Quasi-Experiments I

POLSCI 4SS3

Winter 2023

Course surveys due April 12, 11:59 PM

McMaster University | BRIGHTER WORLD

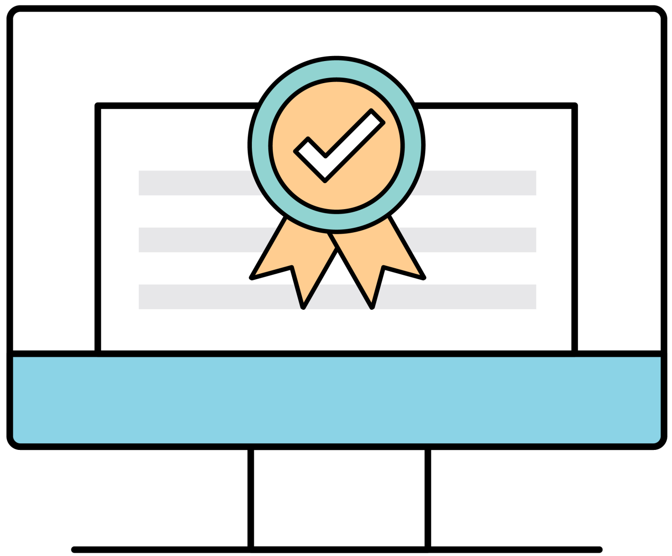
Now Open! Student Course Experience Surveys

It's time to share your feedback on your learning experience.

Here are a few quick tips to make the most impact with your comments.

- Think constructively about your course(s).
- Be respectful.
- Be specific and provide a reasonable amount of information.
- Consider what's working and what's not working.

Get started on your course surveys: **mcmaster.bluera.com/mcmaster**



Announcements

- Two more weeks left!
- Final projects due **April 21**
- Groups need to meet with instructor one more time before **April 19**
- Extra office hour times **April 13-19**
- Every group member needs to be in **at least one** group meeting to receive group meeting grade

Last time

- Learning from experiments
- Good to understand what works, but not why or where
- Need to think about support factors
- Scaling up, drilling down
- **Today:** Observational causal data strategies

Types of data strategy

	Data strategy	
Inquiry	**Observational**	**Experimental**
Descriptive	Sample survey	List experiment
Causal		Survey/field experiment

Types of data strategy

	Data strategy	
Inquiry	**Observational**	**Experimental**
Descriptive	Sample survey	List experiment
Causal	**Quasi-experiment**	Survey/field experiment

Challenges to causal interpretations

1. Reverse causation

- Instead of Z causing Y, Y causes Z

- **Simultaneity:** Z causes Y and vice versa

> **Example**
>
> Students who are likely to participate enroll in Political Science courses more often

Challenges to causal interpretations

2. Omitted variable bias

- There is an unobserved factor X that explains the relationship between Z and Y

Example

- We believe that more education increases income
- But having smart parents increases both education and income

Challenges to causal interpretations

3. Selection bias

- Individuals *sort* into condition Z in a manner that predicts outcome Y

- Treatment and control are not comparable

Example

- Always-takers are more likely to participate in the TUP program

Challenges to causal interpretations

1. Reverse causation

2. Omitted variable bias

3. Selection bias

- Random assignment avoids this *in expectation*
- Hard to overcome with *observational causal* data strategies
- Need to pretend that we can analyze data as if it was an experiment

Quasi-experiments

- Answer strategies that produce data as-if they were drawn from an experiment
- **Natural experiment:** Random assignment outside of the researcher control
- **Example:** Choosing municipalities at random for auditing
- **Quasi-experiment:** Conditions are assigned in a manner that is **sufficiently orthogonal** to potential outcomes

