150C Causal Inference

Potential Outcomes Model for Causal Inference

Jonathan Mummolo

Stanford University

Purpose, Scope, and Examples

Goal in causal inference is to assess the causal effect of some potential **cause** (e.g. an institution, intervention, policy, or event) on some **outcome**.

Purpose, Scope, and Examples

Goal in causal inference is to assess the causal effect of some potential **cause** (e.g. an institution, intervention, policy, or event) on some **outcome**.

Examples of such research questions include...

What is the effect of:

- political institutions on corruption?

Purpose, Scope, and Examples

Goal in causal inference is to assess the causal effect of some potential **cause** (e.g. an institution, intervention, policy, or event) on some **outcome**.

Examples of such research questions include...

What is the effect of:

- political institutions on corruption?
- voting technology on voting fraud?

Purpose, Scope, and Examples

Goal in causal inference is to assess the causal effect of some potential **cause** (e.g. an institution, intervention, policy, or event) on some **outcome**.

Examples of such research questions include...

What is the effect of:

- political institutions on corruption?
- voting technology on voting fraud?
- incumbency status on vote shares?

Purpose, Scope, and Examples

Goal in causal inference is to assess the causal effect of some potential **cause** (e.g. an institution, intervention, policy, or event) on some **outcome**.

Examples of such research questions include...

What is the effect of:

- political institutions on corruption?
- voting technology on voting fraud?
- incumbency status on vote shares?
- peacekeeping missions on peace?

Purpose, Scope, and Examples

Goal in causal inference is to assess the causal effect of some potential **cause** (e.g. an institution, intervention, policy, or event) on some **outcome.**

Examples of such research questions include...

What is the effect of:

- political institutions on corruption?
- voting technology on voting fraud?
- incumbency status on vote shares?
- peacekeeping missions on peace?
- mass media on voter preferences?

Purpose, Scope, and Examples

Goal in causal inference is to assess the causal effect of some potential **cause** (e.g. an institution, intervention, policy, or event) on some **outcome**.

Examples of such research questions include...

What is the effect of:

- political institutions on corruption?
- voting technology on voting fraud?
- incumbency status on vote shares?
- peacekeeping missions on peace?
- mass media on voter preferences?
- church attendance on turnout?

What Do We Mean by Causal Inference?

As in all statistics, we must begin with a model of the reality we are interested in studying, such as:

$$y_i = \alpha + \tau D_i + X_i\beta + \epsilon_i$$

Key problems with regression **when used as a causal inference strategy**:

What Do We Mean by Causal Inference?

As in all statistics, we must begin with a model of the reality we are interested in studying, such as:

$$y_i = \alpha + \tau D_i + X_i \beta + \epsilon_i$$

Key problems with regression **when used as a causal inference strategy**:

- Endogeneity and omitted variable bias

What Do We Mean by Causal Inference?

As in all statistics, we must begin with a model of the reality we are interested in studying, such as:

$$y_i = \alpha + \tau D_i + X_i \beta + \epsilon_i$$

Key problems with regression **when used as a causal inference strategy**:

- Endogeneity and omitted variable bias
- Misspecified functional form

What Do We Mean by Causal Inference?

As in all statistics, we must begin with a model of the reality we are interested in studying, such as:

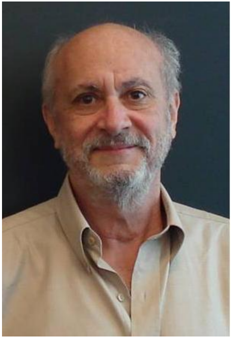
$$y_i = \alpha + \tau D_i + X_i \beta + \epsilon_i$$

Key problems with regression **when used as a causal inference strategy**:

- Endogeneity and omitted variable bias
- Misspecified functional form
- Heterogenous treatment effects (when treatment assignment process unknown)

