3. Inference for the Average Treatment Effect

ISS5096 || ECI

Jaewon ("Jay-one") Yoo
National Tsing Hua University

Where are we? Where are we going?

- Fisher: Use sharp null to fill in science table + permutation test.
  - No way to estimate or infer about "average" effects, just the sharp null.

Where are we? Where are we going?

- Fisher: Use sharp null to fill in science table + permutation test.
  - No way to estimate or infer about "average" effects, just the sharp null.
- Neyman: Use the difference-in-means as an **estimator** of the ATE.
  - No assumptions to fill in the potential outcomes.
  - No exact derivation of the randomization distribution.
  - $\rightsquigarrow$ Asymptotic approximations!

Where are we? Where are we going?

- Fisher: Use sharp null to fill in science table + permutation test.
    - No way to estimate or infer about "average" effects, just the sharp null.
- Neyman: Use the difference-in-means as an **estimator** of the ATE.
    - No assumptions to fill in the potential outcomes.
    - No exact derivation of the randomization distribution.
    - $\rightsquigarrow$ Asymptotic approximations!
- What's common: the focus on **randomization** as generating variation in estimators.

Social Pressure and Turnout (Gerber et al. 2008. Am. Poli. Sci. Rev.)

- Experimental study where each household for 2006 Michigan primary election was randomly assigned to one of 4 conditions:
  - Control: no mail.
  - Civic Duty: a mail saying voting is your civic duty.
  - Hawthorne: a “we’re watching you” message.
  - Neighbors: a mail naming-and-shaming the receipient for not voting (social pressure).
- Sample size: 180,000 households
- Outcome: whether household members voted or not.

Neighborhood Social Pressure Message

Dear Registered Voter:

WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?

Why do so many people fail to vote? We've been talking about the problem for years, but it only seems to get worse. This year, we're taking a new approach. We're sending this mailing to you and your neighbors to publicize who does and does not vote.

The chart shows the names of some of your neighbors, showing which have voted in the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not.

DO YOUR CIVIC DUTY — VOTE!

MAPLE DR		Aug 04	Nov 04	Aug 06
9995	JOSEPH JAMES SMITH	Voted	Voted	______
9995	JENNIFER KAY SMITH		Voted	______
9997	RICHARD B JACKSON		Voted	______
9999	KATHY MARIE JACKSON		Voted	______
9999	BRIAN JOSEPH JACKSON		Voted	______

Source: Gerber, Greene, Larimer (2008) [https://doi.org/10.1017/S000305540808009X]

“You are being studied” (Hawthorne) Message

Dear Registered Voter:

YOU ARE BEING STUDIED!

Why do so many people fail to vote? We've been talking about this problem for years, but it only seems to get worse.

This year, we're trying to figure out why people do or do not vote. We'll be studying voter turnout in the August 8 primary election.

Our analysis will be based on public records, so you will not be contacted again or disturbed in any way. Anything we learn about your voting or not voting will remain confidential and will not be disclosed to anyone else.

DO YOUR CIVIC DUTY — VOTE!

Source: Gerber, Greene, Larimer (2008) [https://doi.org/10.1017/S000305540808009X]

Standard Empirical Analysis

TABLE 2. Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Experimental Group				
	Control	Civic Duty	Hawthorne	Self	Neighbors
Percentage Voting	29.7%	31.5%	32.2%	34.5%	37.8%
N of Individuals	191,243	38,218	38,204	38,218	38,201

- Typical reporting of the Neighbors vs Control effect:

$$\text{estimate} = \frac{1}{n_1}\sum_{i=1}^{n} D_i Y_i - \frac{1}{n_0}\sum_{i=1}^{n}(1 - D_i)Y_i \approx 8.1$$

$$\text{standard error} = \sqrt{\frac{\widehat{\sigma}_1^2}{n_1} + \frac{\widehat{\sigma}_0^2}{n_0}} \approx 0.27$$

$$95\%\ \text{CI} = [\text{est} - 1.96 \cdot SE, \text{est} + 1.96 \cdot SE] \approx [7.57, 8.63]$$

Standard Empirical Analysis

TABLE 2. Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Experimental Group				
	Control	Civic Duty	Hawthorne	Self	Neighbors
Percentage Voting	29.7%	31.5%	32.2%	34.5%	37.8%
N of Individuals	191,243	38,218	38,204	38,218	38,201

- Typical reporting of the Neighbors vs Control effect:

$$\text{estimate} = \frac{1}{n_1}\sum_{i=1}^{n} D_i Y_i - \frac{1}{n_0}\sum_{i=1}^{n}(1 - D_i)Y_i \approx 8.1$$

$$\text{standard error} = \sqrt{\frac{\widehat{\sigma}_1^2}{n_1} + \frac{\widehat{\sigma}_0^2}{n_0}} \approx 0.27$$

$$95\%\ \text{CI} = [\text{est} - 1.96 \cdot SE, \text{est} + 1.96 \cdot SE] \approx [7.57, 8.63]$$

- Q: Can this analysis be justified by randomization?

