] = \sum_{x \in \mathcal{X}} x \cdot P(X = x|Y = y)$$

Conditional Expectation (Continuous Variable):

- For a continuous random variable X with conditional density $f(x|y)$:

$$\mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} x \cdot f(x|y)\, dx$$

Example: Bernoulli(0,1)

- $X \in \{0, 1\}$ with $P(X = 1|Y = y) = p(y)$
- $\mathbb{E}[X|Y = y] = 1 \cdot p(y) + 0 \cdot (1 - p(y)) = p(y)$

Conditional Expectation

Conditional Expectation (Discrete Variable):

- For a discrete random variable X with conditional probability $P(X = x|Y = y)$:

$$\mathbb{E}[X|Y = y] = \sum_{x \in \mathcal{X}} x \cdot P(X = x|Y = y)$$

Example: Bernoulli(0,1)

- $X \in \{0, 1\}$ with $P(X = 1|Y = y) = p(y)$
- $\mathbb{E}[X|Y = y] = 1 \cdot p(y) + 0 \cdot (1 - p(y)) = p(y)$

Conditional Expectation (Continuous Variable):

- For a continuous random variable X with conditional density $f(x|y)$:

$$\mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} x \cdot f(x|y)\, dx$$

Example: Normal(0,1)

- Suppose $X \sim N(0, 1)$, and $f(x|y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-y)^2}{2}}$ (a shifted normal)
- $\mathbb{E}[X|Y = y] = y$

Exogenous D

Discrete Outcome with Exogenous D: $D \sim$ Bernoulli(0.5)

i	Y	D
1	3	0
2	8	1
3	5	1
4	7	0
5	4	1
6	9	0

Exogenous D

Discrete Outcome with Exogenous D: $D \sim$ **Bernoulli**(0.5)

i	Y	D
1	3	0
2	8	1
3	5	1
4	7	0
5	4	1
6	9	0

- **Expectation:**

$$\mathbb{E}[Y] = \frac{3+8+5+9+4+7}{6} = 6$$

Discrete Outcome with Exogenous D: $D \sim$ **Bernoulli**(0.5)

i	Y	D
1	3	0
2	8	1
3	5	1
4	7	0
5	4	1
6	9	0

- **Expectation:**

$$\mathbb{E}[Y] = \frac{3+8+5+9+4+7}{6} = 6$$

- **Conditional Expectation $\mathbb{E}[Y|D=1]$:**

$$\mathbb{E}[Y|D=1] = \frac{8+5+4}{3} \approx 6$$

Endogenous D

Endogenous Selection D**:** $D \sim$ **Bernoulli**$(p(Y))$**, where** $p(Y^{High}) > p(Y^{Low})$

i	Y	D
1	3	0
2	8	1
3	5	0
4	9	1
5	4	0
6	7	1

- **Expectation:**

$$\mathbb{E}[Y] = \frac{3+8+5+9+4+7}{6} = 6$$

- **Conditional Expectation $\mathbb{E}[Y|D=1]$:**

$$\mathbb{E}[Y|D=1] = \frac{8+9+7}{3} = 8$$

Law of Iterated Expectations:

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|D]] = \sum_{d \in \mathcal{D}} \mathbb{E}[Y|D = d] \cdot P(D = d) \quad \text{(for discrete D)}$$

i	Y	D
1	3	0
2	8	1
3	5	0
4	9	1
5	4	0
6	7	1

Given $P(D = 1) = 0.5$, the conditional and total expectations are:

$$\mathbb{E}[Y|D = 0] = \frac{3 + 5 + 4}{3} = 4, \quad \mathbb{E}[Y|D = 1] = \frac{8 + 9 + 7}{3} = 8$$

$$\mathbb{E}[Y] = 0.5 \cdot 4 + 0.5 \cdot 8 = 6$$

Content

Potential Outcomes and Parallel Futures

Borges Quote

"Cada vez que un hombre se enfrenta a diversas alternativas, opta por una y elimina las otras; [...] Crea, así, diversos futuros, diversos tiempos, que también proliferan y se bifurcan."
— *Borges, El jardín de los senderos que se bifurcan*