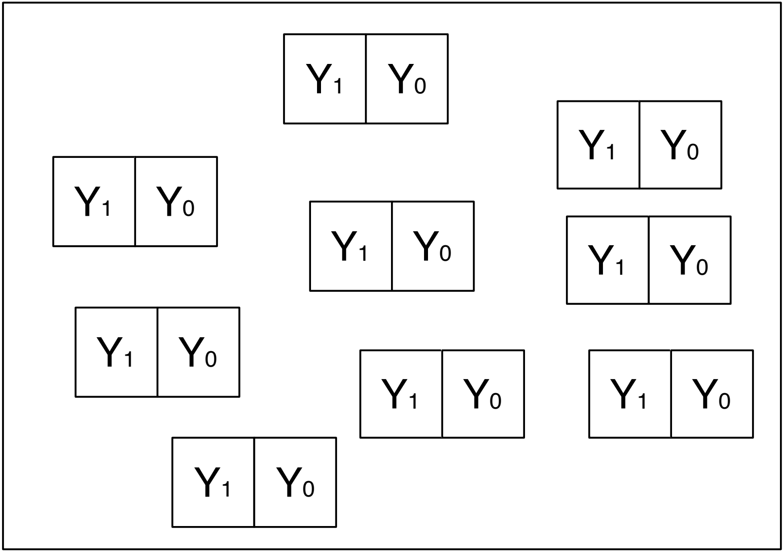
Neyman-Rubin Potential Outcomes Model

Neyman

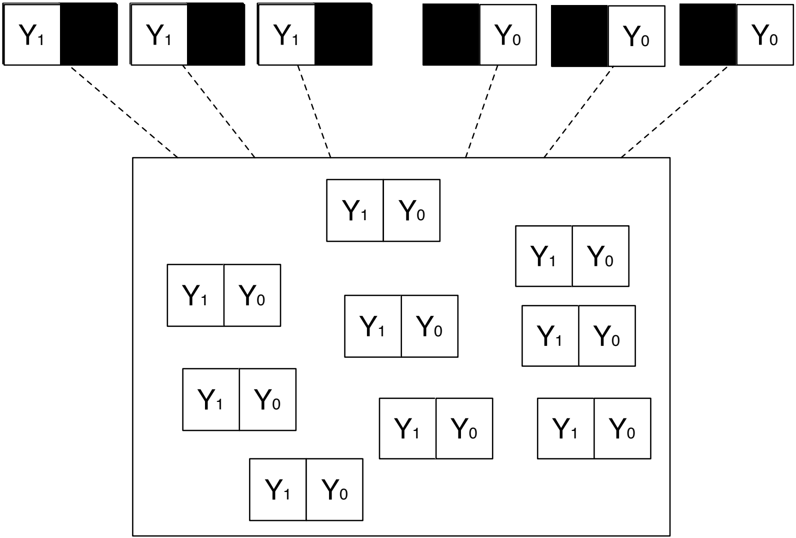


Rubin

Neyman Urn Model



Neyman Urn Model



Causality with Potential Outcomes

Definition (Treatment)

D_i: Indicator of treatment intake for *unit i*

$$D_i = \begin{cases} 1 & \text{if unit } i \text{ received the treatment} \\ 0 & \text{otherwise.} \end{cases}$$

Definition (Outcome)

Y_i: Observed outcome variable of interest for unit i. The treatment occurs temporally before the outcome.

Definition (Potential Outcomes)

Y_{0i} and Y_{1i}: Potential outcomes for unit i

$$Y_{di} = \begin{cases} Y_{1i} & \text{Potential outcome for unit } i \text{ with treatment} \\ Y_{0i} & \text{Potential outcome for unit } i \text{ without treatment} \end{cases}$$

Causality with Potential Outcomes

Definition (Causal Effect)

Causal effect of the treatment on the outcome for unit i is the difference between its two potential outcomes:

