

Lecture 5

Neyman's repeated sampling approach for completely randomized experiments

Outline

- ATE identification under unconfoundedness
- Neyman's repeated sampling approach
 - Motivation
 - Variance calculation
 - CI and hypothesis testing
 - SATE and PATE
- Suggested reading: Imbens and Rubin Chapter 6, Peng's book Chapter 4

Motivation: inference for average effects

Lecture 2: SATE and PATE

We defined average effects for the study units and for a population.

Lectures 3–4: Fisher randomization inference

Tested sharp nulls. Built confidence intervals under a common-effect model.

Today: estimation and uncertainty with heterogeneous effects

Connect causal effects to observed means, then derive estimation uncertainty.

When does random assignment justify the familiar two-sample analysis?

Would paying for attendance improve education?

Teacher-incentive experiment in rural India

An attendance-linked salary may bring teachers to school.
What else would need to happen for students to learn more?

Predict the effect on:

School opening

Students completing a writing test

Writing performance

Pair discussion: Which outcome should respond most directly? Why?

The teacher-incentive comparison

Unit: a single-teacher school in rural India.

Treatment: salary linked to the teacher's attendance.

Control: unchanged salary schedule.

The book analyzes 107 schools with complete data:

53 treated and 54 controls, from an original 113-school sample.

We follow the book's complete-randomization analysis.

Missing-data issues are set aside for this example.

What counts as improved education?

Outcome	School-level measurement
School opening	Proportion of unannounced visits with school open
Test participation	Proportion of students completing the writing test
Score among test takers	Mean writing score among students taking the test
Score with zeros	Mean score with zero assigned to nonparticipants

If the program changes who takes the test, how should we interpret the two score outcomes?

The observed difference in school opening

Group	Schools	Mean open	SD
Control	54	0.58	0.19
Treated	53	0.80	0.13

$$\hat{\tau} = 0.80 - 0.58 = 0.22$$

Why can this difference estimate a causal effect?
How much uncertainty comes from random assignment?

Unconfoundedness and observed means

Recall from Lecture 2

SATE: average effect for the N study units

PATE: average effect in the target population

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

$$\tau_{sp} = E[Y_i(1) - Y_i(0)]$$

Use the population language for a completely randomized experiment.

$$(Y_i(0), Y_i(1)) \perp\!\!\!\perp W_i, \quad 0 < \Pr(W_i = 1) < 1$$

Complete randomization gives unconfoundedness. Consistency gives:

$$Y_i^{obs} = Y_i(W_i)$$

$$E[Y_i^{obs} \mid W_i = w] = E[Y_i(w) \mid W_i = w] = E(Y_i(w))$$

The identity holds for both $w = 0$ and $w = 1$.

Which step uses consistency? Which uses unconfoundedness?

Identification and estimation of ATE

Population identification: an equality of population expectations.

$$\tau_{sp} = E(Y_i(1) - Y_i(0)) = E[Y_i^{obs} | W_i = 1] - E[Y_i^{obs} | W_i = 0]$$

Sample estimator: replace conditional expectations with group means.

$$\hat{\tau} = \overline{Y}_t^{obs} - \overline{Y}_c^{obs}$$

Neyman's next step: fix the study units and average over assignments.

Under complete randomization, is this estimator unbiased for SATE?

Random sampling connects the sample and population targets later.

Complete randomization and the SATE estimator

Fixed: the study units, potential outcomes, treatment counts, and SATE.

Random: assignment W , observed outcomes, group means, and their difference.

$$\Pr(W_i = 1) = \frac{N_t}{N}, \quad N_c = N - N_t, \quad \tau_{fs} = \frac{1}{N} \sum_{i=1}^N [Y_i(1) - Y_i(0)]$$

Observed group means and their difference:

$$\bar{Y}_t^{obs} = \frac{1}{N_t} \sum_{i: W_i=1} Y_i^{obs}, \quad \bar{Y}_c^{obs} = \frac{1}{N_c} \sum_{i: W_i=0} Y_i^{obs}$$

$$\hat{\tau} = \bar{Y}_t^{obs} - \bar{Y}_c^{obs}$$

For SATE, all expectations and variances over W condition on the potential outcomes.

Unbiasedness under random assignment

$$\hat{\tau} = \frac{1}{N_t} \sum_{i=1}^N [W_i Y_i(1)] - \frac{1}{N_c} \sum_{i=1}^N [(1 - W_i) Y_i(0)]$$

Consistency rewrites the estimator in potential outcomes.

$$E_W(\hat{\tau}) = \frac{1}{N_t} \sum_{i=1}^N [E_W(W_i) Y_i(1)] - \frac{1}{N_c} \sum_{i=1}^N [E_W(1 - W_i) Y_i(0)]$$

Take expectation over W .
Potential outcomes are fixed.