Causal Estimand of Interest

- Common estimand in experiments: **sample average treatment effect**

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

Causal Estimand of Interest

- Common estimand in experiments: **sample average treatment effect**

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

- Neyman/our goals:
  - We want to find an estimator that is **unbiased** for the SATE.
  - But, also derive the **sampling variance** for the estimator.

Causal Estimand of Interest

- Common estimand in experiments: **sample average treatment effect**

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

- Neyman/our goals:
  - We want to find an estimator that is **unbiased** for the SATE.
  - But, also derive the **sampling variance** for the estimator.
- Properties of the estimators across repeated samples from:
  - The randomization distribution.
  - The randomization distribution + sampling from the population.

Finite Sample Results

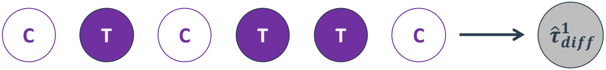
- Setting: **completely randomized experiment**
  - n units, n_1 treated, and n_0 control.
- Natural estimator for the SATE, **difference-in-means**:

$$\widehat{\tau}_{\text{diff}} = \underbrace{\frac{1}{n_1}\sum_{i=1}^{n} D_i Y_i}_{\text{mean among treated}} - \underbrace{\frac{1}{n_0}\sum_{i=1}^{n}(1 - D_i)Y_i}_{\text{mean among control}} \tag{2}$$

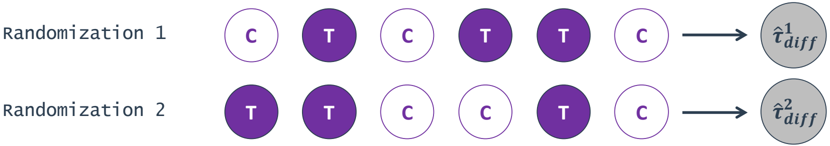
- Conditional on the sample, $\widehat{\tau}_{\text{diff}}$ only varies because of D_i.

