Regression Discontinuity

INFO/STSCI/ILRST 3900: Causal Inference

24 Oct 2023

Learning goals for today

At the end of class, you will be able to:

1. Explain how discontinuities can be exploited for causal identification
2. Understand bias variance trade-off in selecting bandwiths

What is the effect of a certificate?

What is the effect of a certificate?

- Thistlewaite and Campbell (1960) are interested in measuring causal effect of winning a scholarship

What is the effect of a certificate?

- Thistlewaite and Campbell (1960) are interested in measuring causal effect of winning a scholarship
- High school students take Scholarship Qualifying Test (SQT)

What is the effect of a certificate?

- Thistlewaite and Campbell (1960) are interested in measuring causal effect of winning a scholarship
- High school students take Scholarship Qualifying Test (SQT)
- Each state has a specific threshold for score

What is the effect of a certificate?

- Thistlewaite and Campbell (1960) are interested in measuring causal effect of winning a scholarship
- High school students take Scholarship Qualifying Test (SQT)
- Each state has a specific threshold for score
- Students who score above a state specific threshold get **Certificate of Merit**
- Students who score well, but below the threshold get **letter of commendation**

What is the effect of a certificate?

- Thistlewaite and Campbell (1960) are interested in measuring causal effect of winning a scholarship
- High school students take Scholarship Qualifying Test (SQT)
- Each state has a specific threshold for score
- Students who score above a state specific threshold get **Certificate of Merit**
- Students who score well, but below the threshold get **letter of commendation**
- Data contains 5,126 Certificate of Merit winners and 2,848 letters of commendation winners

What is the effect of a certificate?

- CoM winners got $\approx 2.5x$ recognition, published in lists, etc.

What is the effect of a certificate?

- CoM winners got $\approx 2.5x$ recognition, published in lists, etc.
- 6 Months after awards, survey is sent out
  - Other scholarships won
  - Planning to pursue PhD or MD
  - Attitude towards intellectualism

What is the effect of a certificate?

- CoM winners got $\approx 2.5x$ recognition, published in lists, etc.
- 6 Months after awards, survey is sent out
  - Other scholarships won
  - Planning to pursue PhD or MD
  - Attitude towards intellectualism

What is the causal effect of the CoM on various attributes?

What is the effect of a scholarship?

SQT

$\text{CoM} \longrightarrow \text{Other Scholarships}$

What is the effect of a scholarship?

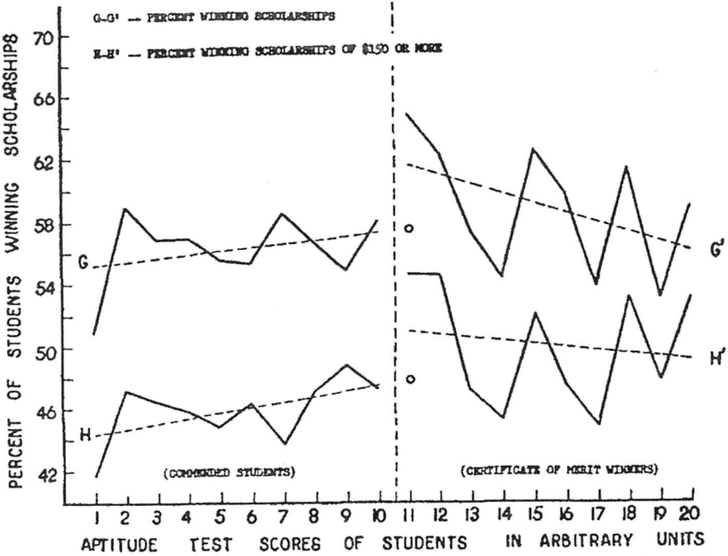


FIG. 2. Regression of success in winning scholarships on exposure determiner.

What is the effect of a scholarship?