Examples of quasi-experiment

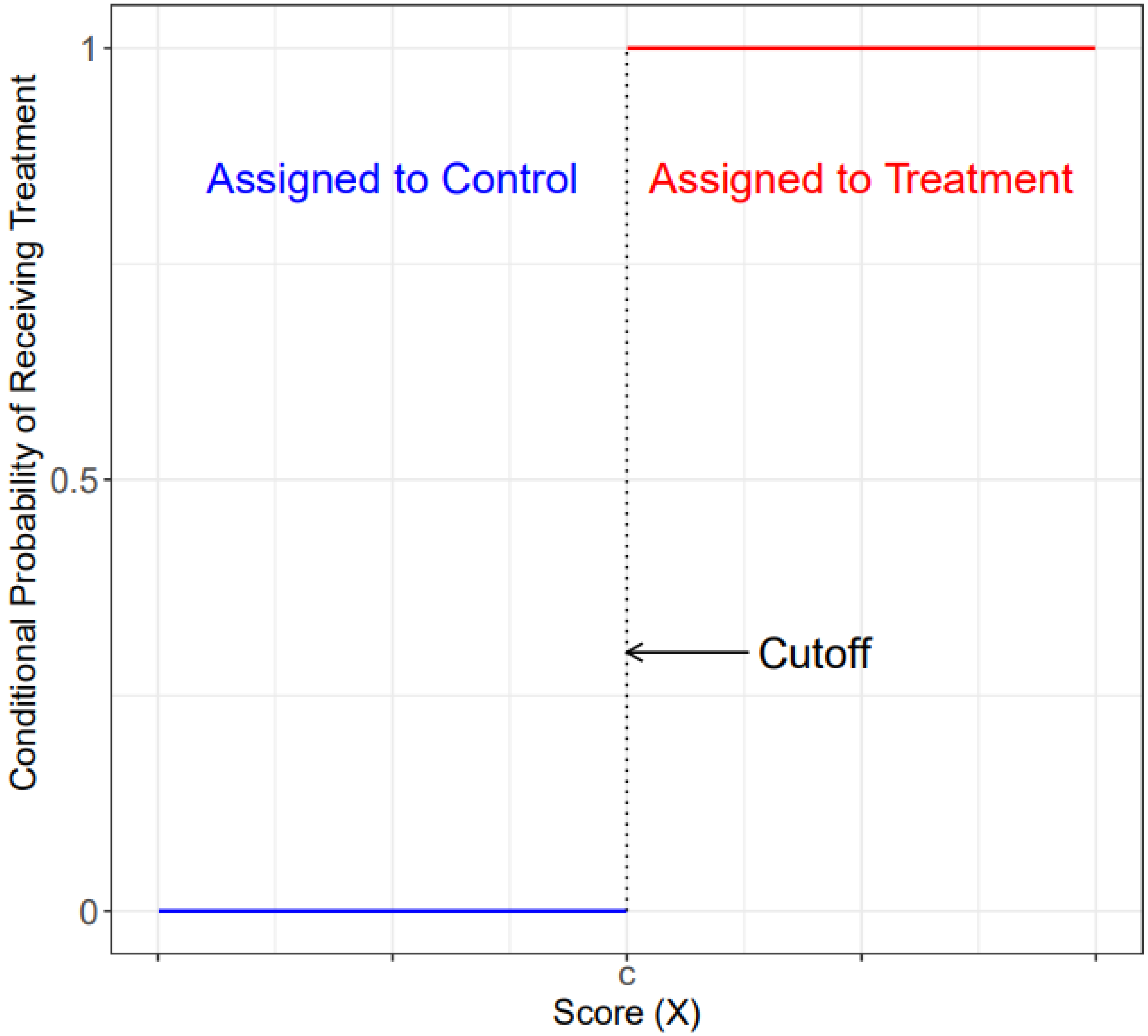
- Benefits eligible only to people under a certain income
- Institutional variation based on arbitrary population thresholds
- Winning an election by a narrow margin
- Being barely inside/outside an administrative border
- Unexpected events

