Social Data Science

SOCIOL 114
Winter 2025

Causal inference: Connections to statistical modeling

Learning goals for today

By the end of class, you will be able to

- connect causal inference (a missing data problem)

  to statistical modeling (predicting missing data)

A running example

I feel confident that I can answer quantitative questions with tools from data science.

- 1 = Agree
- 0 = Disagree

A running example

I feel confident that I can answer quantitative questions with tools from data science.

- $1 =$ Agree
- $0 =$ Disagree

What is the average causal effect of taking this class on confidence in data science skills?

Using potential outcomes

Each Row is a Student in This Class	Outcome under 114	Outcome under no 114
	$Y_1^{\text{Takes 114}}$	$Y_1^{\text{No 114}}$
	$Y_2^{\text{Takes 114}}$	$Y_2^{\text{No 114}}$
	$Y_3^{\text{Takes 114}}$	$Y_3^{\text{No 114}}$
	$Y_4^{\text{Takes 114}}$	$Y_4^{\text{No 114}}$
	$Y_5^{\text{Takes 114}}$	$Y_5^{\text{No 114}}$
	$Y_6^{\text{Takes 114}}$	$Y_6^{\text{No 114}}$

$Y =$ I feel confident that I can answer quantitative questions with tools from data science

Using potential outcomes

Each Row is a Student in This Class	Outcome under 114	Outcome under no 114
1	$Y_1^{\text{Takes 114}}$	?
2	$Y_2^{\text{Takes 114}}$	?
3	$Y_3^{\text{Takes 114}}$	?
4	$Y_4^{\text{Takes 114}}$	?
5	$Y_5^{\text{Takes 114}}$	?
6	$Y_6^{\text{Takes 114}}$	?

$Y =$ I feel confident that I can answer quantitative questions with tools from data science

Using potential outcomes

Each Row is a Student in This Class	Outcome under 114	Outcome under no 114
1	$Y_1^{\text{Takes 114}}$	?
2	$Y_2^{\text{Takes 114}}$	?
3	$Y_3^{\text{Takes 114}}$	?
4	$Y_4^{\text{Takes 114}}$	?
5	$Y_5^{\text{Takes 114}}$	?
6	$Y_6^{\text{Takes 114}}$	?

$Y =$ I feel confident that I can answer quantitative questions with tools from data science

How could we learn about the (?)

Strategy 1: A subgroup with conditional exchangeability

Strategy 1: A subgroup with conditional exchangeability

- Some of the class was on the waitlist
  - some got in
  - others didn't

Strategy 1: A subgroup with conditional exchangeability

Y = I feel confident that I can answer quantitative questions with tools from data science

Each Row is a Student in This Class

$Y_1^{\text{Takes 114}}$	?
$Y_2^{\text{Takes 114}}$	?
$Y_3^{\text{Takes 114}}$	?
$Y_4^{\text{Takes 114}}$	?
?	$Y_5^{\text{No 114}}$
?	$Y_6^{\text{No 114}}$
?	$Y_7^{\text{No 114}}$
?	$Y_8^{\text{No 114}}$

Strategy 1: A subgroup with conditional exchangeability

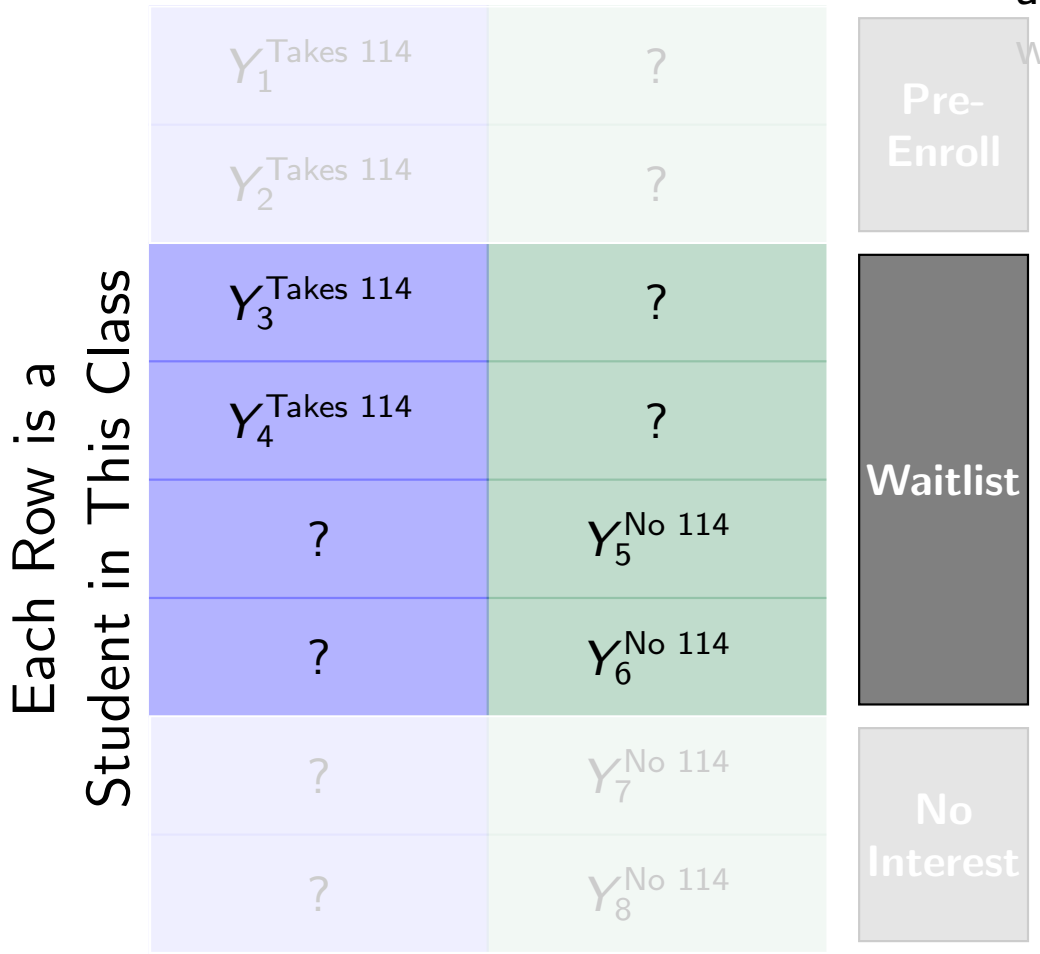
Y = I feel confident that I can answer quantitative questions with tools from data science

