Causal Inference (6.S059/15.C08/17.C08)

Recitation, Week 1.

Topic: Randomization Inference and Designs

Benjamín Muñoz

February 29, 2024

MIT

Massachusetts Institute of Technology

Table of contents

1. PATE and SATE

2. Randomization Inference

3. Hints PSet 2

1/ PATE and SATE

Measures of Uncertainty

Randomization Variance of Observed Difference in Means

Measure of uncertainty of $\hat{\tau}$: $\mathbb{V}(\hat{\tau}) = \frac{1}{N-1}\left(\frac{N_0}{N_1}S_1^2 + \frac{N_1}{N_0}S_0^2 + 2S_{01}\right)$

$$S_1^2 : \text{Variance of the P.O. under Treatment} \rightsquigarrow \widehat{S_1^2} = \frac{1}{N_1 - 1}\sum_{i=1}^{N} D_i(Y_i - \tilde{Y_1})^2$$

$$S_0^2 : \text{Variance of the P.O. under Control} \rightsquigarrow \widehat{S_0^2} = \frac{1}{N_0 - 1}\sum_{i=1}^{N} (1 - D_i)(Y_i - \tilde{Y_0})^2$$

S_{01} :Covariance of Y_{1i} and Y_{0i}

Estimate of the neasure of uncertainty:

$$\widehat{\mathbb{V}(\hat{\tau})} \equiv \frac{\widehat{S_1^2}}{N_1} + \frac{\widehat{S_0^2}}{N_0}$$

PATE and SATE

- Two **estimands** (Quantities of Interest):
  1. **SATE:** Sample average treatment effect

     SATE $= \frac{1}{N} \sum_{i=1}^{n} \{Y_{1i} - Y_{0i}\}$

     Sampling distribution is created by varying treatment assignments.
  2. **PATE:** Population average treatment effect

     PATE $= \mathbb{E}[Y_{1i} - Y_{0i}]$

     Sampling distribution is created by varying treatment assignments and re-drawing the units in.

- One shared **estimator**:

$$\hat{\tau} = \frac{1}{N_1} \sum_{i=1}^{N} D_i Y_i - \frac{1}{N_0} \sum_{i=1}^{N} (1 - D_i) Y_i$$

- When is this an unbiased estimator for SATE? and PATE?

Unbiasedness for SATE

Logic of repeated random treatment assignment ($\mathcal{O} \equiv$ current sample).

$$
\begin{aligned}
\hat{\tau} &= \frac{1}{N_1}\sum_{i=1}^{N} D_i Y_i - \frac{1}{N_0}\sum_{i=1}^{N}(1 - D_i) Y_i \\
\mathbb{E}[\hat{\tau}|\mathcal{O}] &= \frac{1}{N_1}\sum_{i=1}^{N} \mathbb{E}[D_i|\mathcal{O}] Y_i - \frac{1}{N_0}\sum_{i=1}^{N}(1 - \mathbb{E}[D_i|\mathcal{O}]) Y_i \\
&= \frac{1}{N_1}\sum_{i=1}^{N} \mathbb{E}[D_i|\mathcal{O}] Y_{1i} - \frac{1}{N_0}\sum_{i=1}^{N}(1 - \mathbb{E}[D_i|\mathcal{O}]) Y_{0i} \\
&= \frac{1}{N_1}\frac{N_1}{N}\sum_{i=1}^{N} Y_{1i} - \frac{1}{N_0}\frac{N_0}{N}\sum_{i=1}^{N} Y_{0i} \\
&= \frac{1}{N}\sum_{i=1}^{N} Y_{1i} - \frac{1}{N}\sum_{i=1}^{N} Y_{0i} \\
&= \frac{1}{N}\sum_{i=1}^{N}(Y_{1i} - Y_{0i})
\end{aligned}
$$

PATE and SATE: Implications for Uncertainty

- Two-step process:
    1. We randomly sample N units from the population.
    2. We randomly assign treatment to N_1 units.

