CTE | Career & Technical Education RESEARCH NETWORK

Regression Discontinuity Designs: Theory

August 19, 2020

Dr. Scott Davis

Principal Researcher

The work of the CTE Research Network Lead is supported by the Institute of Education Sciences at the U.S. Department of Education with funds provided under the *Carl D. Perkins Career and Technical Education Act* through Grant R305N180005 to the American Institutes for Research (AIR). The work of the Network member projects is supported by the Institute. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education

Introduction

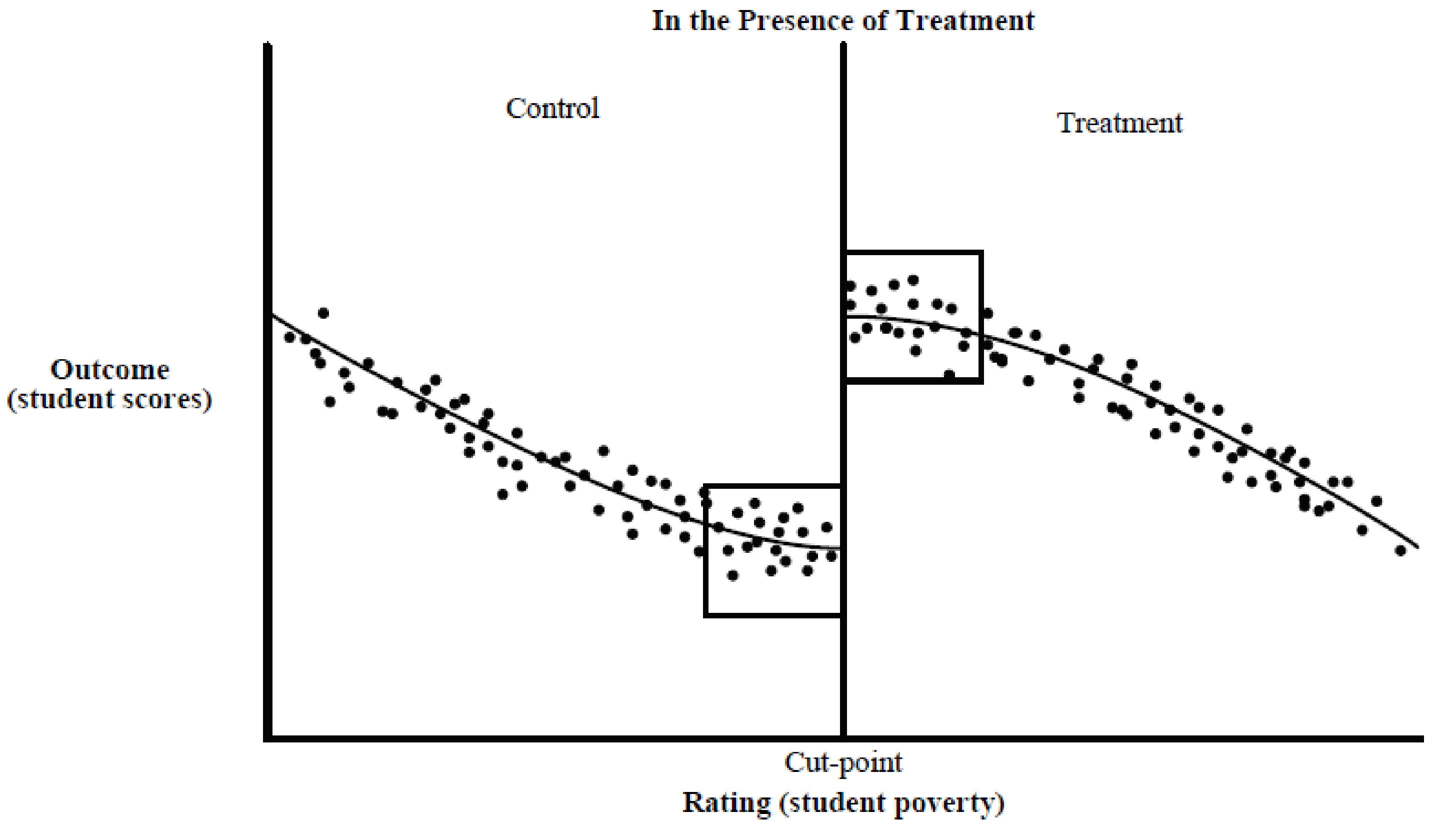
- The purpose is to introduce regression discontinuity design (RDD) and some practical considerations when using it.
- Structure and organization follows:
  - Jacob et al. (2012). *A Practical Guide to Regression Discontinuity*
  - Cattaneo, M., Idrobo, N., & Titiunik, R. (2020). *A Practical Introduction to Regression Discontinuity Designs: Foundations*
  - Cattaneo, M., Idrobo, N., & Titiunik, R. Forthcoming. *A Practical Introduction to Regression Discontinuity Designs: Extensions*
- We will not cover everything.

Source: Placeholder for sources and permissions (if needed).

