Regression Discontinuity

INFO/STSCI/ILRST 3900: Causal Inference

29 Oct 2024

Learning goals for today

Reminder: Project Topic Survey on Ed/Canvas

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

1. Describe examples of when we would use a regression discontinuity design (RDD) to estimate causal effects

2. Explain the smoothness (continuity) assumption for RDD

3. Discuss the difference between the LATE and the ATE

4. Outline the steps it takes to estimate the LATE with a RDD

After today's class, read the following from Huntington-Klein:

▶ 20.1 How Does it Work?

▶ 20.3.4 Dealing with Bandwidths

What is the effect public recognition?

▶ Thistlewaite and Campbell (1960)

▶ Interested in measuring causal effect of receiving public recognition on outcomes such as receiving scholarshops, attitude towards intellectuallism, and career paths

▶ National Merit Scholarship program

▶ High school students take Scholarship Qualifying Test (SQT)

▶ Students who score above specific threshold get **Certificate of Merit (CoM)**

▶ Students who score well, but below the threshold get **Letter of Commendation (LoC)**

What is the effect of public recognition?

- 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?
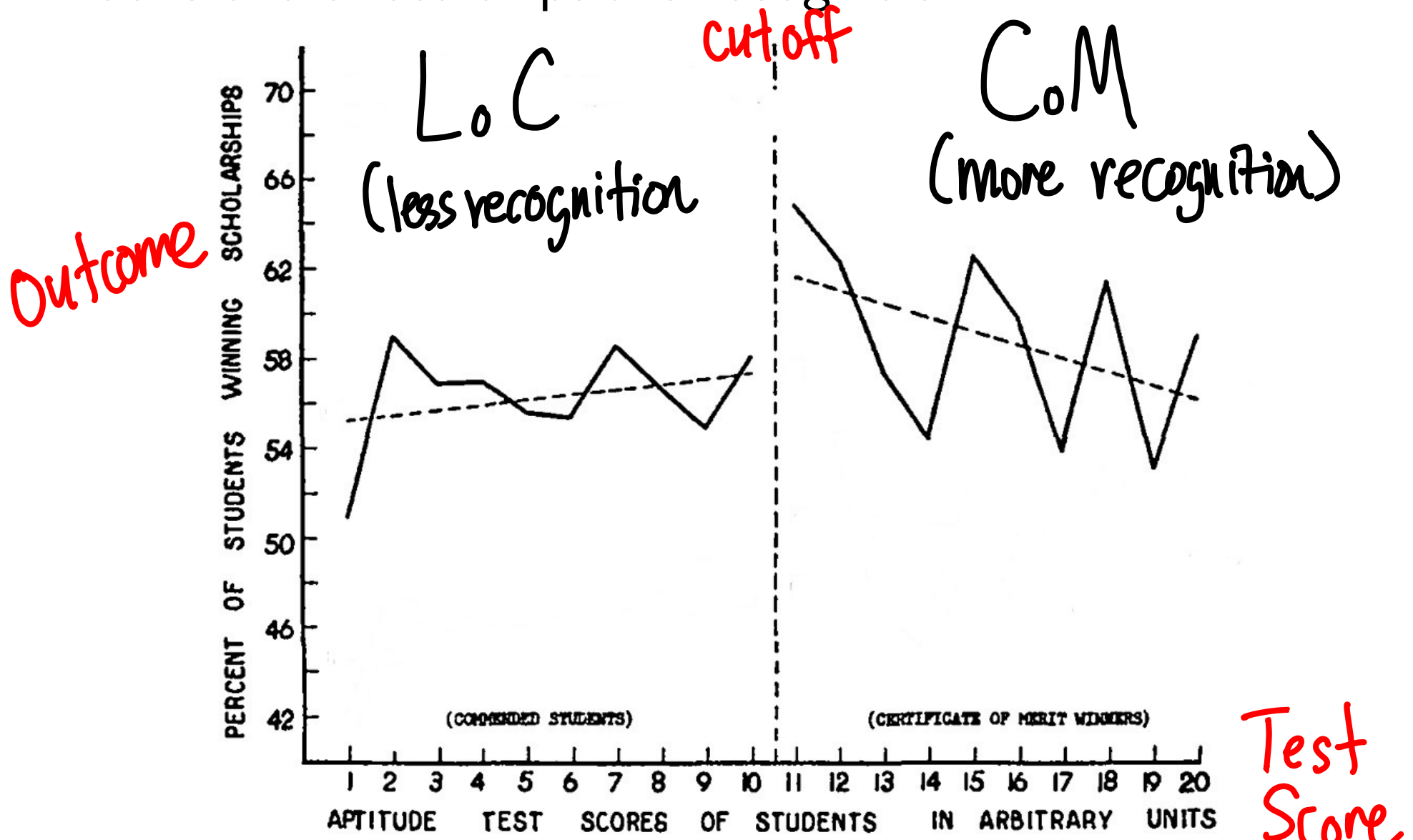
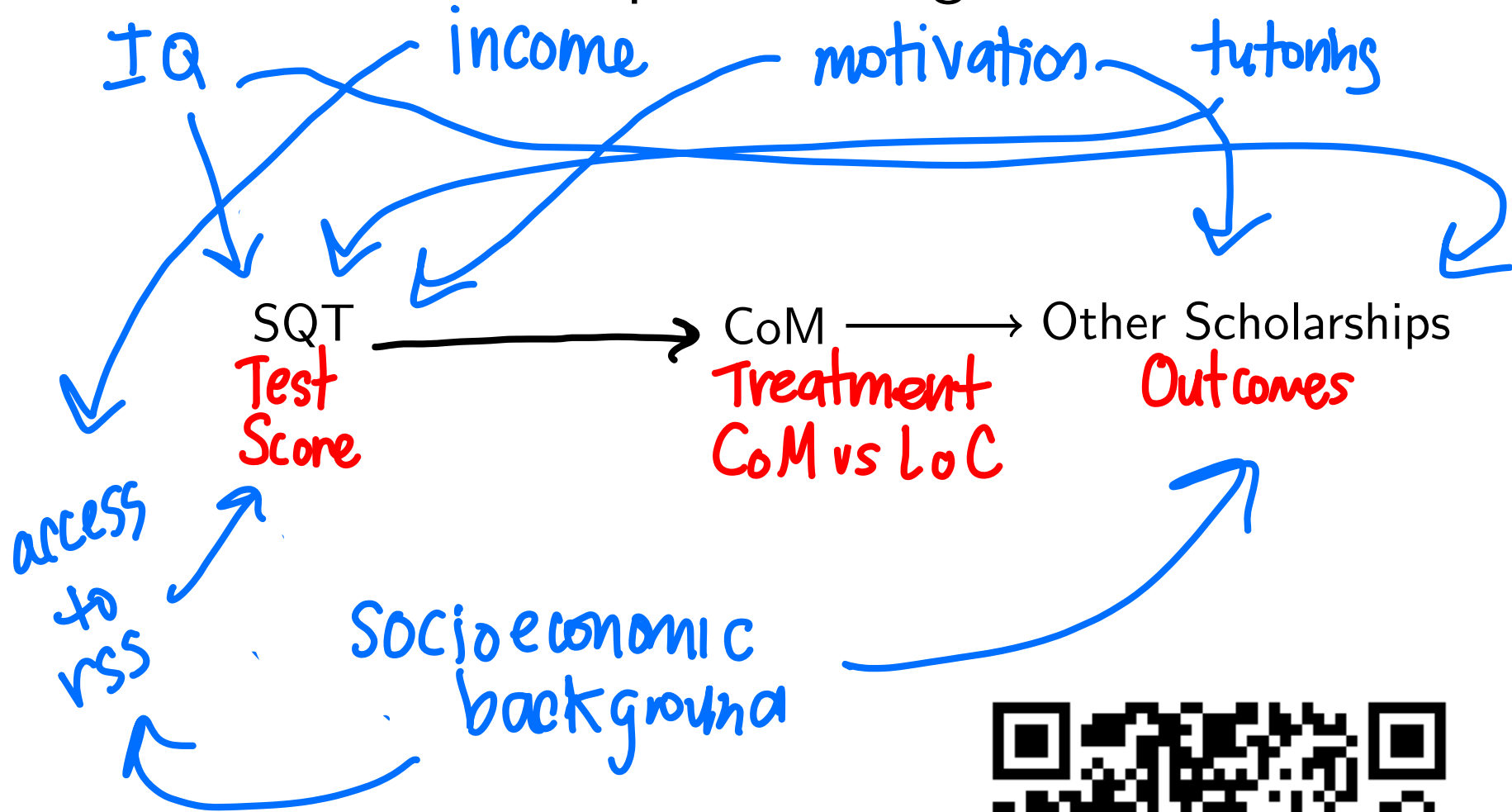# What is the effect of public recognition?

Figure: Plot from Thistlewaite and Campbell (1960)

What is the effect of public recognition?

SQT → CoM → Other Scholarships

Test Score → **Treatment CoM vs LoC** → **Outcomes**

income, motivation, tutoring, IQ, access to rss, socioeconomic background

Join by web:
PollEv.com/causal3900

Positivity assumption

Conditioning on SQT score yields conditional exchangeability.

