Causal Inference

Randomized Experiments

Prof. Jennifer Hill
Fall 2018

Quick statistical reminders

- E[Y] : the *expectation* of Y, or the population average of Y (this may have been called **μ** when you were taught it)
- If we have a random sample from the population, a simple way to estimate this population mean, E[Y], is with the sample mean, $\bar{y}$. Think of estimating the average height of everyone in the US by taking a random sample of people in the US and calculating the mean of their heights.
- E[Y | X] : conditional expectation of Y *given* X, or the population average of Y within a given subpopulation defined by X.
- This is also sometimes an estimand of interest. (Now think of estimating the average height of all women in the US (where X=1 for women) by taking a random sample of women and calculating the sample mean of their heights.)

Notation/definitions

The following definitions will be used

- Treatment, *Z*: A manipulable variable we want to know the effect of (pollution, a vaccine, welfare work requirements, tax increase, child maltreatment)
- Outcome, *Y*: The variable we hypothesize might be affected by the treatment
- Covariates or confounders, *X*: Background variables that we think may be associated *both* with the treatment that people participate in *as well as* the outcome of interest

Counterfactuals

- The most important concept in this class is that of the **counterfactual**
- Most causal inference statisticians define causal effects as comparisons between what *would happen* in two or more different states (one of which will be *factual*, the other(s) *counterfactual*)
- Examples:
  - Blood pressure after taking statins versus blood pressure *after not* taking statins
  - Annual salary if attended college versus annual salary if ended schooling after high school

- We'll talk more in the future about what types of phenomena we can plausibly view as "treatments" or "causal variables"

Potential outcome notation

- We operationalize counterfactuals by using potential outcome notation (Rubin, 1978)
- For a given treatment and control condition, each person can be thought of as having two potential outcomes, Y(0) and Y(1)
  - Y(Z=0) =Y(0) is the outcome if treatment is NOT received
  - Y(Z=1) =Y(1) is the outcome if treatment is received

Unit Causal Effects

- We use these potential outcomes (Y(0), Y(1)) to define a causal effect for subject *i* as a comparison between them, such as, $\tau_i = Y_i(1) - Y_i(0)$
- In theory a causal effect can be defined as any function of the potential outcomes such as

  $$g(Y_i(0), Y_i(1)) = 1 \quad \text{if} \quad Y_i(1) > Y_i(0)$$

  but life gets much more complicated once we start to consider non-linear combinations of potential outcomes
- Note: a unit does not have to be an individual subject. A unit can be a classroom or a family for example.
- **Fundamental Problem of Causal Inference:** we can never observe both $Y_i(0)$ and $Y_i(1)$
- One can think of this as a missing data problem
- So how do we proceed?

The Estimand
(What are we trying to estimate?)

(Average Treatment Effects)

The **estimand**: What we are trying to estimate?

- The **estimand** is *the quantity we are trying to estimate*
- Given the inherent problem in estimating individual level causal effects, in causal inference we often focus on estimating *average* causal effects
- The average treatment effect (ATE) can be rather generically defined as

$$E[Y(1)-Y(0)]$$

Population Average Treatment Effects

- The most popular estimand may be the population treatment effect

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

- However it is not always reasonable to expect that we can estimate PATE, particularly if we don't start with a random sample from the population

Sample Average Treatment Effects

- Sometimes it is more realistic to content ourselves with estimating the *sample average treatment effect* (SATE)

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

- *If we have a random sample from the population*, the implications of the choice of sample versus population estimand have to do with variance estimation (because we have different levels of uncertainty about each of these).
- *If we* do not *have a random sample from the population*, the implications of the choice of sample versus population estimand are deeper (unless the effect is the same for everyone)
- We won't focus on these distinctions in this course.

Estimands

- SATE and PATE are inherently unknown to the *researcher*
- For class assignments you will sometimes be asked to **calculate** SATE or PATE (or other estimands) when you were in *all-seeing* or *statistical god-vision* mode. This is because
  - We need both potential outcomes for everyone in our sample to calculate SATE
  - We need both potential outcomes for everyone in the population to calculate PATE
- The best the *researcher* can do is to use observed Y to **estimate** SATE or PATE

Average Treatment Effects

PATE

$$E[Y(1)] = \sum_{i=1}^{N} Y(1) = 12$$

$$E[Y(0)] = \sum_{i=1}^{N} Y(0) = 8$$

SATE

$$\frac{1}{n}\sum_{i=1}^{n} Y(1) = 11.83$$

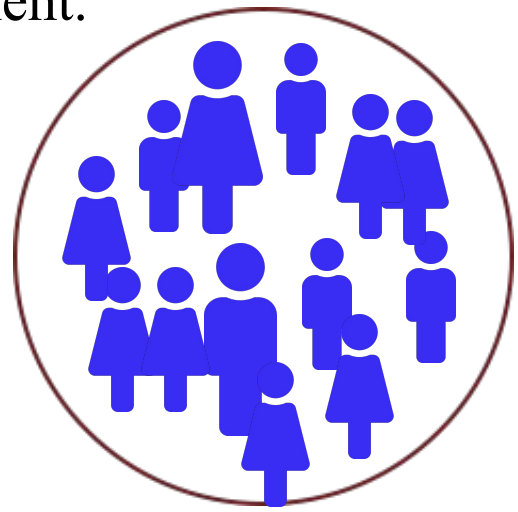
$$\frac{1}{n}\sum_{i=1}^{n} Y(0) = 7.83$$

In order to unbiasedly estimate PATE we need to be able to unbiasedly estimate the two left-hand quantities. In order to unbiasedly estimate SATE we need to be able to unbiasedly estimate the two right-hand quantities.

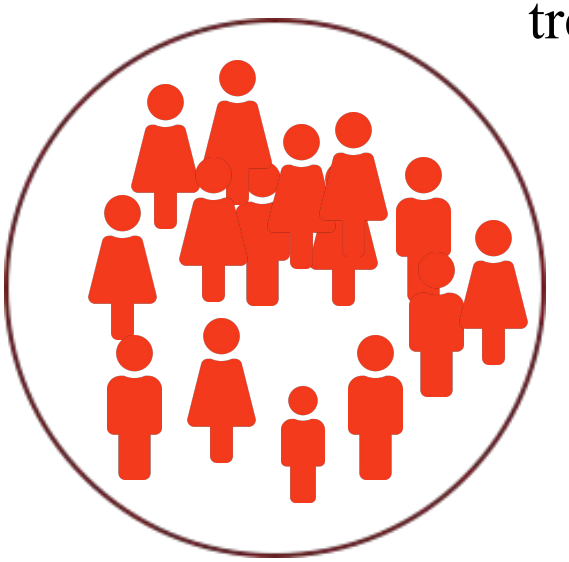
The ingredients of PATE: E[Y(1) - Y(0)]

Imagine two parallel universes, each with N units.

One in which everyone in the population **did not** receive the treatment.



One in which everyone in the population **did** receive the treatment.