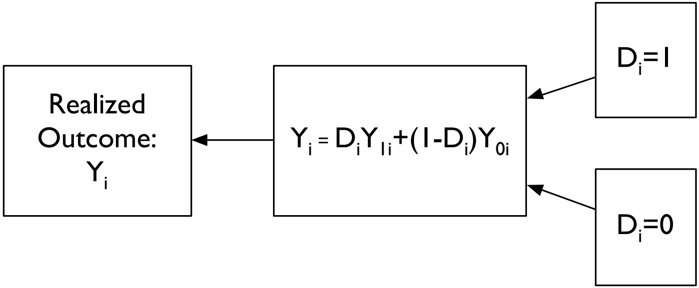
$$\tau_i = Y_{1i} - Y_{0i}$$

Assumption

Observed outcomes are realized as

$$Y_i = D_i \cdot Y_{1i} + (1 - D_i) \cdot Y_{0i} \text{ so } Y_i = \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases}$$

Causal Inference as a Missing Data Problem



Definition (Fundamental Problem of Causal Inference)

We cannot observe both potential outcomes. So how can we calculate $\tau_i = Y_{1i} - Y_{0i}$?

Fundamental Problem of Causal Inference

Imagine a study population with 4 units:

i	D_i	Y_{1i}	Y_{0i}	τ_i
1	1	10	4	6
2	1	1	2	-1
3	0	3	3	0
4	0	5	2	3

Fundamental Problem of Causal Inference

Imagine a study population with 4 units:

i	D_i	Y_{1i}	Y_{0i}	τ_i
1	1	10	4	6
2	1	1	2	-1
3	0	3	3	0
4	0	5	2	3

What do we observe?

Fundamental Problem of Causal Inference

Imagine a study population with 4 units:

i	D_i	Y_{1i}	Y_{0i}	τ_i
1	1	10	4	6
2	1	1	2	-1
3	0	3	3	0
4	0	5	2	3

What do we observe?

i	D_i	Y_{1i}	Y_{0i}	τ_i	Y_i
1	1	10	?	?	10
2	1	1	?	?	1
3	0	?	3	?	3
4	0	?	2	?	2

Causal inference is difficult because it involves missing data.

Causal Inference as a Missing Data Problem

How can we calculate $\tau_i = Y_{1i} - Y_{0i}$?

- Homogeneity is one solution:
  - If $\{Y_{1i}, Y_{0i}\}$ is constant across individuals, then cross-sectional comparisons will recover τ_i
  - If $\{Y_{1i}, Y_{0i}\}$ is constant across time, then before and after comparisons will recover τ_i

Causal Inference as a Missing Data Problem

How can we calculate $\tau_i = Y_{1i} - Y_{0i}$?

- Homogeneity is one solution:
  - If $\{Y_{1i}, Y_{0i}\}$ is constant across individuals, then cross-sectional comparisons will recover τ_i
  - If $\{Y_{1i}, Y_{0i}\}$ is constant across time, then before and after comparisons will recover τ_i

In social phenomena, unfortunately, homogeneity is very rare.

Other Assumptions

Assumption

Observed outcomes are realized as

$$Y_i = D_i \cdot Y_{1i} + (1 - D_i) \cdot Y_{0i}$$

Other Assumptions

Assumption

Observed outcomes are realized as

$$Y_i = D_i \cdot Y_{1i} + (1 - D_i) \cdot Y_{0i}$$

- Embedded in this formulation is the assumption that potential outcomes for unit i are unaffected by treatment assignment for unit j.
- Assumption known by several names:
  - **S**table **U**nit **T**reatment **V**alue **A**ssumption (SUTVA)
  - No interference
  - Individualized Treatment Response
- Examples: vaccination, fertilizer on plot yield, communication

Potential Outcomes with Interference

Let $\mathbf{D} = \{D_i, D_j\}$ be the set of vectors of treatment assignments for two units i (me) and j (you).

How many elements in $\mathbf{D}$?

