STA 235H - Potential Outcomes II

Fall 2021

McCombs School of Business, UT Austin

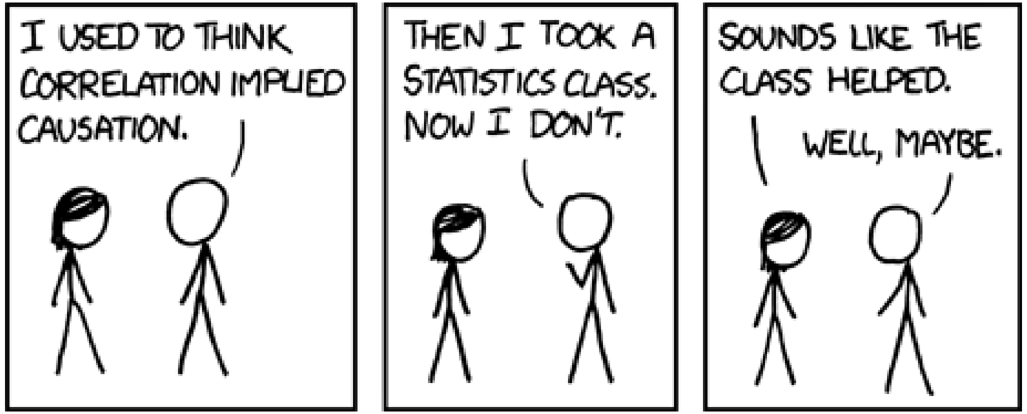
Housekeeping

Homework 2 will be posted on Thursday

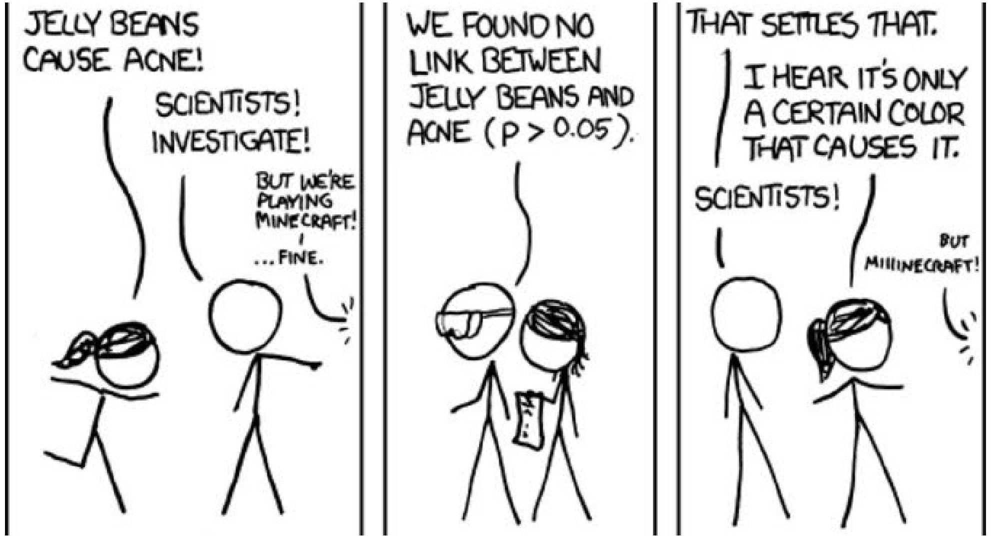
- HW 1 Answer key is posted on the course website (review it!).
- Added additional resources on our website (regression and intro to causal inference).
- Student suggestion: Class recordings will be available two weeks before the midterm.
    - Trial run: Check if attendance drops significantly.

Last week

- Started talking about Causal Inference.
  - Potential outcomes framework: What are potential outcomes? How we identify a counterfactual? What is an estimand?
  - Fundamental problem of causal inference: Ignorability assumption
  - Sources of bias: Selection bias and heterogenous return to treatment bias.



Today



- Finish with potential outcomes framework:
  - Examples related to causal inference.
  - Confounder vs. Collider.
- Introduction to Randomized Controlled Trials (RCTs)

Let's look back at some math

$$\mu = E[Y_i]$$

What does the previous equation mean?

"μ is the expected value of the observed outcome Y"

$$\mu_1 = E[Y_i | Z = 1]$$

What does the previous equation mean?