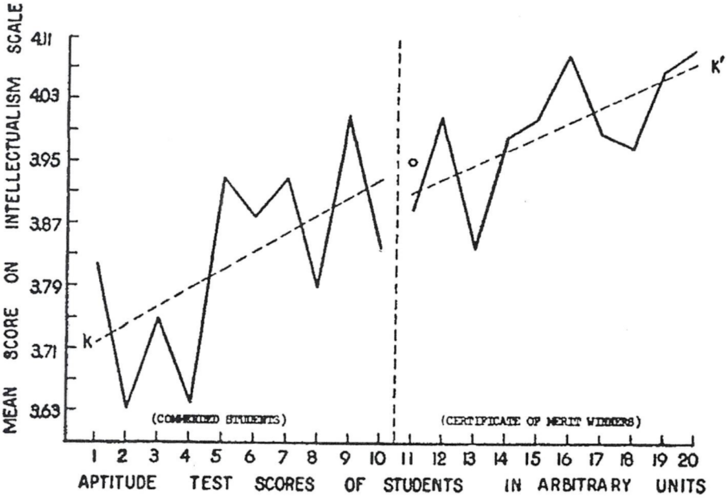


Fig. 4. Regression of attitudes toward intellectualism on exposure determiner.

Overlap assumption

Overlap assumption

- Conditioning on SQT score yields conditional exchangeability

Overlap assumption

- Conditioning on SQT score yields conditional exchangeability
- Try to apply previous methods:

Overlap assumption

- Conditioning on SQT score yields conditional exchangeability
- Try to apply previous methods:
  - Standardization:

$$\mathrm{E}(Y^{a=1}) = \sum_s \mathrm{E}(Y \mid Score = s, A = 1) P(\mathrm{Score} = s)$$

$$\mathrm{E}(Y^{a=0}) = \sum_s \mathrm{E}(Y \mid Score = s, A = 0) P(\mathrm{Score} = s)$$

Overlap assumption

- Conditioning on SQT score yields conditional exchangeability
- Try to apply previous methods:
  - Standardization:

$$E(Y^{a=1}) = \sum_s E(Y \mid Score = s, A = 1) P(\text{Score} = s)$$
$$E(Y^{a=0}) = \sum_s E(Y \mid Score = s, A = 0) P(\text{Score} = s)$$

  - Matching:

Overlap assumption

- Conditioning on SQT score yields conditional exchangeability
- Try to apply previous methods:
  - Standardization:

$$\mathrm{E}(Y^{a=1}) = \sum_s \mathrm{E}(Y \mid Score = s, A = 1)P(\mathrm{Score} = s)$$

$$\mathrm{E}(Y^{a=0}) = \sum_s \mathrm{E}(Y \mid Score = s, A = 0)P(\mathrm{Score} = s)$$

  - Matching:
    Match each person who received Certificate of Merit with a "similar" person who received Letter of Recommendation
- In both cases, $P(CoM = 1 \mid Score) = 0$ for some scores

Local average treatment effect

Local average treatment effect

- No overlap implies we can't directly estimate ATE without strong assumptions

Local average treatment effect

- No overlap implies we can't directly estimate ATE without strong assumptions
- Let's aim for an easier target

Local average treatment effect

- No overlap implies we can't directly estimate ATE without strong assumptions
- Let's aim for an easier target
- Average Treatment effect for individuals at the cut-off

$$\textbf{Local } ATE = \mathrm{E}(Y_i^{a=1} \mid Score = c) - \mathrm{E}(Y_i^{a=0} \mid Score = c_0)$$

- Does not tell us about treatment effect for everyone!

The big idea

The big idea

- Treatment of interest depends only on whether a **running variable** is above or below a threshold c

The big idea

- Treatment of interest depends only on whether a **running variable** is above or below a threshold c
  CoM only depends on above or below score threshold
- Assume $\mathrm{E}(Y^{a=1} \mid R = r)$ and $\mathrm{E}(Y^{a=0} \mid R = r)$ varies smoothly

The big idea

- Treatment of interest depends only on whether a **running variable** is above or below a threshold c
  CoM only depends on above or below score threshold
- Assume $\mathrm{E}(Y^{a=1} \mid R = r)$ and $\mathrm{E}(Y^{a=0} \mid R = r)$ varies smoothly
  The average **potential outcomes** for score $= 99.9$ is very close to the average for score $= 100.1$
- Above the the cut-off $\mathrm{E}(Y^{a=1} \mid R = r) = \mathrm{E}(Y \mid R = r)$
- Below the the cut-off $\mathrm{E}(Y^{a=0} \mid R = r) = \mathrm{E}(Y \mid R = r)$