Regression discontinuity designs

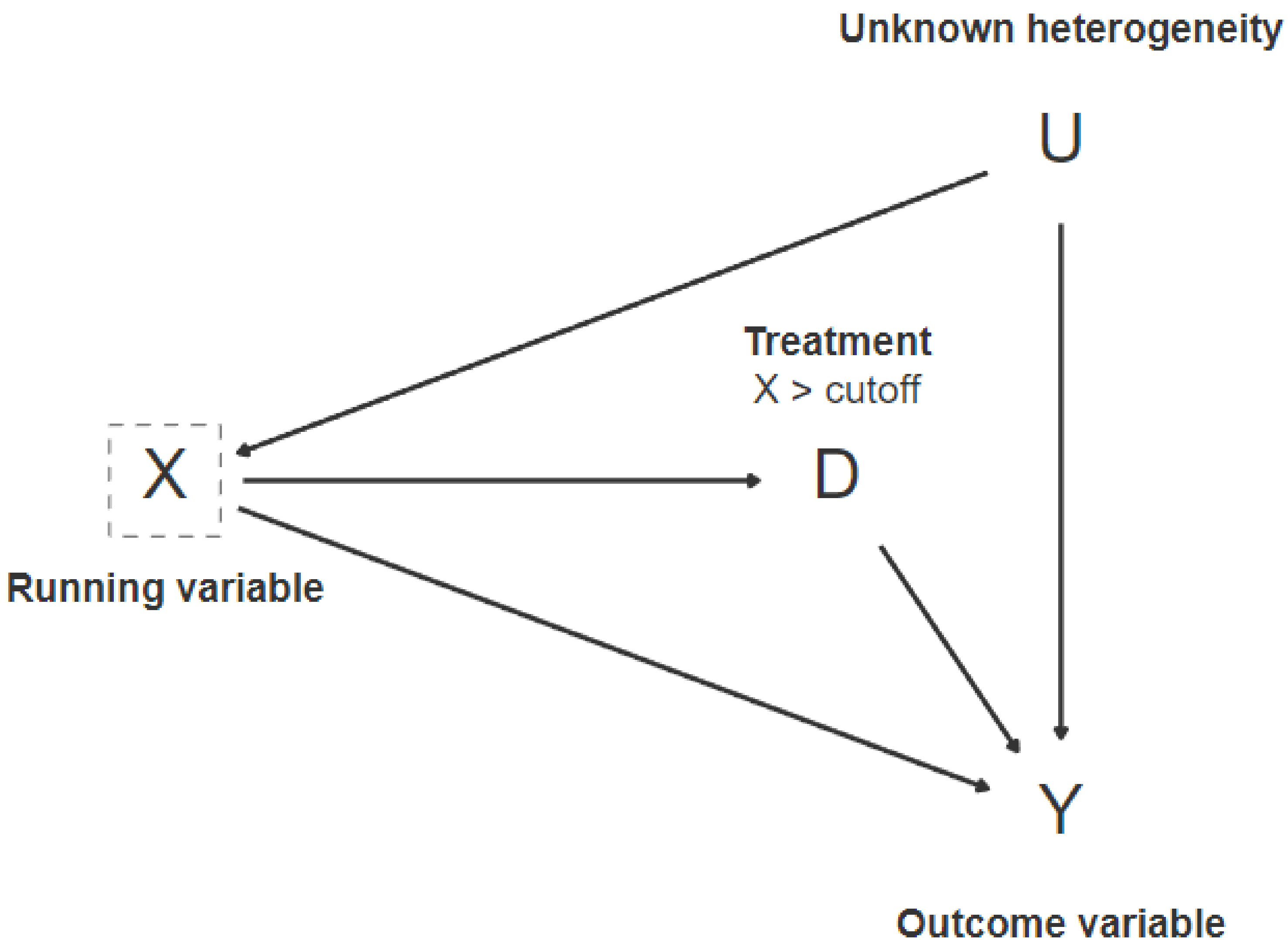
- Three ingredients:

1. Score `(running variable)`
2. Cutoff `(threshold)`
3. Treatment `(at least two conditions)`

Visual representation



As a graph



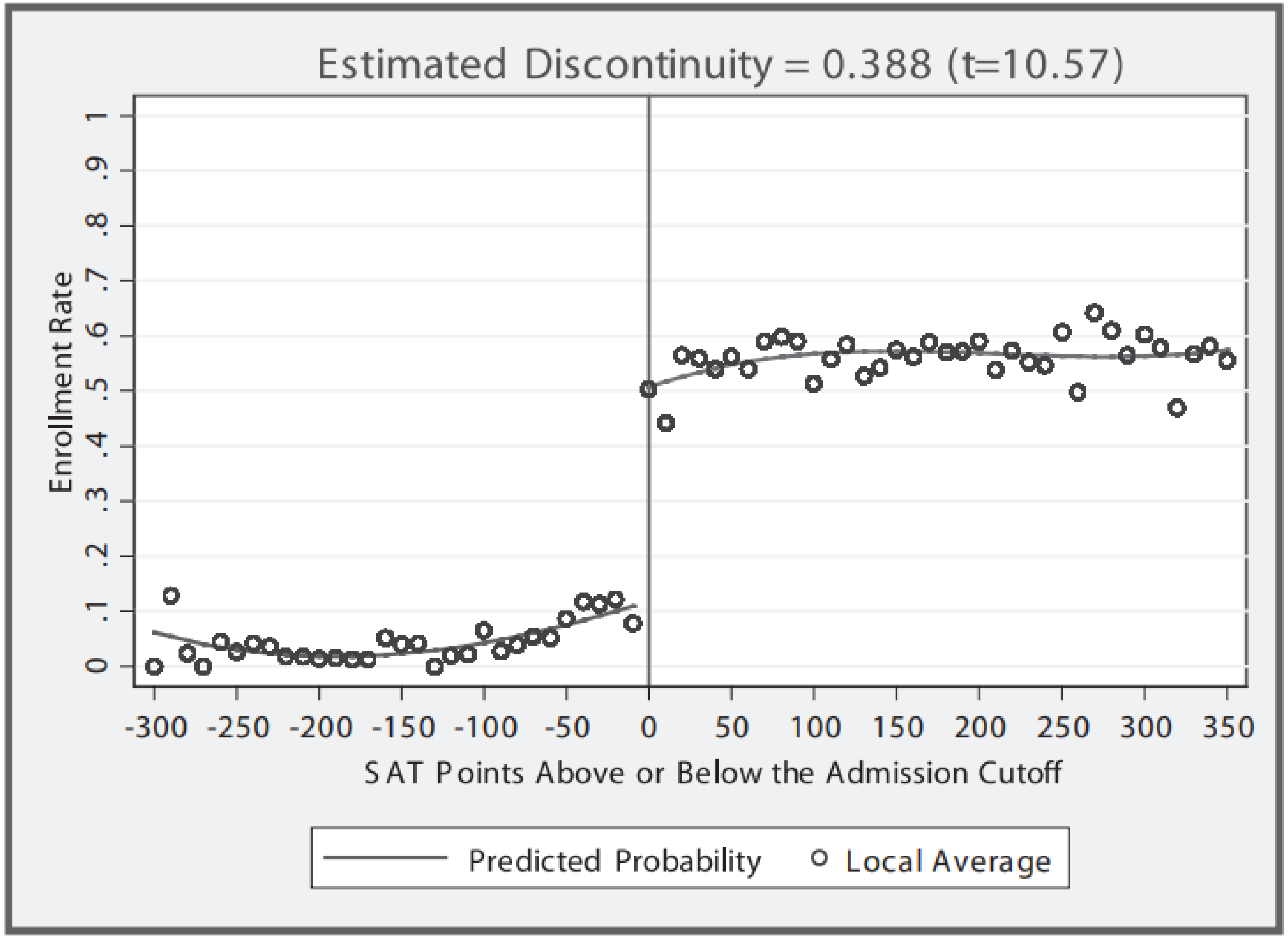
Unknown heterogeneity
U
Treatment
X > cutoff
X
D
Running variable
Y
Outcome variable

