In-person session 8

March 5, 2026

PMAP 8521: Program evaluation
Andrew Young School of Policy Studies

Plan for today

do()

***p*-values and confidence intervals**

RCTs

Matching and IPW

do()

Expectations

$$\mathrm{E}(\cdot) \qquad \mathbf{E}(\cdot) \qquad \mathbb{E}(\cdot)$$

Basically a fancy way of saying "average"

andhs.co/potential-outcomes

Causal effects with potential outcomes

Potential outcomes notation:

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

or alternatively with $\mathbf{E}$

$$\delta = \mathbf{E}[Y_i(1) - Y_i(0)]$$

Causal effects with do()

Pearl notation:

$$\delta = \mathbf{E}[Y_i \mid \text{do}(X = 1) - Y_i \mid \text{do}(X = 0)]$$

or more simply

$$\delta = \mathbf{E}[Y_i \mid \text{do}(X)]$$

$$\mathbf{E}[Y_i \mid \mathrm{do}(X)]$$

$$=$$

$$\mathbf{E}[Y_i(1) - Y_i(0)]$$

We can't see this

$$\mathbf{E}[Y_i \mid \text{do}(X)] \quad \text{or} \quad \mathbf{E}[Y_i(1) - Y_i(0)]$$

So we find the average causal effect (ACE)

$$\hat{\delta} = \mathbf{E}[Y_i \mid X = 1] - \mathbf{E}[Y_i \mid X = 0]$$

The average population-level change in y when *directly intervening* (or doing) x

$$\mathbf{E}(y \mid \mathrm{do}(x)) \neq \mathbf{E}(y \mid x)$$

Causation

The average population-level change in y when accounting for *observed* x

Correlation

p-values and confidence intervals

In the absence of *p*-values, I'm confused about how we report... significance?

Imbens and *p*-values

Nobody really cares about *p*-values

Decision makers want to know a number or a range of numbers—some sort of effect and uncertainty

Nobody cares how likely a number would be in an imaginary null world!

Imbens's solution

Report point estimates and some sort of range

> "It would be preferable if reporting standards emphasized confidence intervals or standard errors, and, even better, Bayesian posterior intervals."

Point estimate

The single number you calculate (mean, coefficient, etc.)

Uncertainty

A range of possible values

Population parameter

Truth = Greek letter

An single unknown number that is true for the entire population

Proportion of left-handed students at GSU

Median rent of apartments in Atlanta

Proportion of red M&Ms produced in a factory

Treatment effect of your program

Samples and estimates

We take a sample and make a guess

This single value is a *point estimate*

(This is the Greek letter with a hat)

Nets and confidence intervals

How confident are we that the sample picked up the population parameter?

Confidence interval is a net

We can be X% confident that our net is picking up that population parameter

If we took 100 samples, at least 95 of them would have the true population parameter in their 95% confidence intervals

> A city manager wants to know the true average property value of single-owner homes in her city. She takes a random sample of 200 houses and builds a 95% confidence interval. The interval is ($180,000, $300,000).

We're 95% confident that the interval ($180,000, $300,000) captured the true mean value

WARNING

It is way too tempting to say "We're 95% sure that the population parameter is X"

People do this all the time! People with PhDs!

YOU will do this too

Nets

**If you took lots of samples,
95% of their confidence intervals
would have the single true value in them**

Frequentism

This kind of statistics is called "frequentism"

The population parameter θ is fixed and singular while the data can vary

$$P(\text{Data} \mid \theta)$$

You can do an experiment over and over again; take more and more samples and polls

Frequentist confidence intervals

"We are 95% confident that this net captures the true population parameter"

~~"There's a 95% chance that the true value falls in this range"~~

Bayesian statistics

Rev. Thomas Bayes

$$P(\theta \mid \text{Data})$$

$$P(\text{H} \mid \text{E}) = \frac{P(\text{H}) \times P(\text{E} \mid \text{H})}{P(\text{E})}$$

$$P(\mathrm{H} \mid \mathrm{E}) = \frac{P(\mathrm{H}) \times P(\mathrm{E} \mid \mathrm{H})}{P(\mathrm{E})}$$

$$P(\mathrm{Hypothesis} \mid \mathrm{Evidence}) =$$

$$\frac{P(\mathrm{Hypothesis}) \times P(\mathrm{Evidence} \mid \mathrm{Hypothesis})}{P(\mathrm{Evidence})}$$

But the math is too hard!

So we simulate!

