

# 3. Inference for the Average Treatment Effect

ISS5096 || ECI

Jaewon (“Jay-one”) Yoo

National Tsing Hua University

# Roadmap

1. Completely Randomized Experiments
2. Block randomized experiments
3. Neyman Inference in Practice
4. In-class Exercises

# 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 recipient 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:

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

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

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

- Q: Can this analysis be justified by randomization?

# 1/ Completely Randomized Experiments

# Causal Estimand of Interest

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

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

- 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**

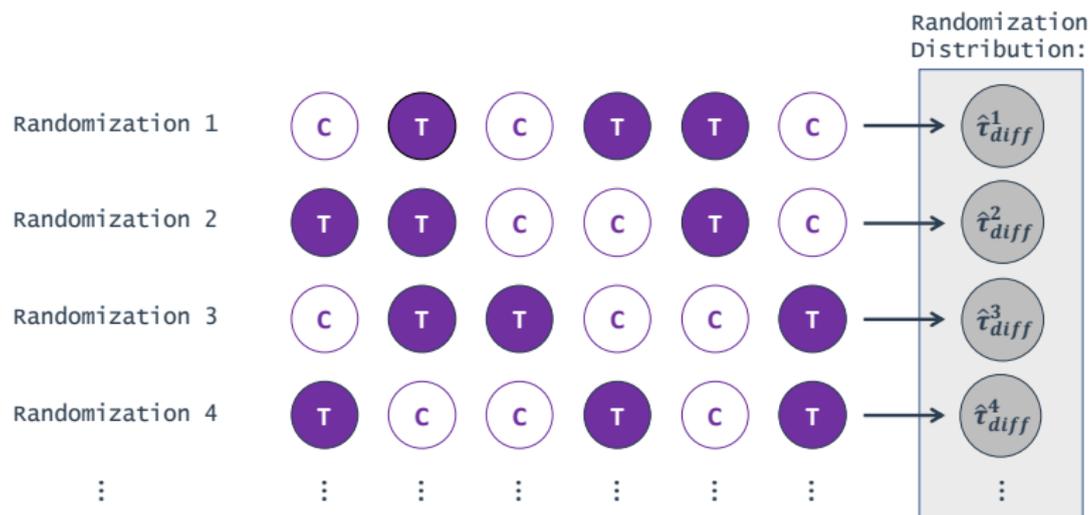
- $n$  units,  $n_1$  treated, and  $n_0$  control.

- Natural estimator for the SATE, **difference-in-means**:

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

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

# 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  $\tau_{\text{fs}}$
- Let  $\mathbf{PO} = \{Y(1), Y(0)\}$  be the potential outcomes:

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

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

# Finite-Sample Sampling Variance

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

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

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

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

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

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

- None of these are directly observable!

# Finite-Sample Sampling Variance

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

- If the treatment effects are constant across units, then  $S_{\tau_i}^2 = 0$ .
  - $\rightsquigarrow$  Sampling variance is largest when treatment effects are constant.
- Intuition looking at two-unit samples:

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

	$i = 1$	$i = 2$	Avg.
$Y_i(0)$	-10	10	0
$Y_i(1)$	10	-10	0
$\tau_i$	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  $S_1^2$ :

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

- Here,  $\bar{Y}_0 = (1/n_0) \sum_{i=1}^n (1 - D_i) Y_i$  and  $\bar{Y}_1 = (1/n_1) \sum_{i=1}^n D_i Y_i$ .
- But what about  $S_{\tau_i}^2$ ?

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

- 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)
- Last term is the **covariance** between potential outcomes:

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

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

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

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

# Conservative Variance Estimation

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

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

- Notice that  $\widehat{V}$  is biased for  $\mathbb{V}(\widehat{\tau}_{\text{diff}}|\mathbf{PO})$ , but that bias is always positive!
- Leads to **'conservative' inferences**:
  - Standard errors,  $\sqrt{\widehat{V}}$ , will be at least as big as they should be.
  - Confidence intervals using  $\sqrt{\widehat{V}}$  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.
- Formulate confidence intervals in the usual way:

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

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

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

# Population Estimands

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

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

# 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**.
- It turns out that the sampling variance of the estimator is simply:

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

- Here,  $\sigma_0^2$  and  $\sigma_1^2$  are the population-level variances of  $Y_i(0)$  and  $Y_i(1)$ .
- The variance of  $\tau_i$  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.