Example: Hoekstra (2019)

- **Question:** Effect of higher education on earnings
- **Challenge:** Selection bias
- **Solution:** Focus on attendance at US state flagship university among 28-33 year olds
- **Outcome:** Earnings
- **Score:** SAT test scores
- **Cutoff:** Admission cutoff
- **Treatment:** Attending flagship university

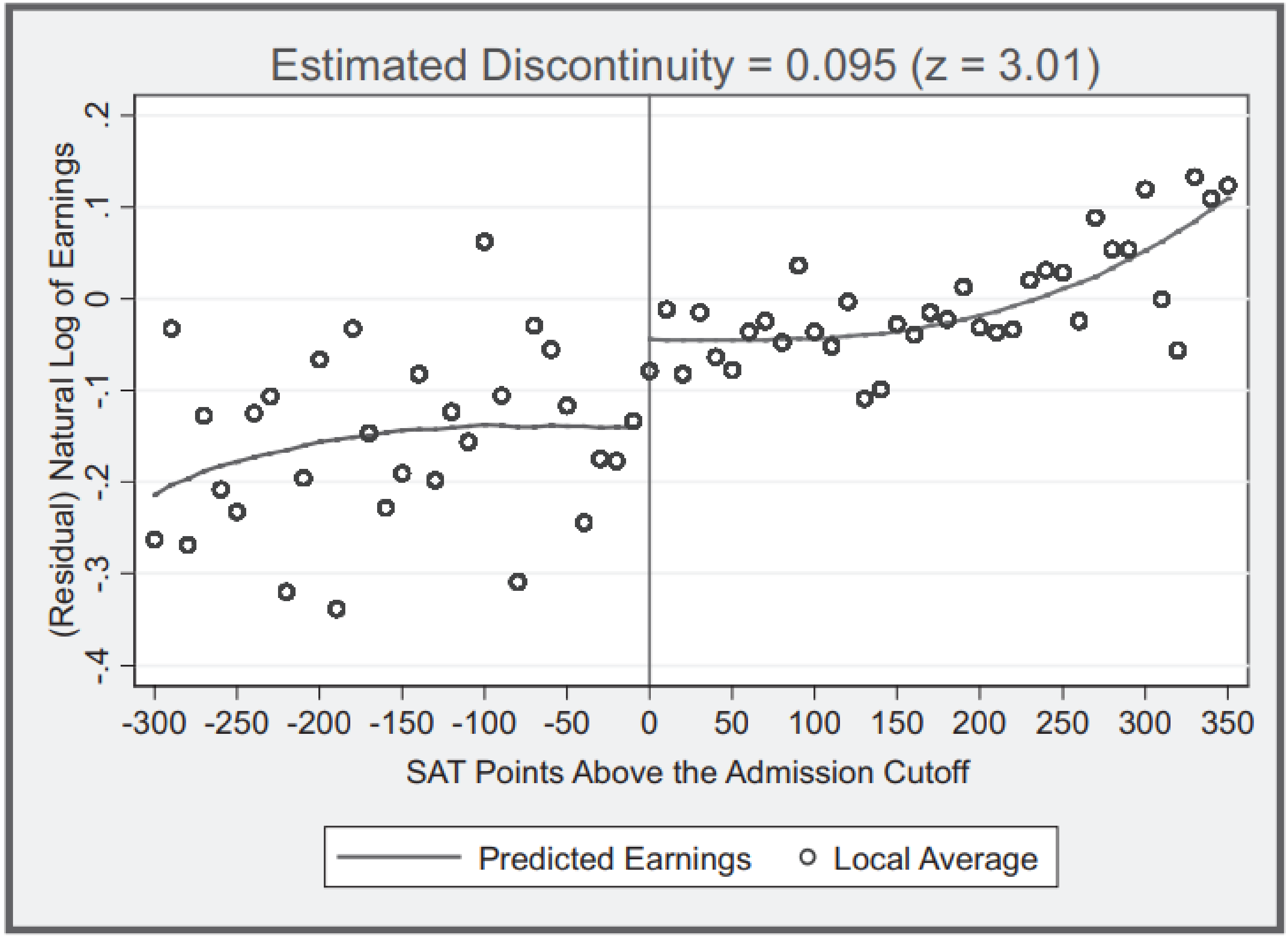
