

# Lecture 5

## Neyman's repeated sampling approach for completely randomized experiments



# Outline

- Neyman's repeated sampling approach
  - Motivation
  - Variance calculation
  - CI and hypothesis testing
- Fisher VS Neyman
- Suggested reading: Imbens and Rubin Chapter 6, Peng's book Chapter 4

# Motivation

- Limitations of the Fisher's randomization inference
  - Do not allow heterogeneity of causal effects across individuals
  - Do not have inference for the population
- Completely randomized experiments: can we use two sample test?
- Neyman's approach
  - Allow heterogeneity of causal effects across individuals
  - Focus on estimation and inference for the average treatment effect: either just for the  $N$  samples or for the whole population (PATE)
  - Repeated sampling: randomization distribution of assignment vector  $\mathbf{W}$ , and sampling generated by drawing from the population units if inferring PATE

# Example: Duflo-Hanna-Ryan teacher-incentive experiment

- Conducted in rural India, designed to study the effect of financial incentives on teacher performance
- In total  $N = 107$  single-teacher schools, 53 schools are randomly chosen and are given a salary that's tied to their attendance
- One outcome: open (proportion of times the school is open during a random visit)

**Table 6.1. Summary Statistics for Duflo-Hanna-Ryan Teacher-Incentive Observed Data**

	Variable	Control ( $N_c = 54$ )		Treated ( $N_t = 53$ )		Min	Max
		Average	(S.D.)	Average	(S.D.)		
Pre-treatment	pctprewritten	0.19	(0.19)	0.16	(0.17)	0.00	0.67
Post-treatment	open	0.58	(0.19)	0.80	(0.13)	0.00	1.00
	pctpostwritten	0.47	(0.19)	0.52	(0.23)	0.05	0.92
	written	0.92	(0.45)	1.09	(0.42)	0.07	2.22
	written_all	0.46	(0.32)	0.60	(0.39)	0.04	1.43

# Example: Duflo-Hanna-Ryan teacher-incentive experiment

Standard two-sample test:

$$\hat{\tau}^{\text{dif}} = 0.80 - 0.58 = 0.22$$

$$s.e. = \sqrt{\frac{0.19^2}{54} + \frac{0.13^2}{53}} \approx 0.032$$

$$95\% CI: [0.22 - 1.96 * 0.032, 0.22 + 1.96 * 0.032]$$

**Table 6.1. Summary Statistics for Duflo-Hanna-Ryan Teacher-Incentive Observed Data**

	Variable	Control ( $N_c = 54$ )		Treated ( $N_t = 53$ )		Min	Max
		Average	(S.D.)	Average	(S.D.)		
Pre-treatment	pctprewritten	0.19	(0.19)	0.16	(0.17)	0.00	0.67
Post-treatment	open	0.58	(0.19)	0.80	(0.13)	0.00	1.00

- This calculation ignores the randomization procedure of the treatment assignment
- Can we justify this standard difference-in-means analysis from the randomization perspective?

# Estimation of the sample average treatment effect

- Causal estimand:  $\text{SATE} = \tau_{\text{fs}} = \frac{1}{N} \sum_{i=1}^N \{Y_i(1) - Y_i(0)\}$  for the sampled  $N$  units
- Difference-in-means estimator:

$$\hat{\tau}^{\text{dif}} = \bar{Y}_{\text{t}}^{\text{obs}} - \bar{Y}_{\text{c}}^{\text{obs}}$$

where  $\bar{Y}_{\text{c}}^{\text{obs}} = \frac{1}{N_{\text{c}}} \sum_{i:W_i=0} Y_i^{\text{obs}}$  and  $\bar{Y}_{\text{t}}^{\text{obs}} = \frac{1}{N_{\text{t}}} \sum_{i:W_i=1} Y_i^{\text{obs}}$ .

- Under complete randomization (random  $W$ ) and treat the potential outcomes as fixed (fixed  $\mathbf{Y}(0) = \{Y_i(0), i = 1, \dots, N\}$  and  $\mathbf{Y}(1) = \{Y_i(1), i = 1, \dots, N\}$ ), this estimator is unbiased

$$\begin{aligned}\mathbb{E}_W [\hat{\tau}^{\text{dif}} | \mathbf{Y}(0), \mathbf{Y}(1)] &= \frac{1}{N} \sum_{i=1}^N \left( \frac{\mathbb{E}_W[W_i] \cdot Y_i(1)}{N_{\text{t}}/N} - \frac{\mathbb{E}_W[1 - W_i] \cdot Y_i(0)}{N_{\text{c}}/N} \right) \\ &= \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0)) = \tau_{\text{fs}}.\end{aligned}$$