## **2/** Block randomized experiments

Cf. Imbens & Rubin (2015), *Causal Inference for Statistics, Social, and Biomedical Sciences*,  
Chs. 9–10

# Block randomized experiments

- Basic idea: run completely randomized experiments within strata defined by covariates.
- Main motivation: **more efficient** than standard design (ie, lower SEs)
- George Box: “Block what you can and randomize what you cannot.”
- We will compare variance of blocked designs to complete randomization.
  - Some confusion in the literature: can blocking hurt?
  - Care needed: comparison depends on sample assumptions (Pashley & Miratrix, 2021, JEBS)

# Simple two block example

- GOTV mailer experiment:
  - We have  $n$  households with registered voters.
  - Complete randomization: choose  $n_1$  households to get mailers.
  - Outcome,  $Y_i$ : turnout in election.
- What if we have data from the voter file: **previous turnout**.
  - Create blocks:  $V_i = 1$  if voted in last election,  $V_i = 0$  otherwise.
  - $n_v$  is the number of previous voters,
  - $n_{nv} = n - n_v$  is the number of previous nonvoters.
- SATEs within blocks defined by  $V_i$ :

$$\tau_{v, fs} = \frac{1}{n_v} \sum_{i:V_i=1} \{Y_i(1) - Y_i(0)\} \quad \tau_{nv, fs} = \frac{1}{n_{nv}} \sum_{i:V_i=0} \{Y_i(1) - Y_i(0)\}$$

- Decomposing the overall SATE by block:

$$\tau_{fs} = \underbrace{\left( \frac{n_v}{n_v + n_{nv}} \right)}_{\text{fraction voters}} \tau_{v, fs} + \underbrace{\left( \frac{n_{nv}}{n_v + n_{nv}} \right)}_{\text{fraction nonvoters}} \tau_{nv, fs}$$

# Block randomized design

- **Block/stratified randomized experiment:**
  - Completely randomized experiment in each block.
  - Choose  $n_{1,v}$  voters to be treated,  $n_{0,v} = n_v - n_{1,v}$  control.
  - Choose  $n_{1,nv}$  nonvoters to be treated,  $n_{0,nv} = n_{nv} - n_{1,nv}$  control.
- Probability of treatment in each group called the **propensity score:**
  - Prob. of treatment for voters:  $\mathbb{P}(D_i = 1 \mid V_i = 1) = p_v = n_{1,v}/n_v$
  - Prob. of treatment for nonvoters:  $\mathbb{P}(D_i = 1 \mid V_i = 0) = p_{nv} = n_{1,nv}/n_{nv}$
- Blocking ensures balance across blocks:
  - With complete randomization, treatment might be very imbalanced across  $V_j$ .
  - No possibility of “chance” imbalances skewing the estimates.

# Estimators in blocked designs

- Within-strata difference in means:

$$\hat{\tau}_v = \bar{Y}_{1,v} - \bar{Y}_{0,v} = \frac{1}{n_{1,v}} \sum_{i:V_i=1} D_i Y_i - \frac{1}{n_{0,v}} \sum_{i:V_i=1} (1 - D_i) Y_i$$

$$\hat{\tau}_{nv} = \bar{Y}_{1,nv} - \bar{Y}_{0,nv} = \frac{1}{n_{1,nv}} \sum_{i:V_i=0} D_i Y_i - \frac{1}{n_{0,nv}} \sum_{i:V_i=0} (1 - D_i) Y_i$$

- Unbiased for the within-strata SATEs:  $\mathbb{E}[\hat{\tau}_v \mid \mathbf{PO}] = \tau_v$
- $\rightsquigarrow$  unbiased estimator for the overall SATE:

$$\hat{\tau}_b = \left(\frac{n_v}{n}\right) \hat{\tau}_v + \left(\frac{n_{nv}}{n}\right) \hat{\tau}_{nv}$$

- Equivalent to the regular difference in means if  $p_v = p_{nv} = 1/2$ .
- Otherwise, standard  $\hat{\tau}_{\text{diff}}$  under block design will be **biased**.