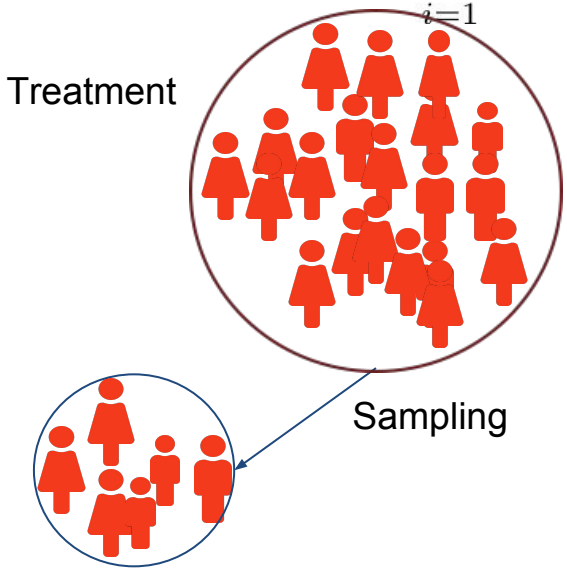
$$E[Y(0)] = \sum_{i=1}^{N} Y(0) = 8$$

$$E[Y(1)] = \sum_{i=1}^{N} Y(1) = 12$$

The ingredients of SATE

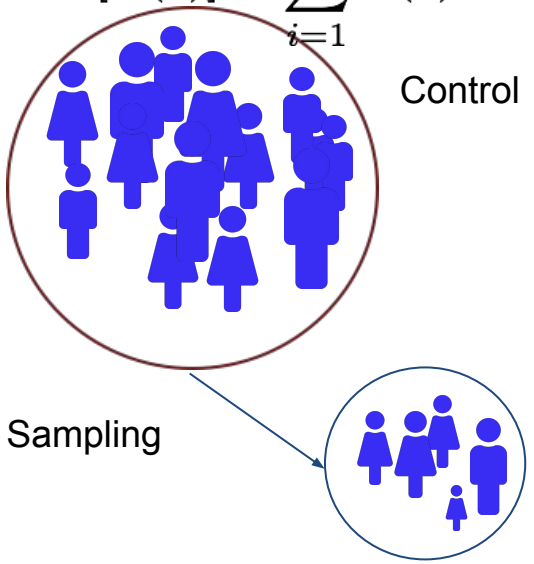
If our treatment group was just a random sample of n units from the full population, then the mean Y(1) would be close to the mean for the population. Same for the control group and mean Y(0).

$$E[Y(1)] = \sum_{i=1}^{N} Y(1) = 12$$



$$\frac{1}{n}\sum_{i=1}^{n} Y(1) = 11.83$$

$$E[Y(0)] = \sum_{i=1}^{N} Y(0) = 8$$



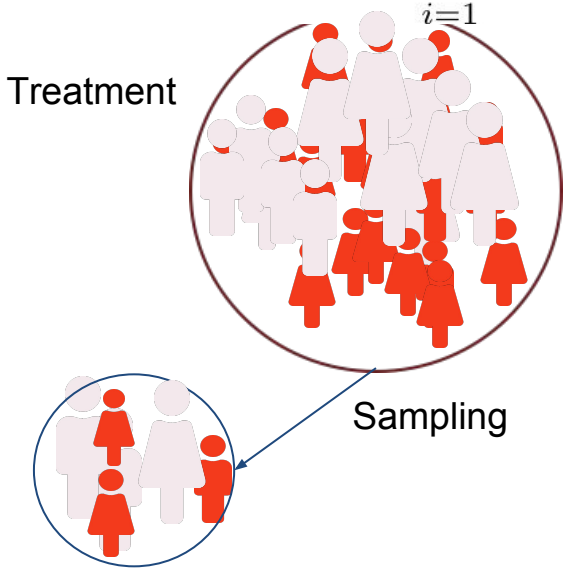
$$\frac{1}{n}\sum_{i=1}^{n} Y(0) = 7.83$$

Under what conditions would you expect a group of treated individuals to have a mean response that resembles either E[Y(1)] or the mean of Y(1) for the sample?

Remember what the researcher sees….

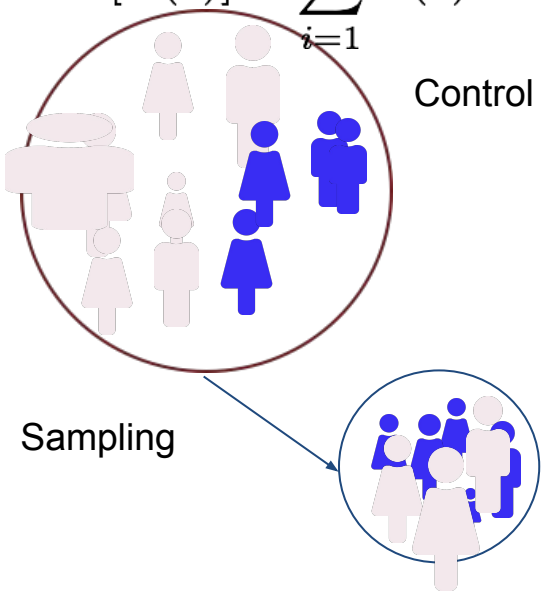
If our treatment group was just a random sample of n units from the full population, then the mean Y(1) would be close to the mean for the population. Same for the control group and mean Y(0).

$$E[Y(1)] = \sum_{i=1}^{N} Y(1) = 12$$



$$\frac{1}{n}\sum_{i=1}^{n} Y(1) = 11.83$$

$$E[Y(0)] = \sum_{i=1}^{N} Y(0) = 8$$



$$\frac{1}{n}\sum_{i=1}^{n} Y(0) = 7.83$$

How can we get unbiased estimates of these quantities????

Randomized experiments

Randomized experiments: the gold standard

- Randomized experiments are the so-called gold standard for answering causal questions because they create two (or more) groups that are virtually identical to each other on average
- So, if each group receives a different treatment and the groups have different outcomes, we can safely attribute these differences in outcomes to the different treatments (rather than any other characteristics of the experimental subjects)

Randomized experiments

- In a randomized experiment, units are assigned to treatments using a known probabilistic rule (such as "assign to treatment group with probability .4").
- Each unit has nonzero probability of being allocated to each treatment.
- Randomized experiments can use quite complicated assignment rules, but we shall focus on three very simple types of randomized experiments
  - the completely randomized experiment,
  - the randomized block experiment, and
  - the matched pairs experiment.

A completely randomized experiment

- In a completely randomized experiment with n study units and 2 treatment groups (i.e. treatment and control)
- n_0 units are assigned to the control group and n_1 units are assigned to the treatment group
- *each unit has the same probability of receiving the treatment as any other unit*
- therefore *all possible treatment allocations are equally likely*
- A common estimate of the treatment effect in this setting would be

$$Y_1 - Y_0$$

(the difference in sample means across treatment groups)

Completely randomized experiments
(more general properties)

- Recall that in a completely randomized experiment, treatments are allocated by a probabilistic mechanism (like a coin toss) without regard for the values Y(0) and Y(1)
- Consequently (in a "large" population),

$$\Pr(Z \mid Y(0), Y(1)) = \Pr(Z)$$

[note that this is stronger than just mean independence]

- Another way of writing this independence property (that we will use over and over) is

$$Z \perp Y(0), Y(1)$$

this means the assignment mechanism is ***ignorable***

- If it is also true that $0 < \Pr(Z) < 1$ we say that the assignment mechanism is ***strongly ignorable***
- One *consequence* of this is that

$$E[Y(0) \mid Z] = E[Y(0)] \quad \text{and} \quad E[Y(1) \mid Z] = E[Y(1)]$$

Estimating causal effects in randomized experiments

- What about the average treatment effect, E[Y(1)-Y(0)] = E[Y(1)] - E[Y(0)]?
- Well in a randomized experiment, if Z stands for treatment assignment.
  - E[Y(1)] = E[Y(1)|Z=1]
  - E[Y(0)] = E[Y(0)|Z=0]
- So we can unbiasedly estimates
  - E[Y(1)] with $\overline{Y}_1$
  - E[Y(0)] with $\overline{Y}_0$

Running example:
What is the effect of job training on wages?

Let's consider an example:

- Suppose we have observed a group of individuals.
- Some individuals were assigned to (and received) a job training program (Z=1),
- others received no intervention (Z==0).
- We have also recorded the
  - age and
  - education level (high school graduate or not)

  of these individuals.
- A year after the program has ended the wage rate of each participant (Y) is recorded.

Observed data

Assume a random sample from the population

$Avg(age)_{Z=1}=23.8$

3

$Avg(age)_{Z=0}=25.1$

7

Person	Treat	Educ.	Age	Y(0)	Y(1)	Y
1	1	1	26	?	14	14
2	1	1	21	?	12	12
3	1	1	30	?	16	16
4	1	1	19	?	12	12
5	1	0	25	?	10	10
6	1	0	22	?	8	8
7	0	1	26	10	?	10
8	0	1	21	8	?	8
9	0	1	42	16	?	16
10	0	1	15	2	?	2
11	0	0	26	6	?	6
12	0	0	21	4	?	4

What information are we missing if we want to **calculate** SATE?

