Section 3

Inference for the Average Treatment Effect

Sooahn Shin

GOV 2003

Sept 23, 2021

Overview

- Logistics:
  - **Pset 3 released!** Due at 11:59 pm (ET) on Sept 29

- Today's topics:
  1. Neyman's approach to completely randomized experiments
  2. Derivation of finite-sample sampling variance
  3. A short review of blocked design

Neyman's approach to completely randomized experiments

Fisher and Neyman

- Design-based inference:
  - Fisher: treatments assigned randomly
  - Neyman: treatments assigned randomly + n samples chosen randomly from a superpopulation

- Fisher: permutation test with sharp null hypothesis
  - Fill in all values of the missing potential outcomes
  - Derive the exact randomization distribution of statistics
  - Limitations:
    - Does not allow heterogeneous treatment effects
    - Does not allow population-level inference

- Neyman: difference-in-means as an estimator of the ATE
  - Inference relies on asymptotic approximation
  - Obtain unbiased estimator ($\hat{\tau}$)
  - Construct an interval estimator for the causal estimand
    - $\rightsquigarrow$ unbiased/conservative estimator ($\widehat{\mathbb{V}}(\hat{\tau})$) for the sampling variance of the estimator ($\mathbb{V}(\hat{\tau})$)

Estimands and difference-in-means estimator

- n samples chosen randomly from a superpopulation
- Sample Average Treatment Effect:

$$\text{SATE} = \frac{1}{n}\sum_{i=1}^{n}[Y_i(1) - Y_i(0)] = \tau_{\text{fs}}$$

- Population Average Treatment Effect:

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

- Difference-in-means estimator:

$$\widehat{\tau}_{\text{diff}} = \frac{1}{n_1}\sum_{i=1}^{n} D_i Y_i - \frac{1}{n_0}\sum_{i=1}^{n}(1 - D_i)Y_i$$

Difference-in-means: Analytical results

- Difference-in-means estimator:

$$\widehat{\tau}_{\text{diff}} = \frac{1}{n_1}\sum_{i=1}^{n} D_i Y_i - \frac{1}{n_0}\sum_{i=1}^{n}(1 - D_i) Y_i$$

- Bias of the DiM estimator:
  - $\widehat{\tau}_{\text{diff}}$ unbiased for SATE:

$$\mathbb{E}_D[\widehat{\tau}_{\text{diff}} | \mathbf{O}] = \tau_{\text{fs}}$$

    - $\rightsquigarrow$ See lecture slides p.9 for derivation
  - $\widehat{\tau}_{\text{diff}}$ unbiased for PATE:

$$\mathbb{E}[\widehat{\tau}_{\text{diff}}] = \mathbb{E}[\mathbb{E}_D[\widehat{\tau}_{\text{diff}} | \mathbf{O}]] = \mathbb{E}[\tau_{\text{fs}}] = \tau$$

Difference-in-means: Analytical results

- Difference-in-means estimator:

$$\widehat{\tau}_{\text{diff}} = \frac{1}{n_1}\sum_{i=1}^{n} D_i Y_i - \frac{1}{n_0}\sum_{i=1}^{n}(1 - D_i) Y_i$$

- Sampling variance of the DiM estimator:
  - At finite-sample level:

$$\mathbb{V}_D(\widehat{\tau}_{\text{diff}} \mid \mathbf{O}) = \frac{S_0^2}{n_0} + \frac{S_1^2}{n_1} - \frac{S_{\tau_i}^2}{n},$$

    - S_0^2 and S_1^2 are the in-sample variances of $Y_i(0)$ and $Y_i(1)$, respectively. Last term is the in-sample variation of the individual treatment effects.
    - Will derive this shortly.
    - None of these are directly observable!
    - Also, can rewrite this as:

$$\mathbb{V}_D(\widehat{\tau}_{\text{diff}} \mid \mathbf{O}) = \frac{1}{n}\left(\frac{n_1}{n_0} S_0^2 + \frac{n_0}{n_1} S_1^2 + 2S_{01}\right)$$