# Sampling variance of blocking estimator

- Each block is a completely randomized experiment so we have:

$$\mathbb{V}(\widehat{\tau}_v | \mathbf{PO}) = \frac{S_{1,v}^2}{n_{1,v}} + \frac{S_{0,v}^2}{n_{0,v}} - \frac{S_{\tau_i,v}^2}{n_v}$$

- $S_{d,v}^2$  are the within-block sample variances of the potential outcomes
- Finite sample variance of the blocked estimator:

$$\mathbb{V}(\widehat{\tau}_b | \mathbf{PO}) = \left(\frac{n_v}{n}\right)^2 \mathbb{V}(\widehat{\tau}_v | \mathbf{PO}) + \left(\frac{n_{nv}}{n}\right)^2 \mathbb{V}(\widehat{\tau}_{nv} | \mathbf{PO})$$

- Use the conservative variance estimators from each strata:

$$\widehat{\mathbb{V}}_b = \left(\frac{n_v}{n}\right)^2 \left(\frac{\widehat{\sigma}_{1,v}^2}{n_{1,v}} + \frac{\widehat{\sigma}_{0,v}^2}{n_{0,v}}\right) + \left(\frac{n_{nv}}{n}\right)^2 \left(\frac{\widehat{\sigma}_{1,nv}^2}{n_{1,nv}} + \frac{\widehat{\sigma}_{0,nv}^2}{n_{0,nv}}\right)$$

- $\widehat{\sigma}_{d,v}^2$  are the within-strata **observed outcome variances**

# General blocking notation

- Blocks,  $j \in \{1, \dots, J\}$ .
  - Block indicator  $B_i = j$  if  $i$  is in block  $j$ .
  - Sizes:  $n_j > 2$  and proportions  $w_j = n_j/n$ .
  - Number treated in each block:  $n_{1,j}$  and  $n_{0,j} = n_j - n_{1,j}$
- Within-block estimators:

$$\hat{\tau}_j = \frac{1}{n_{1,j}} \sum_{i:B_i=j} D_i Y_i - \frac{1}{n_{0,j}} \sum_{i:B_i=j} (1 - D_i) Y_i, \quad \widehat{V}(\hat{\tau}_j) = \frac{\widehat{\sigma}_{1,j}^2}{n_{1,j}} + \frac{\widehat{\sigma}_{0,j}^2}{n_{0,j}}$$

- Aggregate blocking estimators:

$$\hat{\tau}_b = \sum_{j=1}^J w_j \hat{\tau}_j, \quad \widehat{V}(\hat{\tau}_b) = \sum_{j=1}^J w_j^2 \widehat{V}(\hat{\tau}_j)$$

# Efficiency of blocking

- Efficiency of block versus CR depends on the sampling scheme.
  - Usually blocking will be more efficient/lower variance, but not always.
- Finite sample difference in sampling variances:

$$\mathbb{V}(\widehat{\tau}_{\text{CR}} \mid \mathbf{PO}) - \mathbb{V}(\widehat{\tau}_b \mid \mathbf{PO}) = \frac{1}{n-1} [B - W]$$

- Difference can be positive or negative (blocking can hurt or help)
  - $B$  = between-block variation in outcome means.
  - $W$  = within-block estimation variability.
  - High  $B$ , low  $W$   $\rightsquigarrow$  blocking is more efficient.
  - **Blocking is better when outcomes vary a lot across blocks, not much within blocks** (blocks are predictive of outcome, so usually the case)

# How to block

- Discrete covariates  $\rightsquigarrow$  blocks by unique combinations.
- Alternative: create blocks by creating homogeneous groups in  $\mathbf{X}$ .
  - Choose distance metric such as Mahalanobis distance:

$$M(\mathbf{x}_i, \mathbf{x}_k) = \sqrt{(\mathbf{x}_i - \mathbf{x}_k)^\top \widehat{\mathbf{V}}(\mathbf{X})^{-1} (\mathbf{x}_i - \mathbf{x}_k)}$$

- Difficult/impossible to find optimal blocks in general, but “greedy” algorithms exist.
- More on this later in matching/weighting estimators!

