Business Experimentation and Causal Methods

Prof. Fradkin

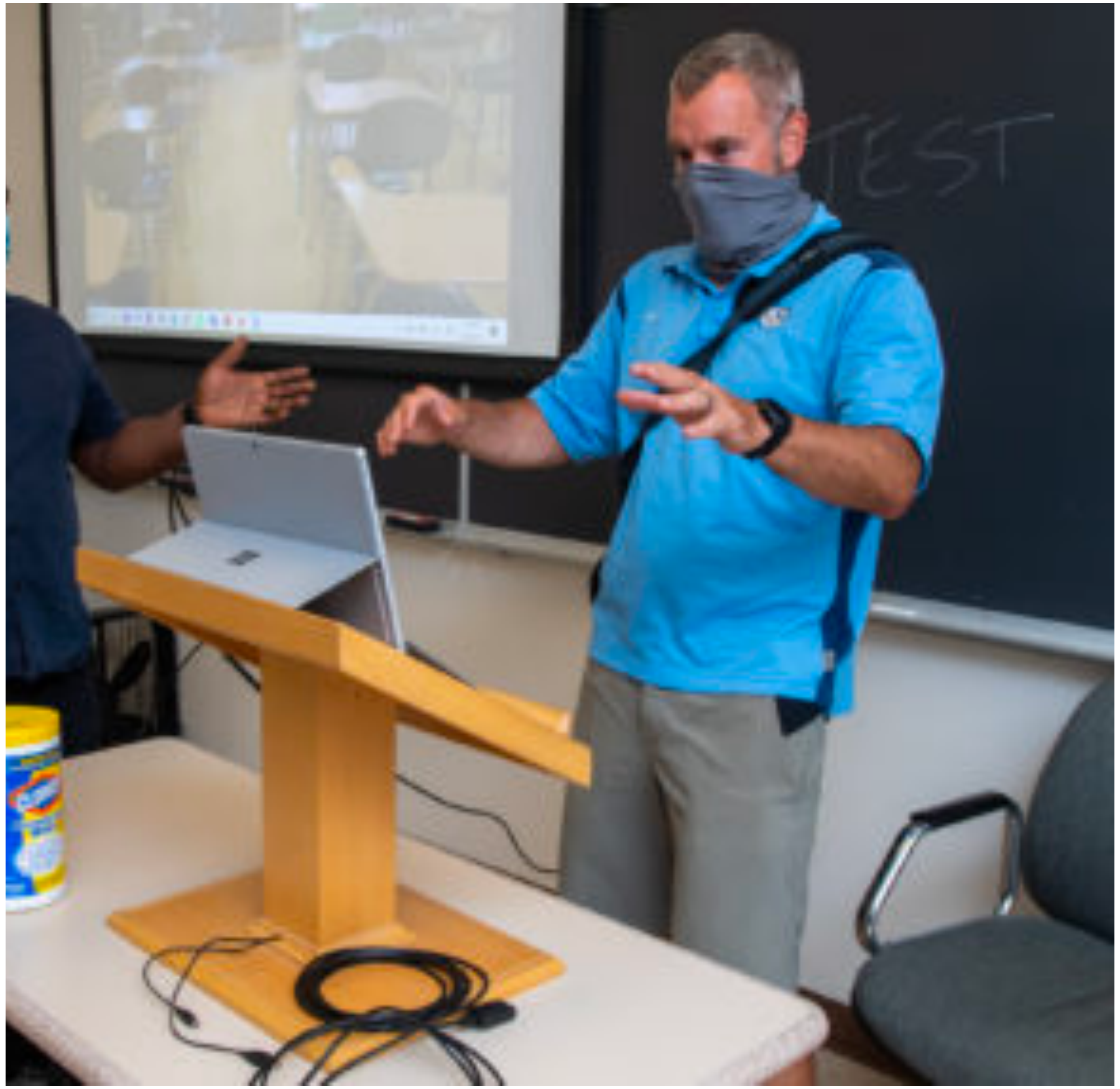
Topic: Potential Outcomes and Selection Bias

Outline

1. **Example of a causal problem**
2. Potential outcomes
3. Average treatment effects
4. Selection bias