Repeated Samples/Randomizations

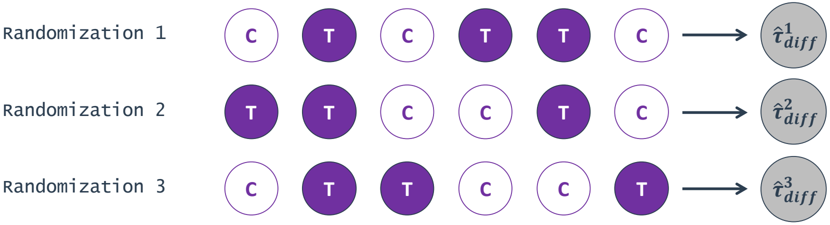
Randomization 1



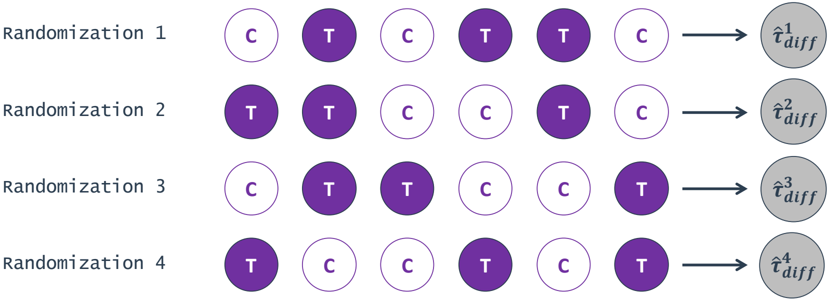
Repeated Samples/Randomizations



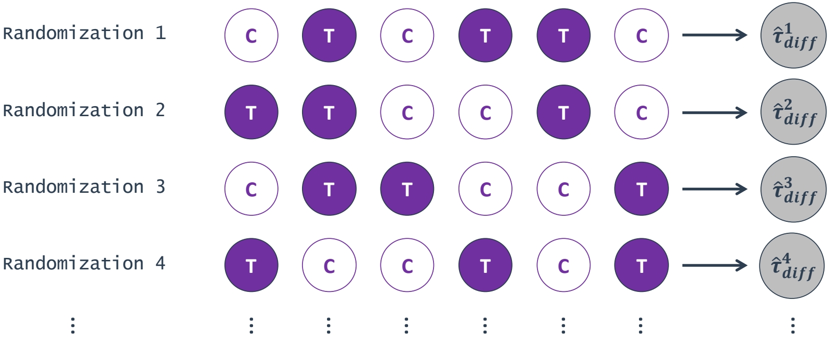
Repeated Samples/Randomizations



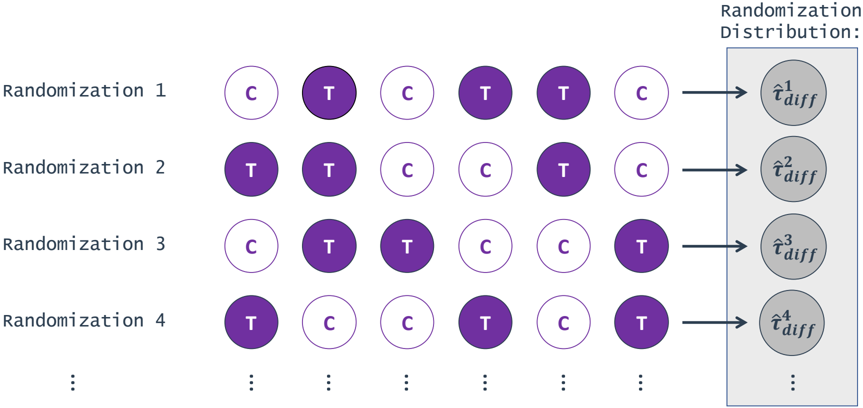
Repeated Samples/Randomizations



Repeated Samples/Randomizations



Repeated Samples/Randomizations



- Randomization distribution = **sampling distribution** of this estimator.

Finite-Sample Properties

- How does $\widehat{\tau}_{\text{diff}}$ vary across randomizations?
- Key properties of the randomization distribution we'd like to know:
  - **Unbiasedness:** is the mean of the randomization distribution equal to the true SATE (i.e., the estimand of our interest)?
  - **Sampling variance:** variance of the randomization distribution?
- Use these properties to construct confidence intervals and conduct tests.

Unbiasedness

- In a completely randomized experiment, $\widehat{\tau}_{\text{diff}}$ is unbiased for τ_{fs}
- Let $\mathbf{PO} = \{Y(1), Y(0)\}$ be the potential outcomes:

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

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

Unbiasedness

- In a completely randomized experiment, $\widehat{\tau}_{\text{diff}}$ is unbiased for τ_{fs}
- Let $\mathbf{PO} = \{Y(1), Y(0)\}$ be the potential outcomes:

$$\begin{aligned}
\mathbb{E}_D[\widehat{\tau}_{\text{diff}}|\mathbf{PO}] &= \frac{1}{n_1}\sum_{i=1}^{n}\mathbb{E}_D[D_iY_i|\mathbf{PO}] - \frac{1}{n_0}\sum_{i=1}^{n}\mathbb{E}_D[(1-D_i)Y_i|\mathbf{PO}] \\
&= \frac{1}{n_1}\sum_{i=1}^{n}\mathbb{E}_D[D_iY_i(1)|\mathbf{PO}] - \frac{1}{n_0}\sum_{i=1}^{n}\mathbb{E}_D[(1-D_i)Y_i(0)|\mathbf{PO}]
\end{aligned}$$

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

Unbiasedness

- In a completely randomized experiment, $\widehat{\tau}_{\text{diff}}$ is unbiased for τ_{fs}
- Let **PO** $= \{Y(1), Y(0)\}$ be the potential outcomes:

$$
\begin{aligned}
\mathbb{E}_D[\widehat{\tau}_{\text{diff}}|\mathbf{PO}] &= \frac{1}{n_1}\sum_{i=1}^{n}\mathbb{E}_D[D_iY_i|\mathbf{PO}] - \frac{1}{n_0}\sum_{i=1}^{n}\mathbb{E}_D[(1-D_i)Y_i|\mathbf{PO}] \\
&= \frac{1}{n_1}\sum_{i=1}^{n}\mathbb{E}_D[D_iY_i(1)|\mathbf{PO}] - \frac{1}{n_0}\sum_{i=1}^{n}\mathbb{E}_D[(1-D_i)Y_i(0)|\mathbf{PO}] \\
&= \frac{1}{n_1}\sum_{i=1}^{n}\mathbb{E}_D[D_i|\mathbf{PO}]Y_i(1) - \frac{1}{n_0}\sum_{i=1}^{n}\mathbb{E}_D[(1-D_i)|\mathbf{PO}]Y_i(0) \\
&= \frac{1}{n}\sum_{i=1}^{n}Y_i(1) - Y_i(0) \quad = \tau_{\text{fs}}
\end{aligned}
$$