Potential Outcomes and Parallel Futures

Borges Quote

"Cada vez que un hombre se enfrenta a diversas alternativas, opta por una y elimina las otras; [...] Crea, así, diversos futuros, diversos tiempos, que también proliferan y se bifurcan."
— *Borges, El jardín de los senderos que se bifurcan*

Potential Outcomes Framework:

Potential Outcomes and Parallel Futures

Borges Quote

> "Cada vez que un hombre se enfrenta a diversas alternativas, opta por una y elimina las otras; [...] Crea, así, diversos futuros, diversos tiempos, que también proliferan y se bifurcan."
> — *Borges, El jardín de los senderos que se bifurcan*

Potential Outcomes Framework:

- $Y(0)$: The outcome that would occur if the individual does not receive the treatment ($D = 0$).

Potential Outcomes and Parallel Futures

Borges Quote

"Cada vez que un hombre se enfrenta a diversas alternativas, opta por una y elimina las otras; [...] Crea, así, diversos futuros, diversos tiempos, que también proliferan y se bifurcan."
— *Borges, El jardín de los senderos que se bifurcan*

Potential Outcomes Framework:

- $Y(0)$: The outcome that would occur if the individual does not receive the treatment ($D = 0$).
- $Y(1)$: The outcome that would occur if the individual receives the treatment ($D = 1$).

Potential Outcomes and Parallel Futures

Borges Quote

"Cada vez que un hombre se enfrenta a diversas alternativas, opta por una y elimina las otras; [...] Crea, así, diversos futuros, diversos tiempos, que también proliferan y se bifurcan."
— *Borges, El jardín de los senderos que se bifurcan*

Potential Outcomes Framework:

- $Y(0)$: The outcome that would occur if the individual does not receive the treatment ($D = 0$).
- $Y(1)$: The outcome that would occur if the individual receives the treatment ($D = 1$).
- The observed outcome Y_i is:

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

Potential Outcomes and Parallel Futures

Borges Quote

"Cada vez que un hombre se enfrenta a diversas alternativas, opta por una y elimina las otras; [...] Crea, así, diversos futuros, diversos tiempos, que también proliferan y se bifurcan."
— *Borges, El jardín de los senderos que se bifurcan*

Potential Outcomes Framework:

- $Y(0)$: The outcome that would occur if the individual does not receive the treatment ($D = 0$).
- $Y(1)$: The outcome that would occur if the individual receives the treatment ($D = 1$).
- The observed outcome Y_i is:

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

Treatment Effects and the Fundamental Problem of Causal Inference

i	$Y(0)$	$Y(1)$	D	Y	τ_i
1	3	5	0	3	2
2	4	8	1	8	4
3	5	6	0	5	1
4	7	9	1	9	2
5	4	5	0	4	1
6	3	7	1	7	4

Treatment Effects and the Fundamental Problem of Causal Inference

i	$Y(0)$	$Y(1)$	D	Y	τ_i
1	3	5	0	3	2
2	4	8	1	8	4
3	5	6	0	5	1
4	7	9	1	9	2
5	4	5	0	4	1
6	3	7	1	7	4

$$\tau_i = Y_i(1) - Y_i(0)$$

Treatment Effects and the Fundamental Problem of Causal Inference

i	$Y(0)$	$Y(1)$	D	Y	τ_i
1	3	5	0	3	2
2	4	8	1	8	4
3	5	6	0	5	1
4	7	9	1	9	2
5	4	5	0	4	1
6	3	7	1	7	4

$$\tau_i = Y_i(1) - Y_i(0)$$

Fundamental Problem of Causal Inference

We can never observe both potential outcomes $Y(0)$ and $Y(1)$ for the same individual at the same time making it impossible to directly observe the true treatment effect τ_i for any single individual.

