Potential Outcomes and Causal Effects

Xinzhou Guo

HKUST

Spring 2024

(Credited to Zhichao Jiang)

Simpson's paradox

	small stones		large stones	
	success	fail	success	fail
Treatment A	81	6	192	71
Treatment B	234	36	55	25

- Treatment A: open surgical procedures
- Treatment B: a minimally-invasive procedure

- Success rate for small stones: **93%** (81/87) > 87% (234/270)
  Success rate for large stones: **73%** (192/263) > 69% (55/80)
- Overall success rate: 78% (273/350) < **83%** (289/350)
  -- Why and Is treatment A better than treatment B?

Potential outcomes framework (Neyman 1923; Rubin 1974)

相关性不是因果性

- Success rate (A > B) → positive association between stone removal and treatment A
- Association ≠ Causation; The comparison between treatment A and treatment B is about association or causation?
- Causation： comparison between potential outcomes under treatment and control for the same unit(s) → What if xxx?
- Defining causal quantities by potential outcomes requires a thought experiment; neither data nor actual experimentation needed

Potential outcome and observed outcome

- Observed data: treatment Z_i, outcome Y_i
- Potential outcomes: $Y_i(1)$ and $Y_i(0)$
  - categorical: $Y_i(0)$, $Y_i(1)$,..., $Y_i(K-1)$
  - continuous: $Y_i(z)$ for any $z \in R$
  - observed outcome: $Y_i(Z_i) \longrightarrow$ only one potential outcome is observed for each unit

Potential outcome and observed outcome

- Observed data: treatment Z_i, outcome Y_i
- Potential outcomes: $Y_i(1)$ and $Y_i(0)$
  - categorical: $Y_i(0)$, $Y_i(1)$,..., $Y_i(K-1)$
  - continuous: $Y_i(z)$ for any $z \in R$
  - observed outcome: $Y_i(Z_i) \longrightarrow$ only one potential outcome is observed for each unit

Unit i	Z_i	$Y_i(1)$	$Y_i(0)$
1	1	0	1
2	0	0	1
3	1	0	0
4	1	1	1
5	0	1	0

Unit i	Z_i	Y_i
1	1	0
2	0	1
3	1	0
4	1	1
5	0	0

Hidden assumptions on potential outcomes

- The notation of $Y_i(z)$ implies three assumptions
  - no interference between units:

$$Y_i(Z_1, \dots, Z_n) = Y_i(Z_i)$$

  - same version of treatment
  - treatment occurs before outcomes
- Stable Unit Treatment Value Assumption (SUTVA)
  - no interference
  - only one version of treatment

Violation of SUTVA

No interference can be violated in infectious diseases or network experiments. For instance, if some of my friends receive flu shots, my chance of getting the flu decreases even if I do not receive the flu shot; if my friends see an advertisement on Facebook, my chance of buying that product increases even if I do not see the advertisement directly. It is an active research area to study situations with interfering units in modern causal inference literature (e.g., Hudgens and Halloran, 2008).

Same treatment version can be violated for treatments with complex components. For instance, when studying the effect of cigarette smoking on lung cancer, the type of cigarettes may matter; when studying the effect of college education on income, the type and major of college education may matter.

Causal quantity

- Any causal quantity is a function of potential outcomes

$$\log Y_i(1) - \log Y_i(0), \quad \frac{Y_i(1)}{Y_i(0)}, \quad \mathbf{1}\{Y_i(1) > Y_i(0)\}, \quad \text{etc.}$$

Causal quantity

- Any causal quantity is a function of potential outcomes

$$\log Y_i(1) - \log Y_i(0), \quad \frac{Y_i(1)}{Y_i(0)}, \quad \mathbf{1}\{Y_i(1) > Y_i(0)\}, \quad \text{etc.}$$

- A causal effect is defined to be the comparison of the potential outcomes on the same units

Causal quantity

- Any causal quantity is a function of potential outcomes

$$\log Y_i(1) - \log Y_i(0), \quad \frac{Y_i(1)}{Y_i(0)}, \quad \mathbf{1}\{Y_i(1) > Y_i(0)\}, \quad \text{etc.}$$

- A causal effect is defined to be the comparison of the potential outcomes on the same units

- Fundamental problem of causal inference
  → only one potential outcome is observed
  - we never see both $Y_i(1)$ and $Y_i(0)$
  - most features of $Y_i(1) - Y_i(0)$ are not point identified, e.g., $\mathrm{pr}\{Y_i(1) - Y_i(0) \leq 0\}$

Average causal effect

- Individual causal effect: $Y_i(1) - Y_i(0)$ → difficult to estimate

Average causal effect

- Individual causal effect: $Y_i(1) - Y_i(0)$ -> difficult to estimate
- Average causal effect (ACE): $\mathrm{E}\{Y_i(1) - Y_i(0)\}$