First, we can re-write  $\hat{\tau}^{dif}$ :

$$\begin{aligned}
 \hat{\tau}^{dif} &= \bar{Y}_t^{obs} - \bar{Y}_c^{obs} \\
 &= \frac{1}{N_t} \sum_{i:W_i=1} Y_i^{obs} - \frac{1}{N_c} \sum_{i:W_i=0} Y_i^{obs} \\
 &= \frac{1}{N_t} \sum_{i:W_i=1} W_i Y_i(1) - \frac{1}{N_c} \sum_{i:W_i=0} (1 - W_i) Y_i(0) \\
 &= \frac{1}{N} \sum_{i=1}^N \left( \frac{N}{N_t} W_i Y_i(1) - \frac{N}{N_c} (1 - W_i) Y_i(0) \right)
 \end{aligned}$$

And now we can take an expectation:

$$\begin{aligned}
 E(\hat{\tau}^{dif}) &= \frac{1}{N} \sum_{i=1}^N \left( \frac{N}{N_t} E(W_i) Y_i(1) - \frac{N}{N_c} E(1 - W_i) Y_i(0) \right) \\
 &= \frac{1}{N} \sum_{i=1}^N \left( \frac{N}{N_t} \frac{N_t}{N} Y_i(1) - \frac{N}{N_c} \frac{N_c}{N} Y_i(0) \right) \\
 &= \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0)) \\
 &= \tau
 \end{aligned}$$

# Calculate the variance of the estimator

- Causal estimand:  $\text{SATE} = \tau_{fs} = \frac{1}{N} \sum_{i=1}^N \{Y_i(1) - Y_i(0)\}$  for the sampled  $N$  units
- Difference-in-means estimator:  $\hat{\tau}^{\text{dif}} = \bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}}$
- Under complete randomization and fixed potential outcomes, we can also calculate the variance of  $\hat{\tau}^{\text{dif}}$  (if you are interested in the proof, see Appendix A of Chapter 6 in Rubin's book or Section 4.3 in Peng's book)

$$V_W[\hat{\tau}^{\text{dif}} | \mathbf{Y}(0), \mathbf{Y}(1)] = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{S_{ct}^2}{N}$$

where

$$S_c^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i(0) - \bar{Y}(0))^2, \quad \text{and} \quad S_t^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i(1) - \bar{Y}(1))^2$$

Sample variance  
of  $Y_i(0)$  and  $Y_i(1)$

$$\begin{aligned} S_{ct}^2 &= \frac{1}{N-1} \sum_{i=1}^N (Y_i(1) - Y_i(0) - (\bar{Y}(1) - \bar{Y}(0)))^2 \\ &= \frac{1}{N-1} \sum_{i=1}^N (Y_i(1) - Y_i(0) - \tau_{fs})^2. \end{aligned}$$

Sample variance  
of the unit-level  
treatment effects

# Some explanation of the variance

$$V_W[\hat{\tau}^{\text{dif}} | \mathbf{Y}(0), \mathbf{Y}(1)] = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{S_{ct}^2}{N}$$

- Where does the randomness come from? Treatment assignment
  - Potential outcomes are fixed (conditioned on)
- How is it different from the classical variance formula of  $\hat{\tau}^{\text{dif}}$ ?
  - Classical formula treats  $Y_i$  i.i.d. within group and the group indicators  $W_i$  fixed
- How is it different from the setting in Fisher's randomization test?
  - The formula allows for arbitrary treatment effect sizes and heterogeneity
  - This formula only works for completely randomized experiment
- How to estimate these quantities with observed variables?

# Conservative approximation of the variance of the estimator

- $\mathbb{V}_W[\hat{\tau}^{\text{dif}}|Y(0), Y(1)] = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{S_{ct}^2}{N}$

where

$$S_c^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i(0) - \bar{Y}(0))^2, \quad \text{and} \quad S_t^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i(1) - \bar{Y}(1))^2$$

Sample variance  
of  $Y_i(0)$  and  $Y_i(1)$

$$\begin{aligned} S_{ct}^2 &= \frac{1}{N-1} \sum_{i=1}^N (Y_i(1) - Y_i(0) - (\bar{Y}(1) - \bar{Y}(0)))^2 \\ &= \frac{1}{N-1} \sum_{i=1}^N (Y_i(1) - Y_i(0) - \tau_{fs})^2. \end{aligned}$$