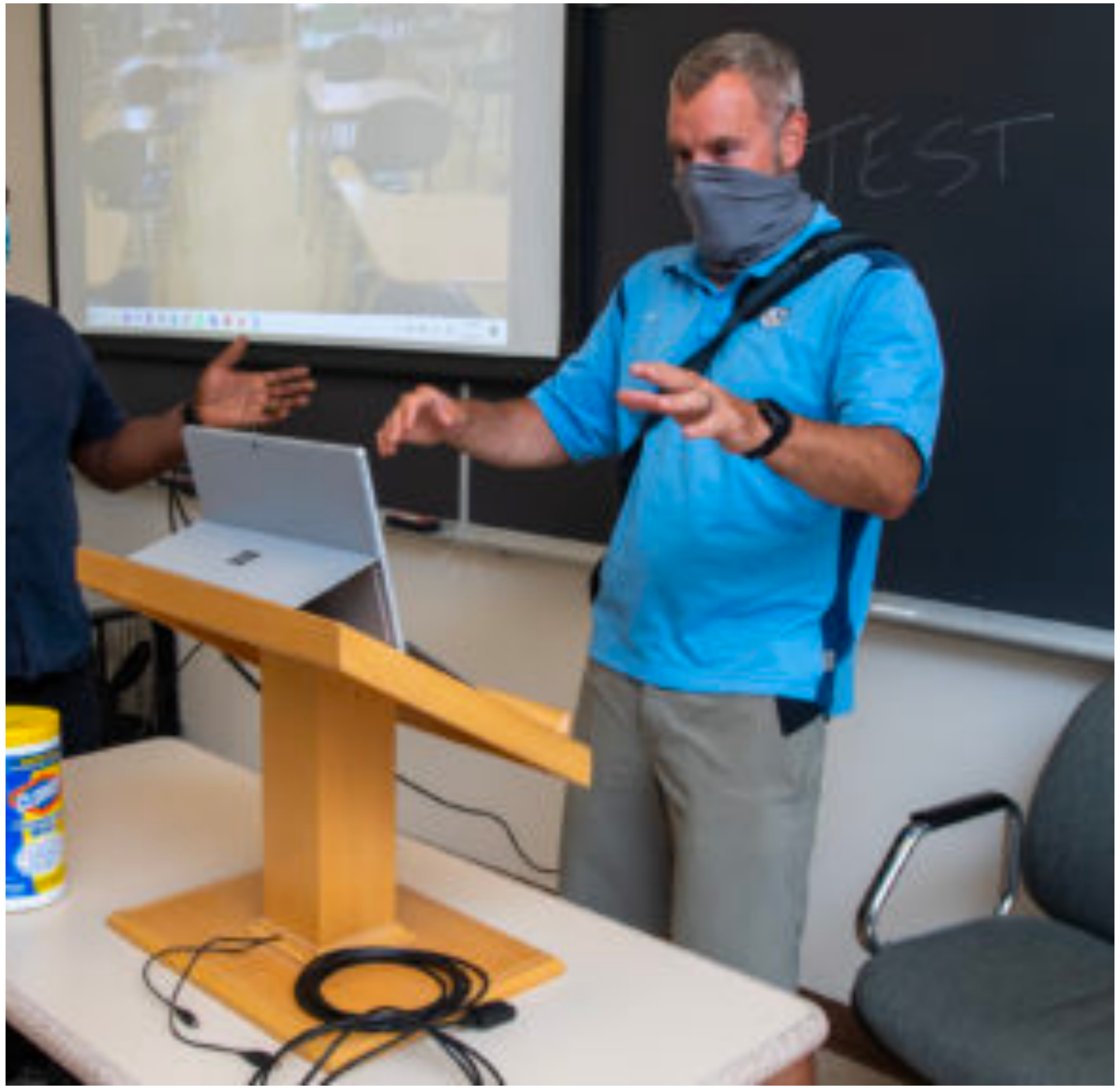
Our Example Causal Problem

- Start with practical question: should you go to class in person or via Zoom?
- What might we care about? Learning, convenience, and safety.
- **Does attending class in-person increase learning?**
- This is close to a causal problem, but… We need to be more precise.



Making a Causal Problem Precise

- What is the treatment exactly?
  - Let’s say attending every class in-person.
- What is the alternative to going to class in-person?
  - Let’s say the alternative is attending Zoom for every class instead.
- What do we mean by learning?
  - Let’s say getting an A vs getting less than an A.



Why do we want this answer?

- If you, as a student, commit to going to every class in-person, will you learn more than if you went to every class via Zoom?
- If you learn a lot more, you might choose to come to class in-person. Otherwise you might not.
- It might also inform your decision to come to class in-person sometimes.

Outline

1. Example of a causal problem
2. **Potential outcomes**
3. Average treatment effects
4. Selection bias

Potential Outcomes: An Illustration

Also known as the Rubin Causal Model

Hypothetical situation:

Person	Grade if in-person	Grade if on Zoom
John	A	A
Mary	B	A
Suraj	B	B
Katerina	A	A
Molly	B	A
Leroy	B	B

We can never observe this in real life!