Unbiasedness

- In a completely randomized experiment, $\widehat{\tau}_{\text{diff}}$ is unbiased for τ_{fs}
- Let $\textbf{PO} = \{Y(1), Y(0)\}$ be the potential outcomes:

$$
\begin{aligned}
\mathbb{E}_D[\widehat{\tau}_{\text{diff}}|\textbf{PO}] &= \frac{1}{n_1}\sum_{i=1}^{n}\mathbb{E}_D[D_iY_i|\textbf{PO}] - \frac{1}{n_0}\sum_{i=1}^{n}\mathbb{E}_D[(1-D_i)Y_i|\textbf{PO}] \\
&= \frac{1}{n_1}\sum_{i=1}^{n}\mathbb{E}_D[D_iY_i(1)|\textbf{PO}] - \frac{1}{n_0}\sum_{i=1}^{n}\mathbb{E}_D[(1-D_i)Y_i(0)|\textbf{PO}] \\
&= \frac{1}{n_1}\sum_{i=1}^{n}\mathbb{E}_D[D_i|\textbf{PO}]Y_i(1) - \frac{1}{n_0}\sum_{i=1}^{n}\mathbb{E}_D[(1-D_i)|\textbf{PO}]Y_i(0) \\
&= \frac{1}{n_1}\sum_{i=1}^{n}\left(\frac{n_1}{n}\right)Y_i(1) - \frac{1}{n_0}\sum_{i=1}^{n}\left(\frac{n_0}{n}\right)Y_i(0) \\
&= \frac{1}{n}\sum_{i=1}^{n}Y_i(1) - Y_i(0) \quad = \tau_{\text{fs}}
\end{aligned}
$$

- Note: number of treated/control units doesn't matter for unbiasedness!

Finite-Sample Sampling Variance

- Sampling variance of the *difference-in-means estimator* is:

$$\mathbb{V}_D(\widehat{\tau}_{\text{diff}}|\mathbf{PO}) = \frac{S_1^2}{n_1} + \frac{S_0^2}{n_0} - \frac{S_{\tau_i}^2}{n} \tag{3}$$

Finite-Sample Sampling Variance

- Sampling variance of the *difference-in-means estimator* is:

$$\mathbb{V}_D(\widehat{\tau}_{\text{diff}}|\mathbf{PO}) = \frac{S_1^2}{n_1} + \frac{S_0^2}{n_0} - \frac{S_{\tau_i}^2}{n} \tag{3}$$

- S_0^2 and S_1^2 are in-sample variances of $Y_i(0)$ and $Y_i(1)$, respectively:

$$S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i(0) - \overline{Y}(0))^2 \qquad S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i(1) - \overline{Y}(1))^2$$

- Here, $\overline{Y}(d) = \frac{1}{n}\sum_{i=1}^{n} Y_i(d)$.

Finite-Sample Sampling Variance

- Sampling variance of the *difference-in-means estimator* is:

$$\mathbb{V}_D(\widehat{\tau}_{\text{diff}}|\mathbf{PO}) = \frac{S_1^2}{n_1} + \frac{S_0^2}{n_0} - \frac{S_{\tau_i}^2}{n} \tag{3}$$

- S_0^2 and S_1^2 are in-sample variances of $Y_i(0)$ and $Y_i(1)$, respectively:

$$S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i(0) - \overline{Y}(0))^2 \qquad S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i(1) - \overline{Y}(1))^2$$

- Here, $\overline{Y}(d) = \frac{1}{n}\sum_{i=1}^{n} Y_i(d)$.
- $S_{\tau_i}^2$ is the in-sample variation from individual treatment effects:

$$S_{\tau_i}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i(1) - Y_i(0) - \tau_{\text{fs}})^2$$

- None of these are directly observable!

Finite-Sample Sampling Variance

$$\mathbb{V}_D(\widehat{\tau}_{\text{diff}}|\mathbf{PO}) = \frac{S_1^2}{n_1} + \frac{S_0^2}{n_0} - \frac{S_{\tau_i}^2}{n} \tag{4}$$

- If the treatment effects are *constant across units*, then $S_{\tau_i}^2 = 0$.
    - $\rightsquigarrow$ Sampling variance is largest when treatment effects are constant.

Finite-Sample Sampling Variance

$$\mathbb{V}_D(\widehat{\tau}_{\text{diff}}|\mathbf{PO}) = \frac{S_1^2}{n_1} + \frac{S_0^2}{n_0} - \frac{S_{\tau_i}^2}{n} \tag{4}$$

- If the treatment effects are <u>*constant across units*</u>, then $S_{\tau_i}^2 = 0$.
    - $\rightsquigarrow$ Sampling variance is largest when treatment effects are constant.