Each Row is a Student in This Class

$Y_1^{\text{Takes 114}}$	?	Pre-Enroll
$Y_2^{\text{Takes 114}}$	?	Pre-Enroll
$Y_3^{\text{Takes 114}}$	?	Waitlist
$Y_4^{\text{Takes 114}}$	?	Waitlist
?	$Y_5^{\text{No 114}}$	Waitlist
?	$Y_6^{\text{No 114}}$	Waitlist
?	$Y_7^{\text{No 114}}$	No Interest
?	$Y_8^{\text{No 114}}$	No Interest

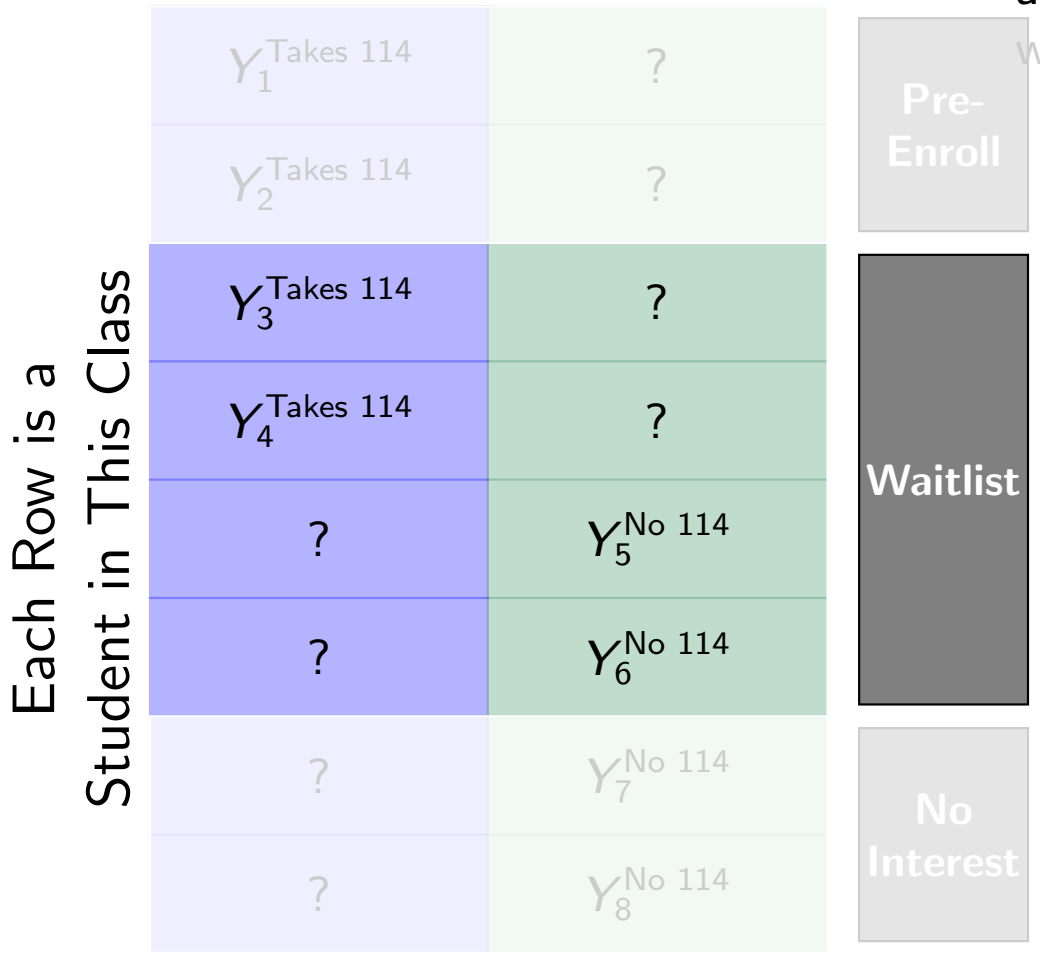
Strategy 1: A subgroup with conditional exchangeability

Y = I feel confident that I can answer quantitative questions with tools from data science



Strategy 1: A subgroup with conditional exchangeability

Y = I feel confident that I can answer quantitative questions with tools from data science