Treatment take-up

FIGURE 1.—FRACTION ENROLLED AT THE FLAGSHIP STATE UNIVERSITY



Results

Figure 2.—Natural Log of Annual Earnings for White Men Ten to Fifteen Years after High School Graduation (Fit with a Cubic Polynomial of Adjusted SAT Score)



How do you get an estimate?

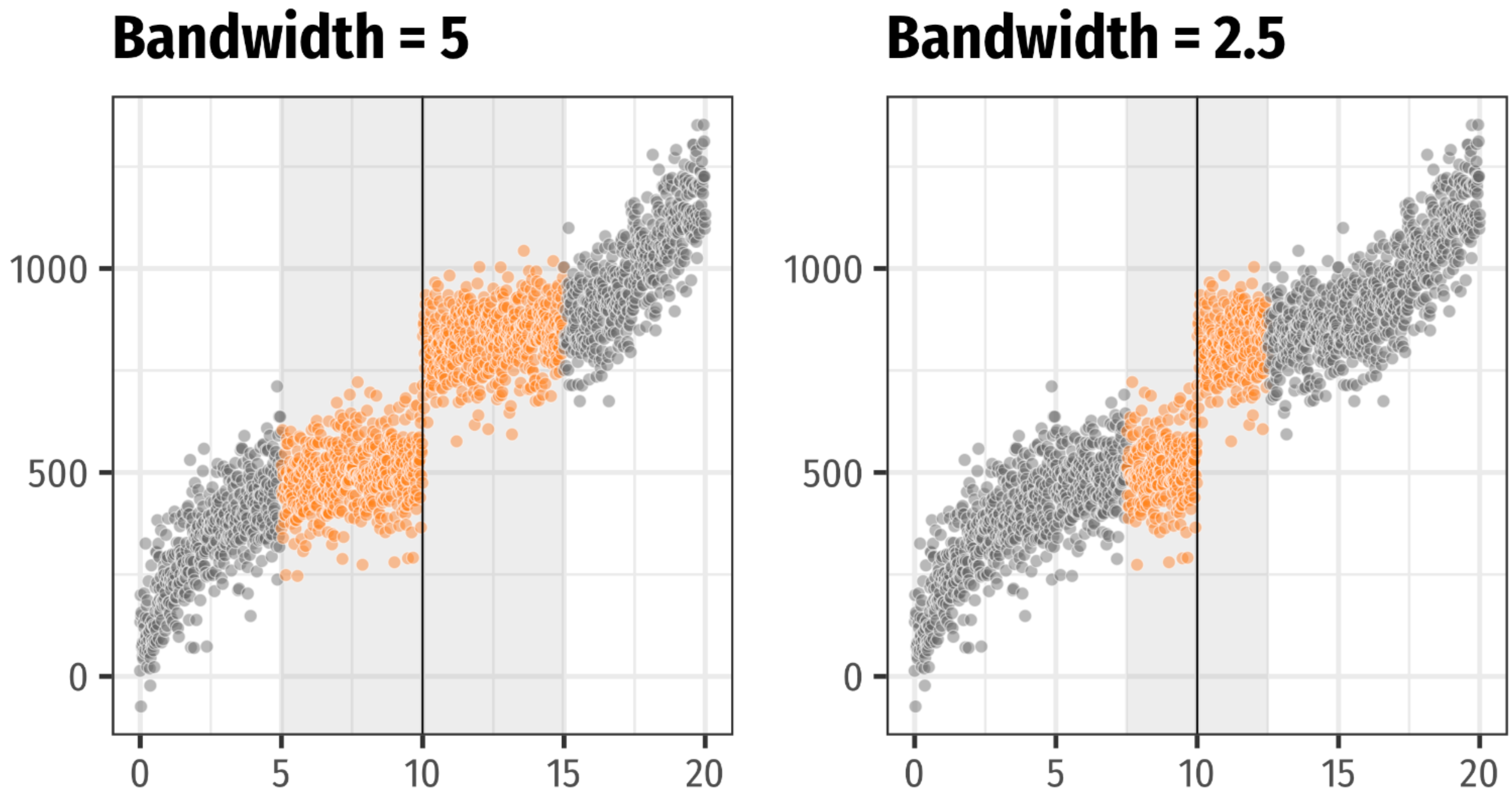
- Two approaches to RDD data:

1. Local randomization
2. Continuity-based

Local randomization

- Potential outcomes are not random because they depend on the score `(and other things)`
- However, around the cutoff, treatment assignment is as good as random
- So we can pretend we have an experiment within a **bandwidth** around the cutoff

Bandwidth tradeoff



A small bandwidth has low bias but high variance. A larger bandwidth has lower variance

RDD assumptions

> ⚠ **1. Continuity**
>
> Score is continuous at the cutoff

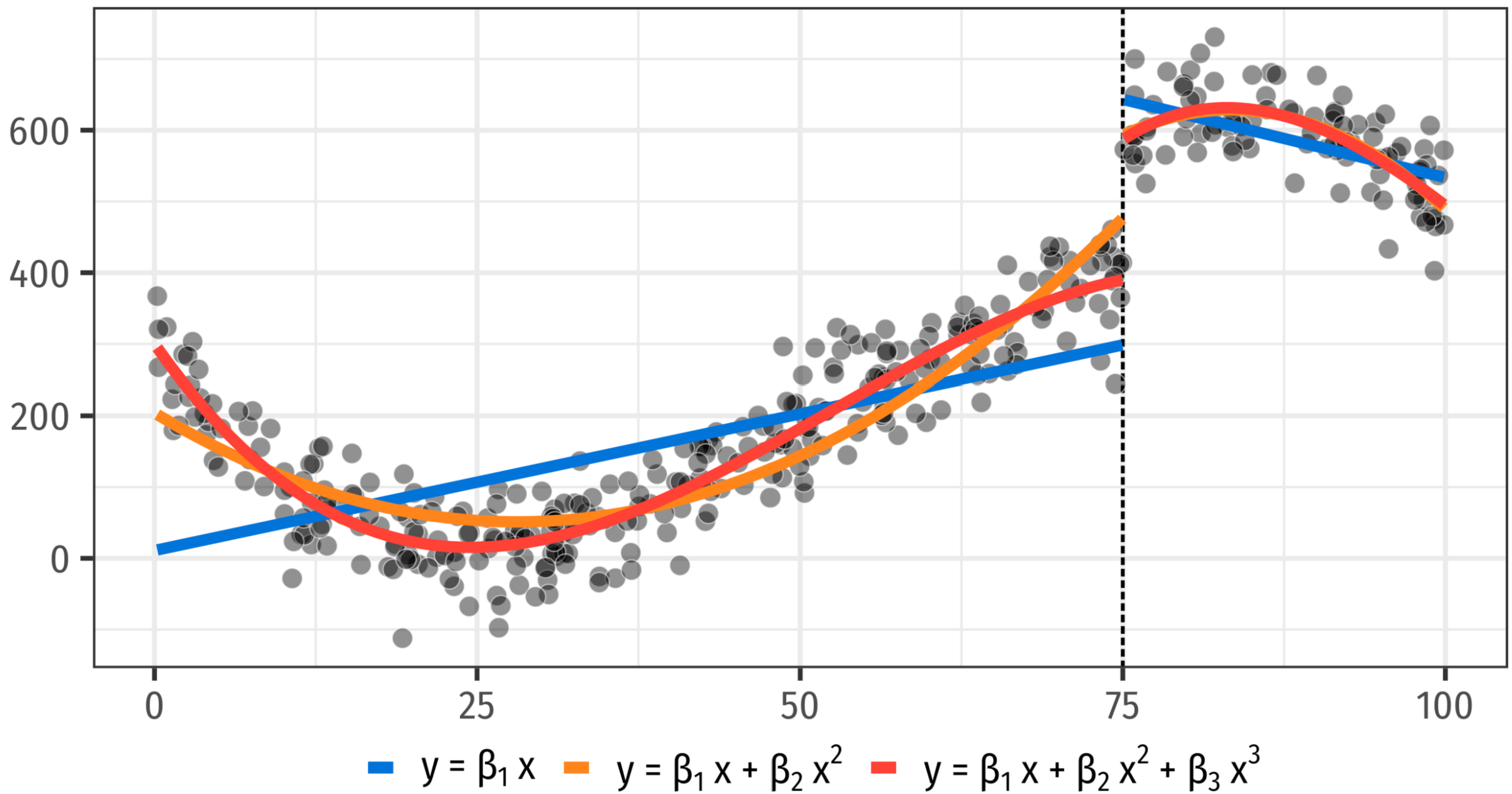
> ⚠ **2. Comparability**
>
> Units are comparable at or around the cutoff