- Intuition looking at two-unit samples:

	$i = 1$	$i = 2$	Avg.
$Y_i(0)$	10	-10	0
$Y_i(1)$	10	-10	0
τ_i	**0**	**0**	$\tau_{\text{fs}} = 0$

	$i = 1$	$i = 2$	Avg.
$Y_i(0)$	-10	10	0
$Y_i(1)$	10	-10	0
τ_i	**20**	**-20**	$\tau_{\text{fs}} = 0$

1. Both have $\tau_{\text{fs}} = 0$, first has constant effects across units.
2. In Case 1, $\widehat{\tau}_{\text{diff}} = 20$ or -20 depending on randomization.
3. In Case 2, $\widehat{\tau}_{\text{diff}} = 0$ in either randomization.

Estimating the Sampling Variance

- We can use sample variances within levels of D_i to estimate S_0^2 and S_1^2:

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

- Here, $\overline{Y}_0 = (1/n_0) \sum_{i=1}^{n} (1 - D_i) Y_i$ and $\overline{Y}_1 = (1/n_1) \sum_{i=1}^{n} D_i Y_i$.

Estimating the Sampling Variance

- We can use sample variances within levels of D_i to estimate S_0^2 and S_1^2:

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

- Here, $\overline{Y}_0 = (1/n_0) \sum_{i=1}^{n} (1 - D_i) Y_i$ and $\overline{Y}_1 = (1/n_1) \sum_{i=1}^{n} D_i Y_i$.

- But what about $S_{\tau_i}^2$?

$$S_{\tau_i}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\underbrace{Y_i(1) - Y_i(0)}_{???} - \tau_{\text{fs}})^2$$

- Q: What to do?

Bounding the Sampling Variance

- First approach: find the worst possible (largest) variance.
- We can rewrite the variance as:

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

  - $\rightsquigarrow$ See Appendix A of Imbens & Rubin, Ch.6 (p.105-107)

Bounding the Sampling Variance

- First approach: find the worst possible (largest) variance.
- We can rewrite the variance as:

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

  - ⇝ See Appendix A of Imbens & Rubin, Ch.6 (p.105-107)
- Last term is the **covariance** between potential outcomes:

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

Bounding the Sampling Variance

- First approach: find the worst possible (largest) variance.
- We can rewrite the variance as:

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

  - $\rightsquigarrow$ See Appendix A of Imbens & Rubin, Ch.6 (p.105-107)
- Last term is the **covariance** between potential outcomes:

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

- We can use the **Cauchy-Schwarz inequality**: $S_{01} \leq S_0 S_1$

$$\mathbb{V}(\widehat{\tau}_{\text{diff}}|\mathbf{PO}) \leq \frac{1}{n}\left(\frac{n_1}{n_0}S_0^2 + \frac{n_0}{n_1}S_1^2 + 2S_0S_1\right) = \frac{n_0 n_1}{n}\left(\frac{S_0}{n_0} + \frac{S_1}{n_1}\right)^2$$

- Upper bound that is only a function of identified parameters.

Conservative Variance Estimation

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

$$\widehat{\mathbb{V}} = \frac{\widehat{\sigma}_1^2}{n_1} + \frac{\widehat{\sigma}_0^2}{n_0}, \qquad \mathbb{E}[\widehat{\mathbb{V}} | \mathbf{PO}] = \frac{S_1^2}{n_1} + \frac{S_0^2}{n_0}$$

- Notice that $\widehat{\mathbb{V}}$ is biased for $\mathbb{V}(\widehat{\tau}_{\text{diff}} | \mathbf{PO})$, but that bias is always positive!
- Leads to **'conservative' inferences**:
  - Standard errors, $\sqrt{\widehat{\mathbb{V}}}$, will be at least as big as they should be.
  - Confidence intervals using $\sqrt{\widehat{\mathbb{V}}}$ will be at least as wide as they should be.
  - Type I error rates will still be correct, power will be lower.
  - Both will be exactly right if treatment effects are constant.

Inference in the Neyman Approach

- If n is large, central limit theorem (CLT) will imply that $\widehat{\tau}_{\text{diff}}$ is approximately normal.

Inference in the Neyman Approach

- If n is large, central limit theorem (CLT) will imply that $\widehat{\tau}_{\text{diff}}$ is approximately normal.
- Formulate confidence intervals in the usual way:

$$\text{CI}^{95}(\tau_{\text{fs}}) = \left(\widehat{\tau}_{\text{diff}} - 1.96 \cdot \widehat{\mathbb{V}}^{1/2},\ \widehat{\tau}_{\text{diff}} + 1.96 \cdot \widehat{\mathbb{V}}^{1/2}\right)$$

Inference in the Neyman Approach

- If n is large, central limit theorem (CLT) will imply that $\widehat{\tau}_{\text{diff}}$ is approximately normal.