Sample variance  
of the unit-level  
treatment effects

- Estimate  $S_c^2$  and  $S_t^2$  by sample variance of observed outcomes

$$S_c^2 = \frac{\sum_{i:W_i=0} (Y_i^{\text{obs}} - \bar{Y}_c^{\text{obs}})^2}{N_c - 1}, \quad S_t^2 = \frac{\sum_{i:W_i=1} (Y_i^{\text{obs}} - \bar{Y}_t^{\text{obs}})^2}{N_t - 1}$$

- $S_{ct}^2$  is not identifiable

- No heterogeneity of treatment effects across individuals  $S_{ct}^2 = 0$
- In general,  $S_{ct}^2 \geq 0$  though the exact value is unknown

# Conservative approximation of the variance of the estimator

- $\mathbb{V}_W[\hat{\tau}^{\text{dif}}|Y(0), Y(1)] = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{S_{ct}^2}{N}$

where

$$S_c^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i(0) - \bar{Y}(0))^2, \quad \text{and} \quad S_t^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i(1) - \bar{Y}(1))^2$$

Sample variance  
of  $Y_i(0)$  and  $Y_i(1)$

$$\begin{aligned} S_{ct}^2 &= \frac{1}{N-1} \sum_{i=1}^N (Y_i(1) - Y_i(0) - (\bar{Y}(1) - \bar{Y}(0)))^2 \\ &= \frac{1}{N-1} \sum_{i=1}^N (Y_i(1) - Y_i(0) - \tau_{fs})^2. \end{aligned}$$

Sample variance  
of the unit-level  
treatment effects

- A conservative estimator of  $\text{Var}_W[\hat{\tau}^{\text{dif}}|Y(0), Y(1)]$

$$\mathbb{V}_W[\hat{\tau}^{\text{dif}}|Y(0), Y(1)] \leq \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} = \mathbb{E}_W \left[ \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} \mid Y(0), Y(1) \right]$$

Neyman's estimator of the variance, same as  
s.e. on slide 5

# Estimation of the population average treatment effect

- Causal estimand:  $\text{PATE} = \tau_{\text{sp}} = \mathbb{E}(Y_i(1) - Y_i(0)) = \mathbb{E}(\text{SATE}) = \mathbb{E}(\tau_{\text{fs}})$
- We assume that  $(Y_i(0), Y_i(1))$  are jointly i.i.d samples from a super population with variance  $\sigma_c^2$  and  $\sigma_s^2$
- We still use difference-in-means estimator:

$$\hat{\tau}^{\text{dif}} = \bar{Y}_{\text{t}}^{\text{obs}} - \bar{Y}_{\text{c}}^{\text{obs}}$$

- $\hat{\tau}^{\text{dif}}$  is still unbiased for  $\tau_{\text{sp}}$ :  $\mathbb{E}(\hat{\tau}^{\text{dif}}) = \mathbb{E}(\mathbb{E}_W[\hat{\tau}^{\text{dif}}|Y(0), Y(1)]) = \mathbb{E}(\tau_{\text{fs}}) = \tau_{\text{sp}}$
- The variance of  $\hat{\tau}^{\text{dif}}$  (variance decomposition formula):
  - Check Wikipedia if you do not know the variance decomposition formula  
[https://en.wikipedia.org/wiki/Law\\_of\\_total\\_variance](https://en.wikipedia.org/wiki/Law_of_total_variance)

$$\mathbb{V}(\hat{\tau}^{\text{dif}}) = \mathbb{E}(\mathbb{V}_W[\hat{\tau}^{\text{dif}}|Y(0), Y(1)]) + \mathbb{V}(\mathbb{E}_W[\hat{\tau}^{\text{dif}}|Y(0), Y(1)])$$

# Variance calculation for the population

$$\mathbb{V}(\hat{\tau}^{\text{dif}}) = \mathbb{E}(\mathbb{V}_W[\hat{\tau}^{\text{dif}}|Y(0), Y(1)]) + \mathbb{V}(\mathbb{E}_W[\hat{\tau}^{\text{dif}}|Y(0), Y(1)])$$

- $\mathbb{V}_W[\hat{\tau}^{\text{dif}}|Y(0), Y(1)] = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{S_{ct}^2}{N}$ , with  
 $\mathbb{E}(S_c^2) = \sigma_c^2, \mathbb{E}(S_t^2) = \sigma_t^2, \mathbb{E}(S_{ct}^2) = \mathbb{V}(Y_i(1) - Y_i(0))$
- $\mathbb{V}(\mathbb{E}_W[\hat{\tau}^{\text{dif}}|Y(0), Y(1)]) = \mathbb{V}(\tau_{\text{fs}}) = \mathbb{V}\left(\frac{1}{N} \sum_{i=1}^N \{Y_i(1) - Y_i(0)\}\right) = \frac{1}{N} \mathbb{V}(Y_i(1) - Y_i(0))$
- So  $\mathbb{V}_W(\hat{\tau}^{\text{dif}}) = \frac{\sigma_c^2}{N_c} + \frac{\sigma_t^2}{N_t}$  **exactly the same as in two-sample testing**
  - In two-sample testing, we assume that observed outcome  $Y_i$  are i.i.d. in the treatment group and  $Y_i$  are i.i.d. in the control group
  - Under complete randomization,  $Y_i = Y_i(W_i)$  are not i.i.d. even with the treatment/control group because  $W_i$  are negatively correlated across  $i$