The big idea

- Treatment of interest depends only on whether a **running variable** is above or below a threshold c
  CoM only depends on above or below score threshold
- Assume $\mathrm{E}(Y^{a=1} \mid R = r)$ and $\mathrm{E}(Y^{a=0} \mid R = r)$ varies smoothly
  The average **potential outcomes** for score $= 99.9$ is very close to the average for score $= 100.1$
- Above the the cut-off $\mathrm{E}(Y^{a=1} \mid R = r) = \mathrm{E}(Y \mid R = r)$
- Below the the cut-off $\mathrm{E}(Y^{a=0} \mid R = r) = \mathrm{E}(Y \mid R = r)$
- Using observed data, estimate, $\mathrm{E}(Y \mid R = r)$ for r closer and closer to the cut-off

The big idea

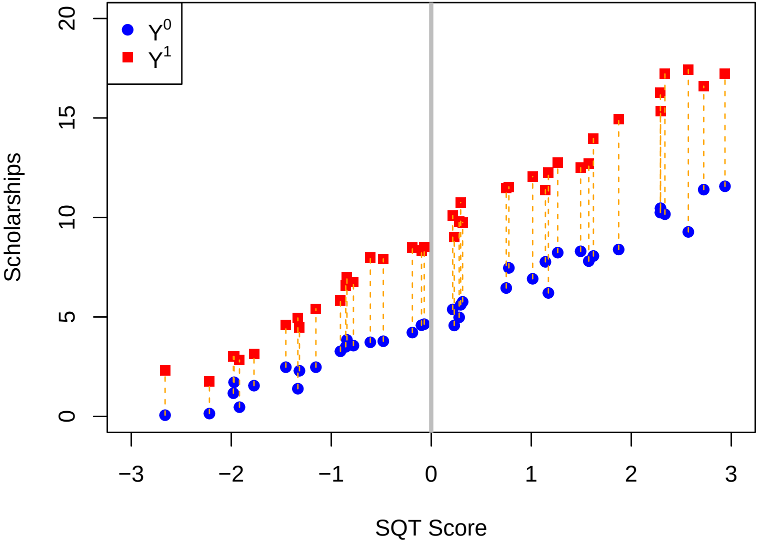
- Treatment of interest depends only on whether a **running variable** is above or below a threshold c

  CoM only depends on above or below score threshold
- Assume $\mathrm{E}(Y^{a=1} \mid R = r)$ and $\mathrm{E}(Y^{a=0} \mid R = r)$ varies smoothly

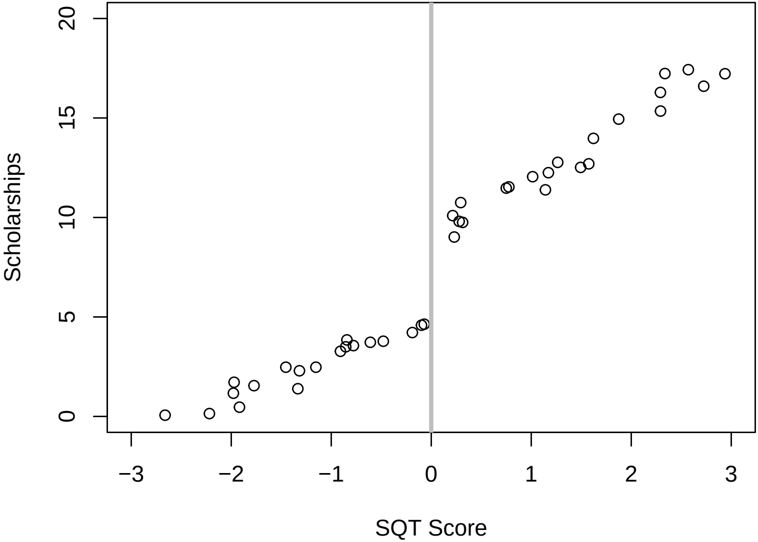
  The average **potential outcomes** for score $= 99.9$ is very close to the average for score $= 100.1$
- Above the the cut-off $\mathrm{E}(Y^{a=1} \mid R = r) = \mathrm{E}(Y \mid R = r)$
- Below the the cut-off $\mathrm{E}(Y^{a=0} \mid R = r) = \mathrm{E}(Y \mid R = r)$
- Using observed data, estimate, $\mathrm{E}(Y \mid R = r)$ for r closer and closer to the cut-off
- Estimate local ATE $\mathrm{E}(Y_i^{a=1} \mid R_i = c) - \mathrm{E}(Y_i^{a=0} \mid X_i = c)$ by