## **3/** Neyman Inference in Practice

# Social Pressure Data

- Basic setup: unblocked Neighbors vs. Control comparison
- Blocked extension: estimate within blocks, then take a weighted average

```
1 > library(infer)
2 > data(social, package = "qss") # also available at "https://bit.ly/4ampWhU"
3 > social <- as_tibble(social); social
4
5 # A tibble: 305,866 × 6
6   sex      yearofbirth primary2004 messages  primary2006 hhsiz
7   <chr>      <int>      <int> <chr>      <int> <int>
8   1 male      1941          0 Civic Duty    0     2
9   2 female   1947          0 Civic Duty    0     2
10  3 male     1951          0 Hawthorne    1     3
11  4 female   1950          0 Hawthorne    1     3
12  5 female   1982          0 Hawthorne    1     3
13  6 male     1981          0 Control      0     3
14  7 female   1959          0 Control      1     3
15  8 male     1956          0 Control      1     3
16  9 female   1968          0 Control      0     2
17 10 male     1967          0 Control      0     2
18 # i 305,856 more rows
19 # i 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!

# Weak Null of No Effect

- Parameter: **population ATE**  $\mu_T - \mu_C$ 
  - $\mu_T = \mathbb{E}[Y_i(1)]$ : Turnout rate in the population if everyone received treatment.
  - $\mu_C = \mathbb{E}[Y_i(0)]$ : Turnout rate in the population if everyone received control.
- Goal: learn about the population difference in mean potential outcomes.
- Usual null hypothesis: no difference in population means (ATE = 0)
  - Null:  $H_0 : \mu_T - \mu_C = 0$
  - Two-sided alternative:  $H_A : \mu_T - \mu_C \neq 0$
- In words: is the difference 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).

```
1 > social <- social |>
2   filter(messages %in% c("Neighbors", "Control"))
3
4 > social <- social |>
5   mutate(turnout = if_else(primary2006 == 1, "Voted", "Didn't Vote"))
6
7 > est_ate <- social |>
8   specify(turnout ~ messages, success = "Voted") |>
9   calculate(stat = "diff in props", order = c("Neighbors", "Control")); est_ate
10
11 Response: turnout (factor)
12 Explanatory: messages (factor)
13 # A tibble: 1 × 1
14   stat
15   <dbl>
16 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:

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

```
1 # Calculate sample sizes for each group
2 > n_treatment <- social |>
3   filter(messages == "Neighbors") |>
4   nrow()
5
6 > n_control <- social |>
7   filter(messages == "Control") |>
8   nrow()
9
10 # Calculate the estimated proportion of voters for each group
11 > p_treatment <- social |>
12   filter(messages == "Neighbors") |>
13   summarize(p = mean(turnout == "Voted")) |>
14   pull(p); p_treatment
15 [1] 0.3779482
16
17 > p_control <- social |>
18   filter(messages == "Control") |>
19   summarize(p = mean(turnout == "Voted")) |>
20   pull(p); p_control
21 [1] 0.2966383
```

# Calculating Neyman Robust Variance

```
1 # For a binary outcome, the sample variance is p * (1 - p)
2 > var_treatment <- p_treatment * (1 - p_treatment)
3 > var_control <- p_control * (1 - p_control)
4
5 # Neyman robust variance estimator for the difference in proportions:
6 > neyman_var <- var_treatment / n_treatment + var_control / n_control
7 > neyman_se <- sqrt(neyman_var)
8
9 # Present the results
10 > cat("Estimated difference in proportions (Neighbors - Control):", pull(est_ate), "\n")
11 Estimated difference in proportions (Neighbors - Control): 0.08130991
12
13 > cat("Neyman robust variance estimate:", neyman_var, "\n")
14 Neyman robust variance estimate: 7.245366e-06
15
16 > cat("Neyman robust standard error:", neyman_se, "\n")
17 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)!

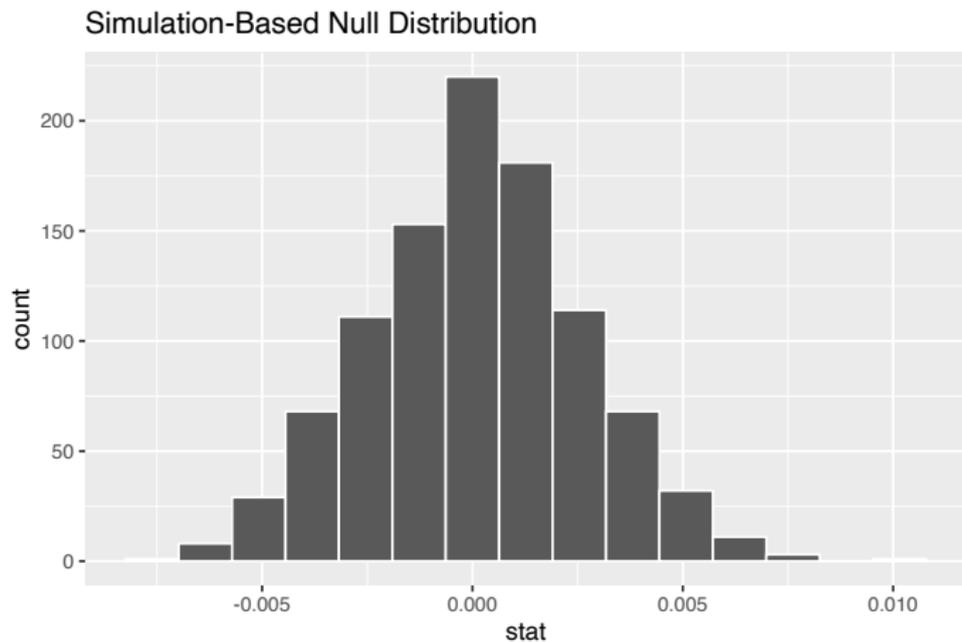
# Calculating the Diff in Props in Each Sample

- Alternative way to estimate the sampling variance / standard error using a simulation-based null distribution:

