Section: Module 3

Average Treatment Effect

Kentaro Nakamura

GOV 2002

September 26th, 2025

Logistics

- Please look at the solution of Problem Set 2
  - Great job everyone
  - I know conditional randomization test is hard, so please take a look at the solution!
  - It is important to understand the idea

Mathematical Tool: Variance

- Variance for random variable X_i is defined as

$$\mathbb{V}[X_i] = \mathbb{E}[(X_i - \mathbb{E}[X_i])^2]$$

- Another useful representation

$$\mathbb{V}[X_i] = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2$$

- Properties
  - $\mathbb{V}[X_i + Y_i] = \mathbb{V}[X_i] + \mathbb{V}[Y_i] + 2\mathrm{Cov}(X_i, Y_i)$
  - If a is constant, $\mathbb{V}[a] = 0$
  - If a is constant, $\mathbb{V}[aX_i] = a^2\mathbb{V}[X_i]$
  - **Law of Total Variance**:

$$\mathbb{V}[X_i] = \mathbb{V}[\mathbb{E}[X_i \mid Y_i]] + \mathbb{E}[\mathbb{V}[X_i \mid Y_i]]$$

Two inference frameworks: Overview

- **within sample inference** (finite-population framework)
  - Given data on n units ($i = 1, \cdots, n$), we are interested in the average treatment effect on that sample
  - Only source of randomness is *treatment assignment*
  - Estimand: Sample Average Treatment Effect (SATE)

$$\tau_{SATE} = \frac{1}{N} \sum_{i=1}^{n} \{Y_i(1) - Y_i(0)\}$$

- **population inference** (super-population framework)
  - Generalizing the inference on the obtained sample to some population of interest
  - Source of randomness is (1) *treatment assignment* and (2) *sampling*
  - Estimand: Population Average Treatment Effect (PATE)

$$\tau_{PATE} = \mathbb{E}[Y_i(1) - Y_i(0)]$$

Two inference frameworks: Overview

Population (e.g., all students in class)

Sampling

Finite Population (e.g., students in section)

Y(1)	Y(0)
10	5
3	7
2	9

Treatment Assignment

Observed Data

T	Y
1	10
0	7
1	2

Estimand and Estimator

- Can you differentiate τ , $\hat{\tau}$, $\mathbb{V}[\hat{\tau}]$, and $\widehat{\mathbb{V}[\hat{\tau}]}$?
- τ : Estimand for Average Treatment Effect
  - **Estimand**: Target quantity of interest

$$\tau_{SATE} = \frac{1}{N}\sum_{i=1}^{n}\{Y_i(1) - Y_i(0)\}, \qquad \tau_{PATE} = \mathbb{E}[Y_i(1) - Y_i(0)]$$

- $\hat{\tau}$: Estimator of Average Treatment effect
  - **Estimator**: something you can calculate from data
  - In the case of SATE and PATE,

$$\hat{\tau}_{SATE} = \hat{\tau}_{PATE} = \frac{1}{n_1}\sum_{i=1}^{n} T_i Y_i - \frac{1}{n_0}\sum_{i=1}^{n}(1 - T_i) Y_i$$

- $\mathbb{V}[\hat{\tau}]$: **Estimand** of Variance of ATE Estimator
- $\widehat{\mathbb{V}[\hat{\tau}]}$: **Estimator** of Variance of ATE Estimator
  - Take a squared root to get standard error

Difference-in-Means Estimator

- Recall that estimand for PATE and SATE is

$$\tau_{SATE} = \frac{1}{N}\sum_{i=1}^{n}\{Y_i(1) - Y_i(0)\}, \qquad \tau_{PATE} = \mathbb{E}[Y_i(1) - Y_i(0)]$$

- Difference-in-Means estimator is

$$\hat{\tau} = \frac{1}{n_1}\sum_{i=1}^{n} T_i Y_i - \frac{1}{n_0}\sum_{i=1}^{n}(1 - T_i)Y_i$$

which is **unbiased** for both PATE and SATE

- $\rightarrow$ Justification for point estimate

Unbiasedness Proof (SATE) - (1)

- For SATE, we need to show $\mathbb{E}[\hat{\tau} \mid \mathcal{O}_n] = \tau_{SATE}$.
- By consistency,

$$\begin{aligned}\mathbb{E}[\hat{\tau} \mid \mathcal{O}_n] &= \mathbb{E}\Big[\frac{1}{n_1}\sum_{i=1}^{n} T_i Y_i - \frac{1}{n_0}\sum_{i=1}^{n}(1 - T_i)Y_i \mid \mathcal{O}_n\Big] \\ &= \mathbb{E}\Big[\frac{1}{n_1}\sum_{i=1}^{n} T_i Y_i(1) - \frac{1}{n_0}\sum_{i=1}^{n}(1 - T_i)Y_i(0) \mid \mathcal{O}_n\Big]\end{aligned}$$