- Two sources of uncertainty:
    - SATE: Uncertainty comes only from treatment assignment.

$$\hat{\tau} = \frac{1}{N_1} \sum_{i=1}^{N} D_i Y_i - \frac{1}{N_0} \sum_{i=1}^{N} (1 - D_i) Y_i$$

    - PATE: It's from treatment assignment and sampling variation.

$$\hat{\tau} = \frac{1}{N_1} \sum_{i=1}^{N} D_i Y_i - \frac{1}{N_0} \sum_{i=1}^{N} (1 - D_i) Y_i$$

PATE and SATE: Implications for Estimation

- When we make inferences from a sample about the population, we need to account for the additional uncertainty.
- What does this mean for our usual t-test variance estimator?
  - It's a <u>biased</u> estimator for $V(\hat{\tau})$ when estimating the SATE.

    $\rightsquigarrow$ It's conservative – on average, it overstates the uncertainty.

$$\widehat{\mathbb{V}(\hat{\tau}|\mathcal{O})} \equiv \frac{S_1^2}{N_1} + \frac{S_0^2}{N_0} \geq \mathbb{V}(\hat{\tau}|\mathcal{O})$$

  - But <u>unbiased</u> when estimating the PATE.

    $\rightsquigarrow$ It's not overstating – PATE has extra uncertainty from sampling.

  - Asymptotics: SATE $\xrightarrow{p}$ PATE (LLN).
  - $\hat{\tau} \xrightarrow{p}$ PATE (LLN).
  - $\hat{\tau} \overset{a}{\sim} \mathcal{N}($ PATE, $\mathbb{V}(\hat{\tau}))$ (CLT).

Unbiasedness for PATE

$$
\begin{aligned}
\mathbb{E}_D\left[\frac{1}{N_1}\sum_{i=1}^{N} D_i Y_i - \frac{1}{N_0}\sum_{i=1}^{N}(1-D_i)Y_i\right] &= \mathbb{E}_{\mathcal{O}}\left[\mathbb{E}_D\left[\frac{1}{N_1}\sum_{i=1}^{N} D_i Y_i - \frac{1}{N_0}\sum_{i=1}^{N}(1-D_i)Y_i \middle| \mathcal{O}\right]\right] \\
&= \mathbb{E}_{\mathcal{O}}\left[\frac{1}{N}\sum_{i=1}^{N}(Y_{1i} - Y_{0i})\right] \\
&= \mathbb{E}_{\mathcal{O}}\left[\frac{1}{N}\sum_{i=1}^{N^p}\{z_i(Y_{1i} - Y_{0i})\}\right] \\
&= \frac{1}{N}\sum_{i=1}^{N^p}\{\mathbb{E}_{\mathcal{O}}[z_i](Y_{1i} - Y_{0i})\} \\
&= \frac{1}{N}\sum_{i=1}^{N^p}\left\{\frac{N}{N^p}(Y_{1i} - Y_{0i})\right\} \\
&= \frac{1}{N^p}\sum_{i=1}^{N^p}(Y_{1i} - Y_{0i}) = \mathbb{E}[Y_{1i} - Y_{0i}]
\end{aligned}
$$

2/ Randomization Inference

Standard Hypothesis Testing

1. Define the null hypothesis for our estimand τ.

   $H_0 : \mathbb{E}[Y_{1i}] = \mathbb{E}[Y_{0i}]$

2. What's the **sampling distribution** of $\hat{\tau}$ look like if H_0 were true?

   s This is the null distribution.