- ▶ Try to apply matching:
  - ▶ Match people who received Certificate of Merit (CoM) with people who received Letter of Commendation (LoC)
  - ▶ Find people who got the same SQT score but who received a different type of recognition
  - ▶ Is there a potential problem here?

Positivity assumption

Conditioning on SQT score yields conditional exchangeability.

▶ Try to apply matching:

  ▶ Match people who received Certificate of Merit (CoM) with people who received Letter of Commendation (LoC)
  ▶ Find people who got the same SQT score but who received a different type of recognition
  ▶ Is there a potential problem here?

[handwritten: 10.5 ||]

People who have the same score also have the same type of recognition!

[handwritten: CoM Test Score]

Recall positivity: $P(A = a | L = \ell) > 0$ for all a and all ℓ

[handwritten: LoC]

Since treatment is assigned completely according to a cutoff, our setting violates the positivity assumption!

Local average treatment effect (LATE)

▶ Without positivity, we can't directly estimate ATE without strong assumptions

▶ Let's aim for an easier target: notice that students around the cutoff do have "similar" scores
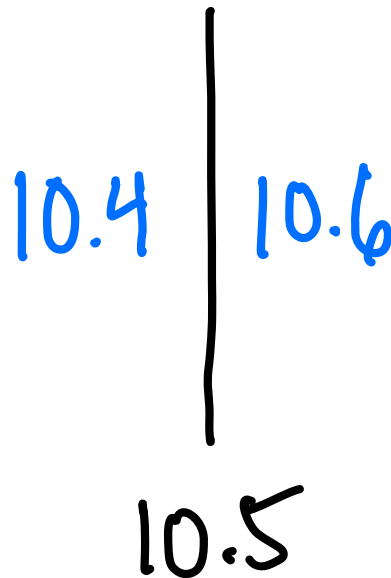
Local average treatment effect (LATE)

10 → 5

c

- ▶ Without positivity, we can't directly estimate ATE without strong assumptions

- ▶ Let's aim for an easier target: notice that students around the cutoff do have "similar" scores

- ▶ Average Treatment Effect for individuals *at the cut-off c*

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

- ▶ Does not tell us about treatment effect for everyone!

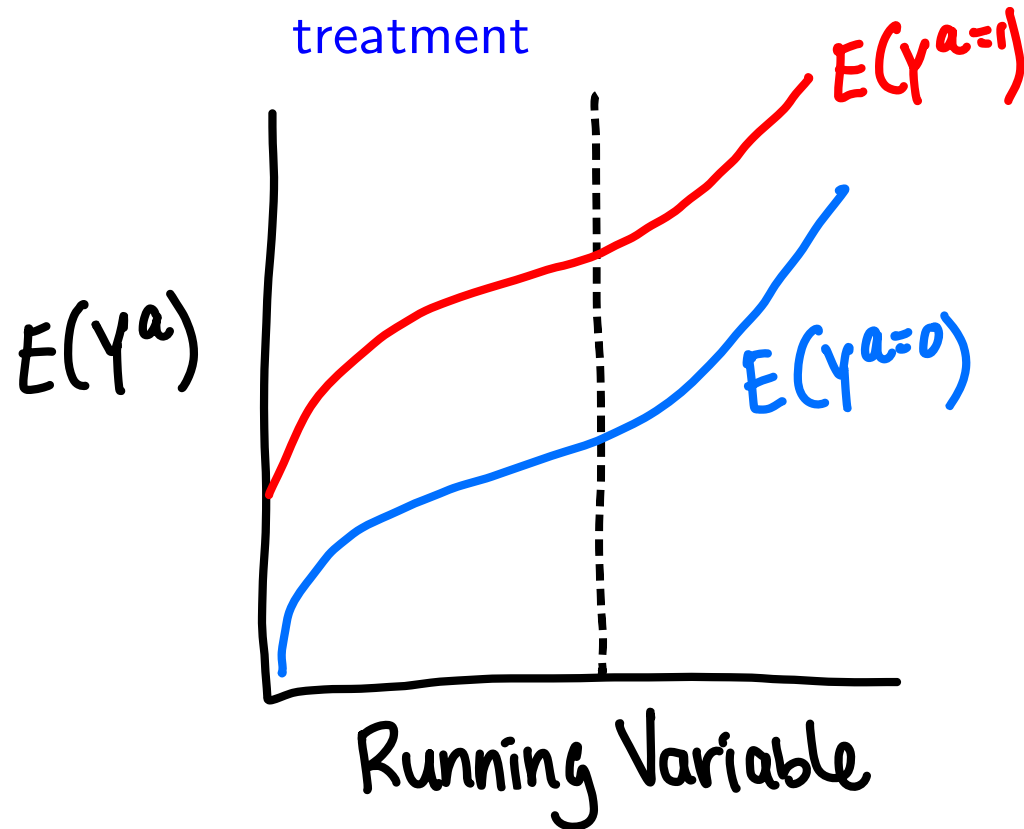# RDD: Requirements (& Assumptions)