(Monte Carlo Markov Chains, or MCMC)

Bayesianism and parameters

In the world of frequentism, there's a fixed population parameter and the data can hypothetically vary

$$P(\text{Data} \mid \theta)$$

In the world of Bayesianism, the data is fixed (you collected it just once!) and the population parameter can vary

$$P(\theta \mid \text{Data})$$

Bayesian credible intervals

(AKA posterior intervals)

"Given the data, there is a 95% probability that the true population parameter falls in the credible interval"

Intervals

Frequentism

There's a 95% probability that the range contains the true value

Probability of the range

Few people naturally think like this

Bayesianism

There's a 95% probability that the true value falls in this range

Probability of the actual value

People *do* naturally think like this!

Thinking Bayesianly

**We all think Bayesianly,
even if you've never heard of Bayesian stats**

Every time you look at a confidence interval, you inherently think that the parameter is around that value, but that's wrong!

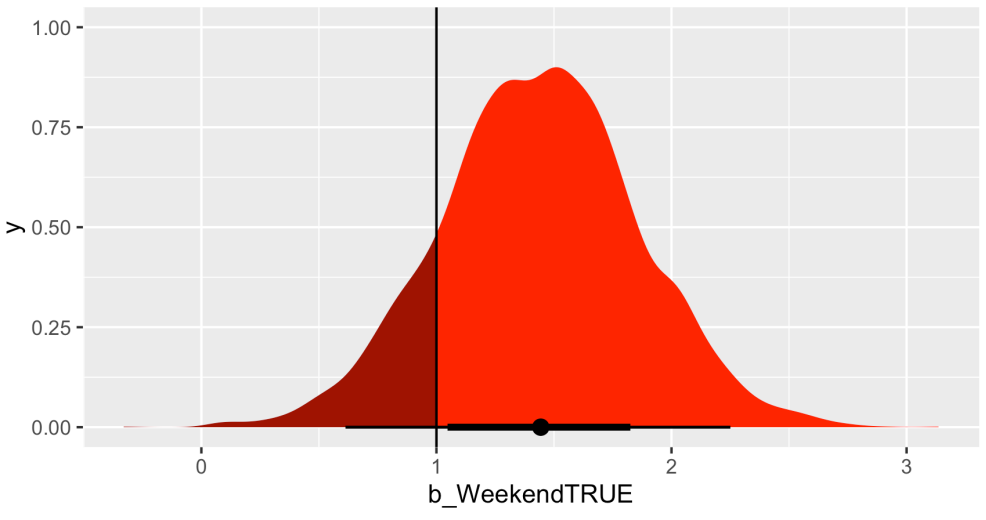
BUT Imbens cites research that that's actually generally okay

Often credible intervals are super similar to confidence intervals

Bayesian inference

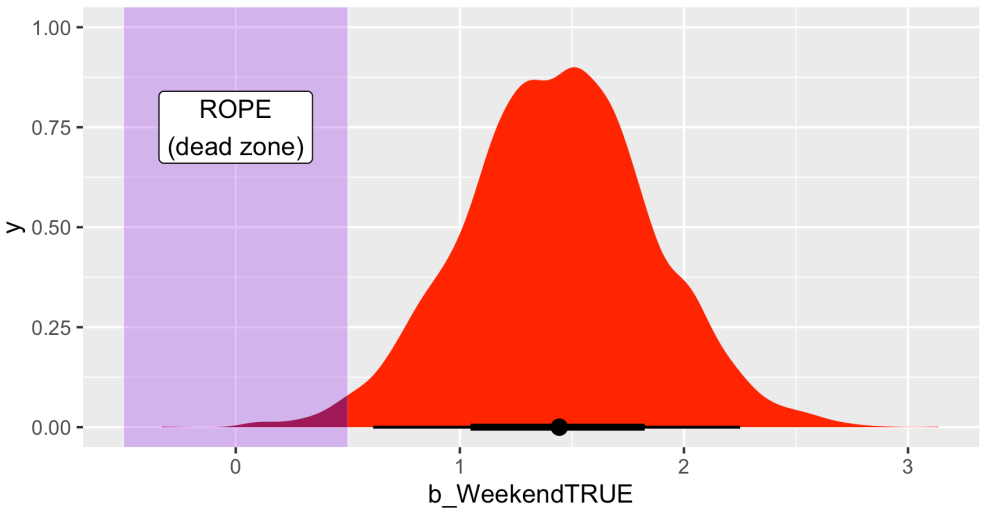
Inference without *p*-values!

