Gov 2002: 2. Randomized Experiments

Matthew Blackwell

September 10, 2015

Where are we (going)?

- Last time: defined potential outcomes and causal estimands/quantities of interest.
- This time: how we identify these quantities in randomized experiments.
- Later: what if randomization only happens conditional on covariates?
- Or, what if we weren't able to randomize?

What is the selection problem?

- First pass at the data: prima facie or naive difference in means:

$$
\begin{aligned}
& E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \\
=& E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0] \qquad \text{(consistency)} \\
=& E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] + E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0] \\
=& \underbrace{E[Y_i(1) - Y_i(0)|D_i = 1]}_{\text{ATT}} + \underbrace{E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]}_{\text{selection bias}}
\end{aligned}
$$

- Naive $=$ ATT $+$ selection bias.
- Selection bias: how different the treated and control groups are in terms of their potential outcome under control.

Selection bias = unidentified ATT

$$\underbrace{E[Y_i(1) - Y_i(0)|D_i = 1]}_{\text{ATT}} + \underbrace{E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]}_{\text{selection bias}}$$

- ATT is unidentified here unless selection bias equals 0.
  - Both ATT and the SB are unobserved.
  - No amount of data will help us distinguish between them.
- Example: effect of negativity on vote shares.
  - Naive estimate: negative candidates do worse than positive candidates.
  - Could mean that the ATT is negative OR the ATT is positive and there is large negative selection bias.
  - SB = candidates that go negative are worse than those who stay positive, even if they ran the same campaigns.
- With an unbounded Y_i, we can't even bound the ATT because, in principle, SB could be anywhere from $-\infty$ to ∞.

Notation

- We'll need some notation for the entire vector of treatment, outcomes, etc:
  - $\mathbf{D} = (D_1, D_2, \ldots, D_N)$
  - $\mathbf{X}, \mathbf{Y}(1)$, and $\mathbf{Y}(0)$ are similarly defined for X_i, $Y_i(1)$, and $Y_i(0)$.

Experiments

- An experiment is a study where assignment to treatment is controlled by the researcher.
  - $p_i = \mathbb{P}[D_i = 1]$ be the probability of treatment assignment probability.
  - p_i is controlled and known by researcher in an experiment.
- A randomized experiment is an experiment with the following properties:

1. Positivity: assignment is probabilistic: $0 < p_i < 1$
   - No deterministic assignment.
2. Unconfoundedness: $\mathbb{P}[D_i = 1|\mathbf{Y}(1), \mathbf{Y}(0)] = \mathbb{P}[D_i = 1]$
   - Treatment assignment does not depend on any potential outcomes.
   - Sometimes written as $D_i \perp\!\!\!\perp (\mathbf{Y}(1), \mathbf{Y}(0))$

Natural experiment

- Natural experiment: experiment where treatment is randomized, but that randomization was not under the control of the researcher.
- Randomization has to be justified in these cases since it wasn't directly implemented.
- Hyde paper on syllabus:
  - election observers were assigned to polling stations "using a method that approximates randomization"

Randomization

- What does randomization (positivity + unconfoundedness) buy us?
  - treatment group is a random sample from the population.
  - control group is a random sample the population.
- $\rightsquigarrow$ sample control mean is unbiased for population control mean:

$$\mathbb{E}[Y_i|D_i = 0] = E[Y_i(0)|D_i = 0] = E[Y_i(0)] = E[Y_i(0)|D_i = 1]$$

- Not the same as the observed outcomes being independent of treatment ($Y_i \perp\!\!\!\perp D_i$)
- Randomization eliminates selection bias:

$$E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0] = E[Y_i(0)] - E[Y_i(0)] = 0$$

Identification by randomization

- Goal: show that we can identify a causal effect under a randomization assumption.
- Use the selection bias result with the naive difference in means:

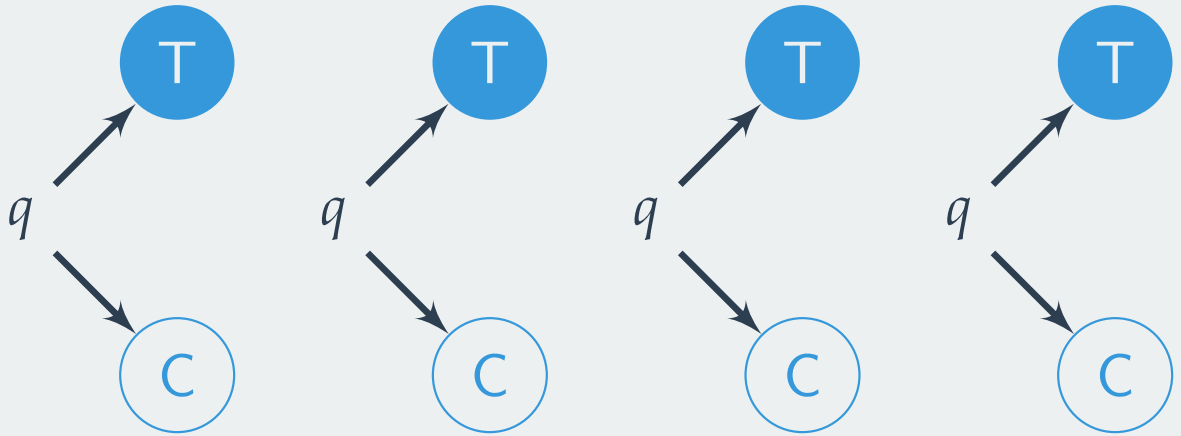
$$
\begin{aligned}
& E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \\
= & \underbrace{E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1]}_{ATT} + \underbrace{0}_{\text{selection bias}} \\
= & E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] \\
= & E[Y_i(1)] - E[Y_i(0)] \qquad \text{(unconfoundedness)}
\end{aligned}
$$

- $E[Y_i(1) - Y_i(0)] = \tau$ is just the ATE.
- Thus, if we can estimate the conditional expectations, $\mathbb{E}[Y_i|D_i = 1]$ and $\mathbb{E}[Y_i|D_i = 0]$, we can estimate the ATE.
- Result: ATE is identified in a randomized experiment.

Types of randomizations/experiments

- Let $N_t = \sum_{i=1}^{N} D_i$ and $N_c = N - N_t$.
- Bernoulli trials:
  - flip coins for each person in the population with probability q
  - $\mathbb{P}[\mathbf{D}] = q^{N_t}(1-q)^{N_c}$
  - Downside: could end up with all treated or all control
- Completely randomized experiment:
  - Randomly sample N_t units from the population to be treated
  - Equal probability of any assignment with $\sum_{i=1}^{N} D_i = N_t$
  - Each possible assignment has probability $\binom{N}{N_t}^{-1}$
  - Each unit has probability $p_i = N_t/N$ of being selected into treatment, but treatment assignment is not independent between units.

Bernoulli assignment



Completely randomized design

- Start with $N = 6$ and say we want to have $N_t = 3$
- Randomly pick 3 from $\{1, 2, 3, 4, 5, 6\}$: $2, 4, 5$
- Not independent: knowing 2 is treated means 3 is less likely to be treated.
- Fixed number of treatment spots induces dependence: $\mathbb{E}[D_i \cdot D_j] \neq \mathbb{E}[D_i]\mathbb{E}[D_j]$

$$\mathbb{E}[D_i \cdot D_j] = \mathbb{P}[D_i = 1] \cdot \mathbb{P}[D_i = 1 | D_i = 1] = \frac{N_t}{N} \frac{N_t - 1}{N - 1}$$

Stratified designs

- Stratified randomized experiment:
  - form J blocks, b_j, $j = 1, \ldots, J$ based on the covariates
  - completely randomized assignment within each block.
  - Randomization depends on the block variable, B_i
  - Conditional unconfoundedness: $D_i \perp\!\!\!\perp (Y_i(1), Y_i(0)) | B_i$.
- Pair randomized experiments:
  - Stratified randomized experiment where each block has 2 units.
  - 1 unit in each pair receives treatment.
  - Extreme version of the stratified/blocked randomized experiment.
  - Also called "matched pair" design
- Both of these seek to remove "bad randomizations" where covariates are related to treatment assignment by chance.

Identification under stratification

- Generally, stratified designs mean that the probability of treatment depends on a covariate, X_i:

$$p_i(x) = \mathbb{P}[D_i = 1 | X_i = x]$$

- Conditional randomization assumptions:
  1. Positivity: $0 < p_i(x) < 1$ for all i and x.
  2. Unconfoundedness: $\mathbb{P}[D_i = 1 | \mathbf{X}, \mathbf{Y}(1), \mathbf{Y}(0)] = \mathbb{P}[D_i = 1 | X_i]$
     - Also written as $D_i \perp\!\!\!\perp (\mathbf{Y}(1), \mathbf{Y}(0)) | X_i$

Stratification and the ATE

- Can we identify the ATE under these stratified designs? Yes!