$A = a$

$Y = Y^a$

Smoothness / Continuity

► Treatment of interest depends only on whether a **running variable** R is above or below a **cutoff** c

CoM only depends on being above or below SQT score cutoff

► Assume $E(Y^{a=1} \mid R = r)$ and $E(Y^{a=0} \mid R = r)$ vary smoothly

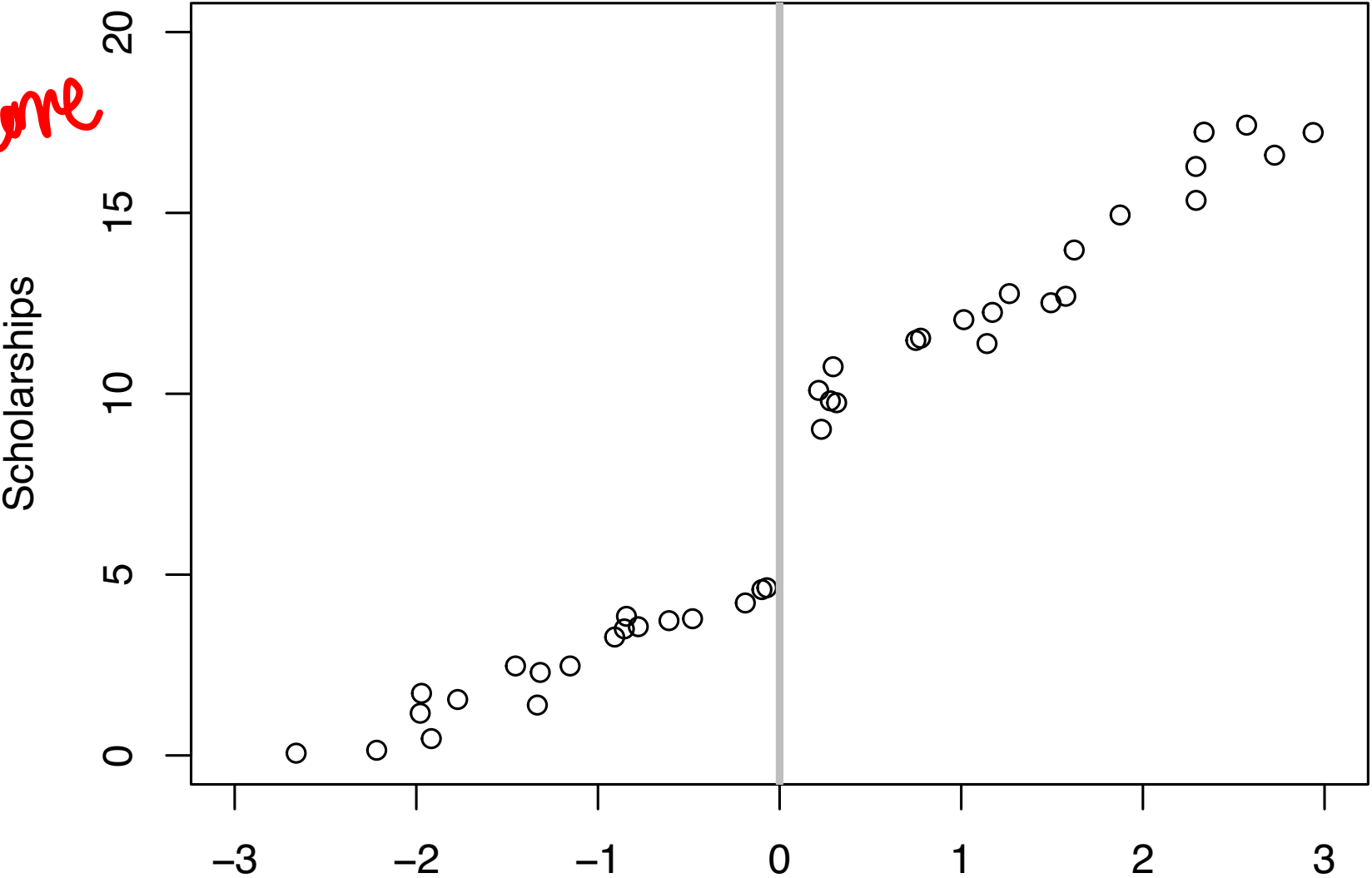The only thing that changes at the cutoff is treatment versus no treatment