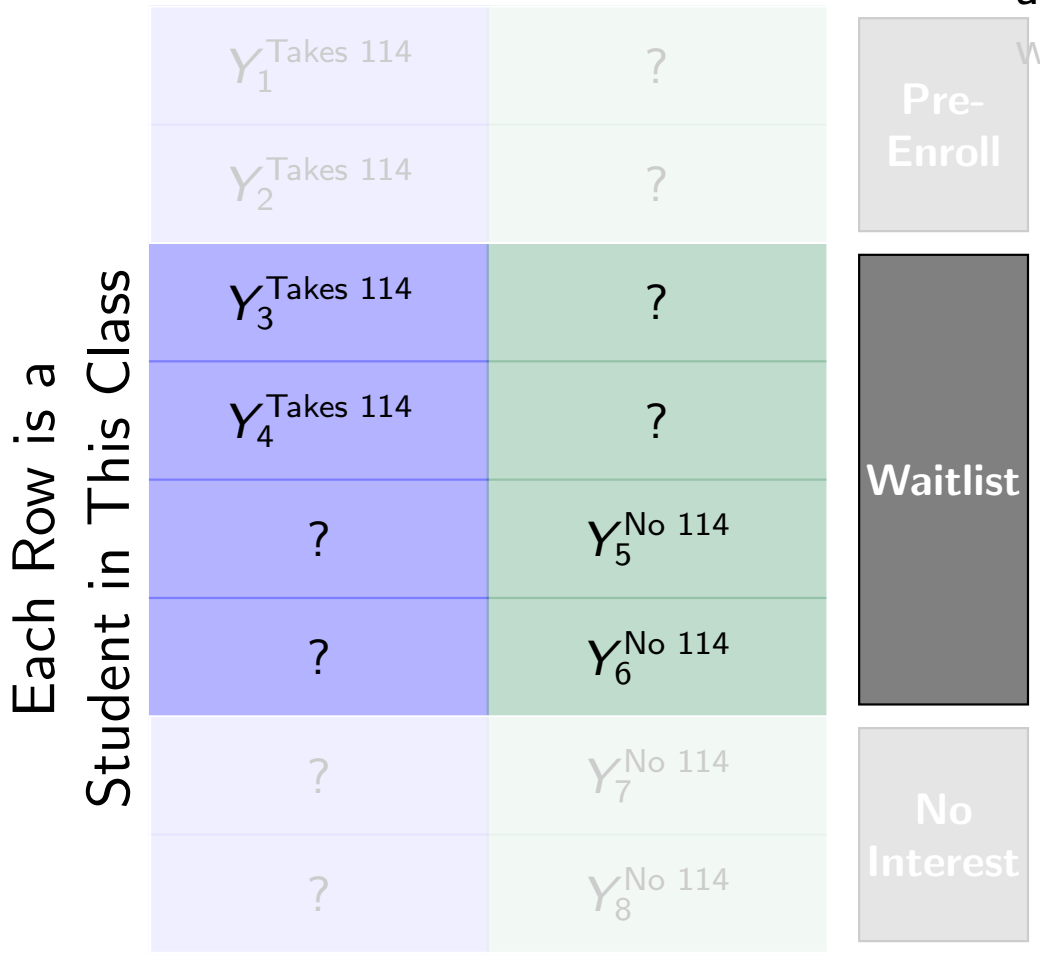


Benefits of strategy

Drawbacks

Strategy 1: A subgroup with conditional exchangeability

Y = I feel confident that I can answer quantitative questions with tools from data science



Benefits of strategy

Credible

Drawbacks

Limited target population

Strategy 2: Adjust for measured confounders

Strategy 2: Adjust for measured confounders

For each of you, we could compare

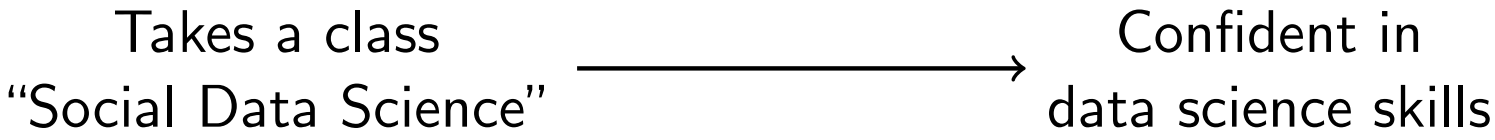
1. your opinion after 114
2. the average opinion of non-114 students who look like you