Basics of RDD

- In some settings, program rules use cutoffs to determine who is eligible to participate in the program.
  - Schools with average test scores below a particular threshold may be required to do things other schools are not required to do.
  - Programs with limited slots may admit participants based on a test score or a composite score.
- Under certain conditions, you can compare outcomes for those just below and just above the cutoff to identify the effect of the program.
- The idea is that people near the cutoff on either side are like one another, but one group is in the program and the other is not.

Visualizing the Intuition Behind RDD



Source: Jacob, R., Zhu, P., Somers, M., & Bloom, H. (2012). *A Practical Guide to Regression Discontinuity*. New York, NY: MDRC. Reprinted with permission of MDRC.

Terms and Validity

- There are some terms used in RDDs that you need to know:
  - **Rating variable** (also called forcing variable or running variable)
  - **Sharp** RDD and **fuzzy** RDD
- RDDs are classified as sharp or fuzzy based on whether everyone on the program side of the cutoff actually participates in the program.
- Internal validity of an RDD depends on several conditions:
  - The program cannot influence the rating variable.
  - The cutoff is independent of the rating variable.
  - There are no other treatments assigned using the same cutoff.

Knowledge Check #1

RDD Scenarios

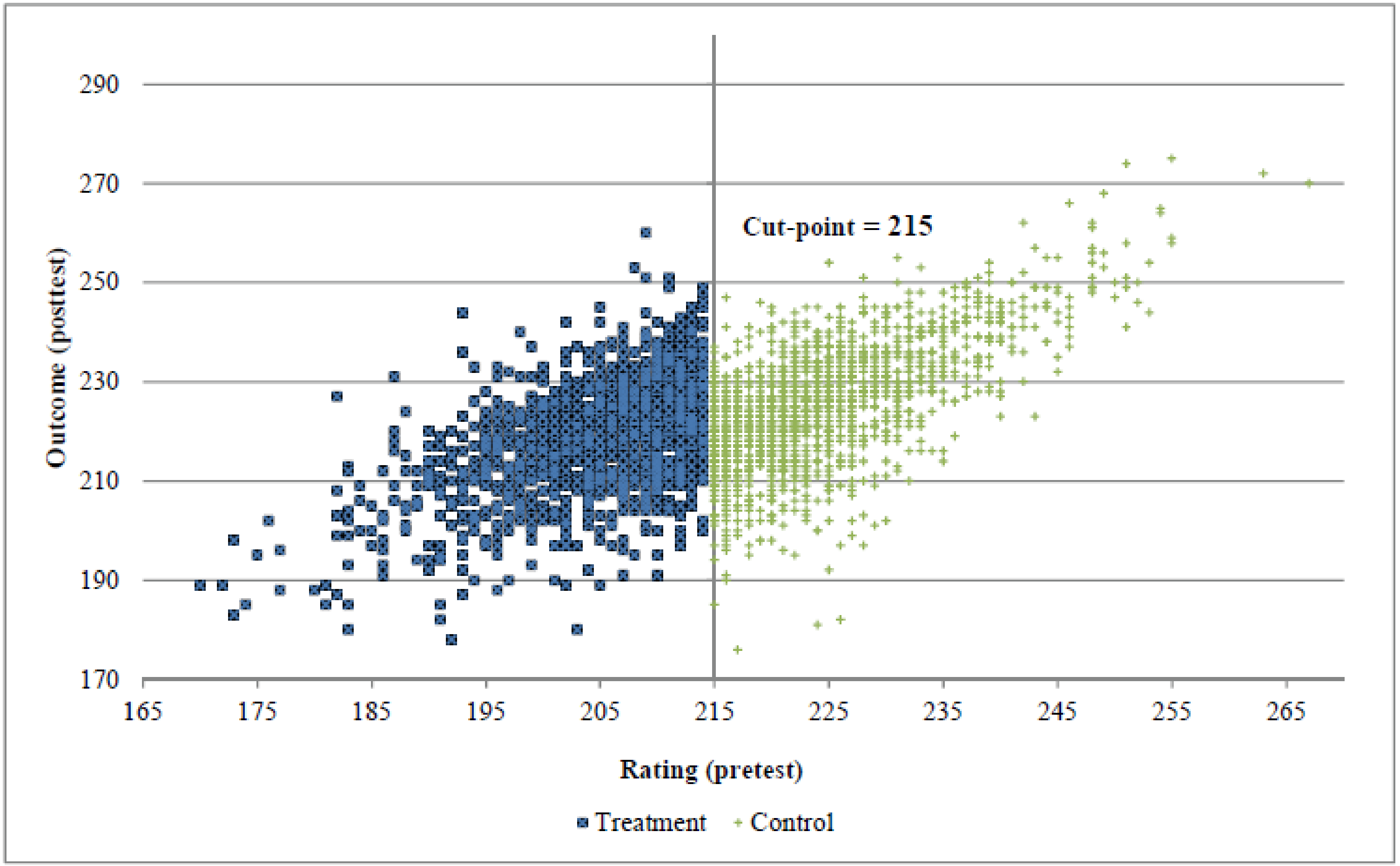
Knowledge Check #1

1. Students at a single high school who want to participate in a work-based learning program are chosen by a lottery for spots in the program.
2. High school students in grade 11 are given a math exam. Students scoring above 75 are eligible to enroll in college-level math courses in their senior year.
3. A high school youth apprenticeship program scores student applications and accepts students to fill 30 open spots by taking the 30 students with the highest-scoring applications.
4. A career readiness-focused soft skills course, open to all interested students, requires that students provide their GPAs when signing up.