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

Equal assignment chances
recover the average over all
 N units.

Where did we use complete randomization?

Did we need the assignment indicators to be independent?

Finite-population variances and covariance

We have established unbiasedness. Now we need the estimator's variance.

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

Spread of control potential outcomes.
Governs the control mean's variance.

$$S_t^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i(1) - \bar{Y}(1))^2$$

Spread of treated potential outcomes.
Governs the treated mean's variance.

$$S_{10} = \frac{1}{N-1} \sum_{i=1}^N [\{Y_i(1) - \bar{Y}(1)\}\{Y_i(0) - \bar{Y}(0)\}]$$

Covariance of potential outcomes.
Governs the covariance of the two group means.

These fixed unobserved quantities describe the fixed potential outcomes of all N study units.

The bars average each potential outcome over all N study units.

The group means vary together

Next: obtain the variances of the group means and their covariance.

Let $p = N_t/N$. For distinct units i and j :

$$\text{Var}_W(W_i) = p(1 - p), \quad \text{Cov}_W(W_i, W_j) = -\frac{p(1 - p)}{N - 1}$$

Fixed treatment counts make assignment indicators dependent.

$$\text{Var}_W(\bar{Y}_t^{obs}) = \left(\frac{1}{N_t} - \frac{1}{N}\right)S_t^2, \quad \text{Var}_W(\bar{Y}_c^{obs}) = \left(\frac{1}{N_c} - \frac{1}{N}\right)S_c^2$$

Each arm samples without replacement from its potential outcomes.

$$\text{Cov}_W(\bar{Y}_t^{obs}, \bar{Y}_c^{obs}) = -\frac{S_{10}}{N}$$

The same partition links the two group means.

For details of a proof, see Appendix A of Chapter 6 in Rubin's book or Section 4.3 in Peng's book

Deriving Neyman's variance formula

The variance of a difference must account for dependence between its two means.

$$\text{Var}_W(\hat{\tau}) = \text{Var}_W(\bar{Y}_t^{obs}) + \text{Var}_W(\bar{Y}_c^{obs}) - 2\text{Cov}_W(\bar{Y}_t^{obs}, \bar{Y}_c^{obs})$$

$$= \frac{S_t^2}{N_t} + \frac{S_c^2}{N_c} - \frac{S_t^2 + S_c^2 - 2S_{10}}{N}$$

Combine the two mean variances and their covariance.

The subtracted numerator is the variance of individual effects:

$$S_{ct}^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i(1) - Y_i(0) - \tau_{fs})^2 = S_t^2 + S_c^2 - 2S_{10}$$

This correction is the variance of individual treatment effects.

$$\text{Var}_W(\hat{\tau}) = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{S_{ct}^2}{N}$$

Exact under complete randomization. Effects may differ across units.

Pair work: what remains unobserved?

No sharp null and no common-effect assumption.

Unit	W_i	$Y_i(0)$	$Y_i(1)$	Observed Y_i
1	1	?	3	3
2	1	?	5	5
3	0	2	?	2
4	0	4	?	4

Can two different potential-outcome tables fit these observations?
Must they have the same variance of individual effects?

Which variance components can we estimate?

Two components we can estimate

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

Unbiased for the variance of control potential outcomes across all N units.

$$s_t^2 = \frac{\sum_{i: W_i=1} (Y_i^{obs} - \bar{Y}_t^{obs})^2}{N_t - 1}, \quad E_W(s_t^2) = S_t^2$$

Unbiased for the variance of treated potential outcomes across all N units.

The remaining obstacle: a fixed, unknown component

$$S_{ct}^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i(1) - Y_i(0) - \tau_{fs})^2 = S_t^2 + S_c^2 - 2S_{10}$$

We never observe both potential outcomes for a unit, so their covariance cannot generally be estimated.

Can we estimate an upper bound of $\text{Var}_W(\hat{\tau})$ instead?

Neyman's conservative variance estimator

An upper bound lets us estimate uncertainty without knowing individual effects.

$$\text{Var}_W(\hat{\tau}) = \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} - \frac{S_{ct}^2}{N} \leq \frac{S_c^2}{N_c} + \frac{S_t^2}{N_t} = V_+$$

Effect variance is nonnegative. Omitting its subtraction gives an upper bound.

$$\hat{V} = \frac{s_c^2}{N_c} + \frac{s_t^2}{N_t}, \quad \text{SE} = \sqrt{\hat{V}}$$

Replace the two population variances by their unbiased sample estimates.

$$E_W(\hat{V}) - \text{Var}_W(\hat{\tau}) = \frac{S_{ct}^2}{N} \geq 0$$

The upward bias is exactly the omitted term. It is zero under constant effects.