What information are we missing if we want to **calculate** PATE?

Observed data

Assume a random sample from the population

$Avg(age)_{Z=1}=23.83$

$Avg(age)_{Z=0}=25.17$

Person	Treat	Educ.	Age	Y(0)	Y(1)	Y
1	1	1	26	?	14	14
2	1	1	21	?	12	12
3	1	1	30	?	16	16
4	1	1	19	?	12	12
5	1	0	25	?	10	10
6	1	0	22	?	8	8
7	0	1	26	10	?	10
8	0	1	21	8	?	8
9	0	1	42	16	?	16
10	0	1	15	2	?	2
11	0	0	26	6	?	6
12	0	0	21	4	?	4

$$\bar{Y}_1 - \bar{Y}_0 = 12 - 7.67 = 4.33$$

Now assuming we are all-seeing (we see all counterfactuals…

τ=4

$Avg(age)_{Z=1}$=23.8

3

$Avg(age)_{Z=0}$=25.1

7

Person	Treat	Educ.	Age	Y(0)	Y(1)	Y
1	1	1	26	10	14	14
2	1	1	21	8	12	12
3	1	1	30	12	16	16
4	1	1	19	8	12	12
5	1	0	25	6	10	10
6	1	0	22	4	8	8
7	0	1	26	10	14	10
8	0	1	21	8	12	8
9	0	1	42	16	20	16
10	0	1	15	2	6	2
11	0	0	26	6	10	6
12	0	0	21	4	8	4

$$\frac{1}{n}\sum_{i=1}^{n} Y(1) - \frac{1}{n}\sum_{i=1}^{n} Y(0) = 11.83 - 7.83$$

Causal inference as a Missing data problem

Because counterfactuals aren't observed for both situations (treatment and control) for anyone, only 50 % of the data is available for causal inference!

One way to think about causal inference is that we need to "impute" or "estimate" everyone's counterfactual potential outcomes:

Y(1) for the controls

Y(0) for the treated

Estimating causal effects in randomized experiments

- What about the average treatment effect, E[Y(1)-Y(0)] = E[Y(1)] - E[Y(0)]?
- Well in a randomized experiment, if Z stands for treatment assignment.
  - E[Y(1)] = E[Y(1)|Z=1]
  - E[Y(0)] = E[Y(0)|Z=0]
- So we can unbiasedly estimates
  - E[Y(1)] with $\overline{Y}_1$
  - E[Y(0)] with $\overline{Y}_0$

Hypothetical example (τ=4)

Person	T	Educ	Age	Y0	Y1	Y
1	1	1	26	10	14	14
2	1	1	21	8	12	12
3	1	1	30	12	16	16
4	1	1	19	8	12	12
5	1	0	25	6	10	10
6	1	0	22	4	8	8
7	0	1	26	10	14	10
8	0	1	21	8	12	8
9	0	1	42	16	20	16
10	0	1	15	2	6	2
11	0	0	26	6	10	6
12	0	0	21	4	8	4

$\text{Avg(age)}_{Z=1}$=23.83

$\text{Avg(age)}_{Z=0}$=25.17

Mean of true potential outcomes for population (requires GV goggles)

$$E[Y(1)] = \frac{1}{N}\sum_{i=1}^{N} Y(1) = 12$$

$$E[Y(0)] = \frac{1}{N}\sum_{i=1}^{N} Y(0) = 8$$

Mean of true potential outcomes for sample (requires GV goggles)

$$\frac{1}{n}\sum_{i=1}^{n} Y_i(1) = 11.83$$

$$\frac{1}{n}\sum_{i=1}^{n} Y_i(0) = 7.83$$

Average outcomes in sample (RESEARCHER VIEW)

$$\bar{Y}_1 = \frac{1}{n_1}\sum_{i=1}^{n} Y_i(1)Z_i = 12$$

$$\bar{Y}_0 = \frac{1}{n_0}\sum_{i=1}^{n} Y_i(0)(1-Z_i) = 7.67$$

Estimating effects using data from randomized experiments

Estimation of treatment effects in the context of randomized experiments

Which of these is a valid estimation strategy for the causal effect of Z on Y?

- Difference in means across randomized groups ✓
- Regression of the outcome on an indicator for randomized treatment assignment ✓
- Regression of the outcome on an indicator for treatment assignment and pre-treatment variables ✓
- Regression of the outcome on an indicator for treatment assignment and a variable representing the level of take-up of the treatment (e.g. number of days of a program that someone chose to attend) ✗

Controlling for post-treatment variables (DON'T!)

- The general rule is DON'T CONTROL FOR POST-TREATMENT VARIABLES!
- However, sometimes important would-be confounders are measured post-treatment and it is unclear whether or not they were affected by the treatment (or whether that association is strong).
- In this case it might be better to control for them than not (there is a trade-off in bias here). This issue is explored nicely in a paper by Rosenbaum (1984, JRSSA)
- We will have a mini-lecture on this later in the semester.

Proceed with caution!

How do we conceptualize the properties of our estimator?

Bias and efficiency

We use randomization or sampling distributions to think about the properties of our estimator.

If the randomization distribution for an estimator is centered on SATE then that estimator is unbiased for SATE. The smaller the variance of that distribution the more efficient the estimator.

If the sampling distribution for an estimator is centered on PATE then that estimator is unbiased for PATE. The smaller the variance of that distribution the more efficient the estimator.

Bias

What exactly does it mean for an estimate to be biased?
A statistic is biased if it is different from the true population parameter being estimated:

$$E[\hat{\theta}] - \theta \neq 0$$

Efficiency

What is efficiency?
Being able to accurately estimate the treatment effect (smallest possible standard errors) with the least amount of observations possible

Estimation of uncertainty

- In traditional statistics courses there is an emphasis on estimating population estimands using (random) samples from that population. Our uncertainty about how close our *estimate* is to the *estimand* (quantity of interest) arises from the fact that we haven't collected data on the entire population.
- One difference in causal inference is that we have uncertainty even if we have collected data on the full sample because we are in essence still missing half of our data (the counterfactual outcomes)

Estimation of uncertainty

Quantifying this uncertainty is what variance estimation is all about. We will not talk much on variance estimation in this course (I will simply tell you what variance estimate to use).

Variance estimates when the estimand is SATE are different from variance estimates when the estimand is PATE.