$E(Y^a)$

$E(Y^{a=1})$

$E(Y^{a=0})$

Running Variable
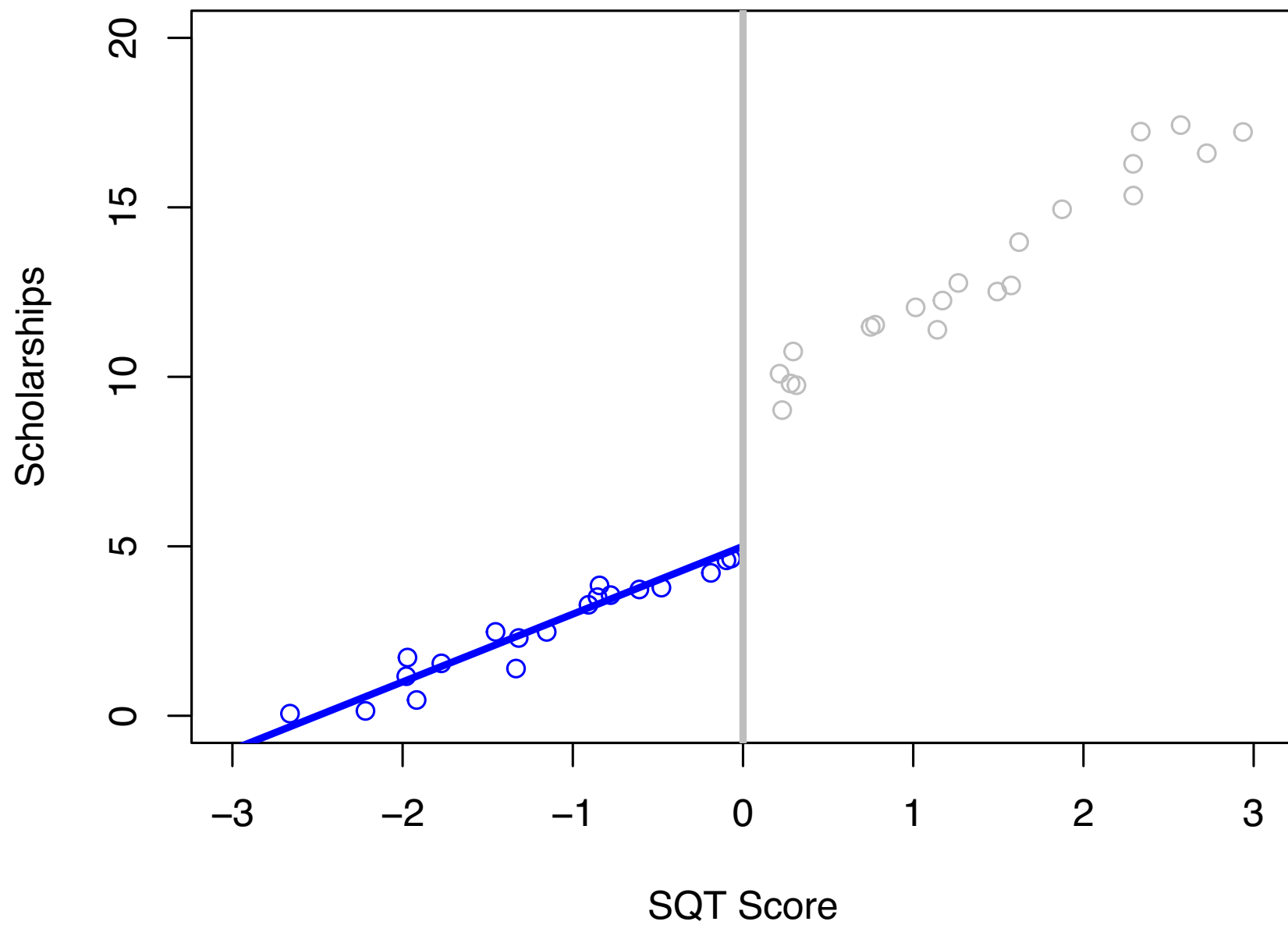