Strategy 2: Adjust for measured confounders

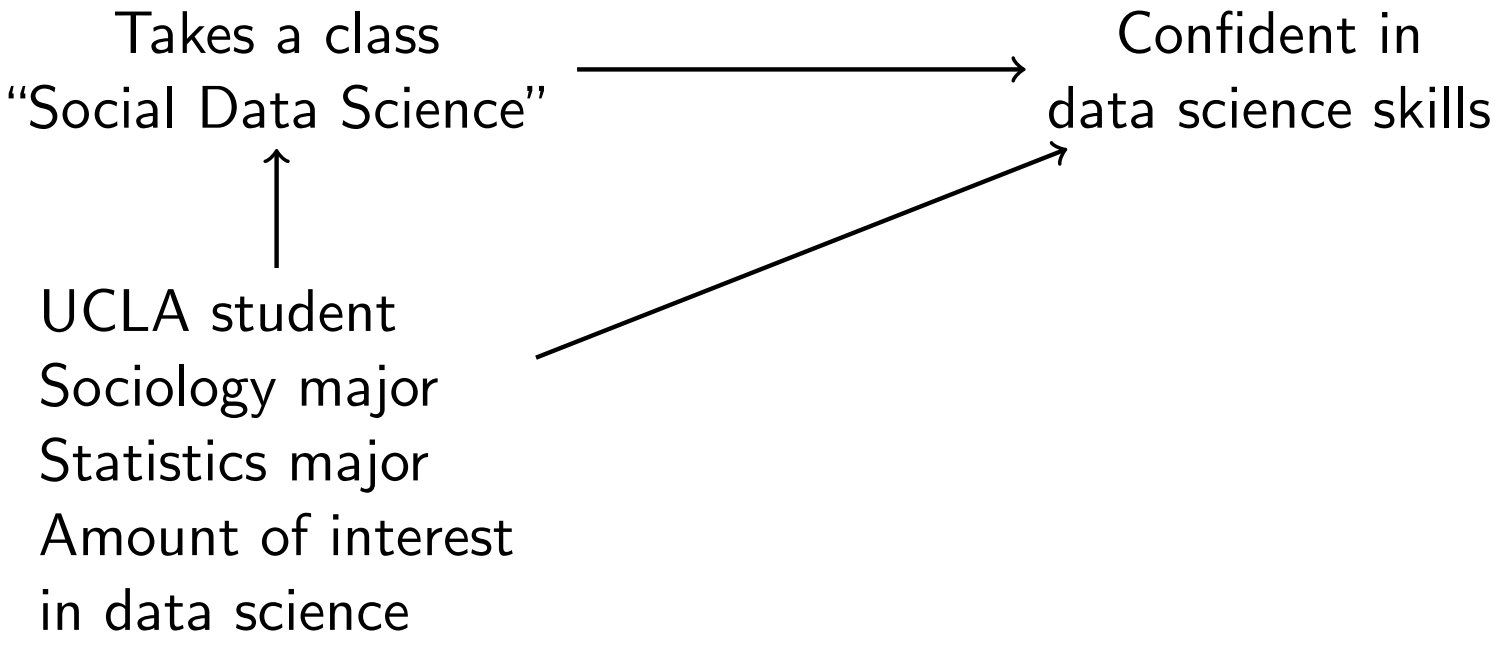
For each of you, we could compare

1. your opinion after 114
2. the average opinion of non-114 students who look like you

Looks like you in what ways? What else belongs in this DAG?



Strategy 2: Adjust for measured confounders



Suppose these are a sufficient adjustment set.

Strategy 2: Adjust for measured confounders

Nonparametric estimation:

For each student in the class, find someone else who

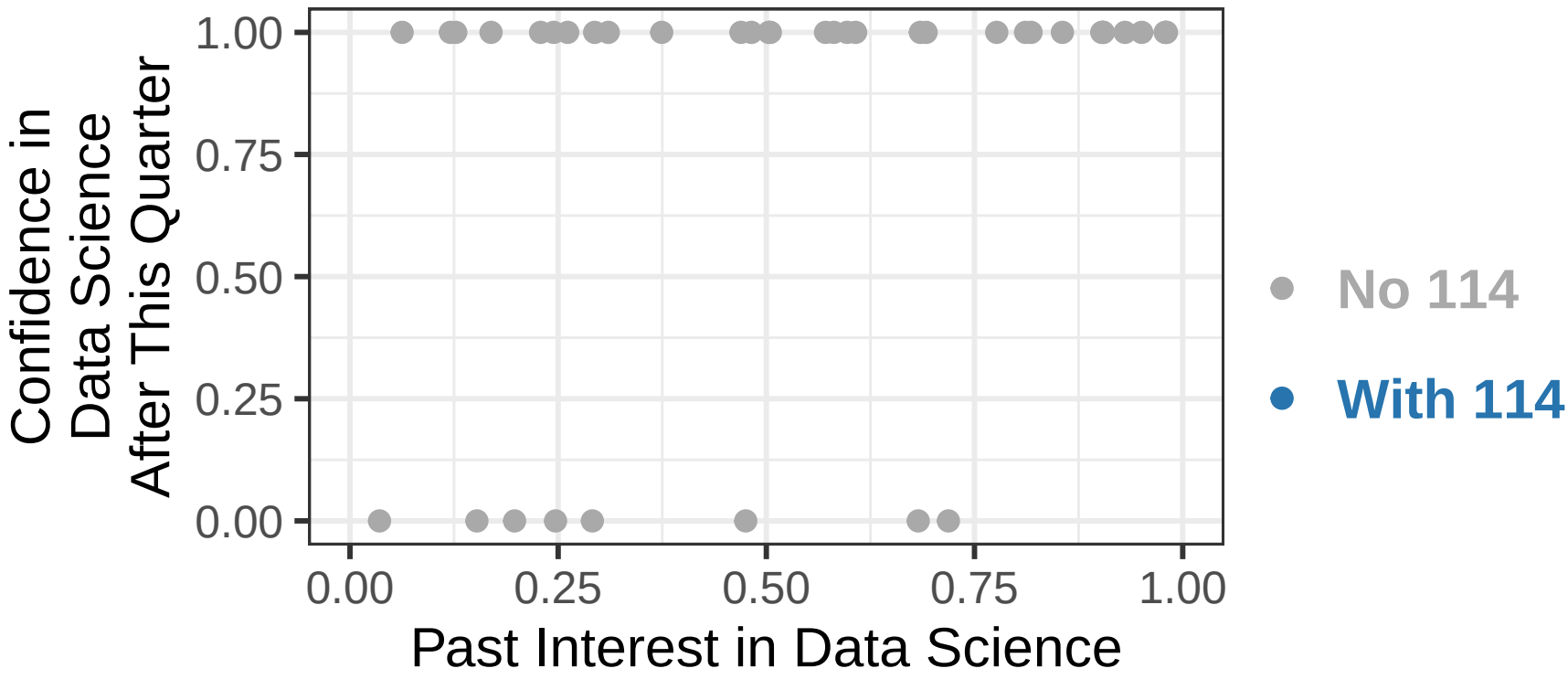
- is a student at UCLA
- shares your major
- is exactly as interested in data science as you are
- but did not take this class