Average Treatment Effect (ATE)

Definition of ATE:

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

Average Treatment Effect (ATE)

Definition of ATE:

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

Computation in the Example:

i	$Y(0)$	$Y(1)$	D	Y	τ_i
1	3	5	0	3	2
2	4	8	1	8	4
3	5	6	0	5	1
4	7	9	1	9	2
5	4	5	0	4	1
6	3	7	1	7	4

Average Treatment Effect (ATE)

Definition of ATE:

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

Computation in the Example:

i	$Y(0)$	$Y(1)$	D	Y	τ_i
1	3	5	0	3	2
2	4	8	1	8	4
3	5	6	0	5	1
4	7	9	1	9	2
5	4	5	0	4	1
6	3	7	1	7	4

$$ATE = \frac{2 + 4 + 1 + 2 + 1 + 4}{6} = 2.34$$

Average Treatment Effect on the Treated (ATT)

Definition of ATT:

$$ATT = \mathbb{E}[Y(1) - Y(0) | D = 1]$$

Definition of ATT:

$$ATT = \mathbb{E}[Y(1) - Y(0) | D = 1]$$

Computation in the Example:

i	$Y(0)$	$Y(1)$	D	Y	τ_i
1	3	5	0	3	2
2	4	8	1	8	4
3	5	6	0	5	1
4	7	9	1	9	2
5	4	5	0	4	1
6	3	7	1	7	4

Average Treatment Effect on the Treated (ATT)

Definition of ATT:

$$ATT = \mathbb{E}[Y(1) - Y(0)|D = 1]$$

Computation in the Example:

i	$Y(0)$	$Y(1)$	D	Y	τ_i
1	3	5	0	3	2
2	4	8	1	8	4
3	5	6	0	5	1
4	7	9	1	9	2
5	4	5	0	4	1
6	3	7	1	7	4

$$ATT = \frac{4 + 2 + 4}{3} = 3.33$$

Average Treatment Effect on the Untreated (ATU)

Definition of ATU:

$$ATU = \mathbb{E}[Y(1) - Y(0)|D = 0]$$

Average Treatment Effect on the Untreated (ATU)

Definition of ATU:

$$ATU = \mathbb{E}[Y(1) - Y(0)|D = 0]$$

Computation in the Example:

i	$Y(0)$	$Y(1)$	D	Y	τ_i
1	3	5	0	3	2
2	4	8	1	8	4
3	5	6	0	5	1
4	7	9	1	9	2
5	4	5	0	4	1
6	3	7	1	7	4

Average Treatment Effect on the Untreated (ATU)

Definition of ATU:

$$ATU = \mathbb{E}[Y(1) - Y(0)|D = 0]$$

Computation in the Example:

i	$Y(0)$	$Y(1)$	D	Y	τ_i
1	3	5	0	3	2
2	4	8	1	8	4
3	5	6	0	5	1
4	7	9	1	9	2
5	4	5	0	4	1
6	3	7	1	7	4

$$ATU = \frac{2+1+1}{3} = 1.33$$

Selection Bias

Naive Comparison:

$$\tau_{naive} = \mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0]$$

Selection Bias

Naive Comparison:

$$\tau_{naive} = \mathbb{E}[Y|D = 1] - \mathbb{E}[Y|D = 0]$$

Naive Comparison Decomposition:

$$\begin{aligned}\tau_{naive} &= \mathbb{E}[Y(1)|D = 1] - \mathbb{E}[Y(0)|D = 0] \\ &= \underbrace{(\mathbb{E}[Y(1)|D = 1] - \mathbb{E}[Y(0)|D = 1])}_{\text{ATT}} + \underbrace{(\mathbb{E}[Y(0)|D = 1] - \mathbb{E}[Y(0)|D = 0])}_{\text{Selection Bias}}\end{aligned}$$

Selection Bias

Naive Comparison:

$$\tau_{naive} = \mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0]$$

Naive Comparison Decomposition:

$$\begin{aligned}\tau_{naive} &= \mathbb{E}[Y(1)|D=1] - \mathbb{E}[Y(0)|D=0] \\ &= \underbrace{(\mathbb{E}[Y(1)|D=1] - \mathbb{E}[Y(0)|D=1])}_{\text{ATT}} + \underbrace{(\mathbb{E}[Y(0)|D=1] - \mathbb{E}[Y(0)|D=0])}_{\text{Selection Bias}}\end{aligned}$$

ATT identification

$$\mathbb{E}[Y(1)|D=1] - \mathbb{E}[Y(0)|D=1] = \mathbb{E}[Y(1) - Y(0)|D=1] = ATT$$

Causal Inference = How to overcome Selection Bias?