"μ_1 is the expected value of observed outcome Y for the treated group (i.e. units for which treatment Z = 1)"

$$\mu_1(0) = E[Y_i(0)|Z = 1]$$

What does the previous equation mean?

"$\mu_1(0)$ is the expected value of the potential outcome Y *under control* (Y(0)) for the treated group (i.e. units for which treatment Z = 1)"

In other words,

"$\mu_1(0)$ is the expected value of the outcome Y for the treated group, *if the treated group had not been treated*"

Let's look at some data

Example: Effect of types of advertising on sales

You want to know whether is more convenient to e-mail or physically mail potential customers to increase your sales.

Example: Effect of types of advertising on sales

You want to know whether is more convenient to **e-mail** or **physically mail** potential customers to increase your sales.

- What is the **treatment**?

Example: Effect of types of advertising on sales

You want to know whether is more convenient to **e-mail** or **physically mail** potential customers to increase your sales.

- What is the **treatment**?
- What is the **causal question** that you want to answer?

Example: Effect of types of advertising on sales

You want to know whether is more convenient to **e-mail** or **physically mail** potential customers to increase your sales.

- What is the **treatment**?
- What is the **causal question** that you want to answer?
- What would the **counterfactual** be?

Looking at some data

- You get some data from a friend in Silicon Valley, who works at a similar company:

% of New Registrations by Type of Campaign

Treatment	Total
E-mail	19%
	(290/1500)
Mail	16%
	(88/550)

Looking at some data

- You get some data from a friend in Silicon Valley, who works at a similar company:

% of New Registrations by Type of Campaign

Treatment	Total
E-mail	**19%**
	(290/1500)
Mail	16%
	(88/550)

Does this mean that e-mailing is more effective in getting new customers?

Looking at some data

- You get some data from a friend in Silicon Valley, who works at a similar company:

% of New Registrations by Type of Campaign

Treatment	Total
E-mail	**19%**
	(290/1500)
Mail	16%
	(88/550)

Does this mean that e-mailing is more effective in getting new customers?

What additional information would you need?

Let's add some covariates

- Your friend now also sends you additional data on whether the individual had ever visited the site:

% of New Registrations by Type of Campaign and Visits to the Website

Treatment	Visited web	Not visited web	Total
E-mail	10%	20%	19%
	10/100	280/1400	(290/1500)
Mail	15%	31%	16%
	77/514	11/36	(88/550)

What seems strange?

The Simpson Paradox

Let's add some covariates

- The majority of the sample that was assigned to "E-mail" had not visited the website before, while the majority of the sample that was sent a mailing had visited the website.

% of the Sample in each Category by Site Visit

Treatment	Visited web	Not visited web
E-mail	6.7%	93.4%
Mail	93.4%	6.5%

Let's add some covariates

- Your friend now also sends you additional data on whether the individual had ever visited the site:

% of New Registrations by Type of Campaign and Visits to the Website

Treatment	Visited web	Not visited web	Total
E-mail	10%	20%	19%
	10/100	280/1400	(290/1500)
Mail	15%	31%	16%
	77/514	11/36	(88/550)

Do we have a confounding problem?

Confounding

Confounder

Variable that is correlated with the treatment AND the outcome which causes a spurious correlation/bias.

Is "Visited the website" a confounder?

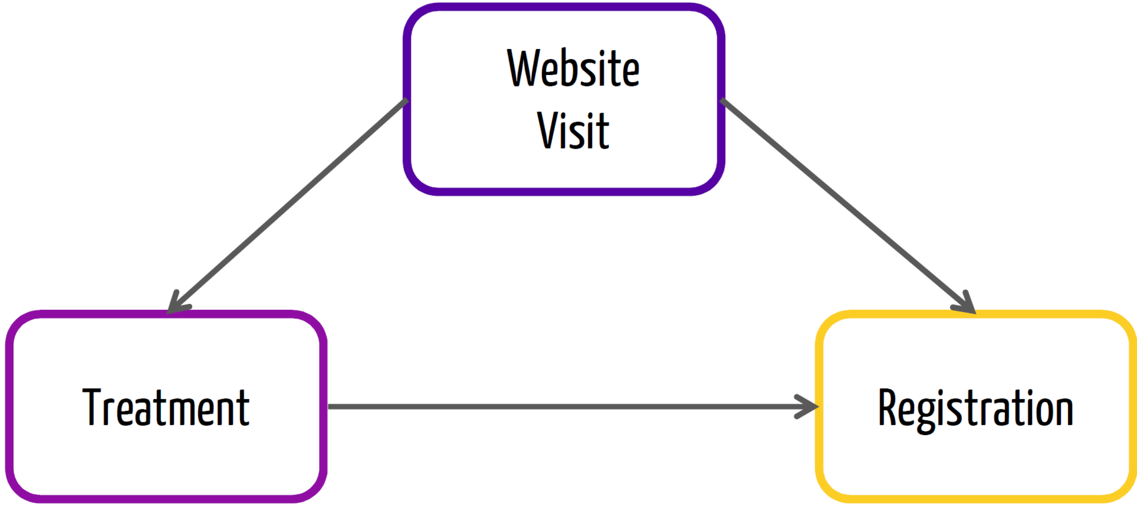
Is "Visited the website" a confounder?