$$\mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}_X\Big\{\mathbb{E}[Y_i(1) - Y_i(0)|X_i]\Big\} \qquad \text{(iterated expectations)}$$

$$= \mathbb{E}_X\Big\{\mathbb{E}[Y_i(1)|X_i] - \mathbb{E}[Y_i(0)|X_i]\Big\}$$

$$= \mathbb{E}_X\Big\{\mathbb{E}[Y_i(1)|D_i = 1, X_i] - \mathbb{E}[Y_i(0)|D_i = 0, X_i]\Big\} \qquad \text{(unconfoundedness)}$$

$$= \mathbb{E}_X\Big\{\mathbb{E}[Y_i|D_i = 1, X_i] - \mathbb{E}[Y_i|D_i = 0, X_i]\Big\} \qquad \text{(consistency)}$$

- ATE is just the average of the within-strata differences in means.
- Identified because the last line is a function of observables.
- The averaging is over the distribution of the strata $\rightsquigarrow$ size of the blocks.

Stratification example

- Stratified by incumbency, where $X_i = 1$ is a Democratic incumbent and $X_i = 0$ is a Democratic challenger.
- Then we have:

$$\begin{aligned}
&\mathbb{E}_X\Big\{\mathbb{E}[Y_i|D_i = 1, X_i] - \mathbb{E}[Y_i|D_i = 0, X_i]\Big\} \\
&= \underbrace{\Big(\mathbb{E}[Y_i|D_i = 1, X_i = 1] - \mathbb{E}[Y_i|D_i = 0, X_i = 1]\Big)}_{\text{diff-in-means for incumbents}} \underbrace{\mathbb{P}[X_i = 1]}_{\text{share of incumbents}} \\
&\quad + \underbrace{\Big(\mathbb{E}[Y_i|D_i = 1, X_i = 0] - \mathbb{E}[Y_i|D_i = 0, X_i = 0]\Big)}_{\text{diff-in-means for challengers}} \underbrace{\mathbb{P}[X_i = 0]}_{\text{share of challengers}}
\end{aligned}$$

- We call this "averaging over X_i"

Effect modification

- Averaging over X_i might hide some interesting variation in the effect size:
  - Effect of negativity might varies by incumbency status?
  - Effect of clientelistic messages varies by gender of recipient?
  - Effect of having daughters varies by gender?
- This means the conditional ATE (CATE) is non-constant:

$$\tau(x) \equiv E[Y_i(1) - Y_i(0)|X_i = x] \neq E[Y_i(1) - Y_i(0)|X_i = x^*] \equiv \tau(x^*)$$

- The difference between $\tau(x)$ and $\tau(x^*)$ might be causal or not.
- Under randomization or stratified randomization, CATE is identified from within-strata difference-in-means (see last slide):

$$\tau(x) = \mathbb{E}[Y_i|D_i = 1, X_i = x] - \mathbb{E}[Y_i|D_i = 0, X_i = x]$$

Estimation and Inference

- Up until now, we've talked about identification.
- Now that we know that the ATE is identified, how will we estimate it?
- Remember: identification first, then estimation.

Samples versus Populations

- Remember the differences between the population, U, of size N, with the PATE:

$$PATE = \tau = E[Y_i(1) - Y_i(0)]$$

- And the sample, S, from the population of size n with the SATE:

$$SATE = \tau_S = \frac{1}{n} \sum_{i \in S} [Y_i(1) - Y_i(0)]$$

- Today, we will focus on the Neyman approach to estimation and inference:
  - derive estimators for these quantites and,
  - derive the properties of these estimators under repeated sampling.
- Next week, we'll discuss an alternative approach proposed by Fisher.