Potential Outcomes with Interference

Let $\mathbf{D} = \{D_i, D_j\}$ be the set of vectors of treatment assignments for two units i (me) and j (you).

How many elements in $\mathbf{D}$?

$$\mathbf{D} = \{(D_i = 0, D_j = 0), (D_i = 1, D_j = 0), (D_i = 0, D_j = 1), (D_i = 1, D_j = 1)\}$$

Potential Outcomes with Interference

Let $\mathbf{D} = \{D_i, D_j\}$ be the set of vectors of treatment assignments for two units i (me) and j (you).

How many elements in $\mathbf{D}$?

$$\mathbf{D} = \{(D_i = 0, D_j = 0), (D_i = 1, D_j = 0), (D_i = 0, D_j = 1), (D_i = 1, D_j = 1)\}$$

How many potential outcomes for unit i?

Potential Outcomes with Interference

Let $\mathbf{D} = \{D_i, D_j\}$ be the set of vectors of treatment assignments for two units i (me) and j (you).

How many elements in $\mathbf{D}$?

$$\mathbf{D} = \{(D_i = 0, D_j = 0), (D_i = 1, D_j = 0), (D_i = 0, D_j = 1), (D_i = 1, D_j = 1)\}$$

How many potential outcomes for unit i?

$$Y_{1i}(\mathbf{D}) = \begin{cases} Y_{1i}(1,1) \\ Y_{1i}(1,0) \end{cases} \quad Y_{0i}(\mathbf{D}) = \begin{cases} Y_{0i}(0,1) \\ Y_{0i}(0,0) \end{cases}$$

Potential Outcomes with Interference

How many causal effects for unit i?

Potential Outcomes with Interference

How many causal effects for unit i?

$$\tau_i(\mathbf{D}) = \begin{cases} Y_{1i}(1,1) - Y_{0i}(0,0) \\ Y_{1i}(1,1) - Y_{0i}(0,1) \\ Y_{1i}(1,0) - Y_{0i}(0,0) \\ Y_{1i}(1,0) - Y_{0i}(0,1) \\ Y_{1i}(1,1) - Y_{1i}(1,0) \\ Y_{0i}(0,1) - Y_{0i}(0,0) \end{cases}$$

Potential Outcomes with Interference

How many causal effects for unit i?

$$\tau_i(\mathbf{D}) = \begin{cases} Y_{1i}(1,1) - Y_{0i}(0,0) \\ Y_{1i}(1,1) - Y_{0i}(0,1) \\ Y_{1i}(1,0) - Y_{0i}(0,0) \\ Y_{1i}(1,0) - Y_{0i}(0,1) \\ Y_{1i}(1,1) - Y_{1i}(1,0) \\ Y_{0i}(0,1) - Y_{0i}(0,0) \end{cases}$$

How many potential outcomes are observed for unit i?

Potential Outcomes with Interference

How many causal effects for unit i?

$$\tau_i(\mathbf{D}) = \begin{cases} Y_{1i}(1,1) - Y_{0i}(0,0) \\ Y_{1i}(1,1) - Y_{0i}(0,1) \\ Y_{1i}(1,0) - Y_{0i}(0,0) \\ Y_{1i}(1,0) - Y_{0i}(0,1) \\ Y_{1i}(1,1) - Y_{1i}(1,0) \\ Y_{0i}(0,1) - Y_{0i}(0,0) \end{cases}$$

How many potential outcomes are observed for unit i?

Since we only observe one of the four potential outcomes, the missing data problem for causal inference is even more severe.

Potential Outcomes with Interference

The No Interference assumption states that unit i's potential outcomes depend on D_i, not $\mathbf{D}$:

$Y_{1i}(1,1) = Y_{1i}(1,0)$ and $Y_{0i}(0,1) = Y_{0i}(0,0)$

This assumption furthermore allows us to define the effect for unit i as $\tau_i = Y_{1i} - Y_{0i}$.