Depends

- Measured before the intervention: Yes → Individuals that have **not** visited the website (VW) don't know you/ might be more willing to try product.
- Measured after the intervention: Don't know → Intervention might have incentivized people to go to the website, and registering also had on effect on traffic.

Collider

Scenario 1: Confounding



Scenario 1: Confounding

Data Generating Process:

- No treatment effect.
- Pr(Registering | Visit) < Pr(Registering | Not Visited)
- Due to data collection, more people in the mailing sample had visited the website than people in the email sample.

Num in Sample and % of New Registrations by Type of Campaign and Visits to the Website

	Not visited	Visited	Registered - NV	Registered - V	Registered - Total
Email	1404	91	0.22	0.12	0.21
Mail	32	523	0.28	0.15	0.15

Note: Simulated data

Scenario 1: Confounding

What happens if we run a **simple model**?

```
summary(lm(y ~ factor(treat), data = confound))
```

- What would you **expect** to see?

Scenario 1: Confounding

What happens if we run a simple model?

```
summary(lm(y ~ factor(treat), data = confound))
```

```
##
## Call:
## lm(formula = y ~ factor(treat), data = confound)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.2107 -0.2107 -0.2107 -0.1532  0.8468
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.21070    0.01023  20.590  < 2e-16 ***
## factor(treat)m -0.05755    0.01967  -2.926  0.00347 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3957 on 2048 degrees of freedom
## Multiple R-squared:  0.004164,	Adjusted R-squared:  0.003677
## F-statistic: 8.563 on 1 and 2048 DF,  p-value: 0.003469
```

Scenario 1: Confounding

What happens if we now **control by whether the person visited the website**?

```
summary(lm(y ~ factor(treat) + visit, data = confound))
```

Scenario 1: Confounding

What happens if we now **control by whether the person visited the website**?

```
summary(lm(y ~ factor(treat) + visit, data = confound))
```

```
##
## Call:
## lm(formula = y ~ factor(treat) + visit, data = confound)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.2532 -0.2172 -0.2172 -0.1470  0.8890
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.21716    0.01046  20.768  < 2e-16 ***
## factor(treat)m  0.03602    0.03786   0.951  0.34154
## visit           -0.10615    0.03673  -2.890  0.00389 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.395 on 2047 degrees of freedom
## Multiple R-squared:  0.00821,    Adjusted R-squared:  0.007241
## F-statistic: 8.473 on 2 and 2047 DF,  p-value: 0.0002165
```