You can’t attend remotely and in-person at the same time

Hypothetical situation:

Person	In-person	Zoom	Treatment Effect (Impact)
John	A	A	-
Mary	B	A	Worse grade
Suraj	B	B	-
Katerina	A	A	-
Molly	B	A	Worse grade
Leroy	B	B	-

It helps us to use numbers: A = 1, B = 0

Person	In-person	Zoom	Treatment Effect
John	1	1	0
Mary	0	1	-1
Suraj	0	0	0
Katerina	1	1	0
Molly	0	1	-1
Leroy	0	0	0

Outline

1. Example of a causal problem
2. Potential outcomes
3. **Average treatment effects**
4. Selection bias

The Average Treatment Effect

- As an individual student, you would like to know your individual treatment effect.
- As the university, you may want to know the average effect across all students.
- If the average effect is positive, then the university may want to mandate in-person attendance and vice versa if negative.

The Average Treatment Effect

- Can calculate the average of the treatment effects.
- Can also calculate the average of the in-person potential outcomes and the average of the zoom potential outcomes.
  - Then take the difference.

Person	In-person	Zoom	Treatment Effect
John	1	1	0
Mary	0	1	-1
Suraj	0	0	0
Katerina	1	1	0
Molly	0	1	-1
Leroy	0	0	0
Average	1/3	2/3	-1/3

The Conditional Average Treatment Effect

- Sometimes we may want to know about the average treatment effect for certain types of people. **This is called the conditional average treatment effect.**

- For example, let's say that we want to find the average treatment effect just for women.

- Then all we have to do is look at the subset of individuals who are women, and calculate the average treatment effect.

- This is the average treatment effect **conditional** on being a woman.

Person	In-person	Zoom	Treatment Effect
Mary	0	1	-1
Katerina	1	1	0
Molly	0	1	-1
Average	1/3	1	-2/3