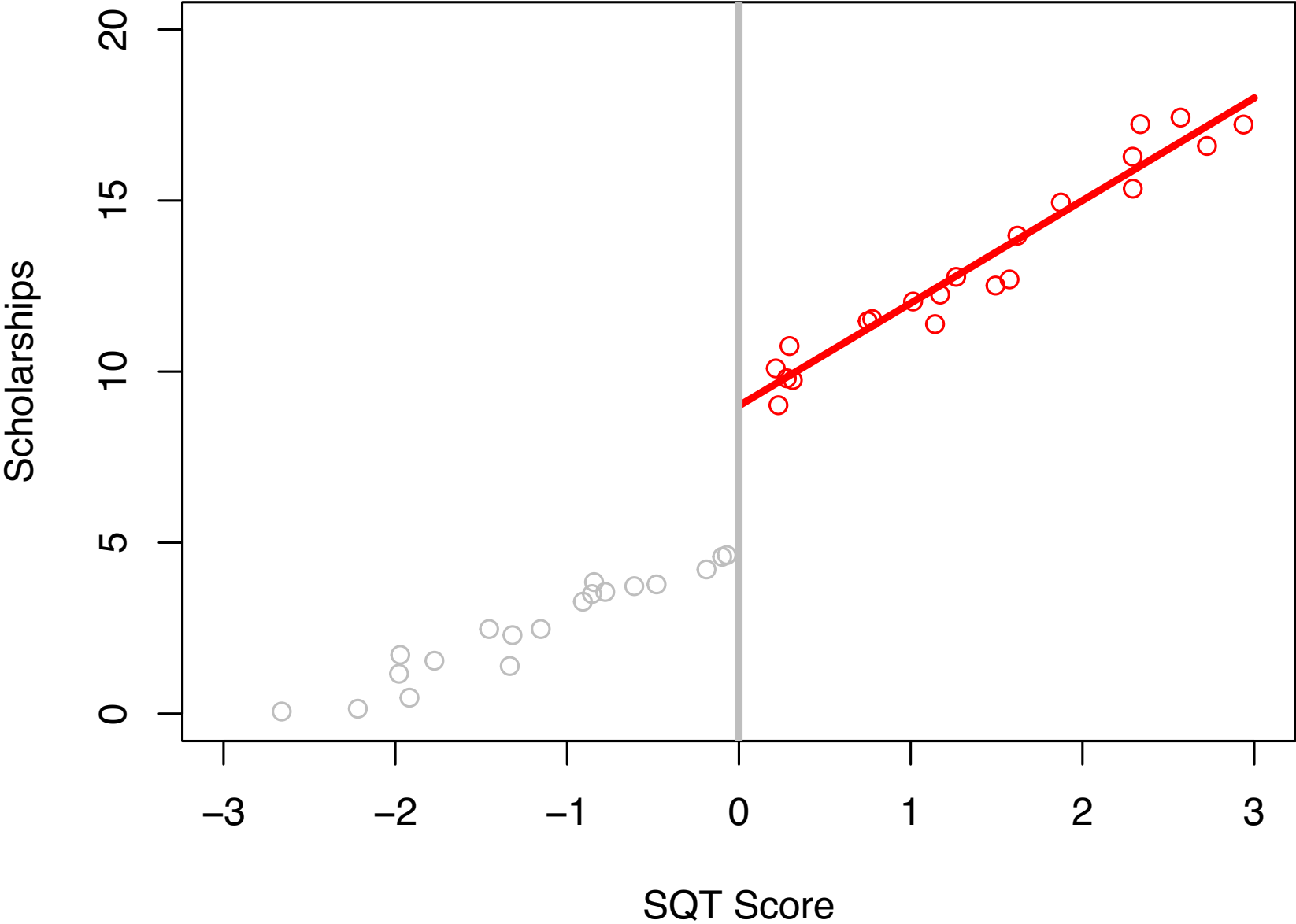
Y
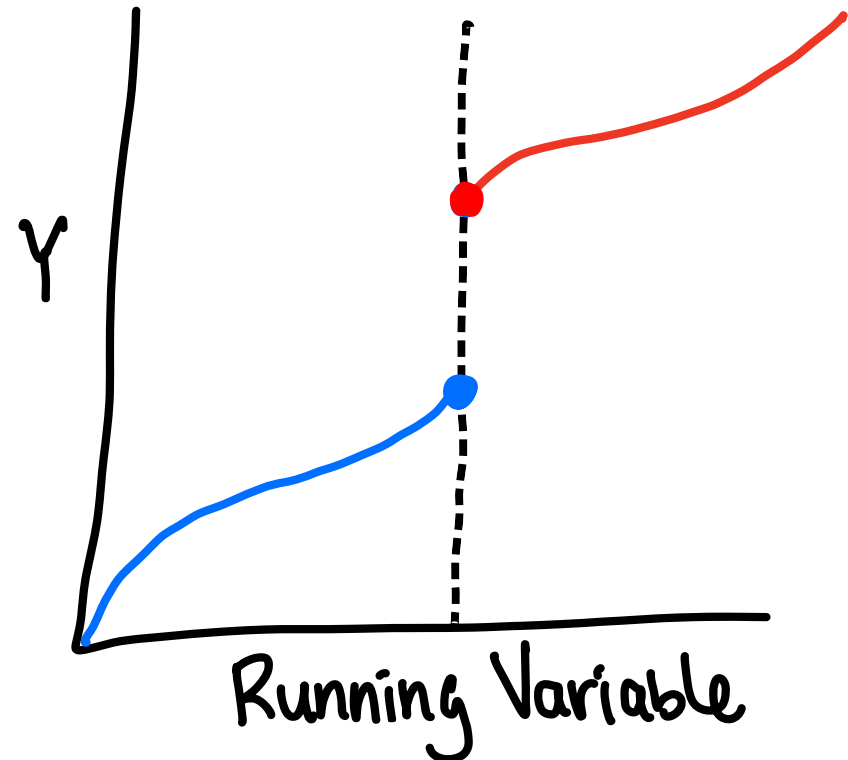