Conservative means upward bias over repeated assignments.
A realized variance estimate can still fall below the true variance.

Normal approximation and confidence intervals

We have an estimate and a standard error. How do we obtain an interval?

$$\frac{\hat{\tau} - \tau_{fs}}{\sqrt{\text{Var}_W(\hat{\tau})}} \Rightarrow N(0,1)$$

Under a finite-population CLT, standardized estimation error is approximately normal.

$$C_{1-\alpha} = [\hat{\tau} - z_{1-\alpha/2}\sqrt{\hat{V}}, \quad \hat{\tau} + z_{1-\alpha/2}\sqrt{\hat{V}}]$$

A computable interval: estimate $\pm$ normal cutoff $\times$ estimated standard error.

$$\liminf_{N \rightarrow \infty} Pr_W(\tau_{fs} \in C_{1-\alpha}) \geq 1 - \alpha$$

Over repeated assignments, asymptotic coverage of SATE is at least $1 - \alpha$.

For 95% confidence, use $z_{0.975} = 1.96$.

Coverage is asymptotic and can exceed 95% when effects vary.

Testing an average-effect null

Is the observed estimate consistent with a proposed average effect?

$$H_0: \tau_{fs} = \tau_0, \quad T = \frac{\hat{\tau} - \tau_0}{\sqrt{\hat{V}}}$$

Measure how many estimated standard errors separate the estimate from the null.

$$p = 2\{1 - \Phi(|T|)\}, \quad \text{reject if } |T| \geq z_{1-\alpha/2}$$

Use a two-sided normal reference to judge how unusual the discrepancy is.

The average-effect null allows positive and negative individual effects.

Sharp null: $Y_i(1) - Y_i(0) = \tau_0$ for every i

A sharp null specifies the effect for every unit.

Which null can impute the missing potential outcomes?

School opening: effect size and uncertainty

What is the effect on school opening, and how precise is the estimate?

$$\hat{t} = 0.80 - 0.58 = 0.22$$

Mean effect: 22 percentage points.

$$SE = \sqrt{\frac{0.19^2}{54} + \frac{0.13^2}{53}} \approx 0.031$$

The SE measures uncertainty from randomized assignment.

$$0.22 \pm 1.96 \times 0.031 \approx [0.16, 0.28]$$

A 95% interval from the normal approximation.

About 22 percentage points more school opening.

Rounded summaries give slightly different endpoints from the book's table.

Does improved attendance extend to student outcomes?

Outcome	Estimate	SE	95% CI
School opening	0.22	0.03	[0.16, 0.28]
Test participation	0.05	0.04	[-0.03, 0.13]
Score among test takers	0.17	0.08	[0.00, 0.34]
Score with zeros	0.14	0.07	[0.00, 0.28]

Which estimates are precise enough to support a clear conclusion?

Does an interval containing zero establish no effect?

Book results, rounded; separate 95% intervals for each outcome.

Pair work: do subgroup effects differ?

- Check if the treatment effect is the same for the subset of schools with 0 proportion of students attending the exam before treatment (N = 40), and for the subset of other schools (N = 67)

Score-with-zeros effect	No baseline test takers N = 40	Some baseline test takers N = 67
Estimate	0.04	0.22
SE	0.07	0.09
95% CI	[-0.10, 0.19]	[0.04, 0.40]

One interval includes zero and the other does not.

Does that establish a difference between subgroup effects?

The difference between subgroup effects

Compare effects directly: no baseline test takers minus some.

Outcome	Difference	SE	95% CI
School opening	0.02	0.06	[-0.10, 0.14]
Test participation	-0.15	0.08	[-0.31, 0.00]
Score among test takers	0.03	0.15	[-0.26, 0.31]
Score with zeros	-0.18	0.12	[-0.41, 0.05]

Different significance decisions do not establish different effects.

Inference for a difference between subgroup effects

Estimate the difference in effects directly to compare the subgroups.