- Because we condition on $\mathcal{O}_n = \{Y_i(1), Y_i(0)\}_{i=1}^{n}$, potential outcomes are not random.

$$= \frac{1}{n_1}\sum_{i=1}^{n} \mathbb{E}[T_i \mid \mathcal{O}_n] Y_i(1) - \frac{1}{n_0}\sum_{i=1}^{n} \mathbb{E}[1 - T_i \mid \mathcal{O}_n] Y_i(0)$$

Unbiasedness Proof (SATE) - (2)

- By complete randomization, we have $Y_i(1), Y_i(0) \perp T_i$. So,

$$= \frac{1}{n_1} \sum_{i=1}^{n} \mathbb{E}[T_i] Y_i(1) - \frac{1}{n_0} \sum_{i=1}^{n} \mathbb{E}[1 - T_i] Y_i(0)$$

- Now, $\mathbb{E}[T_i] = \frac{n_1}{n}$ and $\mathbb{E}[1 - T_i] = \frac{n_0}{n}$. So,

$$\begin{aligned} &= \frac{1}{n_1} \frac{n_1}{n} \sum_{i=1}^{n} Y_i(1) - \frac{1}{n_0} \frac{n_1}{n} \sum_{i=1}^{n} Y_i(0) \\ &= \frac{1}{n} \sum_{i=1}^{n} \{ Y_i(1) - Y_i(0) \} = \tau_{\text{SATE}} \end{aligned}$$

Unbiasedness Proof (PATE)

- We have shown $\mathbb{E}[\hat{\tau} \mid \mathcal{O}_n] = \tau_{SATE}$
- We next show $\mathbb{E}[\hat{\tau}] = \tau_{PATE} = \mathbb{E}[Y_i(1) - Y_i(0)]$
- Recall law of iterated expectation: $\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X]$. Thus,

$$\mathbb{E}[\hat{\tau}] = \mathbb{E}[\mathbb{E}[\hat{\tau} \mid \mathcal{O}_n]] = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{n}\{Y_i(1) - Y_i(0)\}\right]$$

- By linearity of expectation (i.e., $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$),

$$= \frac{1}{N}\sum_{i=1}^{n}\mathbb{E}\left[Y_i(1) - Y_i(0)\right] = \mathbb{E}[Y_i(1) - Y_i(0)] = \tau_{\mathrm{PATE}}$$

Tips for Proof

- Think about which variable is random
  - e.g., In finite-population framework, $Y_i(t)$ is fixed and T_i is random
- Think about the setting
  - e.g., which variables are independent of other vairables?
- Construct your proof step by step
  - Clearly show how you did each transformation in each line
  - Do not use multiple operations in one line
  - e.g., use consistency $\rightarrow$ use randomization $\rightarrow \dots$
- Always law of iterated expectation is your friend

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]]$$

Neyman Variance Estimator

- Recall that we use difference-in-means estimator for both PATE and SATE
- We want to know the variance of the difference-in-means estimator
  - For SATE, $\mathbb{V}[\hat{\tau} \mid \mathcal{O}_n]$ (where $\mathcal{O}_n = \{Y_i(1), Y_i(0)\}_{i=1}^n$)
  - For PATE, $\mathbb{V}[\hat{\tau}]$
- From pre-recorded lecture, we know that

$$\mathbb{V}[\hat{\tau} \mid \mathcal{O}_n] = \frac{1}{n}\left(\frac{n_0}{n_1}S_1^2 + \frac{n_1}{n_0}S_0^2 + 2S_{01}\right)$$

$$\mathbb{V}[\hat{\tau}] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0}$$

where

$$S_t^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i(t) - \bar{Y}(t))^2$$

$$S_{01} = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i(1) - \bar{Y}(1))(Y_i(0) - \bar{Y}(0))$$

Neyman Variance Estimator

- We use Neyman Variance estimator for both PATE and SATE

$$\widehat{\mathbb{V}[\hat{\tau} \mid \mathcal{O}_n]} = \widehat{\mathbb{V}[\hat{\tau}]} = \frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_0^2}{n_0}$$

which is **unbiased** for PATE but **conservative** for SATE.

- Check pre-recorded lecture slide
- Check Review Question 3 for the full derivation of variance.