Use your **match** to infer your $Y_i^{\text{No 114}}$ for people like you:

$$\mathrm{E}(Y^0 \mid \vec{X} = \vec{x}_i) = \underbrace{\mathrm{E}(Y \mid A = 0, \vec{X} = \vec{x}_i)}_{\text{estimated from your match}}$$

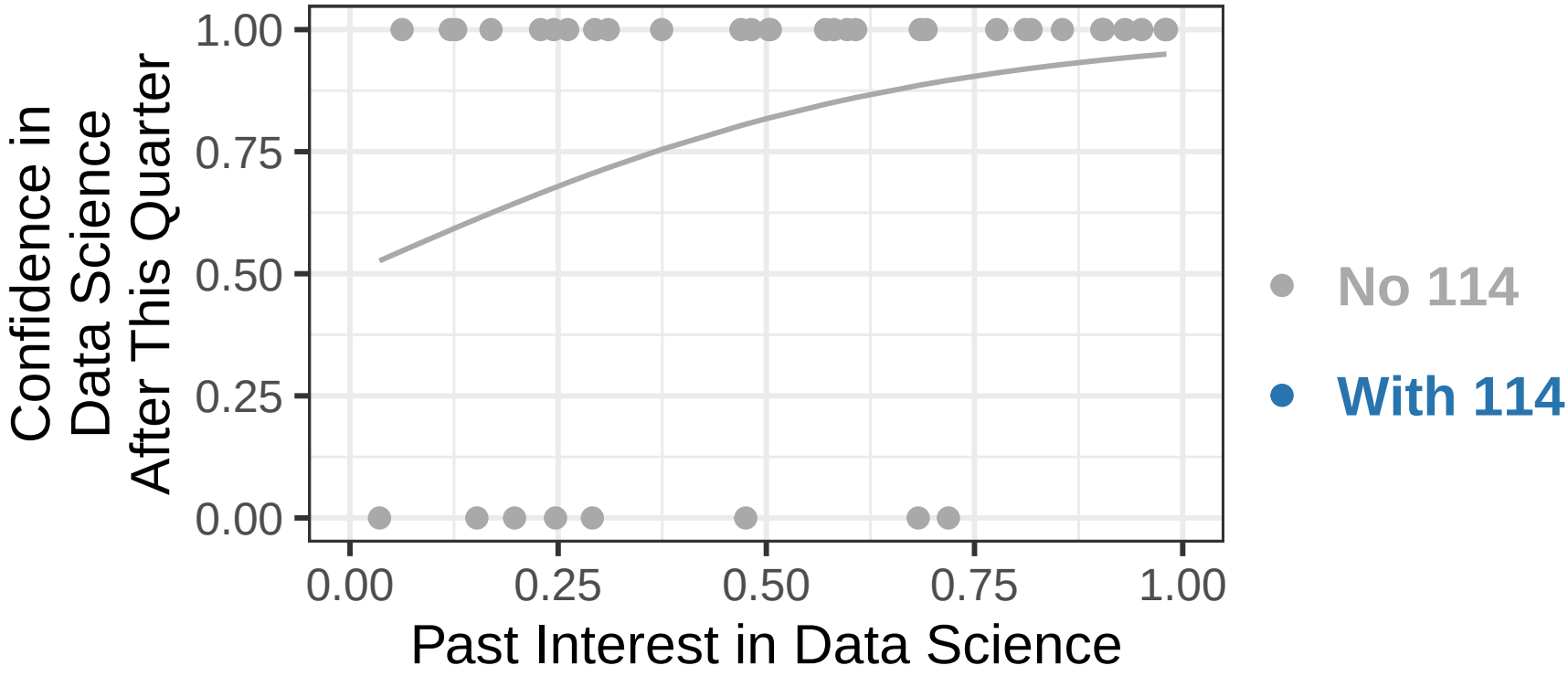
since we have assumed conditional exchangeability given $\vec{X}$.