- These imply no manipulation and no selection bias
- Local randomization is **sufficient** but **not necessary** to satisfy these assumptions

Continuity-based approach

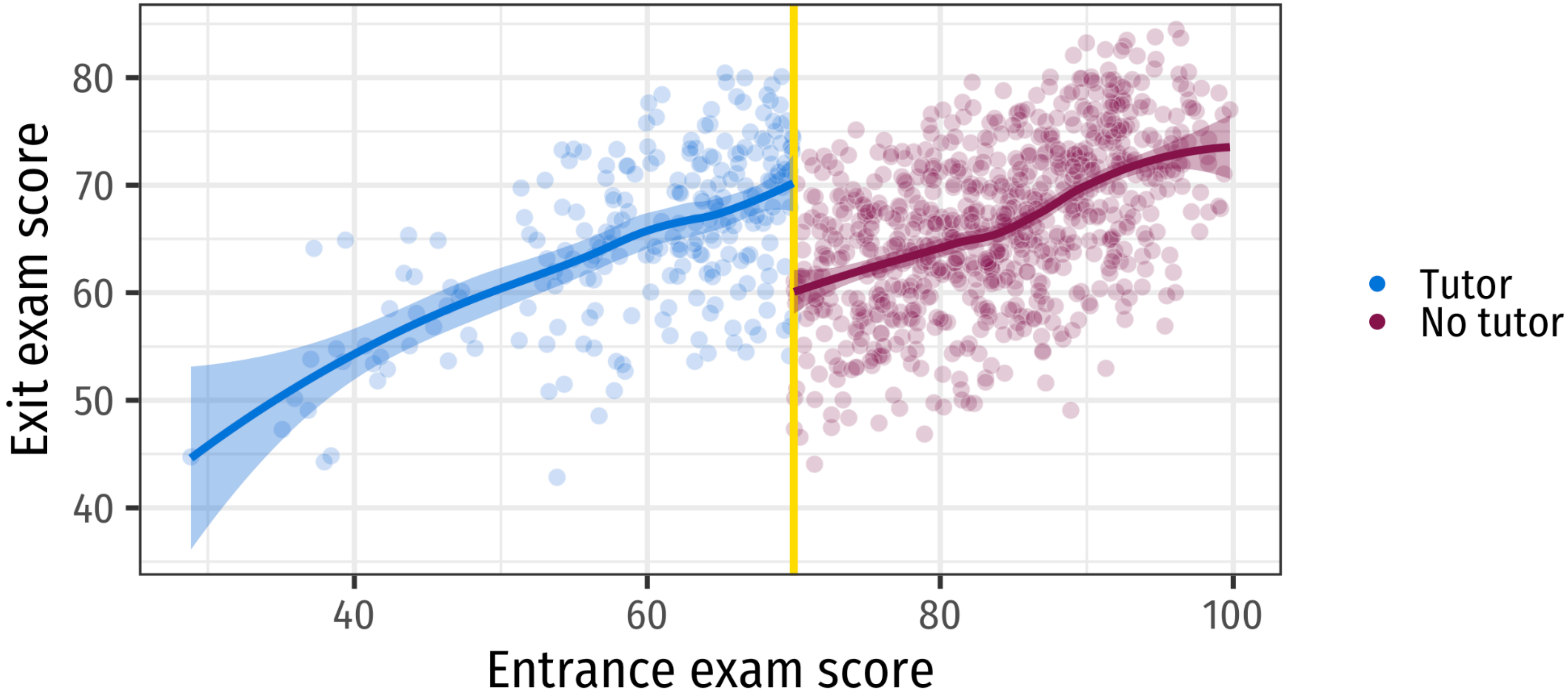
- Treatment assignment is **deterministic at** the cutoff
- But usually too few or no units at the cutoff
- **Task:** Approximate the *gap* at the cutoff as best as possible
- This becomes a **line drawing** problem