Running Variable
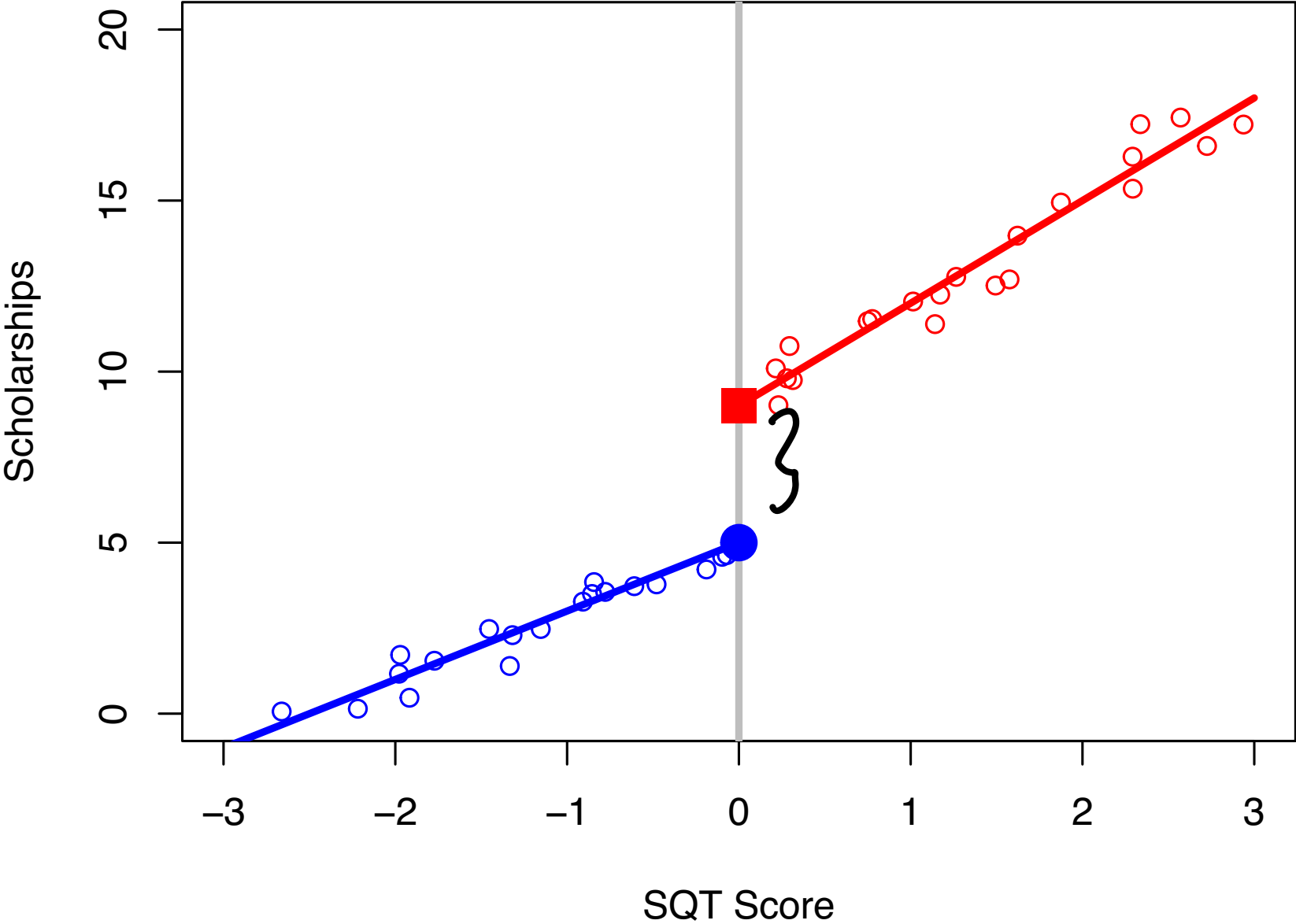
So how do we estimate the LATE?