- Formulate confidence intervals in the usual way:

$$\text{CI}^{95}(\tau_{\text{fs}}) = \left(\widehat{\tau}_{\text{diff}} - 1.96 \cdot \widehat{\mathbb{V}}^{1/2},\ \widehat{\tau}_{\text{diff}} + 1.96 \cdot \widehat{\mathbb{V}}^{1/2}\right)$$

- Testing is very similar to standard normal-approximation tests:

$$H_0 : \frac{1}{n}\sum_{i=1}^{n} Y_i(1) - Y_i(0) = 0 \qquad T = \frac{\widehat{\tau}_{\text{diff}}}{\sqrt{\widehat{\mathbb{V}}}} \overset{a}{\sim} N(0, 1)$$

Population Estimands

- What if we want to make inference to a (super)population?
  - n units are a **simple random sample** from the population.

Population Estimands

- What if we want to make inference to a (super)population?
  - n units are a **simple random sample** from the population.
- New goal: inference for the PATE, $\tau = \mathbb{E}[Y_i(1) - Y_i(0)]$.
  - Average of the SATEs across repeated samples: PATE = $\mathbb{E}$[SATE].

Population Estimands

- What if we want to make inference to a (super)population?
  - n units are a **simple random sample** from the population.
- New goal: inference for the PATE, $\tau = \mathbb{E}[Y_i(1) - Y_i(0)]$.
  - Average of the SATEs across repeated samples: PATE $= \mathbb{E}[\text{SATE}]$.
- Difference-in-means is **unbiased** across repeated samples:

$$\mathbb{E}[\hat{\tau}_{\text{diff}}] = \underbrace{\mathbb{E}[\mathbb{E}_D[\hat{\tau}_{\text{diff}} | \mathbf{PO}]]}_{\text{iterated expectation}} = \underbrace{\mathbb{E}[\tau_{\text{fs}}]}_{\text{SATE unbiasedness}} = \tau$$

Population Sampling Variance

- What about the sampling variance of $\hat{\tau}_{\text{diff}}$ when estimating the PATE?
  - Variation now comes from random sample **and** random assignment.

Population Sampling Variance

- What about the sampling variance of $\hat{\tau}_{\text{diff}}$ when estimating the PATE?
    - Variation now comes from random sample **and** random assignment.
- It turns out that the sampling variance of the estimator is simply:

$$\mathbb{V}(\widehat{\tau}_{\text{diff}}) = \frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1} = \frac{\mathbb{V}(Y_i(0))}{n_0} + \frac{\mathbb{V}(Y_i(1))}{n_1}$$

    - Here, σ_0^2 and σ_1^2 are the population-level variances of $Y_i(0)$ and $Y_i(1)$.

Population Sampling Variance

- What about the sampling variance of $\hat{\tau}_{\text{diff}}$ when estimating the PATE?
  - Variation now comes from random sample **and** random assignment.
- It turns out that the sampling variance of the estimator is simply:

$$\mathbb{V}(\widehat{\tau}_{\text{diff}}) = \frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1} = \frac{\mathbb{V}(Y_i(0))}{n_0} + \frac{\mathbb{V}(Y_i(1))}{n_1}$$

  - Here, σ_0^2 and σ_1^2 are the population-level variances of $Y_i(0)$ and $Y_i(1)$.
- The variance of τ_i term drops out indicating higher variance for PATE than SATE.

Estimating Population Sampling Variance

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

- Notice that the Neyman estimator $\widehat{\mathbb{V}}$ is now unbiased for $\mathbb{V}(\widehat{\tau}_{\text{diff}})$:

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

- Two interpretations of $\widehat{\mathbb{V}}$:
  1. Unbiased estimator for sampling variance of the difference-in-means estimator of the PATE.
  2. Conservative estimator for the sampling variance of the difference-in-means estimator of the SATE.