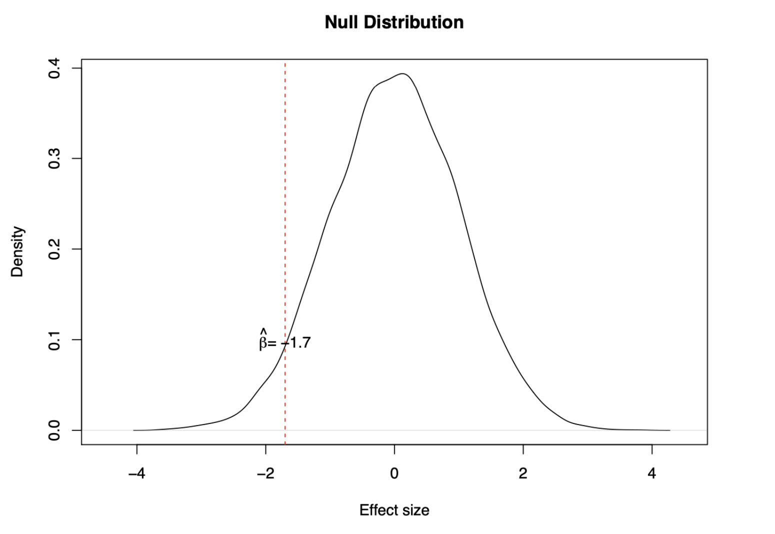
3. If that were true, how likely would we be to obtain our estimate?

   Calculate the area under the null distribution curve that is to the right (and/or left) of our actual estimate.

4. If that probability is lower than some arbitrary level α, reject the null. Otherwise, fail to reject.

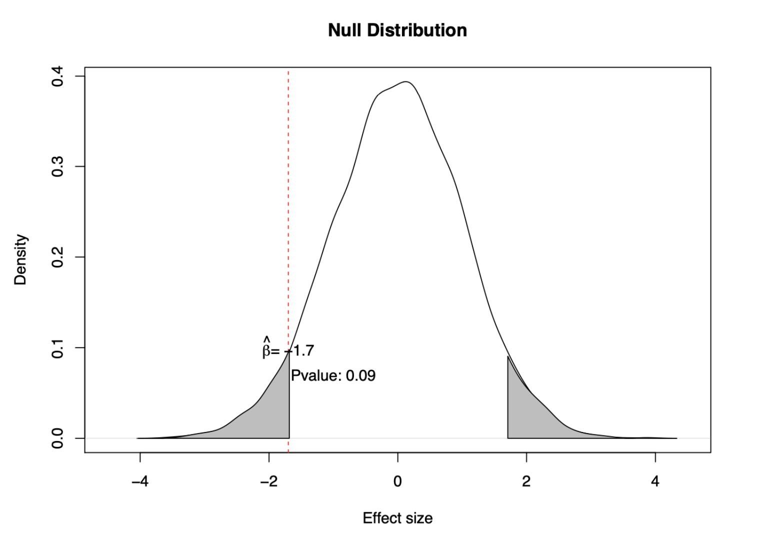
   ⋆ Why does this work? Asymptotic normality.

Example



Null Distribution
Density
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
Effect size
$\hat{\beta}$= −1.7

Example


Null Distribution
Density
Effect size
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
$\hat{\beta}$= −1.7
Pvalue: 0.09

Hypothesis Testing using Randomization Inference

1. Define the sharp null hypothesis.

   $H_0^{sharp} : Y_{1i} = Y_{0i}, \forall$ i

2. What's sampling distribution of $\hat{\tau}$ look like if H_0^{sharp} were true?

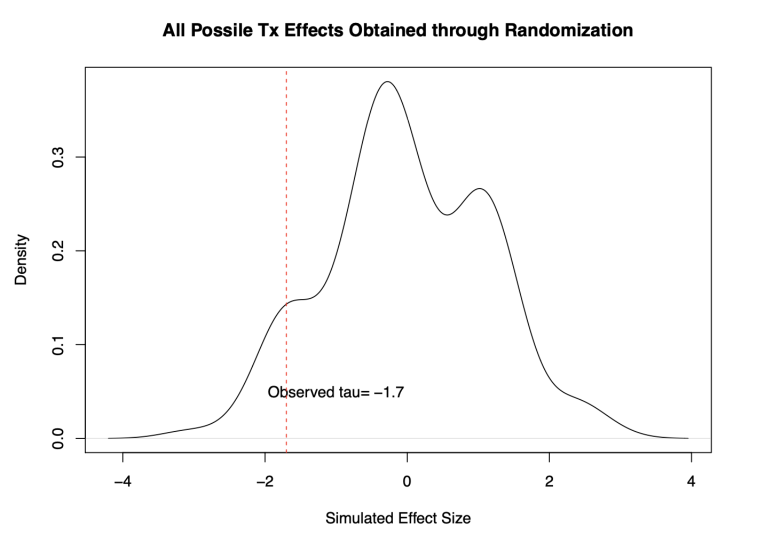
   Assume $Y_{i1} = Y_{i0}$ for all units, simulate all possible treatment assignments (Ω), obtain an estimate of the statistic ($\hat{\tau}(\omega)$, for every $\omega \in \Omega$) using observed outcomes for each, plot them (randomization distribution).