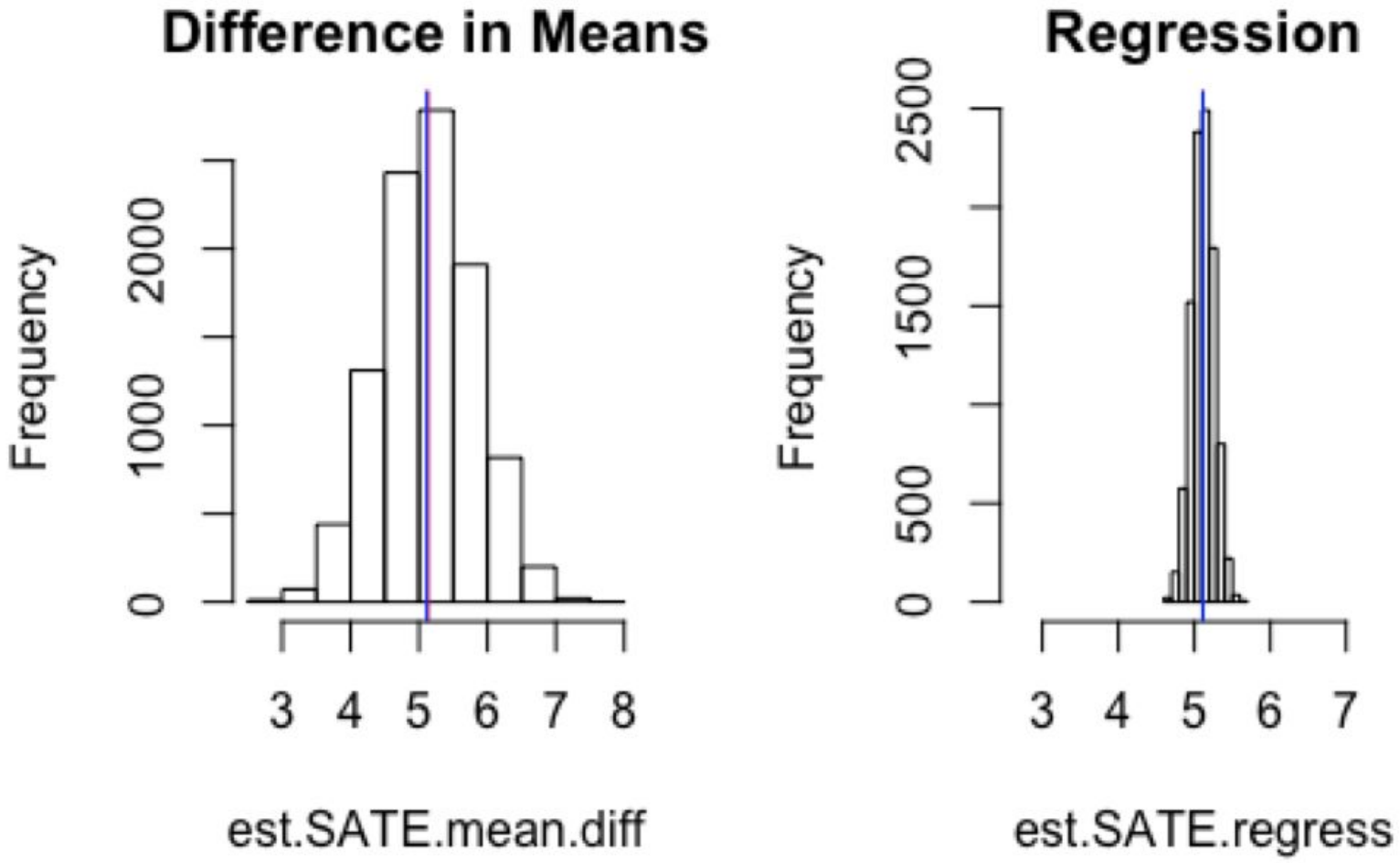
- Missing data for counterfactual outcomes creates uncertainty when estimating either estimand
- For PATE, there is an **additional** source of uncertainty coming from the fact that we only get to see a sample from the population

Sampling Distribution

A sampling distribution additionally reflects our sampling variability. It is a *theoretical probability distribution* of the possible values of some sample statistic (for instance a treatment effect estimate) that would occur if we were to draw all possible samples of a fixed size from a given population and then for each generate that statistic.

Sampling distributions form the basis of classical (frequentist) inference because they are often well-approximated by normal and t-distributions (because of the Central Limit Theorem)

Sampling Distribution



Randomization Distribution

A randomization distribution reflects the uncertainty in treatment effect estimates that manifests due to randomness of the treatment assignment.

To imagine how the god of Statistics would generate such a distribution. She would keep the covariates and potential outcomes fixed. But then she would create new random assignments to treatment. Each time she does this observed y will also change! Then she would record the estimate of the treatment effect using these new data. These estimates can be plotted to visualize the randomization distribution.

Inferences from the randomization distribution are specific to the analysis sample and won't apply to the population.

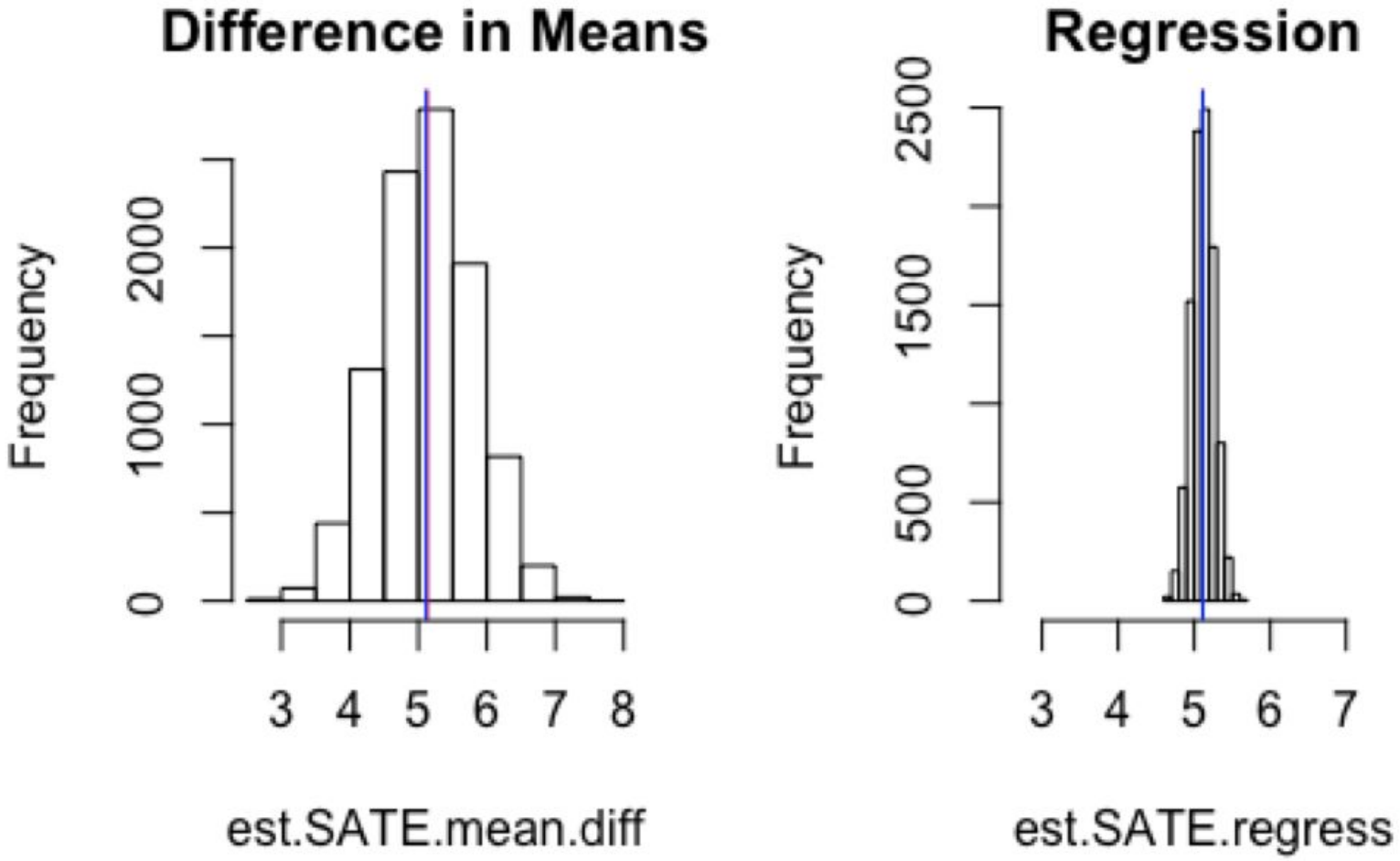
Randomization Distribution: researcher view

How could a researcher mimic this distribution?

Similar to a sampling distribution (above) any given randomization distribution will be specific to a given hypothesis about the treatment effect. This hypothesis will allow the researcher to fill in missing counterfactual outcomes. At that point the same strategy described in the last slide can be used via simulation.

The randomization distribution allows us to determine whether, relative to a given hypothesis about the treatment effect and given the variability among all possible estimates , the one we observed is a common outcome or a rare outcome.

Randomization Distribution



!!Pop-Quiz!!

In estimating causal estimands (ATE),

Which form of inference is best for thinking about properties of the estimator relative to SATE?

Which form of inference is best for thinking about properties of the estimator relative to PATE?

Additional Estimands and Types of Randomized Experiments

Effects within subpopulations

ATE for subpopulations

- The ATE will sometimes vary across subpopulations and in such instances, investigators will often want to look not only at the ATE, but at the ATE within various of these subpopulations.
  - In our job training example, what sub-populations would be interesting to observe causal effects?

- Suppose that X is a variable (or vector of variables) with values x that denote different subpopulations (for instance X=1 for high school educated and X=0 for did not complete high school).