Potential Outcomes with Interference

The No Interference assumption states that unit i's potential outcomes depend on D_i, not $\mathbf{D}$:

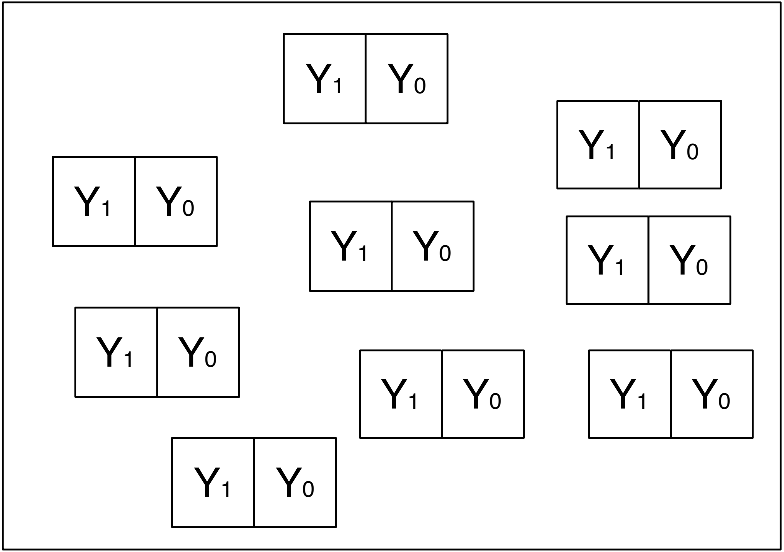
$Y_{1i}(1,1) = Y_{1i}(1,0)$ and $Y_{0i}(0,1) = Y_{0i}(0,0)$

This assumption furthermore allows us to define the effect for unit i as $\tau_i = Y_{1i} - Y_{0i}$.

No interference is an example of an **exclusion restriction**. We rely on outside information to rule out the possibility of certain causal effects (e.g. you taking the treatment has no effect on my potential outcomes).

Note that traditional models like regression also involve an implicit SUTVA assumption (Y_i depends on X_i)

Back to the Neyman Urn Model



Estimands

Because τ_i are unobservable, we shift what we are interested in to:

Definition (Average Treatment Effect (ATE))

$$\tau_{ATE} = \text{Average of all treatment potential outcomes} - \text{Average of all control potential outcomes}$$

or

$$\tau_{ATE} = \frac{1}{N}\sum_{i}^{N} Y_{1i} - \frac{1}{N}\sum_{i}^{N} Y_{0i}$$

or

$$\tau_{ATE} = E[Y_{1i} - Y_{0i}]$$

or

$$\tau_{ATE} = E[\tau_i]$$

Average Treatment Effect

Imagine a study population with 4 units:

i	D_i	Y_{1i}	Y_{0i}	τ_i
1	1	10	4	6
2	1	1	2	-1
3	0	3	3	0
4	0	5	2	3

What is the ATE?

Imagine a study population with 4 units:

i	D_i	Y_{1i}	Y_{0i}	τ_i
1	1	10	4	6
2	1	1	2	-1
3	0	3	3	0
4	0	5	2	3

What is the ATE?

$E[Y_{1i} - Y_{0i}] = 1/4 \times (6 + -1 + 0 + 3) = 2$

Average Treatment Effect

Imagine a study population with 4 units:

i	D_i	Y_{1i}	Y_{0i}	τ_i
1	1	10	4	6
2	1	1	2	-1
3	0	3	3	0
4	0	5	2	3

What is the ATE?

$E[Y_{1i} - Y_{0i}] = 1/4 \times (6 + -1 + 0 + 3) = 2$

Note: Average effect is positive, but τ_i are negative for some units!