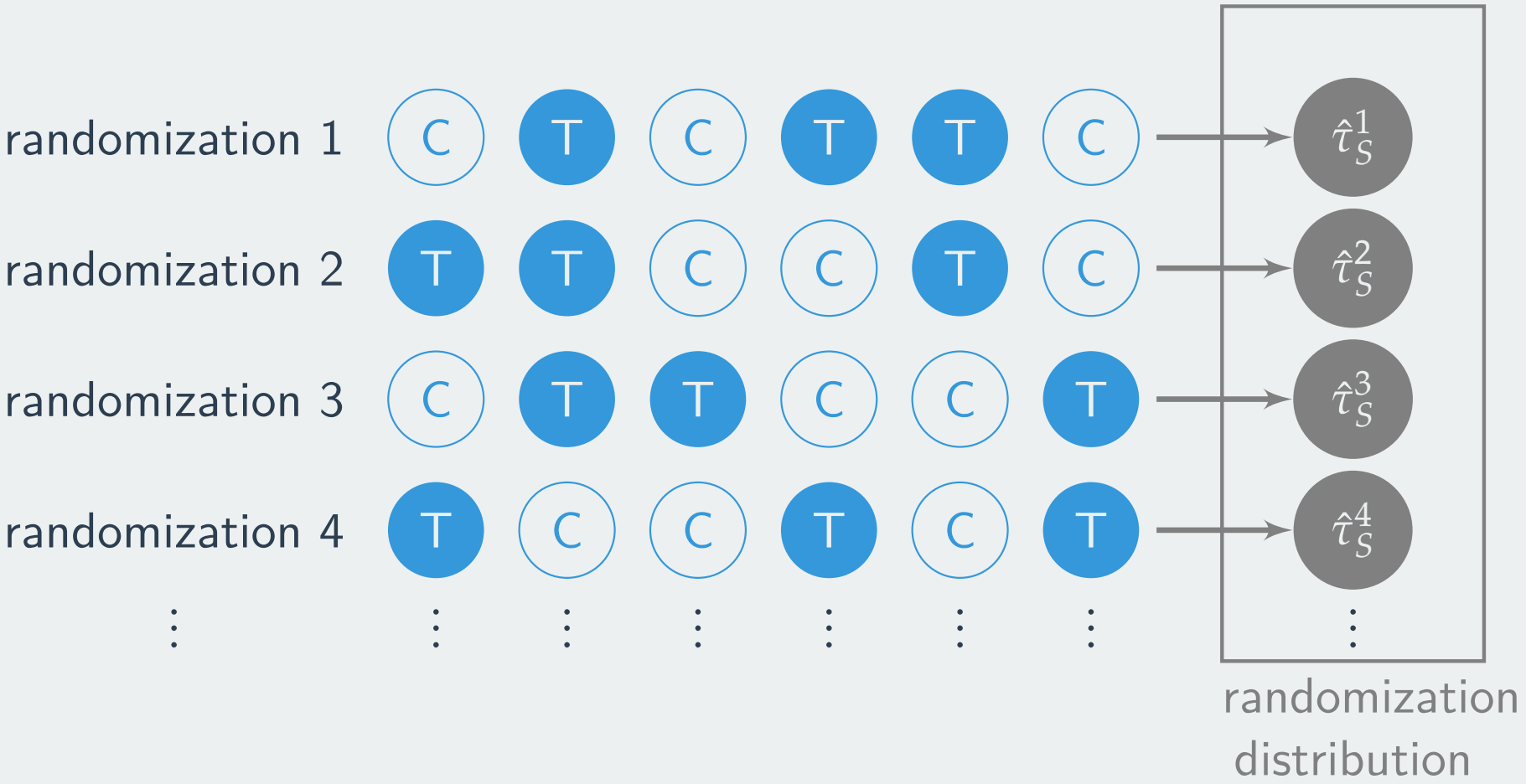
Finite sample results

- Finite sample results take the observed sample as the target of interest.
- Let n_t be the number of treated units in the sample.
- Once we assign some groups to treatment and some to control we do not actually observe $Y_i(1)$ and $Y_i(0)$ and so we cannot actually observe SATE. We can, however, estimate it:

$$\hat{\tau}_S = \underbrace{\frac{1}{n_t}\sum_{i=1}^{n} D_i Y_i}_{\text{mean among treated}} - \underbrace{\frac{1}{n_c}\sum_{i=1}^{n}(1-D_i)Y_i}_{\text{mean among control}}$$

- Conditional on the sample, the only variation in $\hat{\tau}_S$ is from the treatment assignment.
- Unconditionally, there are two sources of variation: the treatment assignment and the sampling procedure.

Repeated samples/randomizations



- Randomization distribution is a special version of the sampling distribution of this estimator.

Finite-sample properties

- What are the properties of $\hat{\tau}_S$ in repeated samples/randomizations? What does the distribution look like?
- Unbiasedness: is the mean of the randomization distribution equal to the true SATE?
- Sampling variance: what is the variance of the randomization distribution?
- Knowing these will allow to construct confidence intervals, conduct tests, etc.