Scenario 1: Confounding

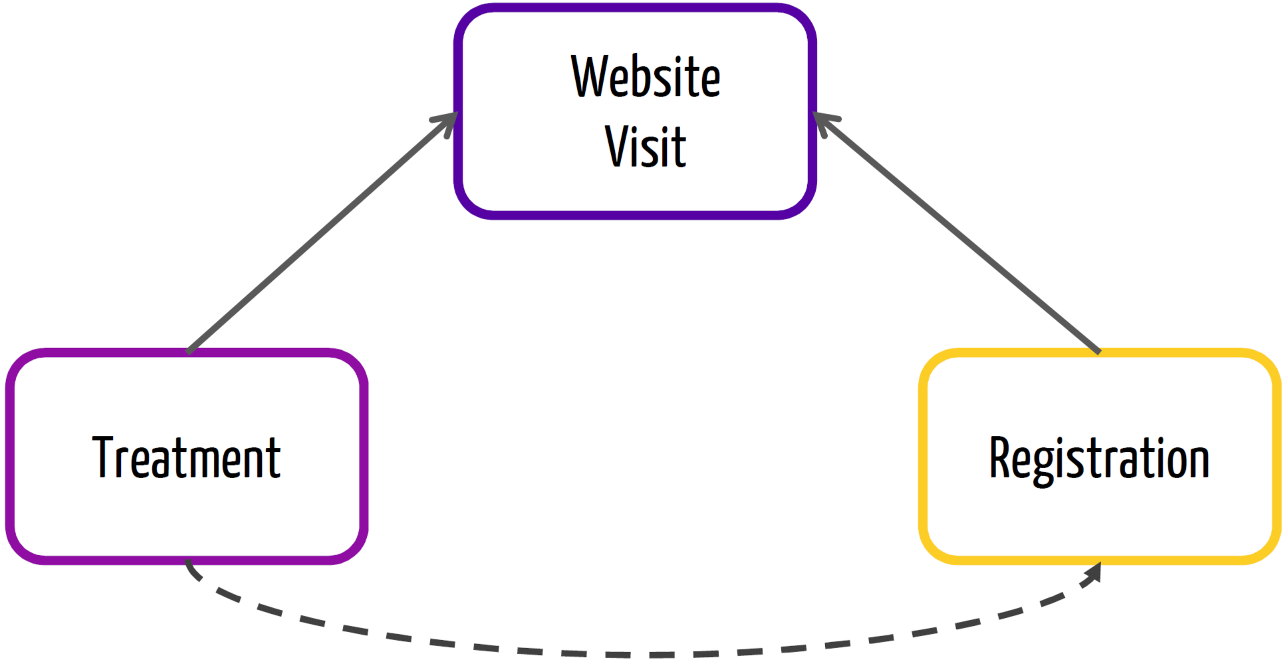
What happens if we now **control by whether the person visited the website**?

```
summary(lm(y ~ factor(treat) + visit, data = confound))
```

```
##
## Call:
## lm(formula = y ~ factor(treat) + visit, data = confound)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.2532 -0.2172 -0.2172 -0.1470  0.8890
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.21716    0.01046  20.768  < 2e-16 ***
## factor(treat)m   0.03602    0.03786   0.951  0.34154
## visit           -0.10615    0.03673  -2.890  0.00389 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.395 on 2047 degrees of freedom
## Multiple R-squared:  0.00821,    Adjusted R-squared:  0.007241
## F-statistic: 8.473 on 2 and 2047 DF,  p-value: 0.0002165
```

- **What conclusions would you make?**

Scenario 2: Collider Bias



Scenario 2: Collider Bias

Data Generating Process:

- No direct treatment effect of mailing over emails.
- Pr(Visit | e-mail) < Pr(Visit | mail)
- People that receive a letter are much more encouraged to visit the website, and people that register are also more likely to visit the website.

Num in Sample and % of New Registrations by Type of Campaign and Visits to the Website

	Not visited	Visited	Registered - NV	Registered - V	Registered - Total
Email	1409	94	0.19	0.11	0.18
Mail	45	502	0.31	0.15	0.16

Note: Simulated data

Scenario 2: Collider Bias

What happens if we now **run a simple model**?

```
summary(lm(y ~ factor(treat), data = collider))
```

- what would you **expect** to see?

Scenario 2: Collider Bias

What happens if we now **run a simple model**?

```
summary(lm(y ~ factor(treat), data = collider))
```

```
##
## Call:
## lm(formula = y ~ factor(treat), data = collider)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.1843 -0.1843 -0.1843 -0.1627  0.8373
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.18430    0.00988  18.654   <2e-16 ***
## factor(treat)m -0.02159    0.01913  -1.129    0.259
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.383 on 2048 degrees of freedom
## Multiple R-squared:  0.0006219,	Adjusted R-squared:  0.0001339
## F-statistic: 1.274 on 1 and 2048 DF,  p-value: 0.2591
```

Scenario 2: Collider Bias

What happens if we now **control by whether the person visited the website**?

```
summary(lm(y ~ factor(treat) + visit, data = collider))
```

Scenario 2: Collider Bias

What happens if we now **control by whether the person visited the website**?

```r
summary(lm(y ~ factor(treat) + visit, data = collider))
```

```
##
## Call:
## lm(formula = y ~ factor(treat) + visit, data = collider)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.2620 -0.1911 -0.1911 -0.1538  0.9171
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.19106    0.01008  18.957   <2e-16 ***
## factor(treat)m   0.07093    0.03449   2.057   0.0398 *
## visit           -0.10819    0.03359  -3.221   0.0013 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3822 on 2047 degrees of freedom
## Multiple R-squared:  0.005661,	Adjusted R-squared:  0.004689
## F-statistic: 5.827 on 2 and 2047 DF,  p-value: 0.002997
```