Naive Estimator and Selection Bias - Example

i	$Y(0)$	$Y(1)$	D	Y	$\tau_i = Y(1) - Y(0)$
1	3	5	0	3	2
2	4	8	1	8	4
3	5	6	0	5	1
4	7	9	1	9	2
5	4	5	0	4	1
6	3	7	1	7	4

Naive Estimator and Selection Bias - Example

i	$Y(0)$	$Y(1)$	D	Y	$\tau_i = Y(1) - Y(0)$
1	3	5	0	3	2
2	4	8	1	8	4
3	5	6	0	5	1
4	7	9	1	9	2
5	4	5	0	4	1
6	3	7	1	7	4

$$\tau_{naive} = \frac{8+9+7}{3} - \frac{3+5+4}{3} = 8 - 4 = 4$$

Naive Estimator and Selection Bias - Example

i	$Y(0)$	$Y(1)$	D	Y	$\tau_i = Y(1) - Y(0)$
1	3	5	0	3	2
2	4	8	1	8	4
3	5	6	0	5	1
4	7	9	1	9	2
5	4	5	0	4	1
6	3	7	1	7	4

$$\tau_{naive} = \frac{8+9+7}{3} - \frac{3+5+4}{3} = 8 - 4 = 4$$

$$Selection\ Bias = \mathbb{E}[Y(0)|D=1] - \mathbb{E}[Y(0)|D=0] = \frac{4+7+3}{3} - 4 = 4.67 - 4 = 0.67$$

Naive Estimator and Selection Bias - Example

i	$Y(0)$	$Y(1)$	D	Y	$\tau_i = Y(1) - Y(0)$
1	3	5	0	3	2
2	4	8	1	8	4
3	5	6	0	5	1
4	7	9	1	9	2
5	4	5	0	4	1
6	3	7	1	7	4

$$\tau_{naive} = \frac{8+9+7}{3} - \frac{3+5+4}{3} = 8 - 4 = 4$$

$$Selection\ Bias = \mathbb{E}[Y(0)|D=1] - \mathbb{E}[Y(0)|D=0] = \frac{4+7+3}{3} - 4 = 4.67 - 4 = 0.67$$

$$\tau_{ATT} = 4 - 0.67 = 3.33$$

Randomization Solves Selection Bias

Naive Estimator:

$$\tau_{naive} = \mathbb{E}[Y(1)|D = 1] - \mathbb{E}[Y(0)|D = 0]$$

Randomization Solves Selection Bias

Naive Estimator:

$$\tau_{naive} = \mathbb{E}[Y(1)|D = 1] - \mathbb{E}[Y(0)|D = 0]$$

If $Y(0)$ is independent of D:

$$\mathbb{E}[Y(0)|D = 0] = \mathbb{E}[Y(0)|D = 1]$$

Randomization Solves Selection Bias

Naive Estimator:

$$\tau_{naive} = \mathbb{E}[Y(1)|D=1] - \mathbb{E}[Y(0)|D=0]$$

If $Y(0)$ is independent of D:

$$\mathbb{E}[Y(0)|D=0] = \mathbb{E}[Y(0)|D=1]$$

Then:

$$\tau_{naive} = \mathbb{E}[Y(1)|D=1] - \mathbb{E}[Y(0)|D=1] = \mathbb{E}[Y(1) - Y(0)|D=1] = ATT$$

Randomization Solves Selection Bias

Naive Estimator:

$$\tau_{naive} = \mathbb{E}[Y(1)|D = 1] - \mathbb{E}[Y(0)|D = 0]$$

If $Y(0)$ is independent of D:

$$\mathbb{E}[Y(0)|D = 0] = \mathbb{E}[Y(0)|D = 1]$$

Then:

$$\tau_{naive} = \mathbb{E}[Y(1)|D = 1] - \mathbb{E}[Y(0)|D = 1] = \mathbb{E}[Y(1) - Y(0)|D = 1] = ATT$$

If D is also independent of $Y(1)$:

$$\mathbb{E}[Y(1) - Y(0)|D = 1] = \mathbb{E}[Y(1) - Y(0)] = ATE$$

Randomization Solves Selection Bias

Naive Estimator:

$$\tau_{naive} = \mathbb{E}[Y(1)|D=1] - \mathbb{E}[Y(0)|D=0]$$

If $Y(0)$ is independent of D:

$$\mathbb{E}[Y(0)|D=0] = \mathbb{E}[Y(0)|D=1]$$

Then:

$$\tau_{naive} = \mathbb{E}[Y(1)|D=1] - \mathbb{E}[Y(0)|D=1] = \mathbb{E}[Y(1) - Y(0)|D=1] = ATT$$