# Construct confidence intervals for $\tau_{fs}$ or $\tau_{sp}$

- We have the same estimator  $\hat{\tau}^{\text{dif}}$  and the same variance approximation of  $\hat{\tau}^{\text{dif}}$

$$\hat{V}(\hat{\tau}^{\text{dif}}) = \frac{s_c^2}{N_c} + \frac{s_t^2}{N_t}$$

no matter we are interested about SATE  $\tau_{fs}$  or PATE  $\tau_{sp}$

- When  $N$  is large enough, we can approximate the distribution of  $\hat{\tau}^{\text{dif}}$  by a normal distribution
- Then the 95% CI for either  $\tau_{fs}$  or  $\tau_{sp}$  is  
 $[\hat{\tau}^{\text{dif}} - 1.96 \times \sqrt{\hat{V}(\hat{\tau}^{\text{dif}})}, \hat{\tau}^{\text{dif}} + 1.96 \times \sqrt{\hat{V}(\hat{\tau}^{\text{dif}})}]$   
same as what we had earlier

# Hypothesis testing for $\tau_{fs}$ or $\tau_{sp}$

- We have the same estimator  $\hat{\tau}^{\text{dif}}$  and the same variance approximation of  $\hat{\tau}^{\text{dif}}$

$$\hat{V}(\hat{\tau}^{\text{dif}}) = \frac{s_c^2}{N_c} + \frac{s_t^2}{N_t}$$

no matter we are interested about SATE  $\tau_{fs}$  or PATE  $\tau_{sp}$

- When  $N$  is large enough, we can approximate the distribution of  $\hat{\tau}^{\text{dif}}$  by a normal distribution
- When can test for the null hypothesis  $H_0: \tau_{fs} = 0$  or  $H_0: \tau_{sp} = 0$

- The t-statistics:  $t = \frac{\hat{\tau}^{\text{dif}}}{\sqrt{\hat{V}(\hat{\tau}^{\text{dif}})}}$

- Under  $H_0$  and when  $N$  is large, we have  $t$  approximately follows a  $N(0, 1)$  distribution
- Two-sided p-value:  $2(1 - \phi(|t|))$

# Application to the Duflo-Hanna-Ryan data

**Table 6.1. Summary Statistics for Duflo-Hanna-Ryan Teacher-Incentive Observed Data**

	Variable	Control ( $N_c = 54$ )		Treated ( $N_t = 53$ )		Min	Max
		Average	(S.D.)	Average	(S.D.)		
Pre-treatment	pctprewritten	0.19	(0.19)	0.16	(0.17)	0.00	0.67
Post-treatment	open	0.58	(0.19)	0.80	(0.13)	0.00	1.00
	pctpostwritten	0.47	(0.19)	0.52	(0.23)	0.05	0.92
	written	0.92	(0.45)	1.09	(0.42)	0.07	2.22
	written_all	0.46	(0.32)	0.60	(0.39)	0.04	1.43

Confidence interval for each of the four outcomes:

$\hat{\tau}$	$\widehat{(\text{s.e.})}$	95% C.I.
0.22	(0.03)	(0.15,0.28)
0.05	(0.04)	(-0.03,0.13)
0.17	(0.08)	(0.00,0.34)
0.14	(0.07)	(0.00,0.28)

# Application to the Duflo-Hanna-Ryan data

## Analysis on two different subgroups:

- Check if the treatment effect is the same for the subset of schools with 0 proportion of students attending the exam before treatment, and for the subset of other schools
  - Conditional on the assignment results of other groups, within each subgroup we still have complete randomization of assignments
    - For more explanations, wait until the later lecture on post-stratification

# Fisher v.s. Neyman

- Like Fisher, Neyman proposed randomization-based inference
- Unlike Fisher,
  - estimands are average treatment effects
  - heterogenous treatment effects are allowed
  - population as well as sample inference is possible
  - asymptotic approximation is required for inference
- Fisher's approach can easily be applied to deal with any randomization mechanism in an experiment, but it can be much harder for Neyman's approach