Probability Mass Function

Definition

Let Y be a discrete random variable. We usually refer to P_Y as the probability mass function (PMF): $P_Y(y) = \Pr(Y = y)$

Probability Mass Function

Definition

Let Y be a discrete random variable. We usually refer to P_Y as the probability mass function (PMF): $P_Y(y) = \Pr(Y = y)$

- Example: Two coin flips where Y is the total number of heads

Probability Mass Function

Definition

Let Y be a discrete random variable. We usually refer to P_Y as the probability mass function (PMF): $P_Y(y) = \Pr(Y = y)$

- Example: Two coin flips where Y is the total number of heads

$$P_Y(y) = \begin{cases} 1/4 & y = 0 \\ 1/2 & y = 1 \\ 1/4 & y = 2 \end{cases}$$

Probability Mass Function

Definition

Let Y be a discrete random variable. We usually refer to P_Y as the probability mass function (PMF): $P_Y(y) = \Pr(Y = y)$

- Example: Two coin flips where Y is the total number of heads

$$P_Y(y) = \begin{cases} 1/4 & y = 0 \\ 1/2 & y = 1 \\ 1/4 & y = 2 \end{cases}$$

- Required restrictions on $P_Y(y)$
  - $P_Y(y) \geq 0$

Probability Mass Function

Definition

Let Y be a discrete random variable. We usually refer to P_Y as the probability mass function (PMF): $P_Y(y) = \Pr(Y = y)$

- Example: Two coin flips where Y is the total number of heads

$$P_Y(y) = \begin{cases} 1/4 & y = 0 \\ 1/2 & y = 1 \\ 1/4 & y = 2 \end{cases}$$

- Required restrictions on $P_Y(y)$
  - $P_Y(y) \geq 0$
  - $\sum_{all\ y} P_Y(y) = 1.$

Probability Mass Function

Definition

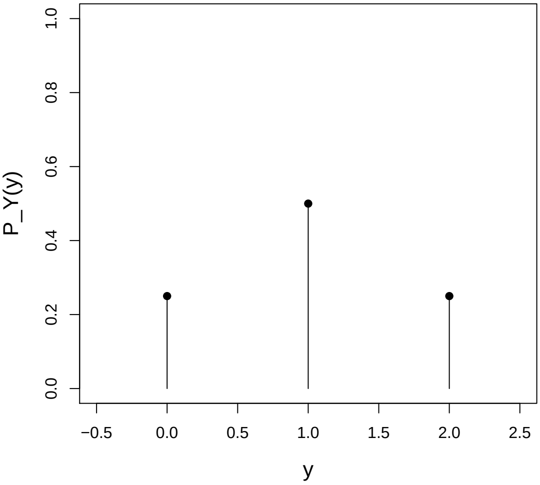
Let Y be a discrete random variable. We usually refer to P_Y as the probability mass function (PMF): $P_Y(y) = \Pr(Y = y)$

- Example: Two coin flips where Y is the total number of heads

$$P_Y(y) = \begin{cases} 1/4 & y = 0 \\ 1/2 & y = 1 \\ 1/4 & y = 2 \end{cases}$$

- Required restrictions on $P_Y(y)$
  - $P_Y(y) \geq 0$
  - $\sum_{all\ y} P_Y(y) = 1.$

PMF Plot



Expected Value Operator

- We are often interested in particular features of the probability distribution of a random variable in the population of interest

Expected Value Operator

- We are often interested in particular features of the probability distribution of a random variable in the population of interest
- For a random variable Y the most widely used location measure is its expected value (or the average), denoted by $E[Y]$

Expected Value Operator

- We are often interested in particular features of the probability distribution of a random variable in the population of interest
- For a random variable Y the most widely used location measure is its expected value (or the average), denoted by $E[Y]$

Definition

For a discrete random variable $E[Y]$ is

$$E[Y] = \sum_{\text{all } y} y \, P_Y(y).$$

where $P_Y(y)$ is the PMF of Y.