Generalizing to a model



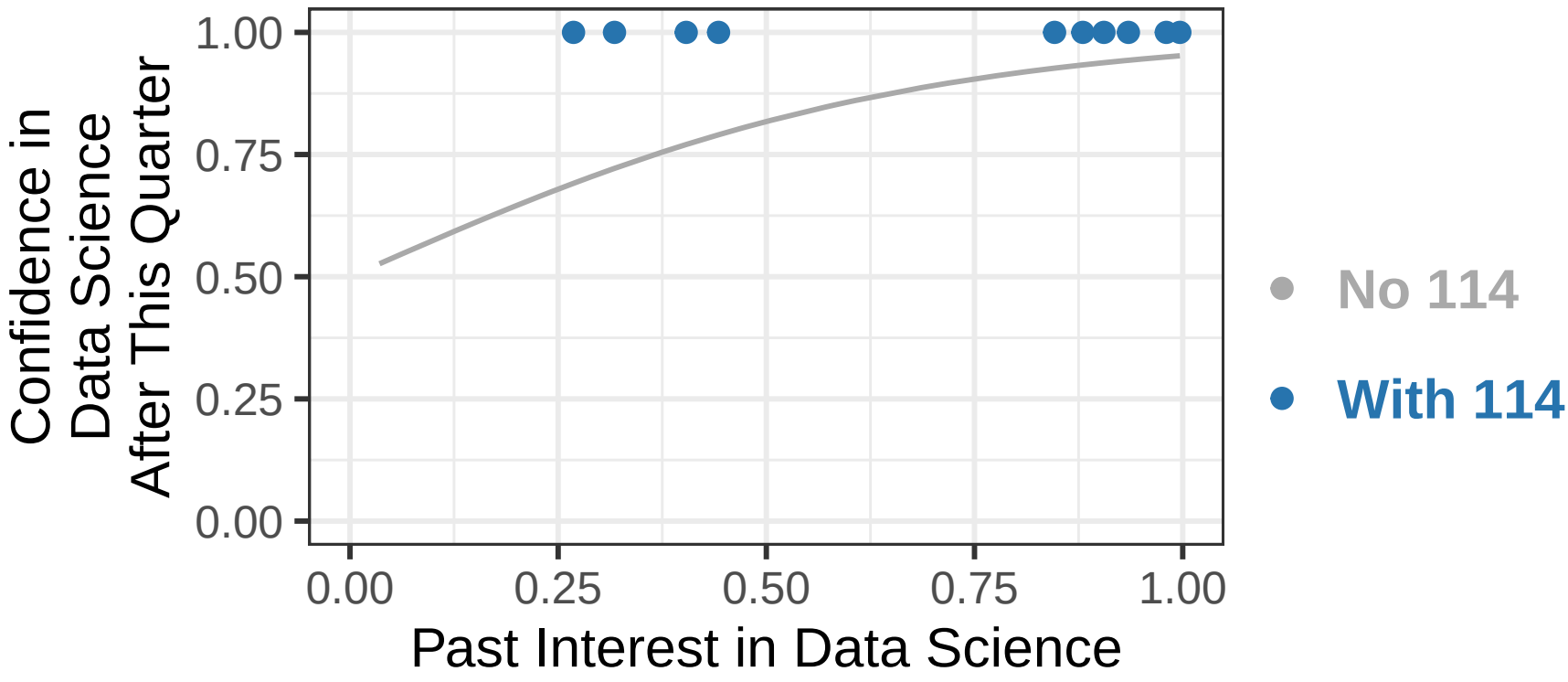
1) Find control units who didn't take this class

Generalizing to a model



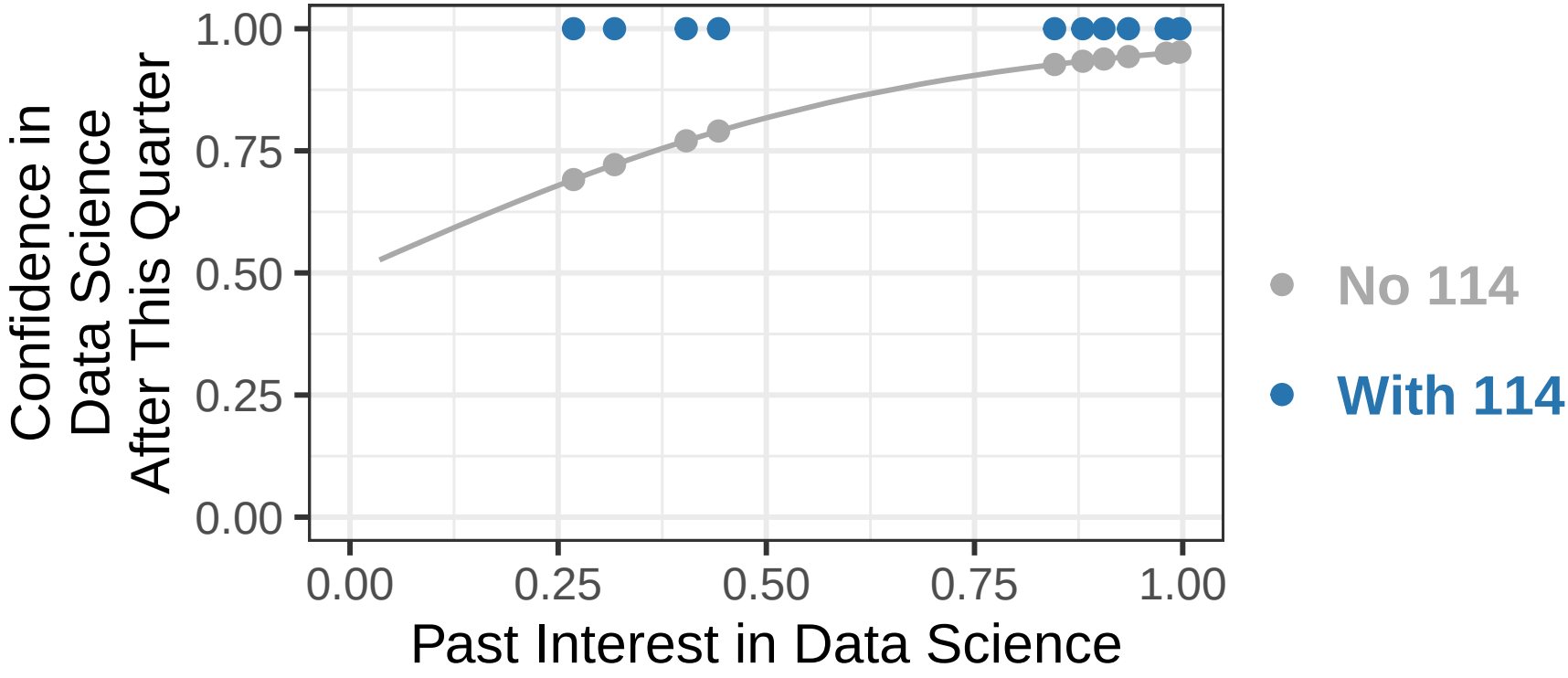
2) Model their outcomes given pre-treatment variables