Unbiasedness

- In a completely randomized experiment, $\hat{\tau}_S$ is unbiased for τ_S
- Let $\mathbf{O} = \{\mathbf{Y}(1), \mathbf{Y}(0)\}$ be the the potential outcomes.

$$\mathbb{E}[\hat{\tau}_S|S, \mathbf{O}] = \frac{1}{n_t}\sum_{i=1}^{n}\mathbb{E}[D_i Y_i|S, \mathbf{O}] - \frac{1}{n_c}\sum_{i=1}^{n}\mathbb{E}[(1-D_i)Y_i|S, \mathbf{O}]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{n}{n_t}\cdot\mathbb{E}[D_i Y_i|S, \mathbf{O}] - \frac{n}{n_c}\cdot\mathbb{E}[(1-D_i)Y_i|S, \mathbf{O}]\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{n}{n_t}\cdot\mathbb{E}[D_i Y_i(1)|S, \mathbf{O}] - \frac{n}{n_c}\cdot\mathbb{E}[(1-D_i)Y_i(0)|S, \mathbf{O}]\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{n}{n_t}\cdot\mathbb{E}[D_i|S, \mathbf{O}]\cdot Y_i(1) - \frac{n}{n_c}\cdot\mathbb{E}[(1-D_i)|S, \mathbf{O}]\cdot Y_i(0)\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{n}{n_t}\cdot\frac{n_t}{n}\cdot Y_i(1) - \frac{n}{n_c}\cdot\frac{n_c}{n}\cdot Y_i(0)\right)$$

$$= \frac{1}{n}\sum_{i\in S} Y_i(1) - Y_i(0) = \tau_S$$

Finite-sample sampling variance

- It turns out that the sampling variance of the difference in means estimator is:

$$\mathbb{V}(\hat{\tau}_S|S) = \frac{S_c^2}{n_c} + \frac{S_t^2}{n_t} - \frac{S_{\tau_i}^2}{n},$$

- S_c^2 and S_t^2 are the in-sample variances of $Y_i(0)$ and $Y_i(1)$, respectively.

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

- Here, $\bar{Y}(d) = (1/n)\sum_{i=1}^{n} Y_i(0)$.
- Last term is the in-sample variation of the individual treatment effects:

$$S_{\tau_i}^2 = \frac{1}{n-1}\sum\left(Y_i(1) - Y_i(0) - \tau_S\right)^2$$

Finite-sample sampling variance

$$\mathbb{V}(\hat{\tau}_S|S) = \frac{S_c^2}{n_c} + \frac{S_t^2}{n_t} - \frac{S_{\tau_i}^2}{n},$$

- If the treatment effects are constant across units, then $S_{\tau_i}^2 = 0$.
- $\rightsquigarrow$ in-sample 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	0

	$i=1$	$i=2$	Avg.
$Y_i(0)$	-10	10	0
$Y_i(1)$	10	-10	0
τ_i	20	-20	0

- Both have $\tau = 0$, first has constant effects.
- In first setup, $\hat{\tau}_S = 20$ or $\hat{\tau}_S = -20$ depending on the randomization.
- In second setup, $\hat{\tau}_S = 0$ in either randomization.

Estimating the sampling variance

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

$$s_c^2 = \frac{1}{n_c - 1} \sum_{i:D_i=0} (Y_i(0) - \bar{Y}_c)^2 \qquad s_t^2 = \frac{1}{n_t - 1} \sum_{i:D_i=1} (Y_i - \bar{Y}_t)^2$$

- Here, $\bar{Y}_c = (1/n_c) \sum_{i=1}^{n} (1 - D_i) Y_i$ and $\bar{Y}_t = (1/n_t) \sum_{i=1}^{n} D_i Y_i$.
- But what about $S_{\tau_i}^2$?

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

- What to do?

Conservative variance estimation

$$\mathbb{V}(\hat{\tau}_S|S) = \frac{S_c^2}{n_c} + \frac{S_t^2}{n_t} - \frac{S_{\tau_i}^2}{n},$$

- We will estimate this quantity with the so-called Neyman (or robust) estimator:

$$\widehat{\mathbb{V}} = \frac{s_c^2}{n_c} + \frac{s_t^2}{n_t},$$

- Notice that $\widehat{\mathbb{V}}$ is biased for $\mathbb{V}$, but that bias is always positive.
- Construct CIs and conduct hypothesis tests as usual.
- 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 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.