$$\underbrace{lim_{x \to c^+} \mathrm{E}(Y \mid X = x)}_{\text{from above the cut-off}} - \underbrace{\lim_{x \to c^-} \mathrm{E}(Y \mid X = x)}_{\text{from below the cut-off}}$$

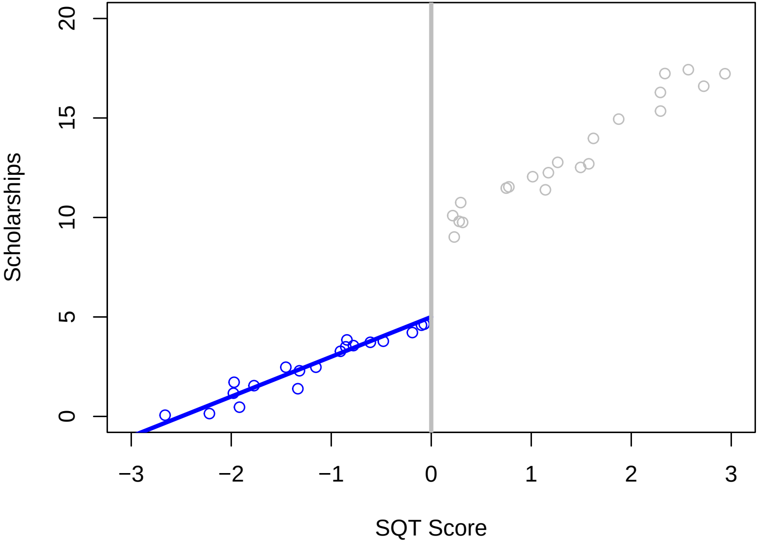
Local average treatment effect



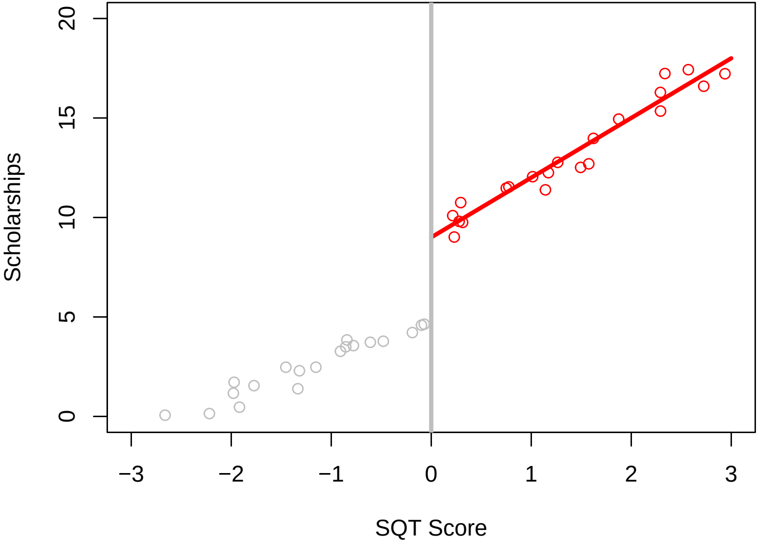
Local average treatment effect



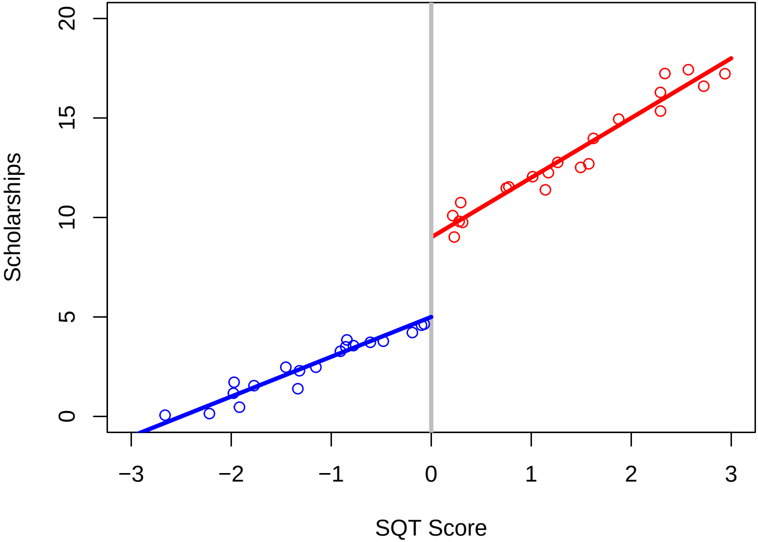
Local average treatment effect



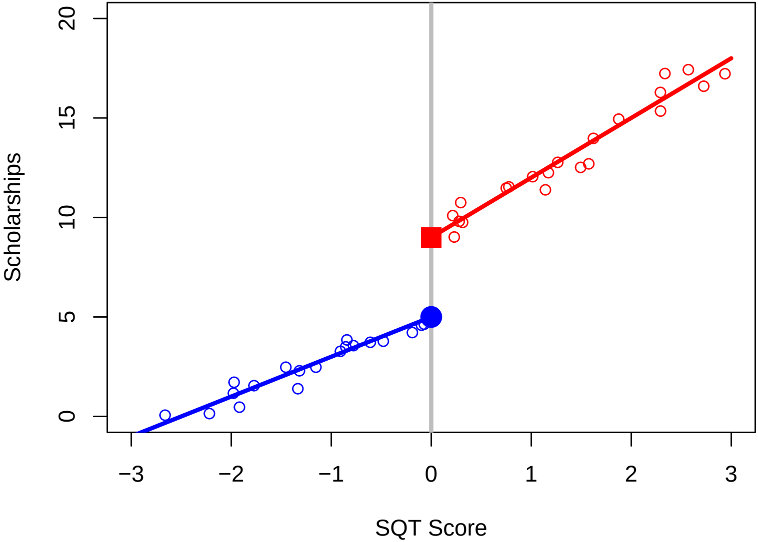
Local average treatment effect