1/ Standard Analysis of Treatment Effects and Variance Estimation

Social Pressure Data

```
> library(infer)
> data(social, package = "qss") # also available at "https://bit.ly/4ampWhU"
> social <- as_tibble(social); social

# A tibble: 305,866 × 6
   sex    yearofbirth primary2004 messages   primary2006 hhsize
   <chr>        <int>       <int> <chr>            <int>  <int>
 1 male          1941           0 Civic Duty           0      2
 2 female        1947           0 Civic Duty           0      2
 3 male          1951           0 Hawthorne            1      3
 4 female        1950           0 Hawthorne            1      3
 5 female        1982           0 Hawthorne            1      3
 6 male          1981           0 Control              0      3
 7 female        1959           0 Control              1      3
 8 male          1956           0 Control              1      3
 9 female        1968           0 Control              0      2
10 male          1967           0 Control              0      2
#  305,856 more rows
#  Use `print(n = ...)` to see more rows
```

- Data source: Kosuke Imai. 2017. Quantitative Social Science: An Introduction. Princeton University Press. ⇝ one of our recommended textbooks!

Two-Sample Hypotheses

- Parameter: **population ATE** $\mu_T - \mu_C$
  - μ_T: Turnout rate in the population if everyone received treatment.
  - μ_C: Turnout rate in the population if everyone received control.
- Goal: learn about the population difference in means.
- Usual null hypothesis: no difference in population means (ATE = 0)
  - Null: $H_0 : \mu_T - \mu_C = 0$
  - Two-sided alternative: $H_0 : \mu_T - \mu_C \neq 0$
- In words: are the differences in sample means just due to random chance?

Calculating the Difference in Proportion

`infer` functions with binary outcome (i.e., voted or not) work best with factor variables.

- Use the argument `success = level` in `specify()` to indicate the level of response/outcome considered as success (voted = 1).

```
> social <- social |>
    filter(messages %in% c("Neighbors", "Control"))

> social <- social |>
    mutate(turnout = if_else(primary2006 == 1, "Voted", "Didn't Vote"))

> est_ate <- social |>
    specify(turnout ~ messages, success = "Voted") |>
    calculate(stat = "diff in props", order = c("Neighbors", "Control")); est_ate

Response: turnout (factor)
Explanatory: messages (factor)
# A tibble: 1 × 1
    stat
   <dbl>
1 0.0813
```

- Our point estimate for the ATE = 0.0813; i.e., diff in proportions between T and C.

Calculating Neyman Robust Variance

- For binary outcomes, recall that $\text{Var}(Y) = p(1-p)$.
- The Neyman variance estimator for the difference in proportions is:

$$\widehat{\mathbb{V}} = \frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_0(1-\widehat{p}_0)}{n_0}.$$

```r
# Calculate sample sizes for each group
> n_treatment <- social |>
    filter(messages == "Neighbors") |>
    nrow()

> n_control   <- social |>
    filter(messages == "Control") |>
    nrow()

# Calculate the estimated proportion of voters for each group
> p_treatment <- social |>
    filter(messages == "Neighbors") |>
    summarize(p = mean(turnout == "Voted")) |>
    pull(p); p_treatment
[1] 0.3779482

> p_control <- social |>
    filter(messages == "Control") |>
    summarize(p = mean(turnout == "Voted")) |>
    pull(p); p_control
[1] 0.2966383
```

Calculating Neyman Robust Variance

```
# For a binary outcome, the sample variance is p * (1 - p)
> var_treatment <- p_treatment * (1 - p_treatment)
> var_control <- p_control * (1 - p_control)

# Neyman robust variance estimator for the difference in proportions:
> neyman_var <- var_treatment / n_treatment + var_control / n_control
> neyman_se  <- sqrt(neyman_var)

# Present the results
> cat("Estimated difference in proportions (Neighbors - Control):", pull(est_ate), "\n")
Estimated difference in proportions (Neighbors - Control): 0.08130991

> cat("Neyman robust variance estimate:", neyman_var, "\n")
Neyman robust variance estimate: 7.245366e-06

> cat("Neyman robust standard error:", neyman_se, "\n")
Neyman robust standard error: 0.002691722
```

- The resulting neyman_var and neyman_se give the estimated variance and standard error.
    - The estimates are now in terms of proportions, we can convert them back to percentages and crosscheck with Table 2 from Gerber et al. (2008)!

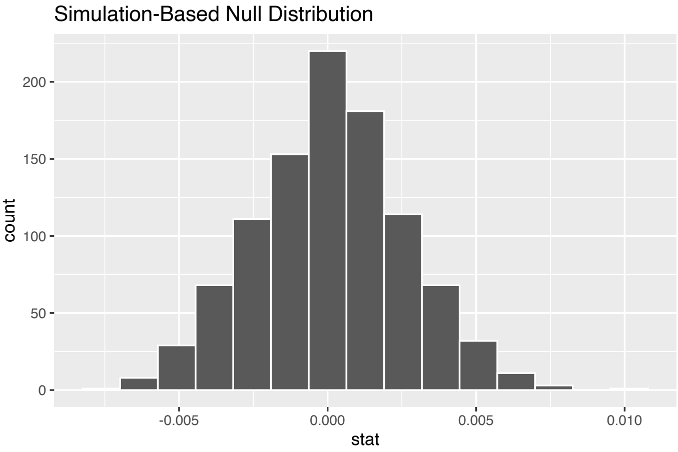
Calculating the Diff in Props in Each Sample