Line drawing: Parametric



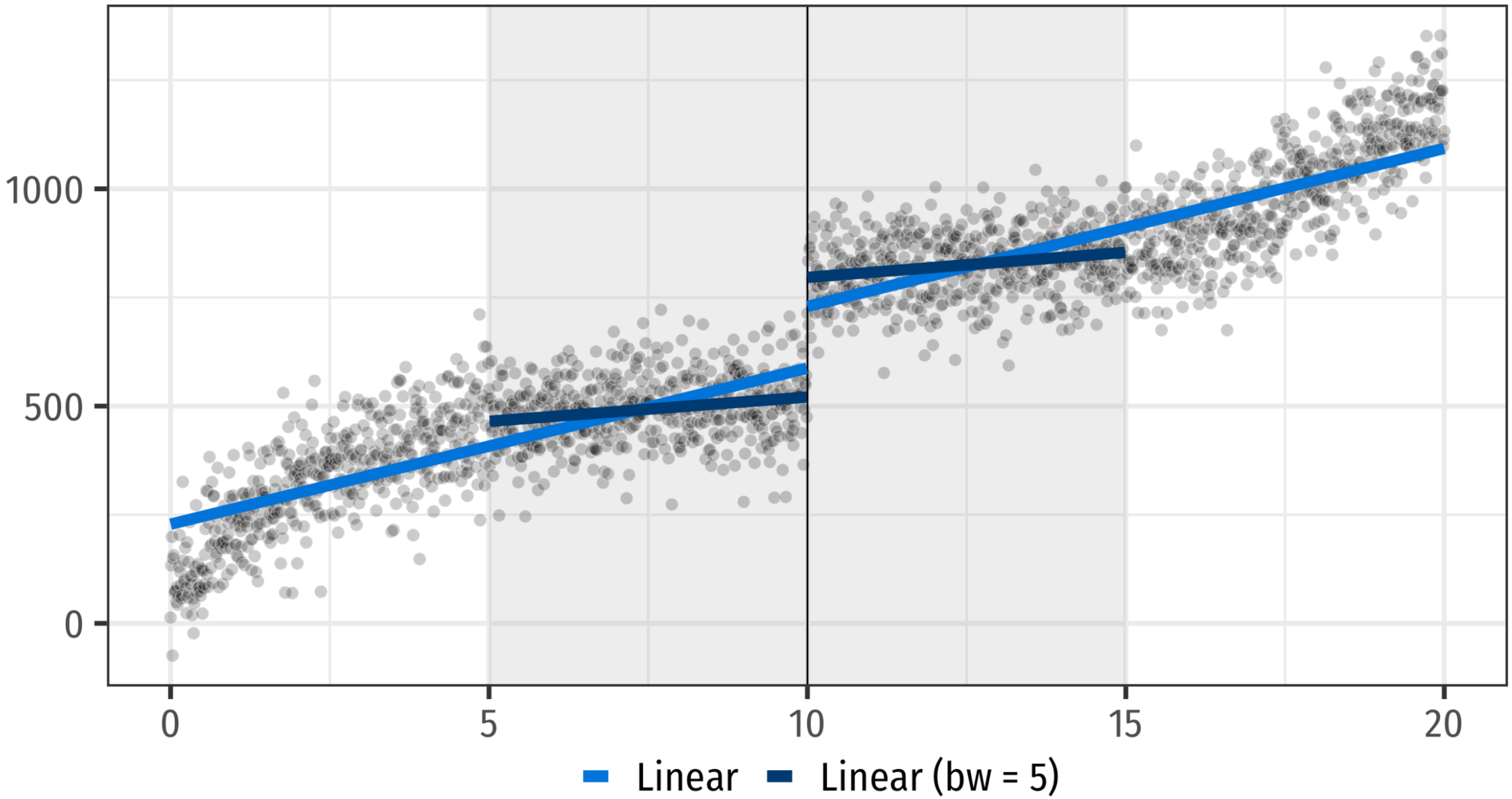
Different functional forms change the size of the gap

Line drawing: Nonparametric



These lines are made by an algorithm that calculates the local average at many windows

Line drawing: Bandwidth



The size of the bandwidth determines the data you use to draw lines

RDDs in balance

Pros

- Intuitive design
- Widely applicable
- Easy to visualize

Cons

- Results are highly *local*
- Scale up? Drill down?
- Too many moving parts

- Software automates most of the decisions
- Ideally, results are consistent under different approaches

Next Week

Quasi-experiments II

Focus on: Difference-in-differences design

Break time!

Lab