Average causal effect

- Individual causal effect: $Y_i(1) - Y_i(0)$ -> difficult to estimate
- Average causal effect (ACE): $E\{Y_i(1) - Y_i(0)\}$
- $E\{Y_i(1)\} \neq E\{Y_i \mid Z_i = 1\}$ (when will they be the same?)
  - $E\{Y_i(1)\}$: average of $Y_i(1)$ for units 1 to 5
  - $E(Y_i \mid Z_i = 1) = E\{Y_i(1) \mid Z_i = 1\}$: average of $Y_i(1)$ for units 1,3,4

Unit i	Z_i	$Y_i(1)$	$Y_i(0)$
1	1	0	1
2	0	0	1
3	1	0	0
4	1	1	1
5	0	1	0

Unit i	Z_i	Y_i
1	1	0
2	0	1
3	1	0
4	1	1
5	0	0

Other causal quantities of interest

- Average treatment effect on the treated (ATT) and on the untreated (ATU): $E\{Y_i(1) - Y_i(0) \mid Z_i = 1\}$, $E\{Y_i(1) - Y_i(0) \mid Z_i = 0\}$
- Heterogeneous effects:
  - conditional average causal effect: $\text{ACE}(\mathbf{x}) = E\{Y_i(1) - Y_i(0) \mid \mathbf{X}_i = \mathbf{x}\}$
  - applications to precision medicine
- Non-additive effects:
  - quantile treatment effects, e.g.,

    $$\text{median}\{Y_i(1) - Y_i(0)\} \quad \text{or} \quad \text{median}\{Y_i(1)\} - \text{median}\{Y_i(0)\}$$

  - odds ratio

    $$\frac{\text{pr}\{Y_i(1) = 1\}/\text{pr}\{Y_i(1) = 0\}}{\text{pr}\{Y_i(0) = 1\}/\text{pr}\{Y_i(0) = 0\}}$$

Causal effect is comparison of potential outcomes

- Let $Z = \mathbf{1}$(Take Aspirin at 3 pm). Which of the following qualifies/qualify as a causal effect?

  A) E(temperature $| Z = 1$) – E(temperature $| Z = 0$)

  B) E(potential pain scale at 4 pm with Aspirin $| Z = 1$) – E(potential pain scale at 4 pm without Aspirin $| Z = 0$)

  C) E(potential pain scale at 2 pm with Aspirin) – E(potential pain scale at 2 pm without Aspirin)

  D) my body temperature after taking Aspirin - my body temperature before taking Aspirin

Causal effects of immutable characteristics

- "No causation without manipulation" (Holland, 1986)
- Immutable characteristics or attributes: gender, race, age, etc.

Causal effects of immutable characteristics

- "No causation without manipulation" (Holland, 1986)
- Immutable characteristics or attributes: gender, race, age, etc.
- Can immutable characteristics have meaningful causal effects?

Causal effects of immutable characteristics

- "No causation without manipulation" (Holland, 1986)
- Immutable characteristics or attributes: gender, race, age, etc.
- Can immutable characteristics have meaningful causal effects?

- Strategies:

Causal effects of immutable characteristics

- "No causation without manipulation" (Holland, 1986)
- Immutable characteristics or attributes: gender, race, age, etc.
- Can immutable characteristics have meaningful causal effects?

- Strategies:
  1. causal effects of perceived characteristics:
     - Causal effect of a job applicant's gender/race on call-back rates (Bertrand and Mullainathan, 2004)

Causal effects of immutable characteristics

- "No causation without manipulation" (Holland, 1986)
- Immutable characteristics or attributes: gender, race, age, etc.
- Can immutable characteristics have meaningful causal effects?

- Strategies:
  1. causal effects of perceived characteristics:
     - Causal effect of a job applicant's gender/race on call-back rates (Bertrand and Mullainathan, 2004)
  2. reinterpretation
     - Causal effect of having a female politician on policy outcomes (Chattopadhyay and Duflo, 2004)

Causal effects of immutable characteristics

- "No causation without manipulation" (Holland, 1986)
- Immutable characteristics or attributes: gender, race, age, etc.
- Can immutable characteristics have meaningful causal effects?