Expectation: Example

Suppose X is a discrete random variable that can take values of 0, 1, and 2. The probability function of X is given by:

$$f_X(x) = \begin{cases} 0.20 & \text{if } x = 0 \\ 0.45 & \text{if } x = 1 \\ 0.35 & \text{if } x = 2 \end{cases}$$

What is $E[X] = \sum_{\text{all } x} x \, f_X(x)$?

Expectation: Example

Suppose X is a discrete random variable that can take values of 0, 1, and 2. The probability function of X is given by:

$$f_X(x) = \begin{cases} 0.20 & \text{if } x = 0 \\ 0.45 & \text{if } x = 1 \\ 0.35 & \text{if } x = 2 \end{cases}$$

What is $E[X] = \sum_{\text{all } x} x \, f_X(x)$?

The expected value of X is:

$$\begin{aligned} E[X] &= 0 \times f_X(0) + 1 \times f_X(1) + 2 \times f_X(2) \\ &= 0 \times 0.20 + 1 \times 0.45 + 2 \times 0.35 \\ &= 1.15 \end{aligned}$$

Properties of Expected value operator

Theorem (Linearity of Expected Value Operator)

Let X be a random variable and a and b be constants. Then for any function $g(x)$ whose expectation exists:

$$E[a\,g(x) + b] = a\,E[g(x)] + b$$

Implies that for any linear function of Y the expected value can be easily evaluated, for example:

- E[Y+ c]= E[Y]+c
- E[Y + X] = E[Y]+E[X]
- E[Y - X] = E[Y]-E[X]
- E[aY] = aE[Y]

Average Treatment Effect

Imagine a study population with 4 units:

i	D_i	Y_{1i}	Y_{0i}	τ_i
1	1	10	4	6
2	1	1	2	-1
3	0	3	3	0
4	0	5	2	3

$E[\tau_i] = 1/4 \times (6 + -1 + 0 + 3) = 2$

Conditional Expectation

Definition (Conditional Expectation Discrete Case)

Let Y and X be discrete RVs, then the conditional expectation of Y given the event $X = x$ is given by:

$$E[Y|X = x] = \sum_{y} y\, P(Y = y|X = x) = \sum_{y} y\, f_{Y|X}(y|x)$$

Conditional Expectation

Definition (Conditional Expectation Discrete Case)

Let Y and X be discrete RVs, then the conditional expectation of Y given the event $X = x$ is given by:

$$E[Y|X = x] = \sum_{y} y\, P(Y = y|X = x) = \sum_{y} y\, f_{Y|X}(y|x)$$

i	D_i	Y_{1i}	Y_{0i}	τ_i
1	1	10	4	6
2	1	1	2	-1
3	0	3	3	0
4	0	5	2	3

$E[Y_1|D = 1] = (10 + 1)/2 = 5.5$

Naive Comparison: Difference in Means

Comparisons between *observed* outcomes of treated and control units can often be misleading.

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

- Bias term unlikely to be 0 in most applications.
- Selection into treatment is often associated with the potential outcomes.

Selection Bias

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

Example: Church Attendance and Political Participation

- Churchgoers are likely to differ from non-churchgoers on a range of background characteristics (e.g. civic duty).
- Given these differences, turnout for churchgoers would be higher than for non-churchgoers even if churchgoers never attended church or church had zero mobilizing effect ($E[Y_{0i}|D_i=1]-E[Y_{0i}|D_i=0]>0$).

Selection Bias

$$
\begin{aligned}
E[Y_i|D_i = 1] - E[Y_i|D_i = 0]& \\
&= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] \\
&= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] + \underbrace{(E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 1])}_{0} \\
&= \underbrace{E[Y_{1i} - Y_{0i}|D_i = 1]}_{\text{ATT}} + \underbrace{\{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]\}}_{\text{BIAS}}
\end{aligned}
$$

Example: Gender Quotas and Redistribution/Representation for Women

- Countries with gender quotas are likely countries where women are politically mobilized.
- Given this difference, policies targeted towards women are more common in quota countries even if these countries had not adopted quotas ($E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] > 0$).