Generalizing to a model



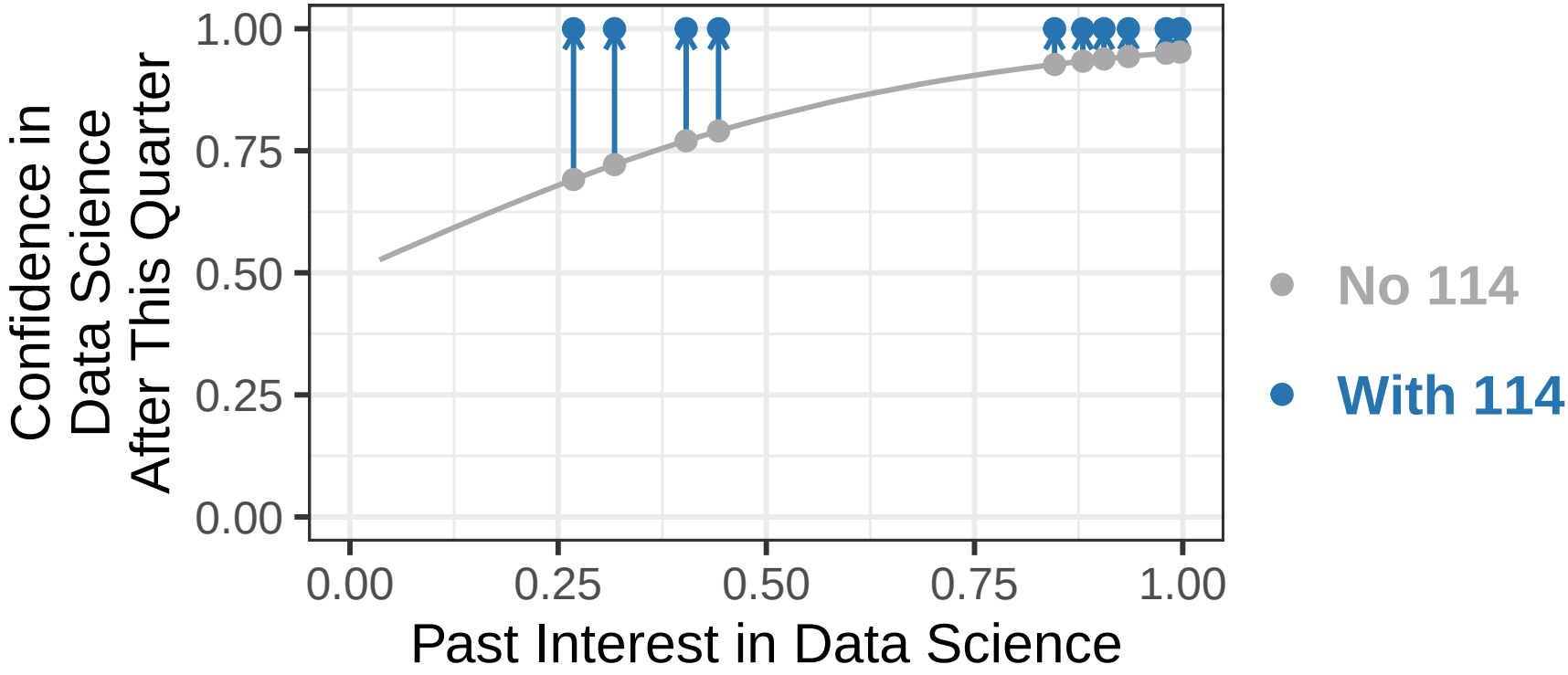
3) Find the treated units of interest

Generalizing to a model



4) Predict their counterfactual outcomes

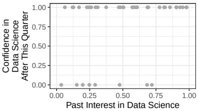
Generalizing to a model



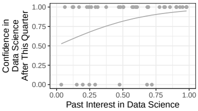
5) Infer causal effect for each person. Average over people

Strategy 2: Generalizing to a model

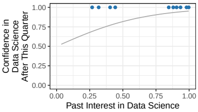
1) Find control units who didn't take this class



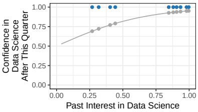
2) Model their outcomes given pre-treatment variables



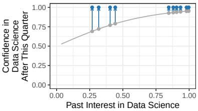
3) Find the treated units of interest



4) Predict their counterfactual outcomes



5) Infer causal effect for each person. Average over people



Summary: Outcome model for causal inference

Each Row is a Student in This Class

Outcome under 114	Outcome under no 114
$Y_1^{\text{Takes 114}}$	?
$Y_2^{\text{Takes 114}}$	?
$Y_3^{\text{Takes 114}}$	?
$Y_4^{\text{Takes 114}}$	?
$Y_5^{\text{Takes 114}}$	?
$Y_6^{\text{Takes 114}}$	?

Summary: Outcome model for causal inference

Each Row is a Student in This Class

$Y_1^{\text{Takes 114}}$	$\hat{Y}_1^{\text{No 114}}$
$Y_2^{\text{Takes 114}}$	$\hat{Y}_2^{\text{No 114}}$
$Y_3^{\text{Takes 114}}$	$\hat{Y}_3^{\text{No 114}}$
$Y_4^{\text{Takes 114}}$	$\hat{Y}_4^{\text{No 114}}$
$Y_5^{\text{Takes 114}}$	$\hat{Y}_5^{\text{No 114}}$
$Y_6^{\text{Takes 114}}$	$\hat{Y}_6^{\text{No 114}}$
Outcome under 114	Outcome under no 114