  We can now model our potential outcomes as:

$$E[Y(1) \mid X = x] \quad \text{and} \quad E[Y(0) \mid X = x]$$

*Where X=x specifies the potential outcomes for desired sub-group

ATE for subpopulations

- Then, conditional ATE within subpopulation x (ATE(x)) is defined:

$$E[Y(1) - Y(0) \mid X = x] = E[Y(1) \mid X = x] - E[Y(0) \mid X = x]$$

- This statement means that we are going to estimate the average treatment effect for the population that satisfies X=x
- So, for instance, with X denoting high school education (as above)

$$E[Y(1) - Y(0) \mid X = 0] = E[Y(1) \mid X = 0] - E[Y(0) \mid X = 0],$$

is the average treatment effect for those who did not complete high school

A note on indicator notation

- I(a) is an indicator function that equals 1 whenever the logical expression a is true and 0 otherwise
- The following equivalence holds

$$\sum_{i=1}^{N}(Y_i(1) - Y_i(0))I(X_i = x) = \sum_{i:X_i=x}(Y_i(1) - Y_i(0))$$

Where i: $X_i = x$ means for all observations i such that $X_i = x$

- The next expression

$$\sum_{i=1}^{N} I(X_i = x)$$

is equal to the number of observations for which $X_i = x$

Estimating ATE for subpopulations

- Could separately estimating ATEs within each subpopulation (by difference in means or regression)
- Could run one regression on all the data but include interactions between the indicators for the subpopulations and the treatment indicator

Randomized Block Designs

Randomized Block Experiment

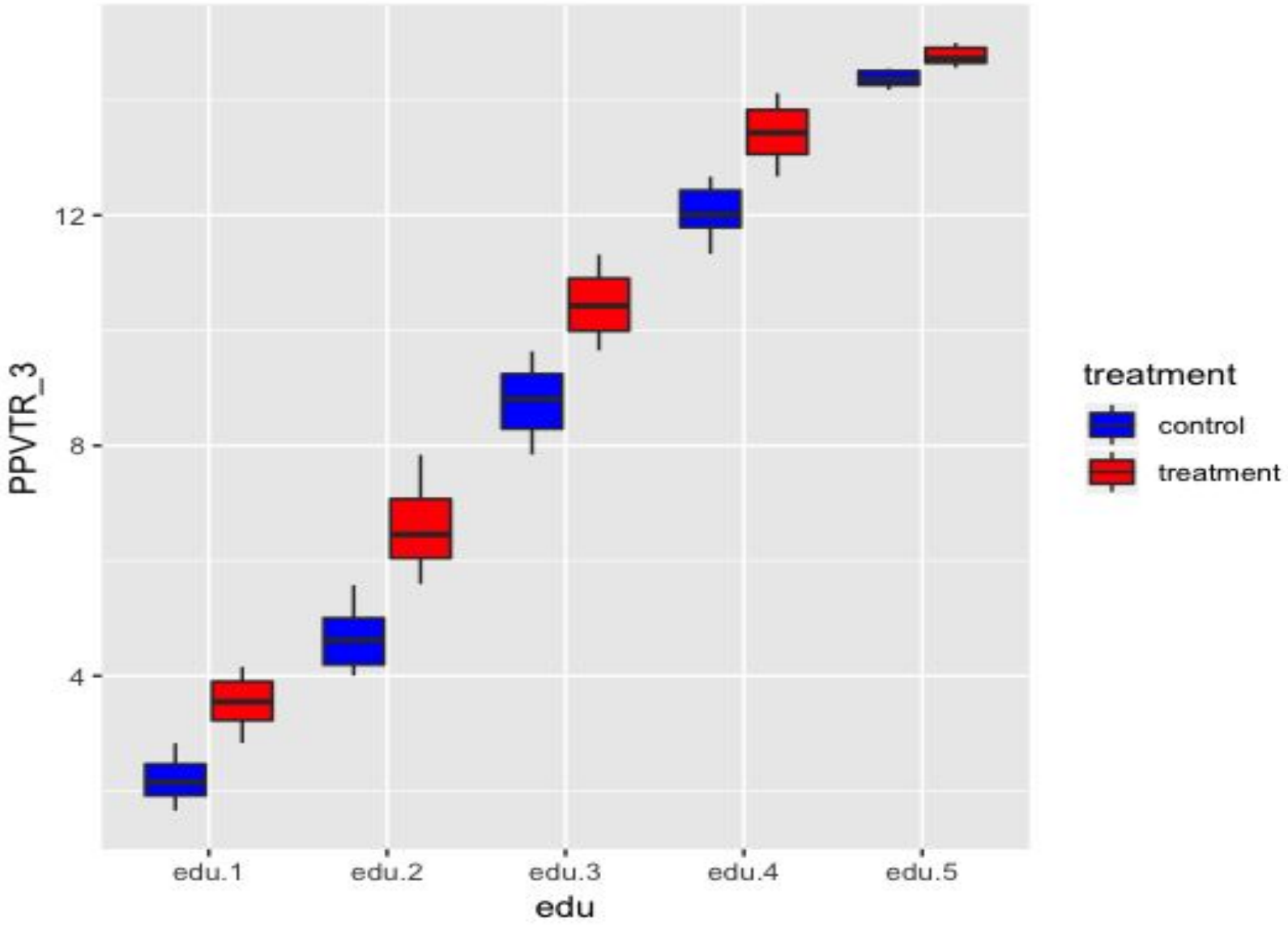
In a randomized block design, Ÿ

- the n study units are first divided into blocks (strata, subpopulations), with n_x units per block
- Ÿthe blocks are typically formed by splitting the units into subgroups of interest. In an experiment involving people, for example, it may be useful to split the subjects by age, race and sex. Ÿ
- within each block n_{1x} units are assigned to the treatment group and n_{0x} to the control group in such a way that every possible allocation is equally likely (so there's a little rand. exp. in each block)

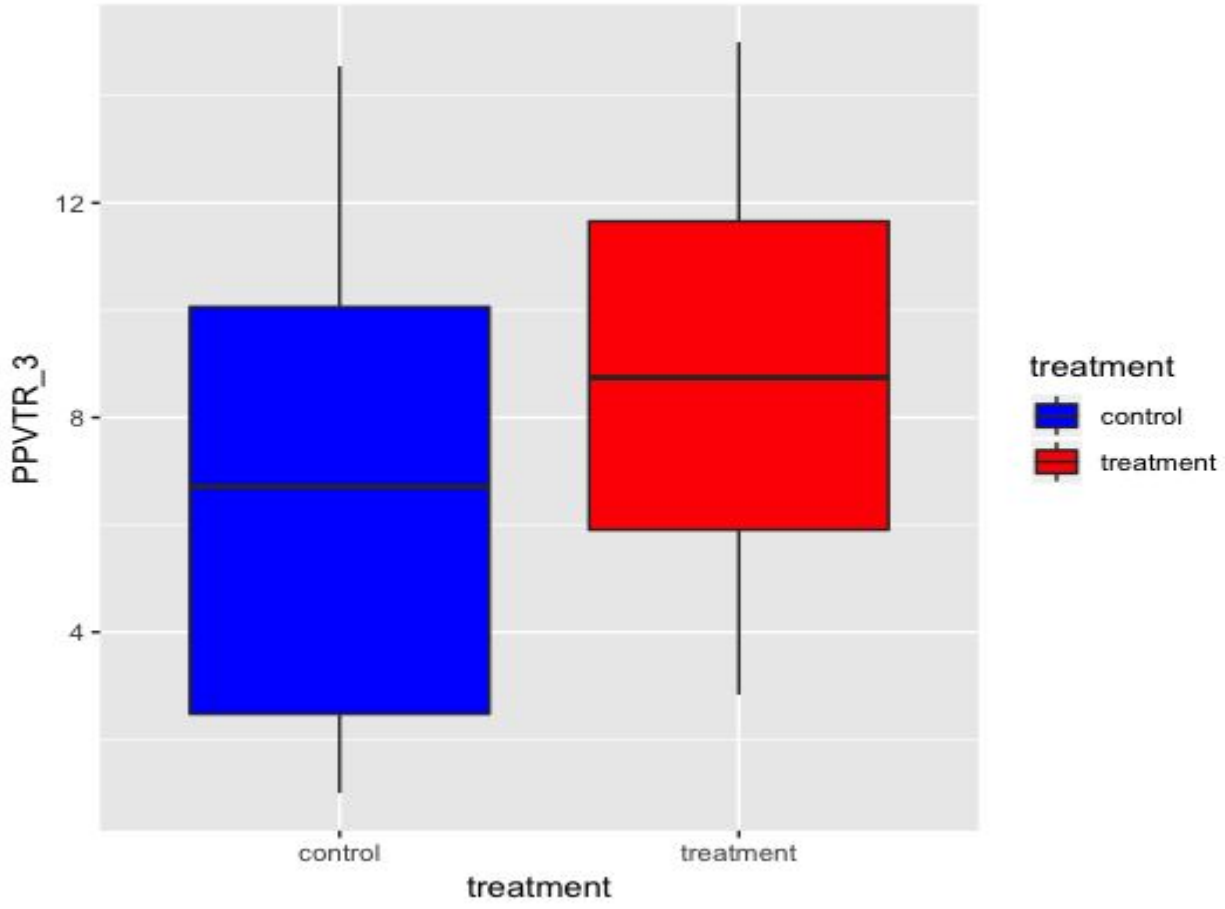
Randomized Block Experiment (cont'd)