Probability of direction



Point shows median value;
thick black bar shows 66% credible interval;
thin black bar shows 95% credible interval

Region of practical equivalence (ROPE)



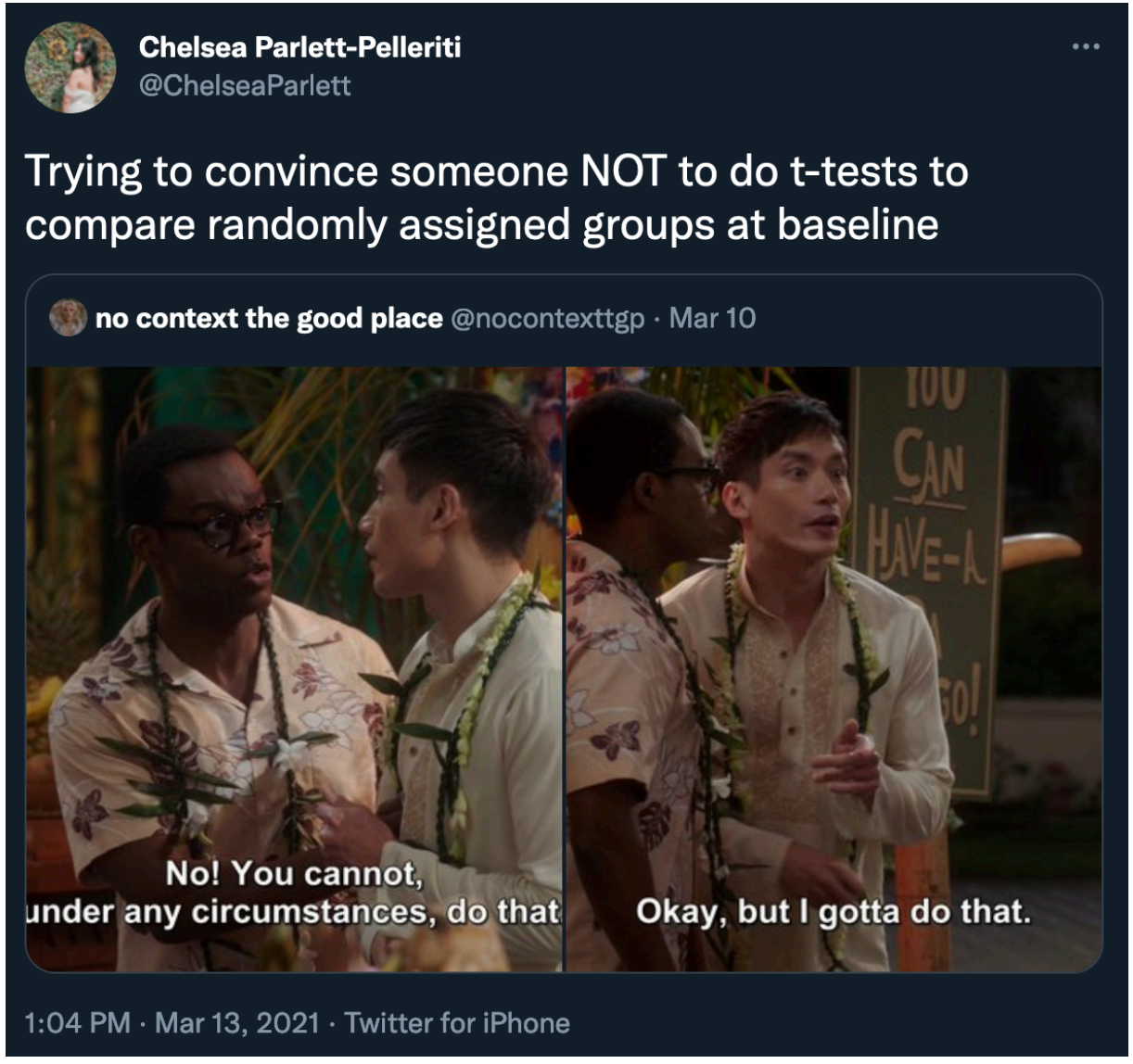
Point shows median value;
thick black bar shows 66% credible interval;
thin black bar shows 95% credible interval