Scenario 2: Collider Bias

What happens if we now **control by whether the person visited the website**?

```r
summary(lm(y ~ factor(treat) + visit, data = collider))
```

```
##
## Call:
## lm(formula = y ~ factor(treat) + visit, data = collider)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.2620 -0.1911 -0.1911 -0.1538  0.9171
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.19106    0.01008  18.957   <2e-16 ***
## factor(treat)m  0.07093    0.03449   2.057   0.0398 *
## visit           -0.10819    0.03359  -3.221   0.0013 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3822 on 2047 degrees of freedom
## Multiple R-squared:  0.005661,    Adjusted R-squared:  0.004689
## F-statistic: 5.827 on 2 and 2047 DF,  p-value: 0.002997
```

What happened here?

Avoiding biases in Causal Inference

- Always check your data!

Avoiding biases in Causal Inference

- Always check your data!
- Assess the plausibility of the ignorability assumption

Avoiding biases in Causal Inference

- Always check your data!
- Assess the plausibility of the ignorability assumption
- The model you have in your head matters!

Avoiding biases in Causal Inference

- Always check your data!
- Assess the plausibility of the ignorability assumption
- The model you have in your head matters!
- Avoid controlling for ex-post variables.

Another Example: Beauty in the Classroom

- Data: Student's evaluations for instructors at UT Austin

```
profs <- read.csv("https://raw.githubusercontent.com/maibennett/sta235/main/exampleSite/content/Classes/Week3/2_Potential(
                  stringsAsFactors = TRUE)

head(profs)
```

```
##   minority age gender credits     beauty eval division native tenure students
## 1      yes  36 female    more  0.2899157  4.3    upper    yes    yes       24
## 2       no  59   male    more -0.7377322  4.5    upper    yes    yes       17
## 3       no  51   male    more -0.5719836  3.7    upper    yes    yes       55
## 4       no  40 female    more -0.6779634  4.3    upper    yes    yes       40
## 5       no  31 female    more  1.5097940  4.4    upper    yes    yes       42
## 6       no  62   male    more  0.5885687  4.2    upper    yes    yes      182
##   allstudents prof
## 1          43    1
## 2          20    2
## 3          55    3
## 4          46    4
## 5          48    5
## 6         282    6
```

Beauty and Evaluations

- **Causal Question:** What is the effect of beauty on teachers evaluations?

Beauty and Evaluations

- Causal Question: What is the effect of beauty on teachers evaluations?

```
summary(lm(eval ~ beauty, data=profs))
```

```
##
## Call:
## lm(formula = eval ~ beauty, data = profs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.80015 -0.36304  0.07254  0.40207  1.10373
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.99827    0.02535 157.727  < 2e-16 ***
## beauty       0.13300    0.03218   4.133 4.25e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5455 on 461 degrees of freedom
## Multiple R-squared:  0.03574,    Adjusted R-squared:  0.03364
## F-statistic: 17.08 on 1 and 461 DF,  p-value: 4.247e-05
```

Clearly not causal

Beauty and Evaluations

- Causal Question: What is the effect of beauty on teachers evaluations?
- What other things could be biasing our estimate?
  - Distinction between what's in our data vs what it's not.

Let's check our data

- Simplify the problem:
  - **Binary Treatment**: Beauty above average (1) vs Below average (0)

```
profs <- profs %>% mutate(treat = as.numeric(beauty > 0),
                          female = 2 - as.numeric(gender),
                          single_credit = as.numeric(credits)-1,
                          upper_div = as.numeric(division)-1,
                          native = as.numeric(native)-1,
                          tenure = as.numeric(tenure)-1,
                          minority = as.numeric(minority)-1)
```

Let's check our data

- Simplify the problem:
    - Binary Treatment: Beauty above average (1) vs Below average (0)

```
profs <- profs %>% mutate(treat = as.numeric(beauty > 0),
                          female = 2 - as.numeric(gender),
                          single_credit = as.numeric(credits)-1,
                          upper_div = as.numeric(division)-1,
                          native = as.numeric(native)-1,
                          tenure = as.numeric(tenure)-1,
                          minority = as.numeric(minority)-1)

library(modelsummary)

covs <- profs %>% select(treat, minority, age, female, single_credit, upper_div,
        native, tenure, students, allstudents)

datasummary_balance(~ treat, data = covs, title = "Balance table", fmt=2, dinm_statistic = "p.value")
```