By grouping the subjects in this way, one can ensure that subjects are "balanced" across groups with respect to these variables.

- This can be of great help in small experiments when the ATE is different for these different types of subjects.
- Note that the probability of being assigned to treatment does not need to be equal across blocks


PPVTR_3
12
8
4
edu.1
edu.2
edu.3
edu.4
edu.5
edu
treatment
control
treatment

Compare to noise in overall comparison (not blocked)



Randomized Block Experiment (RBE!)

- Key properties
- If we use X to define the variable (or vector of variables) that we use to create blocks,

$$Z \perp Y(0), Y(1) \mid X$$

$$[\text{or} \quad p(Y(0),Y(1) \mid X, Z) = p(Y(0),Y(1) \mid X)]$$

- For binary Z, this means that within any given block the distribution of potential outcomes is the same across treatment groups,

$$p(Y(0),Y(1) \mid X, Z=0) = p(Y(0),Y(1) \mid X, Z=1)$$

- It is *not* necessarily true that $Z \perp Y(0), Y(1)$

Data from a randomized experiment

HS	Z	Y(0)	Y(1)	Y	TEi
0	1	4	8	8	4
0	1	3	7	7	4
0	1	2	6	6	4
0	1	3	7	7	4
1	1	9	9	9	0
1	1	11	11	11	0
0	0	2	6	2	4
0	0	4	8	4	4
1	0	9	9	9	0
1	0	8	8	8	0
1	0	11	11	11	0
1	0	10	10	10	0
1	0	11	11	11	0
1	0	9	9	9	0

Questions about the data

1. If you were told that these data arose from a Completely Randomized Experiment, what would be your estimate of the ATE (assuming the simplest possible analysis)?
2. If you were told that these data arose from a Randomized Block Experiment, what would be your estimate of the ATE?
3. What is the ATE? (treating this sample as the population)

Block-specific ATEs: just like subgroup effects

We know that the randomized block experiment is like a bunch of little block-specific randomized experiments.

For each of these we can get an unbiased estimate of the block-specific ATE.

If we label denote the blocking variable as X then these block-specific ATEs are the same as ATE(X)!

Moving from block-specific ATEs to the general ATE

We're usually interested in the ATE (average treatment effect across the whole population).

And often, we also want to see the **ATE(X)** (average treatment effect within subpopulations defined by *X*).

$$\bar{Y}_{1x} - \bar{Y}_{0x}$$

To estimate the ATE, the investigator can use a weighted average of the conditional effects: $\frac{1}{n}\sum_{x} n_x(\bar{Y}_{1x} - \bar{Y}_{0x})$

Matched Pairs Design

Matched Pairs Design

- A paired randomized or matched pairs experiment is just a special case of a randomized block experiment in which $n_{0x} = n_{1x} = 1$
- Subjects are matched on the basis of similarity (in terms of potential confounders) and then a random draw (flip of a coin) decides which pair is in the treatment group and which in the control group
- In this case, $n_x = 2$, leading to an estimate of the ATE as:

$$(2/n)\,\Sigma_x\,(Y_{1x}-Y_{0x})$$

- Once again the goal is greater efficiency (smaller standard errors)

Average Treatment Effect on Treatment

ATT: Average treatment effects on treated is the average effect of treatment on group of individuals who received the treatment.

$$E[Y(1)-Y(0) \mid Z=1] = E[Y(1) \mid Z=1] - E[Y(0) \mid Z=1]$$

ATC: Average treatment effects on control is the average effect of treatment on group of individuals who did not received the treatment.

$$E[Y(1)-Y(0) \mid Z=0] = E[Y(1) \mid Z=0] - E[Y(0) \mid Z=0]$$

Assumptions

List some assumptions

Assumptions for causal inference for randomized experiments

- SUTVA
- Ignorability

Additional assumptions?

- Are randomized experiments then "assumption-free" (assuming the randomization was pristine)
- Actually, we have implicitly been making a major assumption all along. This is the assumption that units do not *interfere* with each other.
- In policy research and in the social sciences, this assumption is often tenuous.

Interference: Example 1

Example 1

- Consider two adjacent plots of land, plot 1 and plot 2. Of interest is the effect of a fertilizer on the amount of crop yield.
- Suppose plot 1 does not get fertilizer. Then we should see $Y_1(0)$. But what if plot 2 gets fertilizer and some of this runs over into plot 1? Then the yield from plot 1 depends on the treatment assigned to plot 2. This means that we must define *four* potential outcomes for unit 1:

  $Y_1(Z_1=0,Z_2=0)$

  $Y_1(Z_1=0,Z_2=1)$

  $Y_1(Z_1=1,Z_2=0)$

  $Y_1(Z_1=1,Z_2=1)$

Other examples of interference

Example 2: Moving To Opportunity.

- Participating families from housing projects are randomly assigned to either receive or not receive a housing voucher to move to the ``suburbs''. Of interest is whether or not the family feels safe. Since there are multiple families from a given project, there may be social interaction effects. Suppose that members of family 1 and family 2 are very close and agree to either move together to the suburbs or not move. Then if family 1 gets a housing voucher the family moves only if family 2 gets a voucher as well... but does not move otherwise.

SUTVA: Stable Unit Treatment Value Assumption

- The SUTVA assumption (Rubin 1980) rules out this type of interference between units.
- Let $\mathbf{Z} = (Z_1, Z_2, \ldots Z_N)'$ denote the treatment assignment for all the units in the population.
  - Let $Y_i(\mathbf{Z})$ denote the outcome for unit *i* when the units are assigned according to $\mathbf{Z}$ and
  - let $Y_i(\mathbf{Z}')$ denote the outcome for unit i when the units are assigned according to $\mathbf{Z}'$.
- We want the response of unit *i* to depend only on the assignment of that unit and not the assignment of other units. SUTVA formalizes this: For all units, we assume $Y_i(\mathbf{Z}) = Y_i(\mathbf{Z}')$ whenever $Z_i = Z'_i$.
- It also assumes that ``treatment'' means the same thing for all study participants.

SUTVA

$$Y_i \begin{pmatrix} Z_1 \\ Z_2 \\ \cdot \\ \cdot \\ \cdot \\ Z_i \\ \cdot \\ \cdot \\ \cdot \\ Z_{n-1} \\ Z_n \end{pmatrix} = Y_i \begin{pmatrix} Z'_1 \\ Z'_2 \\ \cdot \\ \cdot \\ \cdot \\ Z_i \\ \cdot \\ \cdot \\ \cdot \\ Z'_{n-1} \\ Z'_n \end{pmatrix}$$

Where Z_j may or may not equal Z_j' for all $j \neq i$

SUTVA practicalities

- How do we deal with violations of SUTVA in practice?
- Typically the best strategy is to avoid problems by designing studies more carefully. For instance if you are concerned with student-to-student interference then treatments should be applied at the group-level (for instance classroom)
- At the analysis stage there isn't much that can be done. Therefore, throughout this course, we shall make this assumption unless otherwise stated.
- More work is being done in this area every day (typically in the context of studies that have been) and some partial relaxations exist but we won't be able to cover them in this course.