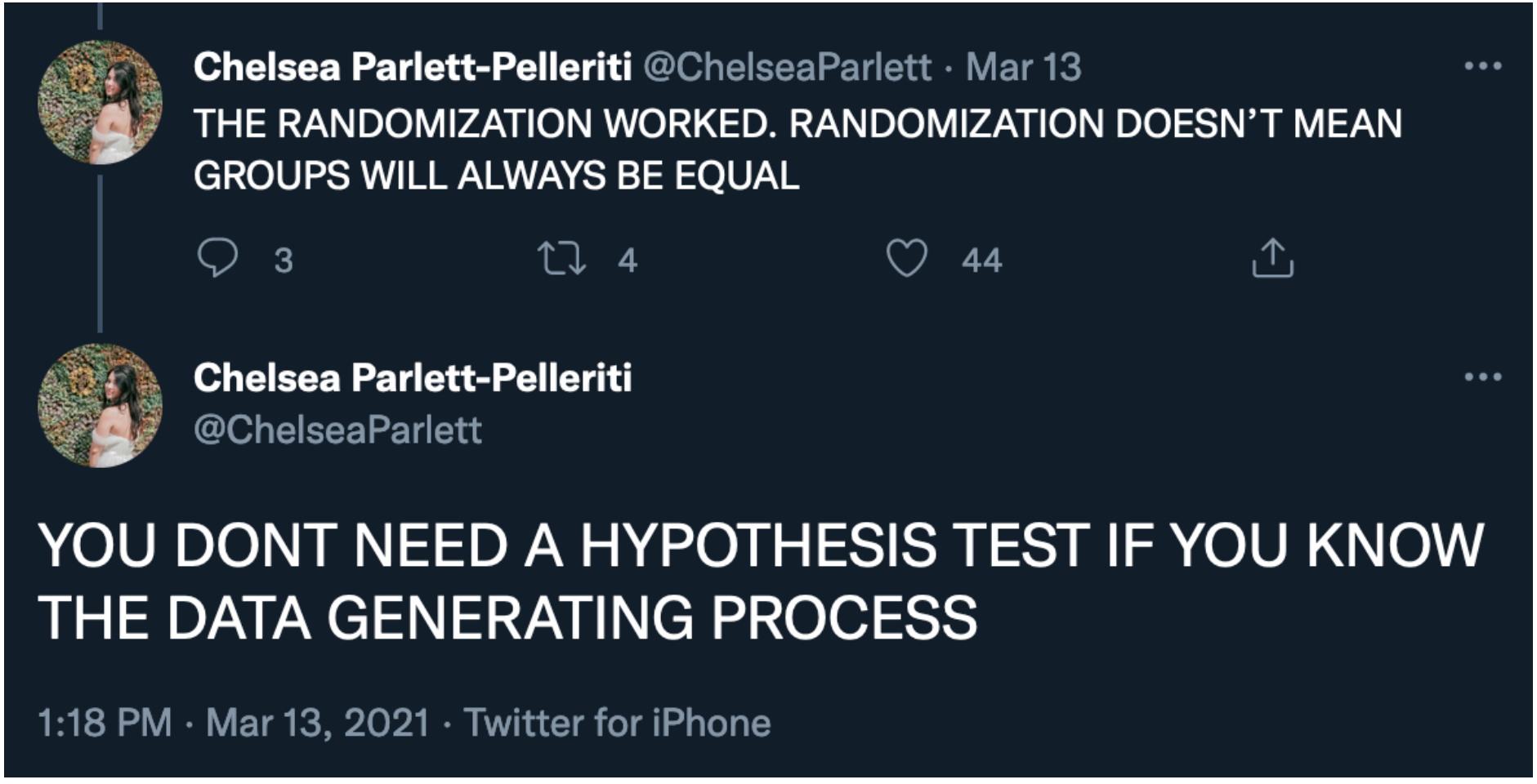
RCTs

Do we really not control for things in an RCT?

Randomness and arrow deletion

Balance tests


Chelsea Parlett-Pelleriti
@ChelseaParlett
Trying to convince someone NOT to do t-tests to compare randomly assigned groups at baseline
no context the good place @nocontexttgp · Mar 10
No! You cannot,
under any circumstances, do that
CAN
HAVE-A
Okay, but I gotta do that.
1:04 PM · Mar 13, 2021 · Twitter for iPhone