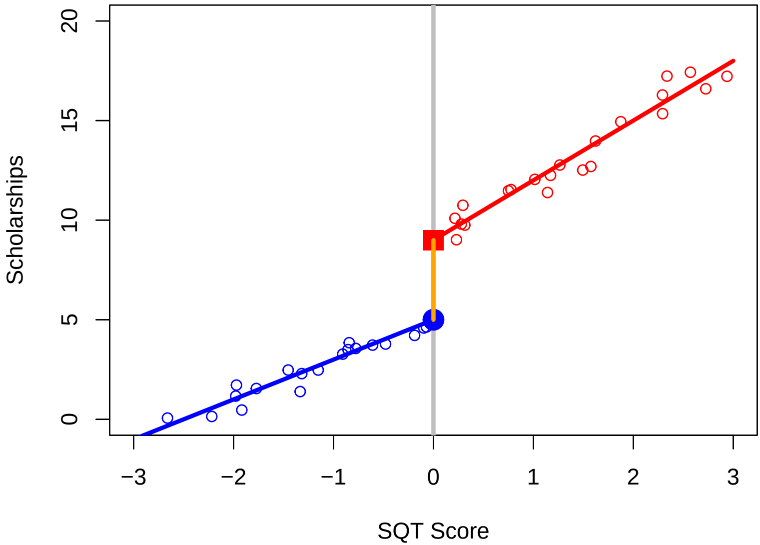
Local average treatment effect



Local average treatment effect



Local average treatment effect



Alternative intuition

- Within a small neighborhood of the cut-off, people are more or less the same

Alternative intuition

- Within a small neighborhood of the cut-off, people are more or less the same
- Ending up above or below the threshold is more or less chance
  Scoring 100.1 vs 99.9 is essentially random