Assumptions when estimating treatment effects using data from a randomized experiment

- Structural assumptions
  - Ignorability (should be satisfied by design)
  - SUTVA
- Parametric assumptions
  - Depends on the model used to estimate the treatment effect
  - Generally speaking any should be fairly robust to departures from the parametric assumptions about the mean structure of the model because of the randomization (and consequent balance/overlap)
  - Error structure (won't address here but possible violations include heteroskedasticity and grouped data structures)

Summary of important concepts

- Counterfactuals
- Potential outcomes
- Causal effects as comparisons of potential outcomes
- Average treatment effects
- Randomization creates independence between potential outcomes and the treatment
- In randomized block experiments the randomization property may only hold within strata that are defined by pre-treatment covariates. It's important to account for the blocked structure when estimating the treatment effect.

Example: IHDP

The Infant Health and Development Program

- Targeted low birth weight, premature infants born in 1985
- 985 infants across 8 sites randomized to the intervention or to a follow-up only group
- Children in the intervention arm received home visits (1st year) and intensive, high-quality, child care (years 2 and 3 of life)

(see e.g. Brooks-Gunn et al., JAMA, 1994, v. 272)

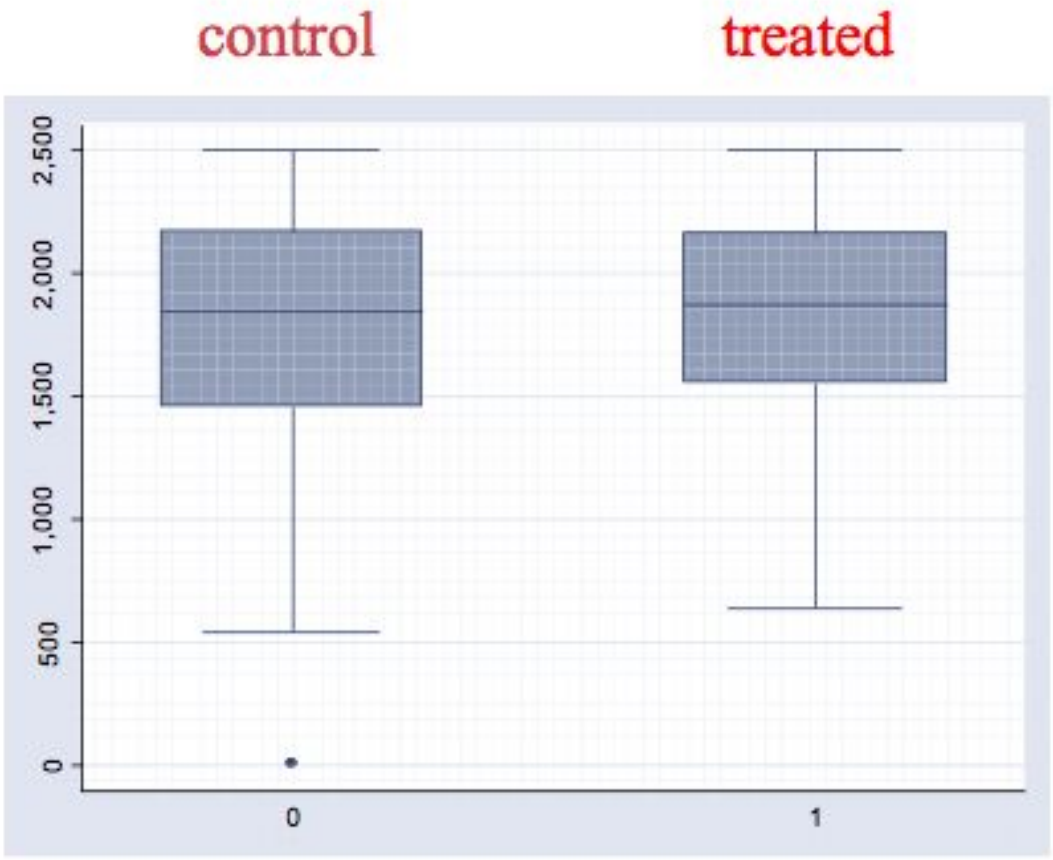
Did the randomization “work” in the IHDP?

What would we expect the **mean** of any given pre-treatment variable to look like for the treatment group relative to the control group?

Variables	FU	IHDP	Variables	FU	IHDP
Mother			*Child*		
Age	25.0	24.7	Birth weight	1787	1816
Black	0.52	0.55	Head circ (birth)	29.5	29.5
Hispanic	0.12	0.09	Sex	0.52	0.50
White	0.36	0.36	Weeks pre-term	7.0	7.0
Married (birth)	0.49	0.43	Birth order	1.9	1.9
< high school	0.37	0.43	Neonatal health	99.6	100.9
High school	0.27	0.28	Twin	0.17	0.19
Some college	0.22	0.17			
College grad	0.13	0.13	*Father*		
Cigarettes (preg)	0.35	0.35	Black	0.52	0.55
Alcohol (preg)	0.13	0.11	Hispanic	0.12	0.10
Drugs (preg)	0.03	0.04	White	0.36	0.35
Worked (preg)	0.59	0.60			
Prenatal care	0.96	0.94			

What would we expect the **distribution** of any given outcome variable to look like for the treatment group relative to the control group?

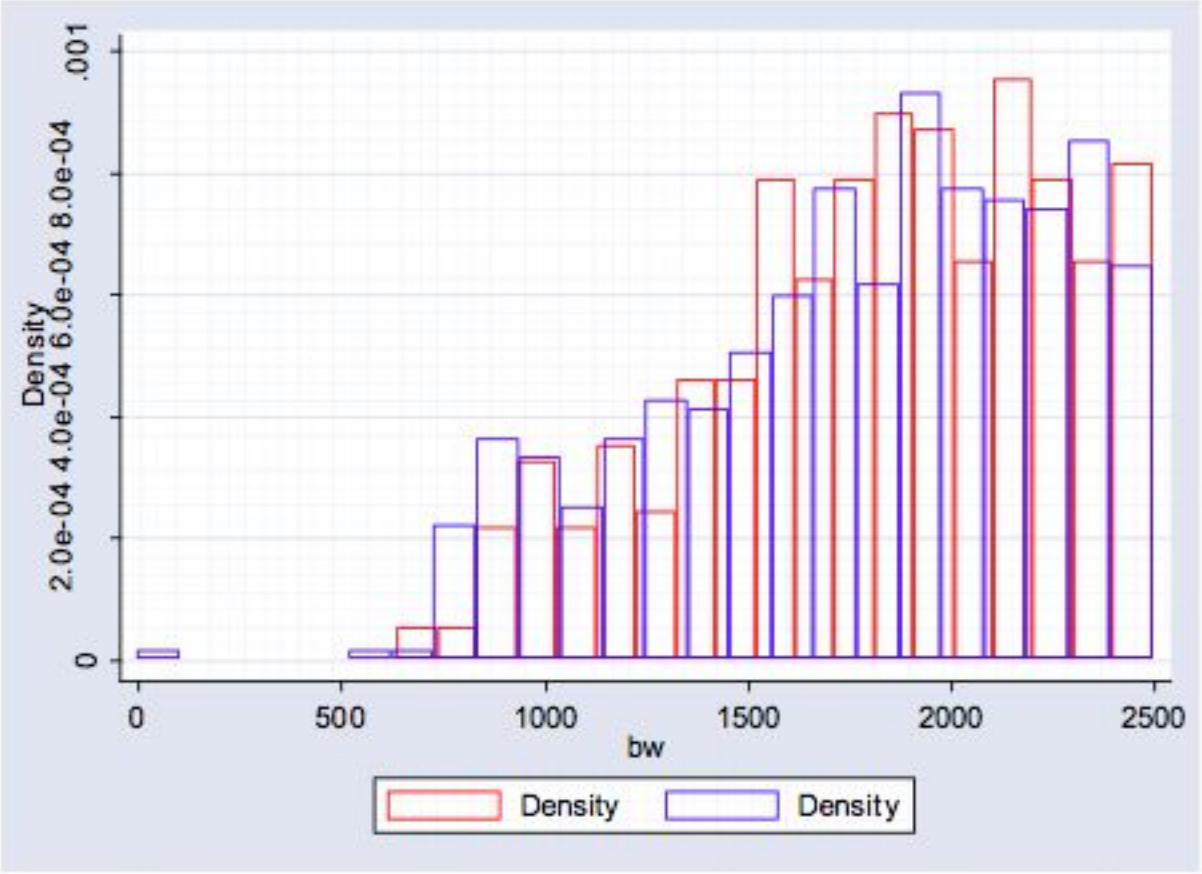
Birthweight in treatment and follow-up-only groups: boxplots



Birthweight in treatment and follow-up-only groups: histograms

control

treated



!!Pop Quiz!!