- Alternative way to estimate the sampling variance/standard error using bootstrap resampling:

```
> null_dist <- social |>
    specify(turnout ~ messages, success = "Voted") |>
    hypothesize(null = "independence") |>
    generate(reps = 1000, type = "permute") |>
    calculate(stat = "diff in props", order = c("Neighbors", "Control")); null_dist

Response: turnout (factor)
Explanatory: messages (factor)
Null Hypothesis: independence
# A tibble: 1,000 × 2
   replicate      stat
       <int>     <dbl>
 1         1  0.00364
 2         2  0.000470
 3         3 -0.000252
 4         4  0.00462
 5         5  0.00176
 6         6  0.000721
 7         7  0.00298
 8         8 -0.00412
 9         9  0.00276
10        10  0.00336
#  990 more rows
#  Use `print(n = ...)` to see more rows
```

Visualize the Null/Reference Distribution

```
> null_dist |>
    visualize()
```



Calculating Robust Variance with Resampling

```
> null_dist |>
    summarise(var(stat), sd(stat))

# A tibble: 1 × 2
  `var(stat)` `sd(stat)`
        <dbl>      <dbl>
1  0.00000678    0.00260
```

- Compare the bootstrap results to the Neyman variance estimator from earlier:

```
> cat("Neyman robust variance estimate:", neyman_var, "\n")
Neyman robust variance estimate: 7.245366e-06

> cat("Neyman robust standard error:", neyman_se, "\n")
Neyman robust standard error: 0.002691722
```

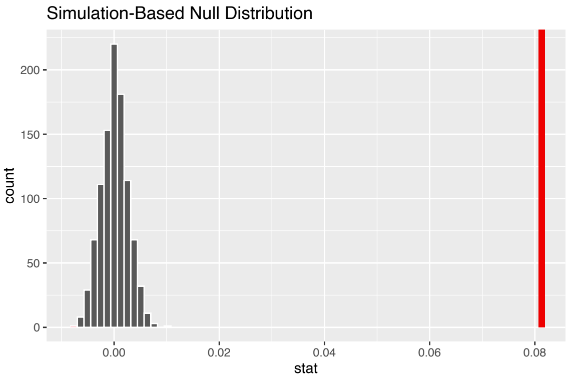
Calculating P-Values

```
> ate_pval <- null_dist |>
    get_p_value(obs_stat = est_ate, direction = "both"); ate_pval

Warning message:
Please be cautious in reporting a p-value of 0. This result is an approximation based on
the number of `reps` chosen in the `generate()` step. See `?get_p_value()` for more information.

# A tibble: 1 × 1
  p_value
    <dbl>
1       0
```

- Theoretically, a true p-value being equal to 0 is impossible (ranges from [0,1]).
  - However, `get_p_value()` may return 0 in some cases due to the simulation-based nature of the `infer` package.

Visualize P-Values

```
> null_dist |>
    visualize() +
    shade_p_value(obs_stat = est_ate, direction = "both")
```



- Red vertical line depicts the observed diff. in proportion (i.e., 0.0813; p.33)

Hypothesis Tests and Confidence Intervals

- There is a deep connection between confidence intervals and tests.

```
> social |>
    specify(turnout ~ messages, success = "Voted") |>
    generate(reps = 1000, type = "bootstrap") |>
    calculate(stat = "diff in props",
              order = c("Neighbors", "Control"))  |>
    get_ci(level = 0.95)

# A tibble: 1 × 2
  lower_ci upper_ci
     <dbl>    <dbl>
1   0.0766   0.0865
```

- Any value outside of a $100 \times (1 - \alpha)\%$ confidence interval would have a p-value less than α (e.g., 0.05) if we tested it as the null hypothesis.
  - 95% CI for the social pressure experiment: [0.0766, 0.0865]
  - $\rightsquigarrow$ p-value for $H_0 : \mu_T - \mu_C = 0$ less than 0.05.

Next Up

- How does the workhorse estimator, OLS, fit into this story?
- Why might we use regression?
  - **Simplicity**: well-known tool that is already very common.
  - **Increased precision**: we may want to consider covariates for more precise effect estimates.

Have a great weekend! :)

Contact Information:
jaewon.yoo@iss.nthu.edu.tw
https://j1yoo4.github.io/