Summary: Outcome model for causal inference

Each Row is a Student in This Class

Outcome under 114	Outcome under no 114
$Y_1^{\text{Takes 114}}$	$\hat{Y}_1^{\text{No 114}}$
$Y_2^{\text{Takes 114}}$	$\hat{Y}_2^{\text{No 114}}$
$Y_3^{\text{Takes 114}}$	$\hat{Y}_3^{\text{No 114}}$
$Y_4^{\text{Takes 114}}$	$\hat{Y}_4^{\text{No 114}}$
$Y_5^{\text{Takes 114}}$	$\hat{Y}_5^{\text{No 114}}$
$Y_6^{\text{Takes 114}}$	$\hat{Y}_6^{\text{No 114}}$

General approach

Summary: Outcome model for causal inference

Each Row is a Student in This Class

Outcome under 114	Outcome under no 114
$Y_1^{\text{Takes 114}}$	$\hat{Y}_1^{\text{No 114}}$
$Y_2^{\text{Takes 114}}$	$\hat{Y}_2^{\text{No 114}}$
$Y_3^{\text{Takes 114}}$	$\hat{Y}_3^{\text{No 114}}$
$Y_4^{\text{Takes 114}}$	$\hat{Y}_4^{\text{No 114}}$
$Y_5^{\text{Takes 114}}$	$\hat{Y}_5^{\text{No 114}}$
$Y_6^{\text{Takes 114}}$	$\hat{Y}_6^{\text{No 114}}$

General approach

1) Define potential outcomes

Summary: Outcome model for causal inference

Each Row is a Student in This Class

Outcome under 114	Outcome under no 114
$Y_1^{\text{Takes 114}}$	$\hat{Y}_1^{\text{No 114}}$
$Y_2^{\text{Takes 114}}$	$\hat{Y}_2^{\text{No 114}}$
$Y_3^{\text{Takes 114}}$	$\hat{Y}_3^{\text{No 114}}$
$Y_4^{\text{Takes 114}}$	$\hat{Y}_4^{\text{No 114}}$
$Y_5^{\text{Takes 114}}$	$\hat{Y}_5^{\text{No 114}}$
$Y_6^{\text{Takes 114}}$	$\hat{Y}_6^{\text{No 114}}$

General approach

1) Define potential outcomes
2) Define target population

Summary: Outcome model for causal inference

Each Row is a Student in This Class

Outcome under 114	Outcome under no 114
$Y_1^{\text{Takes 114}}$	$\hat{Y}_1^{\text{No 114}}$
$Y_2^{\text{Takes 114}}$	$\hat{Y}_2^{\text{No 114}}$
$Y_3^{\text{Takes 114}}$	$\hat{Y}_3^{\text{No 114}}$
$Y_4^{\text{Takes 114}}$	$\hat{Y}_4^{\text{No 114}}$
$Y_5^{\text{Takes 114}}$	$\hat{Y}_5^{\text{No 114}}$
$Y_6^{\text{Takes 114}}$	$\hat{Y}_6^{\text{No 114}}$

General approach

1) Define potential outcomes
2) Define target population
3) Make causal assumptions

Summary: Outcome model for causal inference

Each Row is a Student in This Class

Outcome under 114	Outcome under no 114
$Y_1^{\text{Takes 114}}$	$\hat{Y}_1^{\text{No 114}}$
$Y_2^{\text{Takes 114}}$	$\hat{Y}_2^{\text{No 114}}$
$Y_3^{\text{Takes 114}}$	$\hat{Y}_3^{\text{No 114}}$
$Y_4^{\text{Takes 114}}$	$\hat{Y}_4^{\text{No 114}}$
$Y_5^{\text{Takes 114}}$	$\hat{Y}_5^{\text{No 114}}$
$Y_6^{\text{Takes 114}}$	$\hat{Y}_6^{\text{No 114}}$

General approach

1) Define potential outcomes
2) Define target population
3) Make causal assumptions
4) Model unobserved outcomes

Summary: Outcome model for causal inference

Each Row is a Student in This Class

Outcome under 114	Outcome under no 114
$Y_1^{\text{Takes 114}}$	$\hat{Y}_1^{\text{No 114}}$
$Y_2^{\text{Takes 114}}$	$\hat{Y}_2^{\text{No 114}}$
$Y_3^{\text{Takes 114}}$	$\hat{Y}_3^{\text{No 114}}$
$Y_4^{\text{Takes 114}}$	$\hat{Y}_4^{\text{No 114}}$
$Y_5^{\text{Takes 114}}$	$\hat{Y}_5^{\text{No 114}}$
$Y_6^{\text{Takes 114}}$	$\hat{Y}_6^{\text{No 114}}$

General approach

1) Define potential outcomes
2) Define target population
3) Make causal assumptions
4) Model unobserved outcomes
5) Predict them

Summary: Outcome model for causal inference

Each Row is a Student in This Class

Outcome under 114	Outcome under no 114
$Y_1^{\text{Takes 114}}$	$\hat{Y}_1^{\text{No 114}}$
$Y_2^{\text{Takes 114}}$	$\hat{Y}_2^{\text{No 114}}$
$Y_3^{\text{Takes 114}}$	$\hat{Y}_3^{\text{No 114}}$
$Y_4^{\text{Takes 114}}$	$\hat{Y}_4^{\text{No 114}}$
$Y_5^{\text{Takes 114}}$	$\hat{Y}_5^{\text{No 114}}$
$Y_6^{\text{Takes 114}}$	$\hat{Y}_6^{\text{No 114}}$

General approach

1) Define potential outcomes
2) Define target population
3) Make causal assumptions
4) Model unobserved outcomes
5) Predict them
6) Report an average

Learning goals for today

By the end of class, you will be able to

- connect causal inference (a missing data problem)

  to statistical modeling (predicting missing data)