If D is also independent of $Y(1)$:

$$\mathbb{E}[Y(1) - Y(0)|D=1] = \mathbb{E}[Y(1) - Y(0)] = ATE$$

Conclusion: If $D \perp Y(0), Y(1)$ (Unconfoundedness), then:

$$ATE = ATT = ATU$$

Estimator Results: $\hat{\tau}_{naive}$, ATE, ATT, and ATU

i	$Y(0)$	$Y(1)$	D	Y	$\tau_i = Y(1) - Y(0)$
1	3	5	0	3	2
2	4	8	1	8	4
3	5	6	1	6	1
4	7	9	0	7	2
5	4	5	1	5	1
6	3	7	0	3	4

Estimator Results: $\hat{\tau}_{naive}$, ATE, ATT, and ATU

i	$Y(0)$	$Y(1)$	D	Y	$\tau_i = Y(1) - Y(0)$
1	3	5	0	3	2
2	4	8	1	8	4
3	5	6	1	6	1
4	7	9	0	7	2
5	4	5	1	5	1
6	3	7	0	3	4

$$\hat{\tau}_{naive} = \mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0] = 6.33 - 4.33 = 2$$

Estimator Results: $\hat{\tau}_{naive}$, ATE, ATT, and ATU

i	$Y(0)$	$Y(1)$	D	Y	$\tau_i = Y(1) - Y(0)$
1	3	5	0	3	2
2	4	8	1	8	4
3	5	6	1	6	1
4	7	9	0	7	2
5	4	5	1	5	1
6	3	7	0	3	4

$$\hat{\tau}_{naive} = \mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0] = 6.33 - 4.33 = 2$$

$$ATT = \mathbb{E}[Y(1) - Y(0)|D=1] = 2$$

Estimator Results: $\hat{\tau}_{naive}$, ATE, ATT, and ATU

i	$Y(0)$	$Y(1)$	D	Y	$\tau_i = Y(1) - Y(0)$
1	3	5	0	3	2
2	4	8	1	8	4
3	5	6	1	6	1
4	7	9	0	7	2
5	4	5	1	5	1
6	3	7	0	3	4

$$\hat{\tau}_{naive} = \mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0] = 6.33 - 4.33 = 2$$

$$ATT = \mathbb{E}[Y(1) - Y(0)|D=1] = 2$$

$$ATU = \mathbb{E}[Y(1) - Y(0)|D=0] = 2.67$$

Estimator Results: $\hat{\tau}_{naive}$, ATE, ATT, and ATU

i	$Y(0)$	$Y(1)$	D	Y	$\tau_i = Y(1) - Y(0)$
1	3	5	0	3	2
2	4	8	1	8	4
3	5	6	1	6	1
4	7	9	0	7	2
5	4	5	1	5	1
6	3	7	0	3	4

$$\hat{\tau}_{naive} = \mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0] = 6.33 - 4.33 = 2$$

$$ATT = \mathbb{E}[Y(1) - Y(0)|D=1] = 2$$

$$ATU = \mathbb{E}[Y(1) - Y(0)|D=0] = 2.67$$

$$ATE = \mathbb{E}[Y(1) - Y(0)] = 2.33$$

References I

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American Economic Review*, 93(1):113–132.

Acemoglu, D., Johnson, S., and Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(5):1369–1401.

Angrist, J. D. and Krueger, A. B. (1991). Does compulsory schooling affect schooling and earnings? *Quarterly Journal of Economics*, 106(4):979–1014.

Banerjee, A. and Duflo, E. (2011). *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. PublicAffairs.

Bauer, P. C. (2015). Negative experiences and trust: A causal analysis of the effects of victimization on generalized trust. *European Sociological Review*, 31(4):397–417.

References II

Card, D. (1990). The impact of the mariel boatlift on the miami labor market. *Industrial and Labor Relations Review*, 43(2):245–257.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68.

Fisher, R. A. (1935). *The Design of Experiments*. Oliver and Boyd, Edinburgh, UK.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science*, 5(4):465–472.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688.

References III

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.