Difference-in-means: Analytical results

- Difference-in-means estimator:

$$\widehat{\tau}_{\text{diff}} = \frac{1}{n_1}\sum_{i=1}^{n} D_i Y_i - \frac{1}{n_0}\sum_{i=1}^{n}(1 - D_i) Y_i$$

- Sampling variance of the DiM estimator:
  - At population level:

$$\mathbb{V}(\widehat{\tau}_{\text{diff}}) = \frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1}$$

    - σ_0^2 and σ_1^2 are the population-level variances of $Y_i(1)$ and $Y_i(0)$, respectively.
    - Will derive this in pset 3.
    - None of these are directly observable! $\rightsquigarrow$ obtain an estimator for this

Difference-in-means: Analytical results

- Usual variance estimator is Neyman (or robust) estimator:

$$\widehat{\mathbb{V}} = \frac{\widehat{\sigma}_0^2}{n_0} + \frac{\widehat{\sigma}_1^2}{n_1}$$

  - $\widehat{\sigma}_d^2$ are the sample variances within each group $d \in \{0, 1\}$.

$$\widehat{\sigma}_d^2 = \frac{1}{n_d - 1} \sum_{i=1}^{n} \mathbb{I}\{D_i = d\} \left(Y_i - \overline{Y}_d\right)^2$$

  - $\mathbb{E}\left[\widehat{\mathbb{V}} \mid \mathbf{O}\right] = \frac{S_1^2}{n_1} + \frac{S_0^2}{n_0}$ and $\mathbb{E}\left[\widehat{\mathbb{V}}\right] = \mathbb{E}\left[\mathbb{E}\left[\widehat{\mathbb{V}} \mid \mathbf{O}\right]\right] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0}$
- Estimating sampling variance:
  - At finite-sample level: Neyman estimator is conservative on average

$$\mathbb{V}_D(\widehat{\tau}_{\text{diff}} \mid \mathbf{O}) \leq \frac{S_1^2}{n_1} + \frac{S_0^2}{n_0} = \mathbb{E}\left[\widehat{\mathbb{V}} \mid \mathbf{O}\right]$$

  - At population level: Neyman estimator is unbiased

$$\mathbb{V}(\widehat{\tau}_{\text{diff}}) = \frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1} = \mathbb{E}\left[\widehat{\mathbb{V}}\right]$$

Difference-in-means: Analytical results

- Recap
  - Difference-in-means estimator ($\widehat{\tau}_{\text{diff}}$)
  - Bias of the DiM estimator
    - unbiased for both SATE ($= \tau_{\text{fs}}$) and PATE ($= \tau$)
  - Sampling variance of the DiM estimator
    - $\rightsquigarrow$ unobservable for both finite-sample ($\mathbb{V}_D(\widehat{\tau}_{\text{diff}} \mid \mathbf{O})$)
    - and population level ($\mathbb{V}(\widehat{\tau}_{\text{diff}})$)
  - Introduce Neyman (or robust) estimator
    - Conservative for finite-sample sampling variance
    - Unbiased for population sampling variance

Derivation of finite-sample sampling variance

A short review of blocked design

- Setup:
  - Group units into J blocks; randomize treatment within each block
  - Apply Neyman's analysis to each block $j = 1, \ldots, J$
  - Use weighted average of block estimates and variances, with weights $w_j = n_j/n$
- Motivation: gain in efficiency
  - Unbiasedness still holds:

$$\mathbb{E}[\hat{\tau}_{\text{block}}|\mathbf{O}] = \mathbb{E}[\hat{\tau}|\mathbf{O}] = \text{SATE}.$$

  - Lower population sampling variance: $\mathbb{V}(\hat{\tau}) \geq \mathbb{V}_{\text{block}}(\hat{\tau}_{\text{block}})$
- Example: Student Teacher Achievement Ratio (STAR) project
  - Analyze the relationship between kindergarten class size and student achievement.
  - Within each school, classes were randomized into small (13-17 students) or regular-size (22-25 students).