```
1 > null_dist <- social |>
2   specify(turnout ~ messages, success = "Voted") |>
3   hypothesize(null = "independence") |>
4   generate(reps = 1000, type = "permute") |>
5   calculate(stat = "diff in props", order = c("Neighbors", "Control")); null_dist
6
7 Response: turnout (factor)
8 Explanatory: messages (factor)
9 Null Hypothesis: independence
10 # A tibble: 1,000 × 2
11   replicate    stat
12   <int>      <dbl>
13 1         1  0.00364
14 2         2  0.000470
15 3         3 -0.000252
16 4         4  0.00462
17 5         5  0.00176
18 6         6  0.000721
19 7         7  0.00298
20 8         8 -0.00412
21 9         9  0.00276
22 10        10  0.00336
23 # i 990 more rows
24 # i Use `print(n = ...)` to see more rows
```

# Visualize the Null/Reference Distribution

```
1 > null_dist |>  
2   visualize()
```



# Calculating Robust Variance with Resampling

```
1 > null_dist |>
2   summarise(var(stat), sd(stat))
3
4 # A tibble: 1 × 2
5   `var(stat)` `sd(stat)`
6     <dbl>      <dbl>
7 1 0.00000678  0.00260
```

- Compare the resampling-based variance and standard error to the Neyman estimates from earlier:

```
1 > cat("Neyman robust variance estimate:", neyman_var, "\n")
2 Neyman robust variance estimate: 7.245366e-06
3
4 > cat("Neyman robust standard error:", neyman_se, "\n")
5 Neyman robust standard error: 0.002691722
```

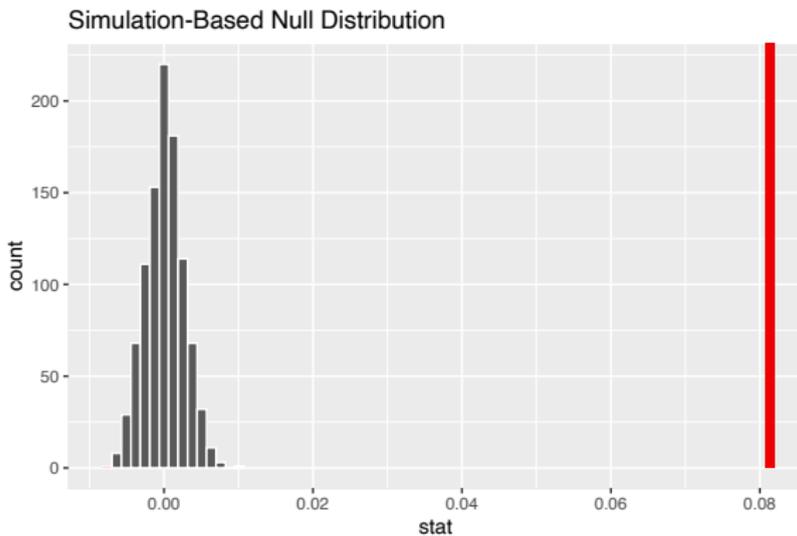
# Calculating P-Values

```
1 > ate_pval <- null_dist |>
2   get_p_value(obs_stat = est_ate, direction = "both"); ate_pval
3
4 Warning message:
5 Please be cautious in reporting a p-value of 0. This result is an approximation based on
6 the number of `reps` chosen in the `generate()` step. See `?get_p_value()` for more information.
7
8 # A tibble: 1 × 1
9   p_value
10  <dbl>
11 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

```
1 > null_dist |>  
2   visualize() +  
3   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.