So how do we estimate the LATE?

So how do we estimate the LATE?

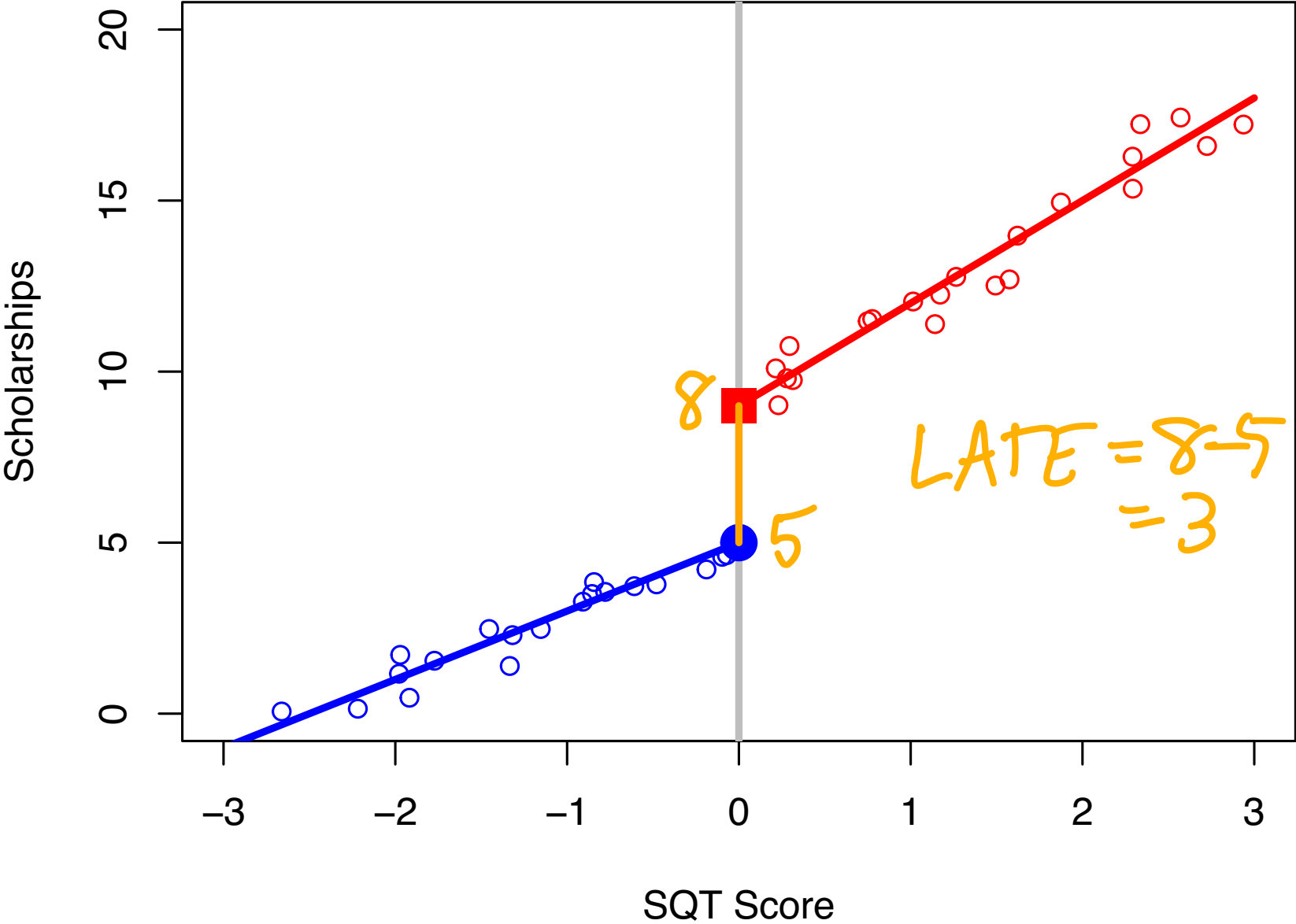# So how do we estimate the LATE?

So how do we estimate the LATE?

Wrap-up

In discussion section tomorrow...

- ► Bandwidths: How close to the cut-off do people need to be?
- ► Activity: Estimate the LATE with an RDD design in R

Discontinuities turn up in lots of places...

- ► Government benefits based on income requirements
- ► Healthcare decisions based on diagnostic test
- ► Policing policies based on jurisdiction lines
- ► ... what other examples can you think of?

What questions do you have?

Learning goals for today

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

1. Describe examples of when we would use a regression discontinuity design (RDD) to estimate causal effects
2. Explain the smoothness (continuity) assumption for RDD
3. Discuss the difference between the LATE and the ATE
4. Outline the steps it takes to estimate the LATE with a RDD

After today's class, read the following from Huntington-Klein:

► 20.1 How Does it Work?
► 20.3.4 Dealing with Bandwidths