How do we estimate causal effects using data from a randomized experiment?

How do we estimate causal effects using data from a randomized experiment?

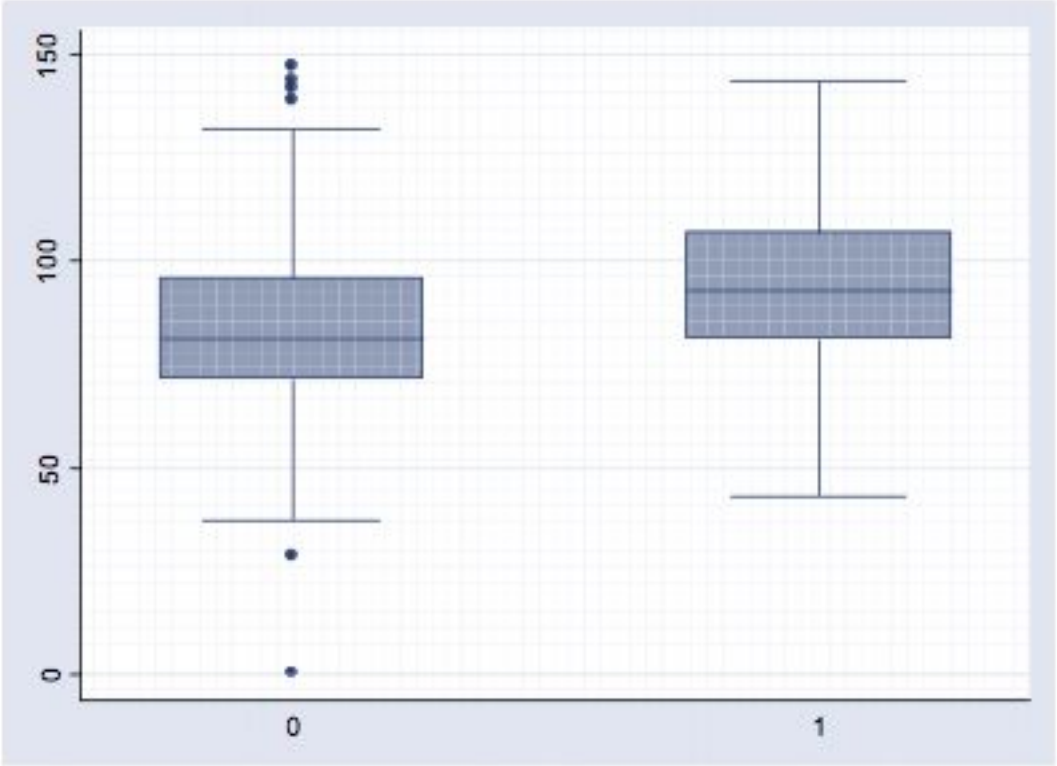
- Difference in means
- Regression with an indicator for treatment (but nothing else)
- Regression with an indicator for treatment plus other pre-treatment variables

Estimating treatment effects in a randomized experiment

- Difference in means: With sufficiently large samples the randomization creates treatment and control groups are the same on average (both in terms of observed *and unobserved* characteristics). Thus all we have to do to unbiasedly estimate a treatment effect is to compare mean outcomes across groups (e.g. a difference in means)

- Coefficient from a regression (or other model): With small sample sizes, or generally if we want more precision, we might choose to estimate treatment effects in the context of a regression model so we can further adjust for remaining (unsystematic) differences across groups in other background characteristics

Year 3 IQ in each treatment group



Results: age 3 test scores

Running a regression of age 3 IQ scores on the treatment indicator and a handful of pre-treatment covariates yields a treatment effect estimate of 6.4 with standard error of 1.2.

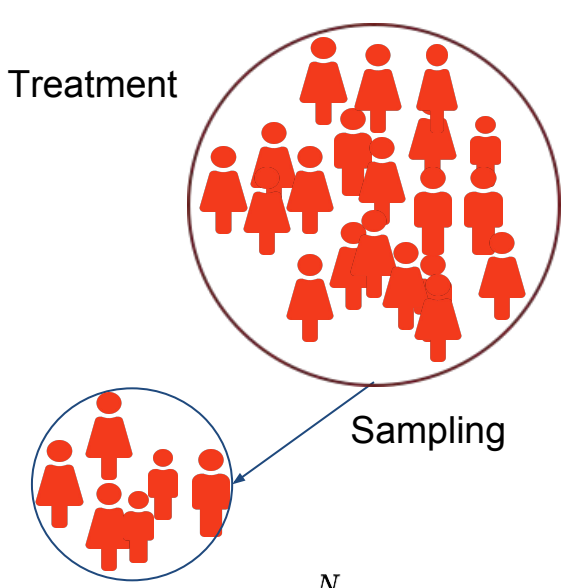
*this from an analysis correcting for missing data issues

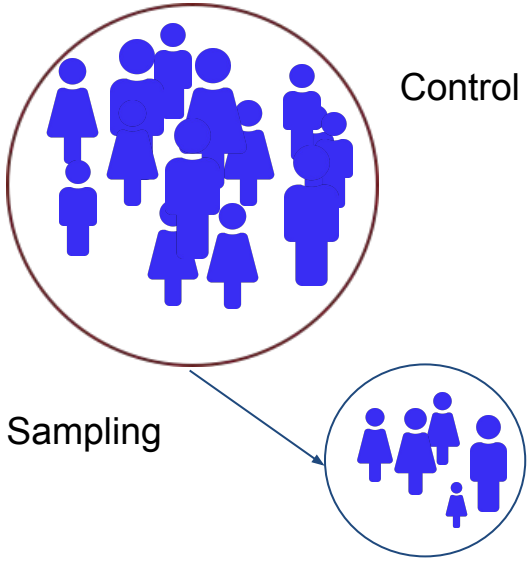
Conclusions

- Randomized experiments are the gold standard for causal inference (though not without flaws)
- Without randomized experiments we need to another way to control for differences between those who receive the treatment and those who do not
- Simple methods such as regression may not be sufficient!
- More complicated methods (e.g. propensity score matching) may also not be sufficient, but may allow us to make slightly less heroic assumptions
- Crucial to be clear about the assumptions were are making with each analysis strategy and to try to minimize the strength of these assumptions

Drawing a sample from the population

Assuming the random sample, the potential outcomes of the selected sample will be similar (if not exactly) like that of the population.





$$E[Y(1)|Z=1] = \frac{1}{N_0}\sum_{i=1}^{N} Y_i(1)(Z_i) = 12$$

$$E[Y(0)|Z=1] = \frac{1}{N_0}\sum_{i=1}^{N} Y_i(1)(Z_i) = 8$$

Same potential outcomes

$$E[Y(1)|Z=0] = \frac{1}{N_0}\sum_{i=1}^{N} Y_i(1)(1-Z_i) = 12$$

$$E[Y(0)|Z=0] = \frac{1}{N_0}\sum_{i=1}^{N} Y_i(1)(1-Z_i) = 8$$