Let's check our data

- Simplify the problem:
  - **Binary Treatment**: Beauty above average (1) vs Below average (0)

```
profs <- profs %>% mutate(treat = as.numeric(beauty > 0),
                          female = 2 - as.numeric(gender),
                          single_credit = as.numeric(credits)-1,
                          upper_div = as.numeric(division)-1,
                          native = as.numeric(native)-1,
                          tenure = as.numeric(tenure)-1,
                          minority = as.numeric(minority)-1)

library(modelsummary)

covs <- profs %>% select(treat, minority, age, female, single_credit, upper_div,
         native, tenure, students, allstudents)

datasummary_balance(~ treat, data = covs, title = "Balance table", fmt=2, dinm_statistic = "p.value")
```

Let's check our data

```r
datasummary_balance(~ treat, data = covs, title = "Balance table", fmt=2, dinm_statistic = "p.value")
```

Balance table

	0		1	
	Mean	Std. Dev.	Mean	Std. Dev.
minority	-1.88	0.33	-1.84	0.37
age	50.56	9.44	45.12	9.44
female	0.37	0.48	0.50	0.50
single_credit	0.08	0.27	0.03	0.16
upper_div	0.62	0.49	0.72	0.45
native	-1.05	0.22	-1.07	0.26
tenure	-1.21	0.41	-1.23	0.42
students	30.98	27.91	44.96	61.36
allstudents	47.27	49.84	66.85	100.48

Let's check our data... now with a Love Plot!

```r
# Reads a user-written function to generate a loveplot
source("https://raw.githubusercontent.com/maibennett/sta235/main/exampleSite/content/Classes/Week3/2_PotentialOutcomes/coc

treat_id <- profs %>% mutate(id = seq(1, nrow(profs))) %>% filter(treat==1) %>% pull(-1)
control_id <- profs %>% mutate(id = seq(1, nrow(profs))) %>% filter(treat==0) %>% pull(-1)

loveplot_balance(covs, treat_id, control_id, v_line = 0.05, format = TRUE)
```

Is it enough to control?

- We can use the covariates we have on our dataset to **control for those group differences**.

	Model 1	Model 2
(Intercept)	3.998***	4.070***
	(0.025)	(0.245)
beauty	0.133***	0.141***
	(0.032)	(0.033)
minority		-0.072
		(0.077)
age		-0.003
		(0.003)
gendermale		0.221***
		(0.053)
divisionupper		-0.094+
		(0.056)
native		0.253*
		(0.110)
tenure		-0.145*
		(0.062)
allstudents		0.000
		(0.000)
Num.Obs.	463	463
F	17.085	7.193

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Beauty coeff. is consistent across models

Other covariates also matter

Is it enough to control?

- We can use the covariates we have on our dataset to control for those group differences.

Is that enough?

What other variable could be confounding our effect?

If I told you professors in the treatment group are taller than the ones in the control group, is height a confounder?

What about self-esteem?

Answering the question

How would you answer this question?
Design a study!

Can you "randomize" beauty?

Main takeaway points

Causal Inference is hard

Main takeaway points

Causal Inference is hard

- Think about the **causal problem**

Main takeaway points

Causal Inference is hard

- Think about the **causal problem**
- Always **look at your data**

Main takeaway points

Causal Inference is hard

- Think about the causal problem
- Always look at your data
- Check validity of assumptions (*Is ignorability plausible? Am I controlling for the right covariates?*)

Main takeaway points

Causal Inference is hard

- Think about the causal problem
- Always look at your data
- Check validity of assumptions (*Is ignorability plausible? Am I controlling for the right covariates?*)
- Most of this chapter will be spent on looking for exogeneous variation to make the ignorability assumption happen.

References

- Angrist, J. & S. Pischke. (2015). "Mastering Metrics". *Chapter 1*.
- Cunningham, S. (2021). "Causal Inference: The Mixtape". *Chapter 4: Potential Outcomes Causal Model*.
- Neil, B. (2020). "Introduction to Causal Inference". *Fall 2020 Course*