Alternative intuition

- Within a small neighborhood of the cut-off, people are more or less the same
- Ending up above or below the threshold is more or less chance
  Scoring 100.1 vs 99.9 is essentially random
- Conditional exchangeability holds for people very close to the cut-off

Alternative intuition

- Within a small neighborhood of the cut-off, people are more or less the same
- Ending up above or below the threshold is more or less chance
  Scoring 100.1 vs 99.9 is essentially random
- Conditional exchangeability holds for people very close to the cut-off
- Conditional exchangeability does not hold for people further from the cut-off

Discontinuities in the wild

Discontinuities turn up in lots of places...

- in flagship state program which requires certain test score

Discontinuities in the wild

Discontinuities turn up in lots of places...

- in flagship state program which requires certain test score
- Government benefits which require means testing

Discontinuities in the wild

Discontinuities turn up in lots of places...

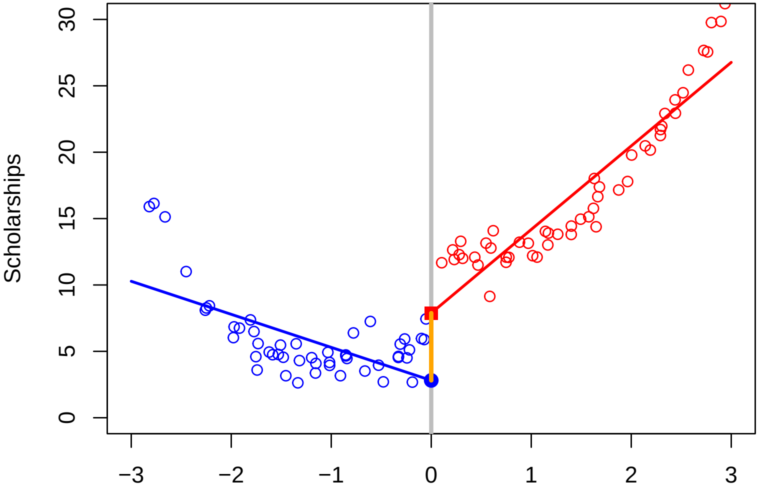
- in flagship state program which requires certain test score
- Government benefits which require means testing
- Healthcare decisions based on diagnostic test
- ...

Non-linear settings

What if $E(Y^{a=1} \mid X)$ is non-linear?

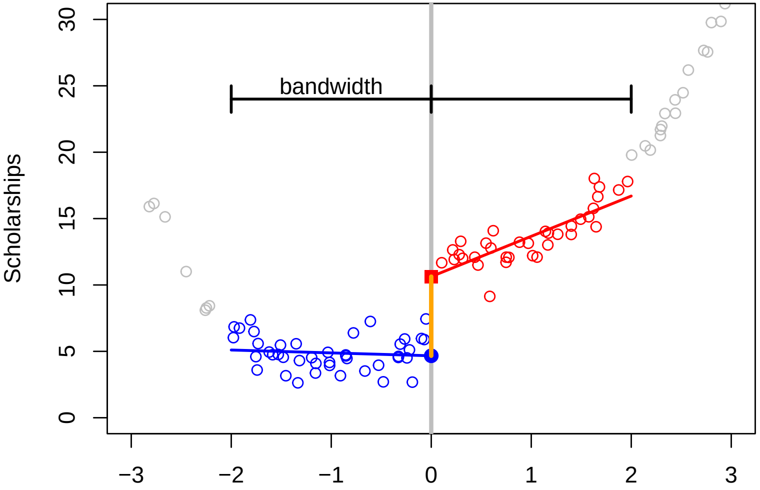
Non-linear settings

What if $E(Y^{a=1} \mid X)$ is non-linear?