Condition on treated counts in the two pre-treatment subgroups A and B.

$$\delta_{fs} = \tau_A - \tau_B, \quad \hat{\delta} = \hat{\tau}_A - \hat{\tau}_B$$

Target and estimate: the effect in subgroup A minus the effect in subgroup B.

$$\hat{V}_g = \frac{s_{c,g}^2}{N_{c,g}} + \frac{s_{t,g}^2}{N_{t,g}}, \quad g \in \{A, B\}$$

Estimate each subgroup's uncertainty using its own treatment and control counts.

$$SE(\hat{\delta}) = \sqrt{\hat{V}_A + \hat{V}_B}, \quad \text{CI: } \hat{\delta} \pm 1.96 \sqrt{\hat{V}_A + \hat{V}_B}$$

Given subgroup counts, assignments are independent across groups, so variances add.

Why do we condition on subgroup counts, and use baseline subgroups?

PATE: the sampling model and unbiasedness

How can an estimate for the study units support a population effect?

Fixed: population P and PATE. Random: sampled units and assignment.

$(Y_i(0), Y_i(1))$ i. i. d. $\sim P$, $\tau_{sp} = E\{Y_i(1) - Y_i(0)\}$ Potential-outcome pairs vary across samples.

Let $\mathcal{F}$ denote one sampled set of potential outcomes. Randomize W within it.

$E(\hat{\tau}) = E\{E_W(\hat{\tau}|\mathcal{F})\} = E(\tau_{fs}) = \tau_{sp}$ Average over assignments first, then over which units were sampled.

The mean difference is unbiased for PATE under this sampling model.

Random assignment alone does not justify a target-population claim.

Total variance: assignment and sampling

Population inference adds sampling variation to randomization variation.

$$\text{Var}(\hat{\tau}) = E\{\text{Var}_W(\hat{\tau}|\mathcal{F})\} + \text{Var}\{E_W(\hat{\tau}|\mathcal{F})\}$$

Variance decomposition formula
Separate the two sources of randomness.

The first term averages the assignment variance across samples.

$$E(S_c^2) = \sigma_c^2, \quad E(S_t^2) = \sigma_t^2, \quad E(S_{ct}^2) = \sigma_\tau^2$$

Average the fixed-sample variance components across samples.

The second term is variation in sample average effects.

$$\text{Var}(\tau_{fs}) = \text{Var}\left\{\frac{1}{N} \sum_{i=1}^N [Y_i(1) - Y_i(0)]\right\} = \frac{\sigma_\tau^2}{N}$$

Samples have different SATEs when individual effects vary.

$\sigma_c^2 = \text{Var}\{Y(0)\}$, $\sigma_t^2 = \text{Var}\{Y(1)\}$, and $\sigma_\tau^2 = \text{Var}\{Y(1) - Y(0)\}$.

Why the heterogeneity terms cancel

Adding sampling variation changes the variance target.

$$\text{Var}(\hat{t}) = \frac{\sigma_c^2}{N_c} + \frac{\sigma_t^2}{N_t} - \frac{\sigma_\tau^2}{N} + \frac{\sigma_\tau^2}{N}$$

The assignment term subtracts effect variation. Sampling adds it back.

$$= \frac{\sigma_c^2}{N_c} + \frac{\sigma_t^2}{N_t}$$

The total variance is the sum of the two marginal contributions.

$$\text{E}(\hat{V}) = \frac{\sigma_c^2}{N_c} + \frac{\sigma_t^2}{N_t} = \text{Var}(\hat{t})$$

The usual variance estimator is unbiased for this total variance.

The familiar variance formula is justified for PATE under i.i.d. sampling.

The same standard error, different justification

$$\hat{V} = \frac{s_c^2}{N_c} + \frac{s_t^2}{N_t}, \quad \text{CI} = \hat{t} \pm 1.96 \times \sqrt{\hat{V}}$$

Same numerical procedure. Its justification depends on the target and sampling model.

Target	Variance estimator	Large-sample interval
SATE	Conservative in expectation under heterogeneous effects	Generally conservative
PATE under i.i.d. sampling	Unbiased for unconditional variance	Approximately nominal coverage

The arithmetic matches an unequal-variance two-sample analysis.
The target and sampling assumptions determine its coverage.

Fisher and Neyman in this course

	Fisher approach taught earlier	Neyman approach today
Main question	Is a sharp null compatible with the data?	What is the average effect and its uncertainty?
Missing outcomes	Imputed under the sharp null	Remain unobserved
Effect heterogeneity	Heterogeneous alternatives allowed. Earlier intervals used a common effect.	Allowed for average-effect inference
Justification	Finite-sample randomization test under the sharp null	Exact mean and variance. Large-sample intervals and tests.

Both analyses must follow the experiment's assignment mechanism.

What supports the conclusion?

1. In repeated randomization, what changes and what stays fixed?
2. Does a positive average effect mean every school benefits?
3. Why is the estimated variance conservative for SATE?
4. What additional assumption supports PATE inference?