Matched-Pair Design: Review

- **Motivation**
  - Blocking improve efficiency $\rightarrow$ Can we keep blocking?
- **Procedure**
  - Create $J = n/2$ similar units
  - Randomize treatment assignment within each pair
- **Estimator**

$$\hat{\tau}_{\text{pair}} = \frac{1}{J}\sum_{j=1}^{J} W_i(Y_{1j} - Y_{2j})$$

where

- Y_{1j} is the first outcome in j-th pair
- Y_{2j} is the second outcome in j-th pair
- $W_i = 1$ if first unit received treatment and $W_i = -1$ if second unit received treatment

Matched-Pair Design: Variance

- **Variance**

$$\mathbb{V}(\hat{\tau}_{\text{pair}}) = \underbrace{\frac{\sigma_1^2}{J} + \frac{\sigma_2^2}{J}}_{\text{Var w/o design}} - \underbrace{\frac{2}{J}\text{Cov}(\text{Y}_{1\text{j}}, \text{Y}_{2\text{j}})}_{\text{within-pair covariance}}$$

where "Var w/o design" is the variance under complete randomization.

- In other words,

$$\mathbb{V}(\hat{\tau}) = \frac{\sigma_1^2}{J} + \frac{\sigma_2^2}{J} = \mathbb{V}(\hat{\tau}_{\text{pair}}) + \frac{2}{J}\text{Cov}(\text{Y}_{1\text{j}}, \text{Y}_{2\text{j}})$$

- Notice that $n_1 = n_0 = J$ (within each pair, one treated and one control)

Matched-Pair Design: Variance Estimator

- **Variance Estimator**

$$\widehat{\mathbb{V}(\hat{\tau}_{\text{pair}})} = \frac{1}{J(J-1)} \sum_{j=1}^{J} \left(W_i (Y_{1j} - Y_{2j}) - \hat{\tau}_{\text{pair}} \right)^2$$

- **Question**: How to optimize matching so that we can maximize efficiency?
  - There are so many ways to pair observations
  - Higher within-pair covariance leads to more efficiency gains
  - We want to create pair in which all pairs are similar enough

Optimal Matching: Idea

- First, calculate the distance between any two units
  - We use **Mahalanobis distance**, in which distance between observation i and j is defined as

$$D(X_i, X_j) = \sqrt{(X_i - X_j)^T \ \mathrm{Var}[X]^{-1}(X_i - X_j)}$$

- Intuition: Distance between covariates normalized by their variance
- **Goal**: Find the matching by which we can minimize the sum of distance
- You will see how different algorithm leads to different variance in Question 2

Greedy v.s. Optimal Matching

- Let's think with an example.
- Suppose there are 4 people in the world.

	1	2	3	4
1		1	9	5
2	1		5	9
3	9	5		12
4	5	9	12	

- NOTE: Diagoal element is missing to avoid the matching of the same unit.

Greedy Matching: 2 (a)

- Consider the greedy matching in 2 (a)
  - First, compute the distance between all pairs of health clusters.
  - Based on this distance matrix, select two clusters which are most similar and set them aside as a match

	1	2	3	4
1		1	9	5
2	1		5	9
3	9	5		12
4	5	9	12	

- You first select 1 (matching between unit 1 and 2)

Greedy Matching: 2 (a)

- After matching unit 1 and 2, you cannot use unit 1 or 2 for the future matching
  - For example, matching unit 1 and 3 is infeasible since unit 1 is already matched with unit 2

	1	2	3	4
1		1	9	5
2	1		5	9
3	9	5		12
4	5	9	12	

- In this case, you need to match unit 3 and 4.
- Sum of total distance: $1 + 12 = 13$

Greedy Matching: 2 (b)

- Consider the greedy matching in 2 (b)
  - Randomly select a cluster and then find the cluster which is most similar to it.
  - Set them aside as a matched pair.
- Suppose that you randomly pick up unit 2.
  - The closest for unit 2 is unit 1.
- As you match unit 2 and 1, you cannot use these units for the future matching.

	1	2	3	4
1		1	9	5
2	1		5	9
3	9	5		12
4	5	9	12	

- Sum of total distance: $1 + 12 = 13$

Optimal Matching

- Greedy matching is suboptimal
  - In the previous table, minimum distance is 10 ($(1,4)$ and $(2,3)$)
- **Optimal Matching**: Directly minimize the sum of distance
  - Algorithm: Optimal nonbipartite matching
- Optimization problem is written as

$$\min_{M} \quad \sum_{i=1}^{n} M_{ij} D_{ij}$$
$$\text{s.t.} \quad \sum_{i=1}^{n} M_{ij} = 1, \quad \sum_{j=1}^{n} M_{ij} = 1$$

- Constraint: Each unit is used for matching only once

	1	2	3	4
1		1	9	5
2	1		5	9
3	9	5		12
4	5	9	12	