3. If that were true, how likely would we be to obtain our estimate?

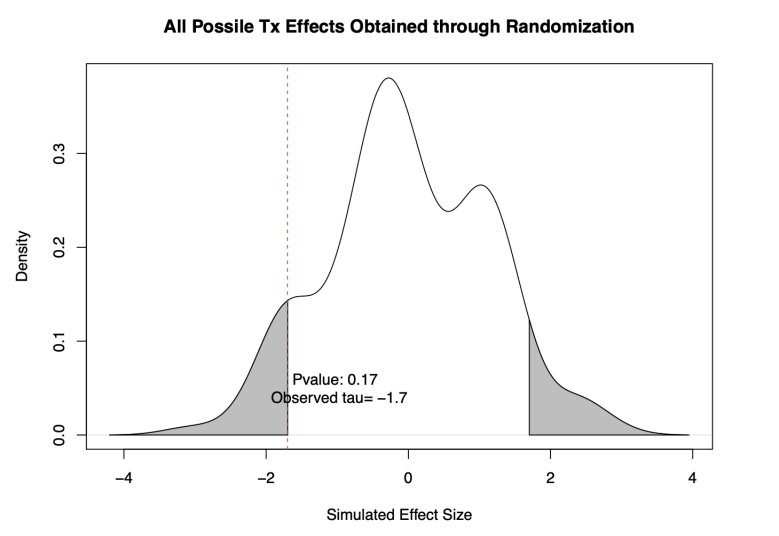
   Compare $\hat{\tau}^{obs}$ with the randomization distribution. Calculate the percentage of estimates from our null ("reference") distribution that are more extreme than our actual estimate. P-value for a two-sided alternative hypothesis $p \equiv Pr(|\hat{\tau}(\omega)| \geq |\hat{\tau}^{obs}|)$

4. If that probability is lower than some arbitrary level α, reject the null. Otherwise, fail to reject.

Example


All Possile Tx Effects Obtained through Randomization
Density
0.0
0.1
0.2
0.3
Observed tau= −1.7
−4
−2
0
2
4
Simulated Effect Size

Example


All Possile Tx Effects Obtained through Randomization
Density
0.0
0.1
0.2
0.3
Pvalue: 0.17
Observed tau= −1.7
−4
−2
0
2
4
Simulated Effect Size

Randomization Inference FAQ

- **Why would I ever do this?**
  - Statistical testing for a small sample (no asymptotic properties).
- **What's the basic intuition?**
  - "Sharp null:" there's zero treatment effect for every single unit.
  - Under the sharp null we know all the Potential Outcomes (fixed but a priori unknown values).
- **How useful is this?**
  - Useful if other tests aren't an option because the sample is small.
  - But rarely interesting – every unit has exactly zero causal effect?

3/ Hints PSet 2

Hints PSet 2

- Many options for the **sharp null** and the **statistic** used in the analysis. Small justifications are enough.
- Always important to be explicit about the alternative hypothesis and the significance level ($\hat{\alpha}$).
- Randomization Inference follows the randomization design.
    - Simple randomization
    - Complete randomization
    - Blocked randomization
    - Clustered randomization
- Key assumption is not an assumption: **physical randomization**. That is not the same as no assumptions (SUTVA).
- Use code available in the Google Colab as template.