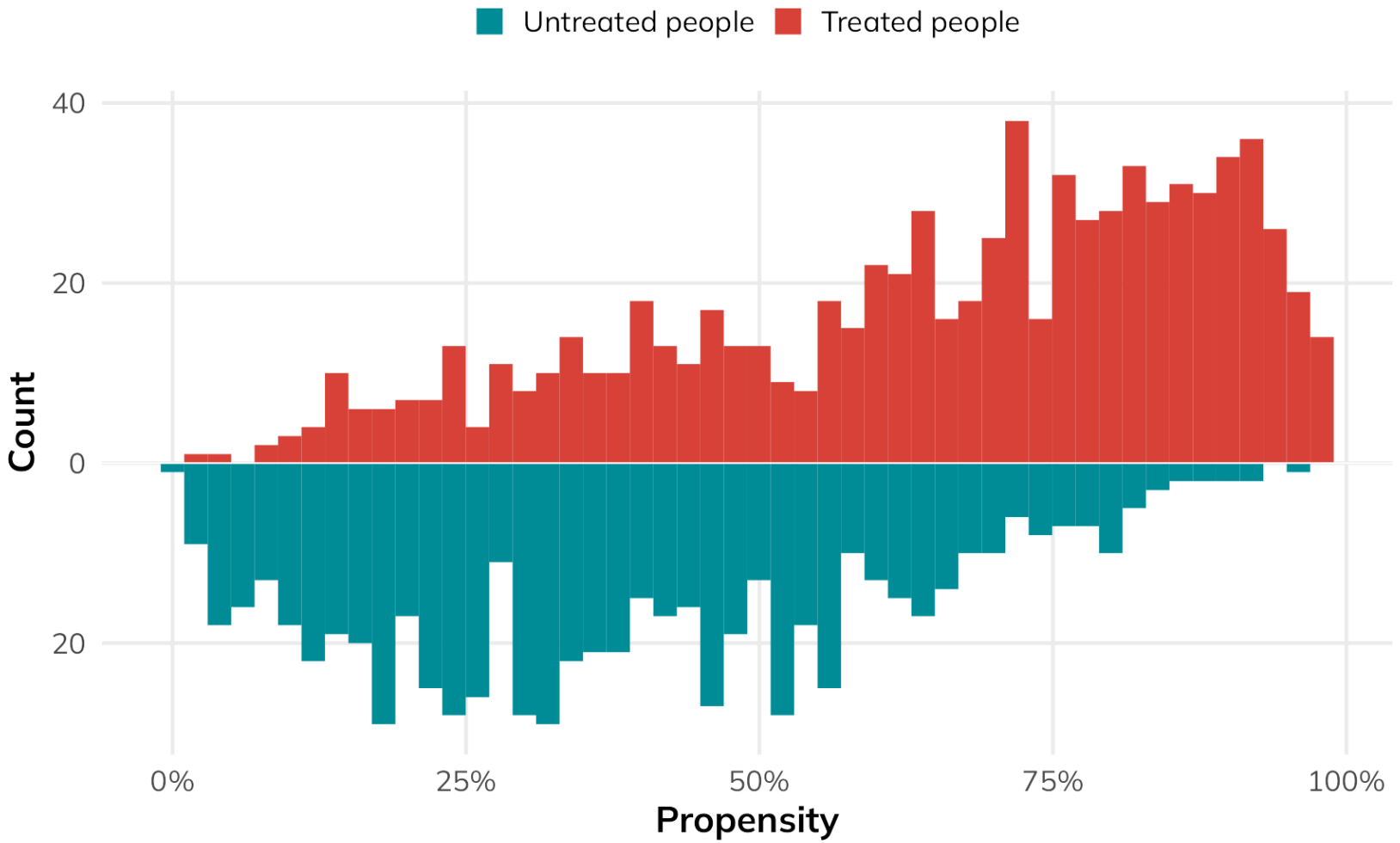

Chelsea Parlett-Pelleriti @ChelseaParlett · Mar 13
THE RANDOMIZATION WORKED. RANDOMIZATION DOESN'T MEAN GROUPS WILL ALWAYS BE EQUAL
3
4
44
Chelsea Parlett-Pelleriti
@ChelseaParlett
YOU DONT NEED A HYPOTHESIS TEST IF YOU KNOW THE DATA GENERATING PROCESS
1:18 PM · Mar 13, 2021 · Twitter for iPhone