```
1 > social |>
2   specify(turnout ~ messages, success = "Voted") |>
3   generate(reps = 1000, type = "bootstrap") |>
4   calculate(stat = "diff in props",
5             order = c("Neighbors", "Control")) |>
6   get_ci(level = 0.95)
7
8 # A tibble: 1 × 2
9   lower_ci upper_ci
10  <dbl>    <dbl>
11  1  0.0766  0.0865
```

- Any value outside of a  $100 \times (1 - \alpha)\%$  confidence interval would have a p-value less than  $\alpha$  (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.

# Extending to Block Estimation

- Recall from Section 2: we can improve precision by **blocking** on a pre-treatment covariate.
- Use `primary2004` (voted in 2004 primary) as the blocking variable:

```
1 > block_stats <- social |>
2   group_by(primary2004, messages) |>
3   summarize(n = n(), p_voted = mean(turnout == "Voted"), .groups = "drop") |>
4   pivot_wider(names_from = messages, values_from = c(n, p_voted)) |>
5   mutate(
6     ate      = p_voted_Neighbors - p_voted_Control,
7     n_block  = n_Neighbors + n_Control,
8     weight   = n_block / sum(n_block)
9   ); block_stats
10
11 # A tibble: 2 × 8
12   primary2004 n_Control n_Neighbors p_voted_Control p_voted_Neighbors ate
13   <dbl>      <int>      <int>          <dbl>          <dbl> <dbl>
14 1           0    114681      22666          0.237          0.306 0.0693
15 2           1     76562      15535          0.386          0.482 0.0965
16 # i 2 more variables: n_block <int>, weight <dbl>
```

- Block-specific ATEs differ: the treatment effect is larger among prior voters (0.097 vs. 0.069).

# Block ATE and Variance Comparison

- Weighted average block ATE:  $\hat{\tau}_{block} = \sum_j \frac{n_j}{N} \hat{\tau}_j$

```
1 > block_ate <- sum(block_stats$ate * block_stats$weight)
2
3 > tibble(
4   estimator = c("Unblocked", "Blocked (by primary2004)"),
5   ate_estimate = sprintf("%.7f", c(pull(est_ate), block_ate))
6 )
7
8 # A tibble: 2 × 2
9   estimator      ate_estimate
10 1 Unblocked      0.0813099
11 2 Blocked (by primary2004) 0.0802257
```

```
1 > blocked_se <- sqrt(sum(block_stats$weight^2 *
2   (block_stats$p_voted_Neighbors * (1 - block_stats$p_voted_Neighbors) / block_stats$n_Neighbors +
3   block_stats$p_voted_Control * (1 - block_stats$p_voted_Control) / block_stats$n_Control)))
4
5 > tibble(
6   estimator = c("Unblocked", "Blocked (by primary2004)"),
7   se = sprintf("%.7f", c(neyman_se, blocked_se)),
8   se_reduction = paste0(round((1 - c(neyman_se, blocked_se) / neyman_se) * 100, 2), "%")
9 )
10
11 # A tibble: 2 × 3
12   estimator      se      se_reduction
13 1 Unblocked      0.0026917 0%
14 2 Blocked (by primary2004) 0.0026482 1.62%
```

- Blocking on prior voting history yields a modest SE reduction ( $\approx 1.6\%$ ).
- $\rightsquigarrow$  Precision gain depends on how predictive the blocking variable is of the outcome.

# 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.

## 4/ In-class Exercises

# Exercise 1: Block ATE with a Different Covariate

- In Section 3, we estimated a block ATE using `primary2004` as the blocking variable.
- **Your turn:** repeat the same analysis using `sex` as the blocking variable.
  1. Compute the block-specific ATEs (i.e., one for each level of `sex`).
  2. Compute the weighted average block ATE.
  3. Compute the blocked Neyman SE and compare it to the unblocked SE.
- **Hint:** You can reuse the code from the `primary2004` example; just replace the blocking variable.

## Exercise 2: When Does Blocking Help?

- Compare your results from Exercise 1 (blocking by sex) to the results from Section 3 (blocking by primary2004).
- Discuss with your neighbor:
  1. Which blocking variable led to a larger SE reduction? Why?
  2. What property should a “good” blocking variable have in relation to the outcome?
  3. Can you think of another variable in this dataset that might be an even better blocking variable? Why?

**Have a great weekend! :)**

*Contact Information:*

[jaewon.yoo@iss.nthu.edu.tw](mailto:jaewon.yoo@iss.nthu.edu.tw)

<https://j1y004.github.io/>