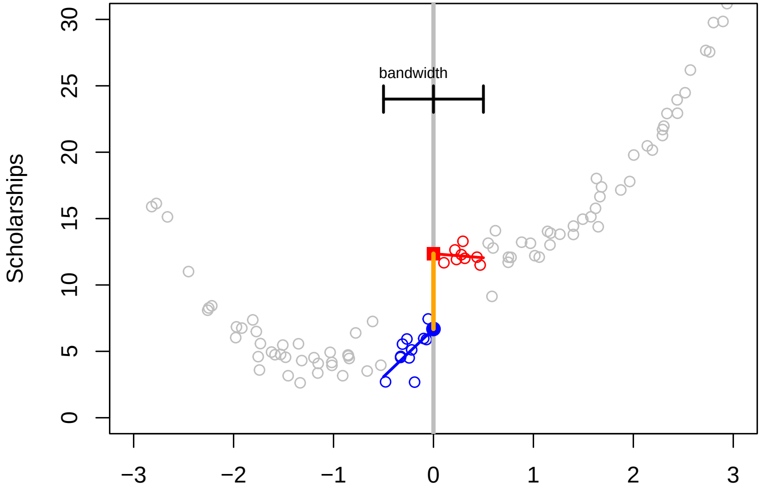
Non-linear settings

What if $E(Y^{a=1} \mid X)$ is non-linear?



Non-linear settings

What if $E(Y^{a=1} \mid X)$ is non-linear?

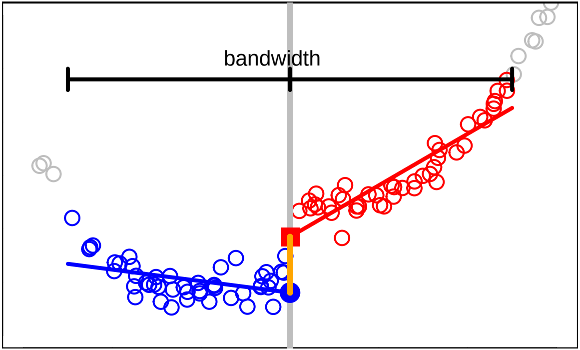


Non-linear settings

How do we choose the bandwidth?

Non-linear settings

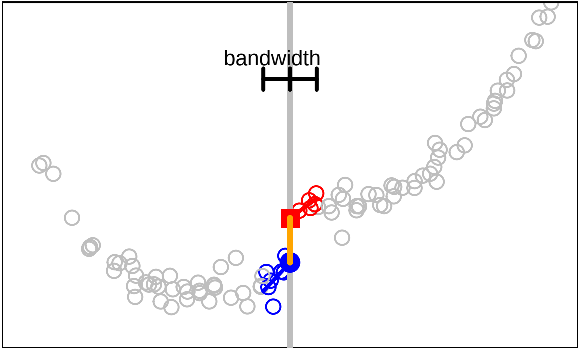
How do we choose the bandwidth?



- Bias: how far off of the truth in infinite data?

Non-linear settings

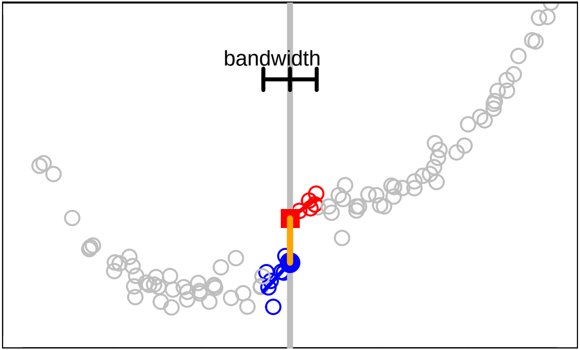
How do we choose the bandwidth?



- Bias: how far off of the truth in infinite data?
- Variance: how much would my estimate change in new sample?

Non-linear settings

How do we choose the bandwidth?



- Bias: how far off of the truth in infinite data?
- Variance: how much would my estimate change in new sample?
- Roughly speaking, bandwidth should be smaller when your data set is larger

Regression discontinuity

Pros:

- Very few assumptions required

Regression discontinuity

Pros:

- Very few assumptions required
- Plausible in many real applications

Cons:

- Can only estimate local ATE, does not generalize well
- Results depend on picking a bandwidth

Learning goals for today

At the end of class, you will be able to:

1. Explain how discontinuities can be exploited for causal identification
2. Understand bias variance trade-off in selecting bandwiths