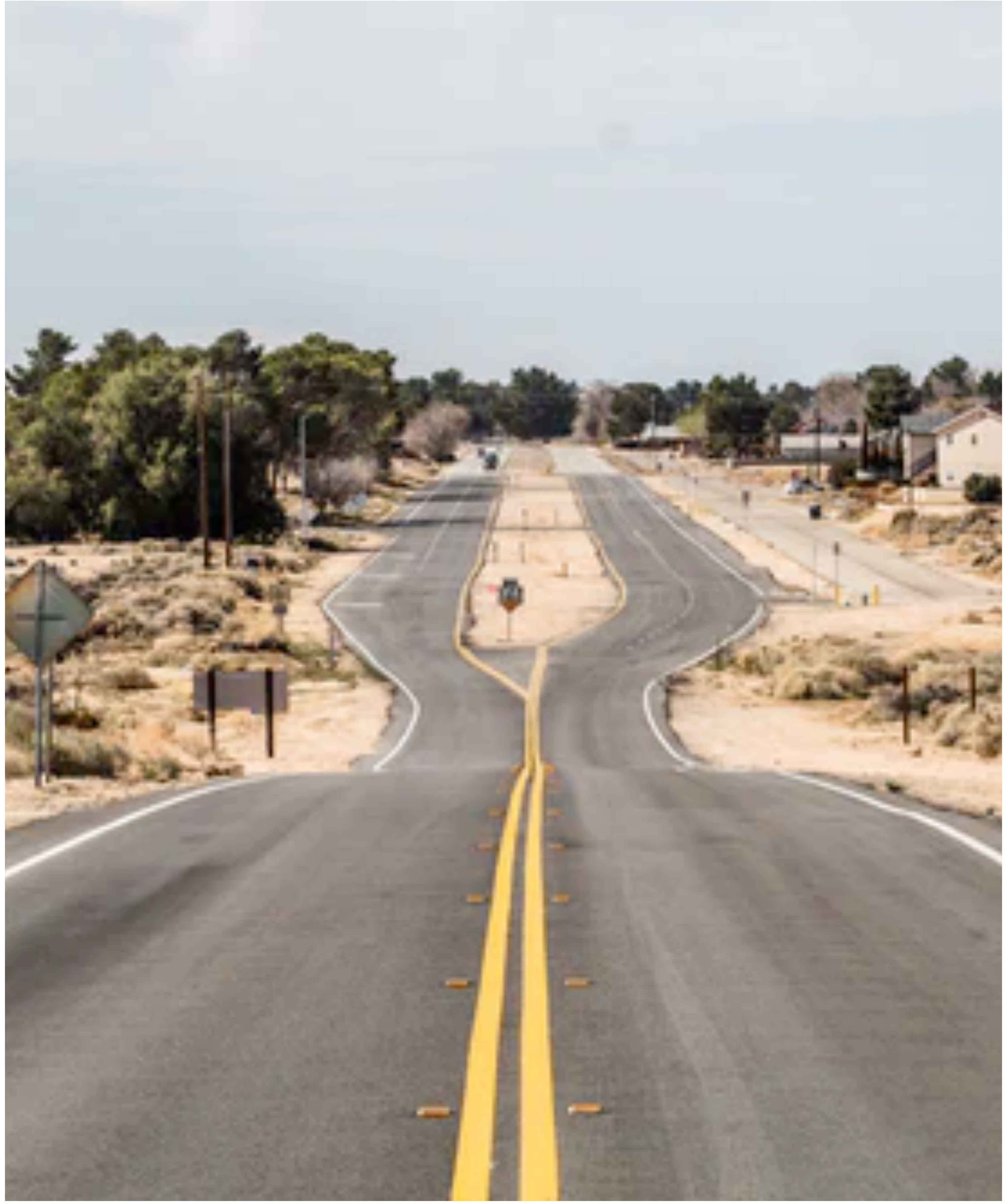
Outline

1. Example of a causal problem
2. Potential outcomes
3. Average treatment effects
4. **Selection bias**

What happens in the real world?

Some students choose to come in-person and others don’t.

Can we compare the grades for those who do come with those who don’t?



John, Katerina, and Molly choose in-person

- Note that we can't see the potential outcome if Zoom for these 3.
- Maybe they chose to come in person because they are **more motivated** and would have **better grades regardless of the treatment.**

Person	In-person	Zoom	Impact
John	1	???	???
Katerina	1	???	???
Molly	0	???	???

Suraj, Leroy, and Mary choose zoom

- Note that we can't see the potential outcome if in-person for these 3.

Person	In-person	Zoom	Impact
Mary	???	1	???
Suraj	???	0	???
Leroy	???	0	???

What do we get if we compare outcomes?

- Take difference in average outcomes between treatment and control group.

$$\frac{2}{3} - \frac{1}{3} = \frac{1}{3}$$

- Recall, the true average treatment effect is:

$$\frac{-1}{3}$$

- Why is there a problem?

Person	In-person	Zoom	Impact
John	1	???	???
Katerina	1	???	???
Molly	0	???	???
Average	2/3	???	???

Person	In-person	Zoom	Impact
Mary	???	1	???
Suraj	???	0	???
Leroy	???	0	???
Average	???	1/3	???

Selection bias

- Difference in potential outcomes between the treatment and control group, if not treated.
- John, Katerina, and Molly all chose to attend class in-person and would have gotten an A if they had used Zoom.
- Mary, Suraj, and Leroy all chose to not attend class in-person and only one of them got an A.
- Difference in potential outcomes if not treated (Zoom) is selection bias.

Person	In-person	Zoom
John	1	1
Katerina	1	1
Molly	0	1

3/3 - 1/3 = 2/3
Selection Bias

Person	In-person	Zoom
Mary	0	1
Suraj	0	0
Leroy	0	0

Difference in outcomes = Impact + Selection Bias

Person	In-person	Zoom	Impact
John	1	1	0
Katerina	1	1	0
Molly	0	1	-1
Average	2/3	1	-1/3

$$\frac{1}{3} = -\frac{1}{3} + \frac{2}{3}$$

Person	In-person	Zoom	Impact
Mary	0	1	-1
Suraj	0	0	0
Leroy	0	0	0
Average	0	1/3	-1/3

But we want impact not selection bias!

Potential outcomes can also be continuous

Person	In-person	Zoom	Treatment Effect
John	100	80	
Mary	80	95	
Suraj	100	100	
Katerina	90	90	
Molly	90	95	
Leroy	95	95	

What are the ATE and ATE estimate if John and Mary are in the treatment, while the rest in the control?

Person	In-person	Zoom	Treatment Effect
John	100	80	
Mary	80	95	
Suraj	100	100	
Katerina	90	90	
Molly	90	95	
Leroy	95	95	

Answers

Person	In-person	Zoom	Treatment Effect
John	100	80	20
Mary	80	95	-15
Suraj	100	100	0
Katerina	90	90	0
Molly	90	95	-5
Leroy	95	95	0

- ATE = 0
- $\widehat{ATE}$ = (100 + 80)/2 - (100 + 90 + 95 + 95)/4
- = 90 - 95 = -5.

Summary

1. Example of a causal problem
2. Potential outcomes
3. Average treatment effects
4. Selection bias

Next Lecture:
Statistics Refresher and Randomization