The Assignment Mechanism

- Since missing potential outcomes are unobservable we must make assumptions to fill them in, i.e. **estimate** missing potential outcomes.

The Assignment Mechanism

- Since missing potential outcomes are unobservable we must make assumptions to fill them in, i.e. **estimate** missing potential outcomes.
- In the causal inference literature, we typically make assumptions about the **assignment mechanism** to do so.

The Assignment Mechanism

- Since missing potential outcomes are unobservable we must make assumptions to fill them in, i.e. **estimate** missing potential outcomes.
- In the causal inference literature, we typically make assumptions about the **assignment mechanism** to do so.

Definition (Assignment Mechanism)

Assignment mechanism is the procedure that determines which units are selected for treatment. Examples include:

- random assignment
- selection on observables
- selection on unobservables

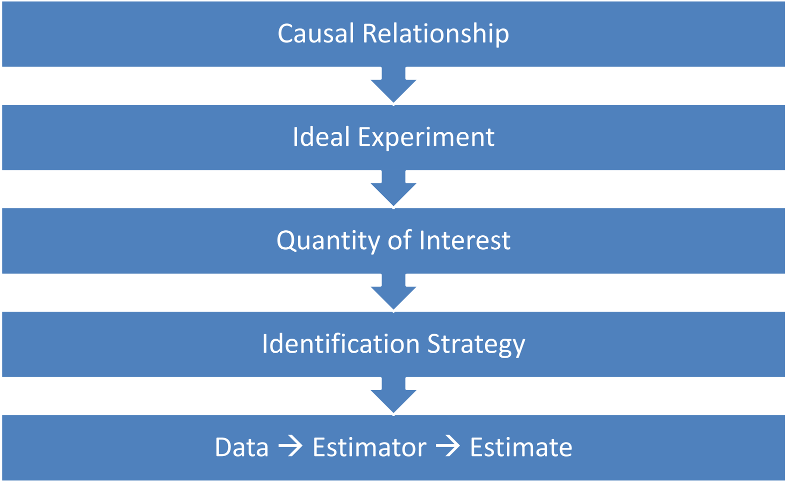
- Most statistical models of causal inference attain identification of treatment effects by restricting the assignment mechanism in some way.

Assignment Mechanism

Imagine a study population with 4 units:

i	$\Pr(D_i = 1)$	D_i	Y_{1i}	Y_{0i}	τ_i
1	?	1	10	4	6
2	?	1	1	2	-1
3	?	0	3	3	0
4	?	0	5	2	3

Causal Inference Workflow



Summing Up: Neyman-Rubin causal model

- Useful for studying the "effects of causes," less so for the "causes of effects."

Summing Up: Neyman-Rubin causal model

- Useful for studying the "effects of causes," less so for the "causes of effects."
- No assumption of homogeneity, allows for causal effects to vary unit by unit.

Summing Up: Neyman-Rubin causal model

- Useful for studying the "effects of causes," less so for the "causes of effects."
- No assumption of homogeneity, allows for causal effects to vary unit by unit.
  - No single "causal effect," thus the need to be precise about the target estimand.

Summing Up: Neyman-Rubin causal model

- Useful for studying the “effects of causes,” less so for the “causes of effects.”
- No assumption of homogeneity, allows for causal effects to vary unit by unit.
  - No single “causal effect,” thus the need to be precise about the target estimand.
- Distinguishes between *observed* outcomes and *potential* outcomes.

Summing Up: Neyman-Rubin causal model

- Useful for studying the “effects of causes,” less so for the “causes of effects.”
- No assumption of homogeneity, allows for causal effects to vary unit by unit.
    - No single “causal effect,” thus the need to be precise about the target estimand.
- Distinguishes between *observed* outcomes and *potential* outcomes.
- Causal inference is a missing data problem: we typically make assumptions about the assignment mechanism to go from descriptive inference to causal inference.