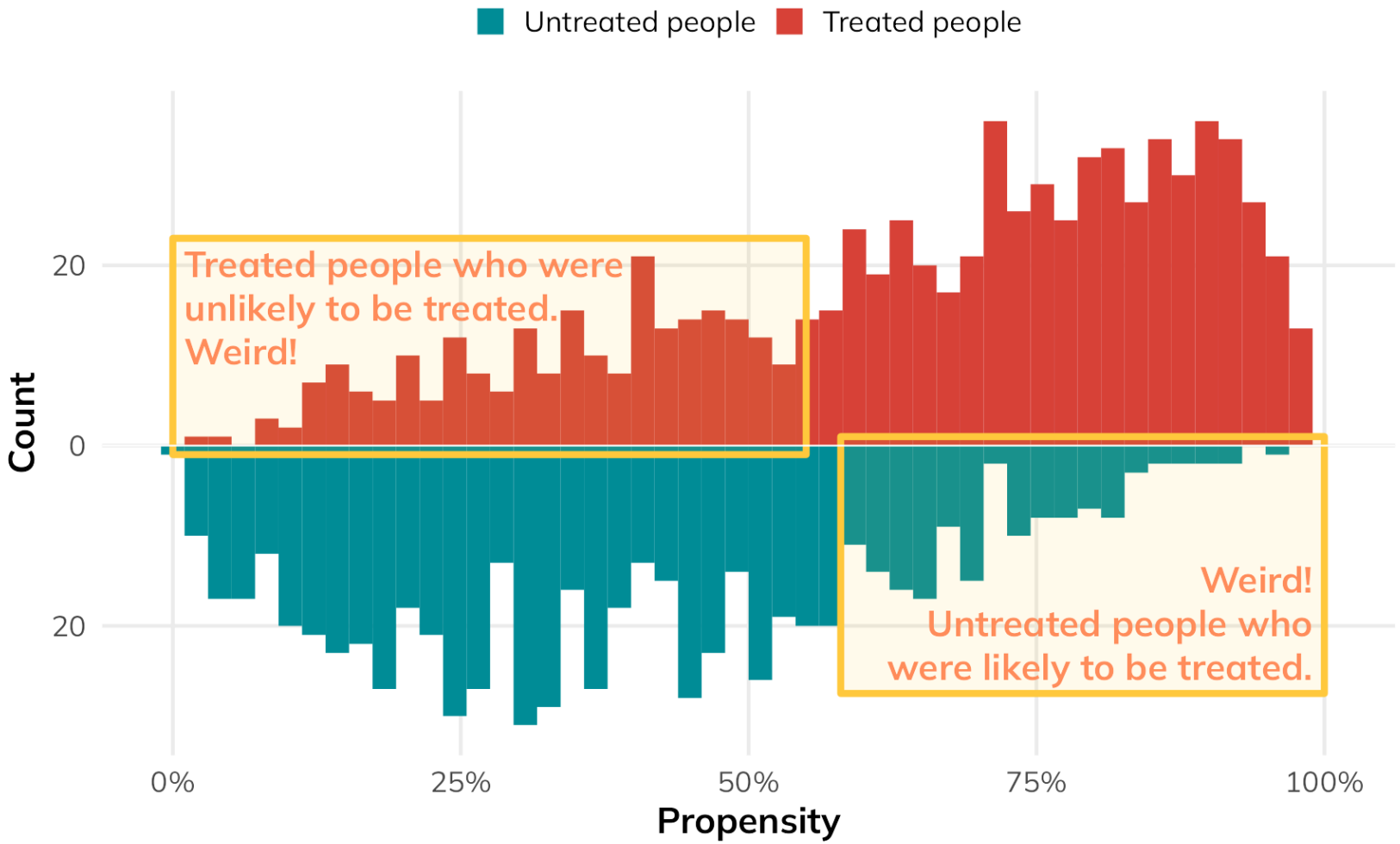
Can you walk through an example of RCTs in class?

Matching and IPW

Can you talk more about propensity scores and "weirdness" weights?

Lecture slide


Untreated people
Treated people
40
20
0
20
Count
0%
25%
50%
75%
100%
Propensity


Untreated people
Treated people
Treated people who were unlikely to be treated. Weird!
Weird! Untreated people who were likely to be treated.
Count
20
0
20
0%
25%
50%
75%
100%
Propensity


Treated people
Untreated people
Treated pseudo-population
Untreated psuedo-population
More weight here
40
0
40
Count
0%
25%
50%
75%
100%
Propensity

Why not just control for confounders instead of doing the whole matching/IPW dance?

Do you have to use logistic regression + OLS for IPW?

No!

Which should we use? Matching or IPW?

Can you walk through an example of IPW and matching in class?