- Strategies:
  1. causal effects of perceived characteristics:
     - Causal effect of a job applicant's gender/race on call-back rates (Bertrand and Mullainathan, 2004)
  2. reinterpretation
     - Causal effect of having a female politician on policy outcomes (Chattopadhyay and Duflo, 2004)
  3. redefinition:
     - Race as a "bundle of sticks": skin color, neighborhood, socio-economic status, etc. (Sen and Wasow, 2016)

Resolving Simpson's paradox

	small stones		large stones	
	success	fail	success	fail
Treatment A	81	6	192	71
Treatment B	234	36	55	25

- Treatment Z_i (1 for A); outcome Y_i (1 for success); covariate X_i (1 for large stones)

Resolving Simpson's paradox

	small stones		large stones	
	success	fail	success	fail
Treatment A	81	6	192	71
Treatment B	234	36	55	25

- Treatment Z_i (1 for A); outcome Y_i (1 for success); covariate X_i (1 for large stones)

- Simpson's paradox:
  - $\hat{\mathrm{E}}(Y_i \mid Z_i = 1, X_i = x) > \hat{\mathrm{E}}(Y_i \mid Z_i = 0, X = x)$ for $x = 0, 1$
  - $\hat{\mathrm{E}}(Y_i \mid Z_i = 1) < \hat{\mathrm{E}}(Y_i \mid Z_i = 0)$

Resolving Simpson's paradox

	small stones		large stones	
	success	fail	success	fail
Treatment A	81	6	192	71
Treatment B	234	36	55	25

- Treatment Z_i (1 for A); outcome Y_i (1 for success); covariate X_i (1 for large stones)

- Simpson's paradox:
  - $\hat{\mathrm{E}}(Y_i \mid Z_i = 1, X_i = x) > \hat{\mathrm{E}}(Y_i \mid Z_i = 0, X = x)$ for $x = 0, 1$
  - $\hat{\mathrm{E}}(Y_i \mid Z_i = 1) < \hat{\mathrm{E}}(Y_i \mid Z_i = 0)$
  - the sign of association may be reversed when adding covariates

Why Association Fail?

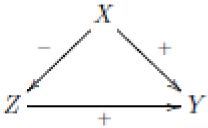


FIGURE 1.1: A diagram for the kidney stone example. The signs indicate the associations of two variables, conditioning on other variables pointing to the downstream variable.

- Patient with larger stones tends to take treatment A
- Patients with smaller stones have higher success probability.

Resolving Simpson's paradox

	small stones		large stones	
	success	fail	success	fail
Treatment A	81	6	192	71
Treatment B	234	36	55	25

- Treatment Z_i (1 for A); outcome Y_i (1 for success); covariate X_i (1 for large stones)
- Can Simpson's paradox happen using ACE instead of success rate?

Resolving Simpson's paradox

	small stones		large stones	
	success	fail	success	fail
Treatment A	81	6	192	71
Treatment B	234	36	55	25

- Treatment Z_i (1 for A); outcome Y_i (1 for success); covariate X_i (1 for large stones)

- Can Simpson's paradox happen using ACE instead of success rate?
  - $\mathrm{E}\{Y_i(1) \mid X_i = x\} > \mathrm{E}\{Y_i(0) \mid X = x\}$ for $x = 0, 1$
  - $\mathrm{E}\{Y_i(1)\} < \mathrm{E}\{Y_i(0)\}$?

Resolving Simpson's paradox

	small stones		large stones	
	success	fail	success	fail
Treatment A	81	6	192	71
Treatment B	234	36	55	25

- Treatment Z_i (1 for A); outcome Y_i (1 for success); covariate X_i (1 for large stones)

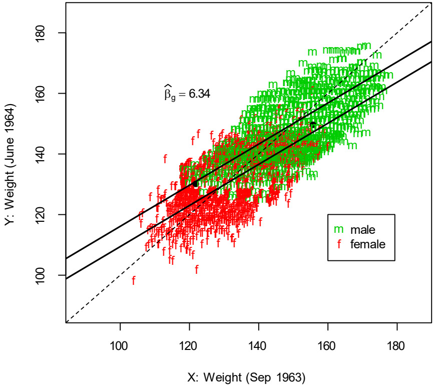
- Can Simpson's paradox happen using ACE instead of success rate?
  - $\mathrm{E}\{Y_i(1) \mid X_i = x\} > \mathrm{E}\{Y_i(0) \mid X = x\}$ for $x = 0, 1$
  - $\mathrm{E}\{Y_i(1)\} < \mathrm{E}\{Y_i(0)\}$?
- Simpson's paradox cannot happen for ACE; Is treatment A better than treatment B?