Population estimands

- Now imagine we want to estimate the PATE, τ.
- Implied DGP: simple random sample (SRS) from the population, then randomized experiment within sample.
  - $\rightsquigarrow$ the sample mean is unbiased for the population mean, $E_S[\tau_S] = \tau$
  - $E_S[\cdot]$ is the expectation over repeated samples from the population.
- How does our difference-in-means estimator do?

$$\mathbb{E}_S[\hat{\tau}_S] = \underbrace{\mathbb{E}_S\left\{E[\hat{\tau}_S|S]\right\}}_{\text{iterated expectations}} = \underbrace{\mathbb{E}_S[\tau_S]}_{\text{SATE unbiasedness}} = \tau$$

- $\hat{\tau}_S$ unbiased for the PATE!

Population sampling variance

- What about the sampling variance of $\hat{\tau}_S$ when estimating the PATE?
- It turns out that the sampling variance of the estimator is simply:

$$\mathbb{V}(\hat{\tau}_S) = \frac{\sigma_c^2}{N_c} + \frac{\sigma_t^2}{N_t},$$

- Here, σ_c^2 and σ_t^2 are the population-level variances of $Y_i(1)$ and $Y_i(0)$.
- The third term drops out $\rightsquigarrow$ higher variance for PATE than SATE.

Estimating pop. sampling variance

$$\mathbb{V}(\hat{\tau}_S) = \frac{\sigma_c^2}{N_c} + \frac{\sigma_t^2}{N_t},$$

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

$$\widehat{\mathbb{V}} = \frac{s_c^2}{n_c} + \frac{s_t^2}{n_t}$$

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

Analyzing experiments with regression?

- Can we just use regression to estimate the ATE in this case?
  - `lm(y ~ d)`?
- Call the coefficient on D_i the regression estimator: $\hat{\tau}_{\text{ols}}$.
- We can justify this using the consistency relationship:

$$
\begin{aligned}
Y_i &= D_i Y_i(1) + (1 - D_i) Y_i(0) \\
&= D_i Y_i(1) + (1 - D_i) Y_i(0) + \mathbb{E}[Y_i(0)] - \mathbb{E}[Y_i(0)] \\
&+ D_i \mathbb{E}[Y_i(1) - Y_i(0)] - D_i \mathbb{E}[Y_i(1) - Y_i(0)] \\
&= \mathbb{E}[Y_i(0)] + D_i \mathbb{E}[Y_i(1) - Y_i(0)] + (Y_i(0) - \mathbb{E}[Y_i(0)]) \\
&+ D_i \cdot ((Y_i(1) - Y_i(0)) - \mathbb{E}[Y_i(1) - Y_i(0)]) \\
&= \alpha + D_i \tau + \epsilon_i
\end{aligned}
$$

- See that $\alpha = E[Y_i(0)]$ and remember that $\tau = \mathbb{E}[Y_i(1) - Y_i(0)]$. And also the residual here is the deviation for the control group plus the treatment effect hetergeneity.

Independent errors

$$\varepsilon_i = (Y_i(1) - Y_i(0)) - \mathbb{E}[Y_i(1) - Y_i(0)]) + D_i \cdot ((Y_i(1) - Y_i(0)) - \mathbb{E}[Y_i(1) - Y_i(0)])$$

- Let's check to see if the errors here are independent of the treatment, which would imply that a regression estimator $\hat{\tau}_{\text{ols}}$ would be unbiased for τ:

$$\begin{aligned}\mathbb{E}[\epsilon_i | D_i = 0] &= \mathbb{E}[Y_i(0) - \mathbb{E}[Y_i(0)] | D_i = 0] \\ &= \mathbb{E}[Y_i(0) | D_i = 0] - \mathbb{E}[Y_i(0)] = 0\end{aligned}$$

- and for $D_i = 1$:

$$\begin{aligned}\mathbb{E}[\epsilon_i | D_i = 1] &= \mathbb{E}[Y_i(1) - \mathbb{E}[Y_i(0)] + \mathbb{E}[Y_i(1) - Y_i(0)] | D_i = 1] \\ &= \mathbb{E}[Y_i(1) | D_i = 1] - \mathbb{E}[Y_i(1)] = 0\end{aligned}$$

- Thus, just using the randomization assumption, we have justified the use of regression.
- No functional form assumptions at all, only consistency.

Including covariates

- Completely randomized design $\rightsquigarrow$ no need to control for covariates.
- Adding covariates won't matter for unbiasedness/consistency.
  - (Not true for stratified designs!)
- Still consistent even if functional form for X_i is misspecified.
- Effects of conditioning on covariates: reduce uncertainty in effect estimates

Next week

- More experiments, this time under Fisherian inference.
- Randomization inference: even fewer assumptions.
- Back to the lady tasting tea!
- Then: regression, matching, etc!