Lord's paradox (Lord, 1967)

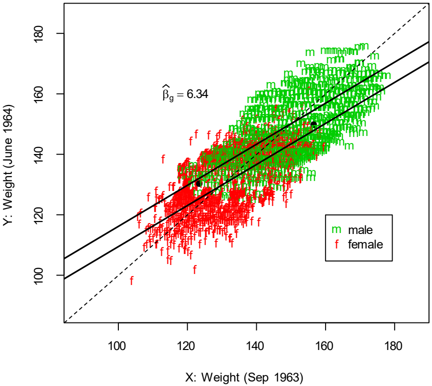
- Question: are the effects of the diet provided in the dining hall different for males and females?
- Data: gender G_i; weight in 1963 X_i; weight in 1964 Y_i

Lord's paradox (Lord, 1967)

- Question: are the effects of the diet provided in the dining hall different for males and females?
- Data: gender G_i; weight in 1963 X_i; weight in 1964 Y_i
- $E(Y_i \mid G_i = 1) = E(X_i \mid G_i = 1) = 150$
- $E(Y_i \mid G_i = 0) = E(X_i \mid G_i = 0) = 130$



Lord's paradox (Lord, 1967)



- Statistician A: average weights unchanged for both males and females
- Statistician B: $Y_i = \beta_0 + \beta_g G_i + \beta_X X_i + E_i$ $\beta_g = 6.34$
- What is the interpretation of β_g
- Who is correct?

Resolving Lord's paradox

- Formulation
  - treatment Z_i (1 for dining)
  - pre-treatment: gender G_i (1 for male); weight in 1963 X_i
  - post-treatment: weight in 1964 Y_i
  - potential outcomes: $Y_i(1)$ and $Y_i(0)$

Resolving Lord's paradox

- Formulation
  - treatment Z_i (1 for dining)
  - pre-treatment: gender G_i (1 for male); weight in 1963 X_i
  - post-treatment: weight in 1964 Y_i
  - potential outcomes: $Y_i(1)$ and $Y_i(0)$
- Causal quantity: $\Delta_g = \mathrm{E}\{Y_i(1) - Y_i(0) \mid G_i = g\}$ for $g = 0, 1$
  - difference between males and females: $\Delta_1 - \Delta_0$

Resolving Lord's paradox

- $\mathrm{E}\{Y_i(1) \mid G_i = g\} = \mathrm{E}(Y_i \mid G_i = g)$, $\mathrm{E}\{Y_i(0) \mid G_i = g\} =???$

Resolving Lord's paradox

- $E\{Y_i(1) \mid G_i = g\} = E(Y_i \mid G_i = g)$, $E\{Y_i(0) \mid G_i = g\}$ =???

- $Y_i(0)$ is missing for all units -> no conclusion without assumptions about $Y_i(0)$ (identifiability issue)

Resolving Lord's paradox

- Statistician A: $Y_i(0) = X_i$ -> $\Delta_1 - \Delta_0 = 0$

Resolving Lord's paradox

- Statistician A: $Y_i(0) = X_i$ -> $\Delta_1 - \Delta_0 = 0$

- Statistician B: $Y_i = \beta_0 + \beta_g G_i + \beta_X X + E_i$

Resolving Lord's paradox

- Statistician A: $Y_i(0) = X_i$ -> $\Delta_1 - \Delta_0 = 0$

- Statistician B: $Y_i = \beta_0 + \beta_g G_i + \beta_X X + E_i$
  - $\mathrm{E}\{Y_i(1) \mid X_i, G_i = g\} = a_g + bX_i$ -> $a_1 - a_0 = \beta_g$
  - $Y_i(0) = a + bX_i$

Resolving Lord's paradox

- Statistician A: $Y_i(0) = X_i$ -> $\Delta_1 - \Delta_0 = 0$

- Statistician B: $Y_i = \beta_0 + \beta_g G_i + \beta_X X + E_i$
  - $E\{Y_i(1) \mid X_i, G_i = g\} = a_g + bX_i$ -> $a_1 - a_0 = \beta_g$
  - $Y_i(0) = a + bX_i$

- Both statisticians' conclusions depend on untestable assumptions

Identification links thought experiment and data

- The target parameters, as defined by potential outcomes, is a function of the unobservables
- Question of identification: what can we learn about this function from the observed data?

- Identification maps assumptions and data to information about target parameters; Which causal quantity is identifiable?
- A parameter is identified if, under the stated assumptions, alternative values of the parameter implies different distributions of observable data
- Identification is a binary property

- In order to achieve identification, assumptions are unavoidable, but we need to figure out what assumptions are plausible in practice

Statistical inference links population and sample

- In practice, we only see a finite sample of the observables
- We do not know the population distribution of data
- Statistical inference: using the sample to infer about the population

- It is useful to separate identification from statistical inference

$$\text{Sample} \quad \underbrace{\longrightarrow}_{\text{statistical inference}} \quad \text{Population} \quad \underbrace{\longrightarrow}_{\text{identification}} \quad \text{Target parameters}$$

- Identification: how much can you learn about the quantities of interest if you had an infinite amount of data?

- We will keep returning to these two steps in the whole semester

Summary

- Causation: comparison of potential outcomes for the same unit(s)
- Causal quantity is a function of potential outcomes

- Fundamental problem of causal inference: only one potential outcome is observed
- Identification links thought experiment and data
- Statistical inference links population and sample