Starting an RDD Analysis With Graphs

- It is good to start your analysis by plotting the data.
- Plot data points with
  - The rating variable on the horizontal axis.
  - The outcome variable on the vertical axis.
- Look for visual evidence of a jump at the cutoff score (though this can be difficult to see).

Graphing the Raw Data

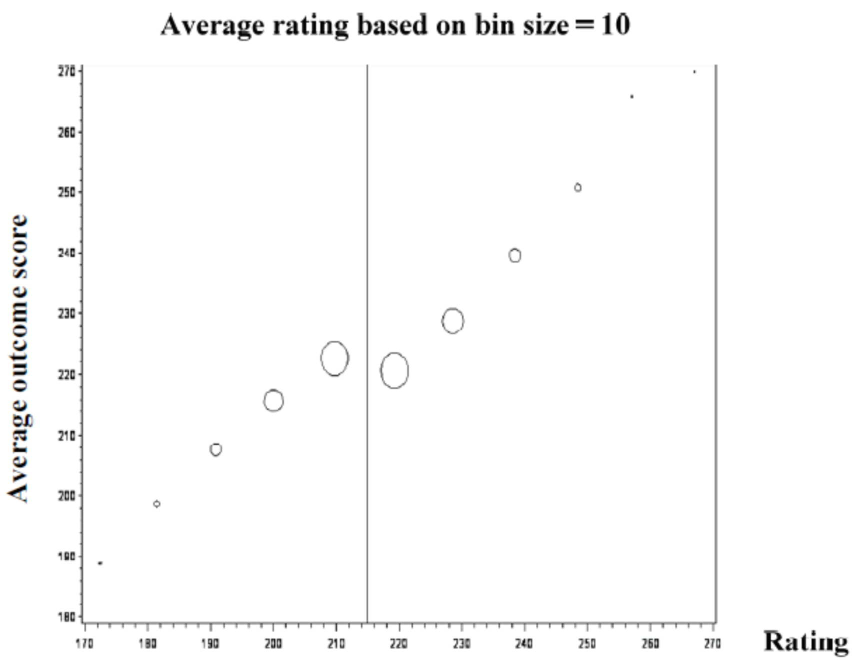


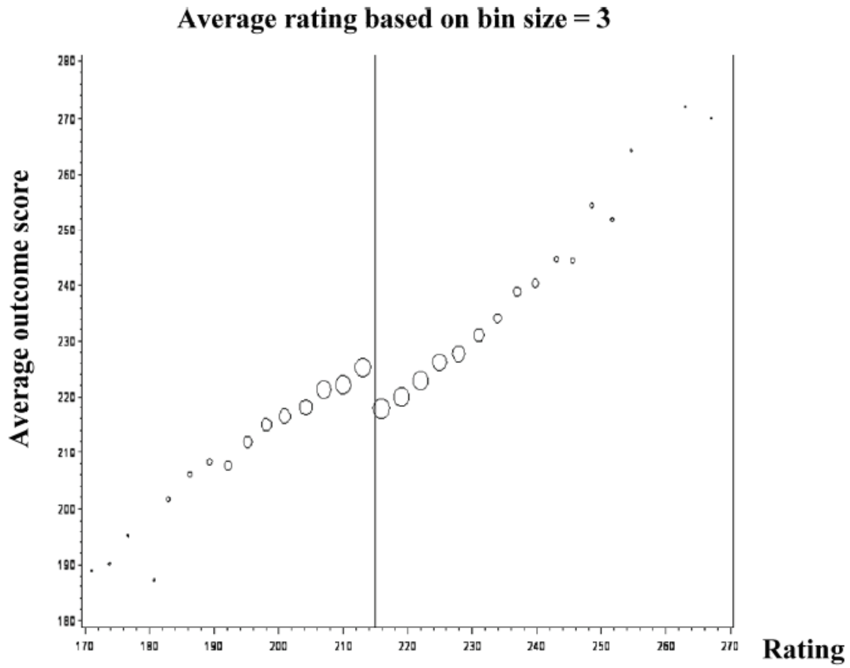
Source: Jacob, R., Zhu, P., Somers, M., & Bloom, H. (2012). *A Practical Guide to Regression Discontinuity*. New York, NY: MDRC. Reprinted with permission of MDRC.

Graphing the Data Using Smoothed Plots

- The next step is to look more closely at the data using smoothed plots.
- Graphing the data is a three-step process:
  - Divide the data into **bins** based on the value of the rating variable.
  - For each bin, calculate the average value of the outcome variable.
  - Plot the data for all the bins, using the midpoint of the bin and the average value of the outcome variable.
- You can use either **evenly-spaced** bins or **quantile-spaced** bins.
- There are ways to identify the optimal number of bins if desired.

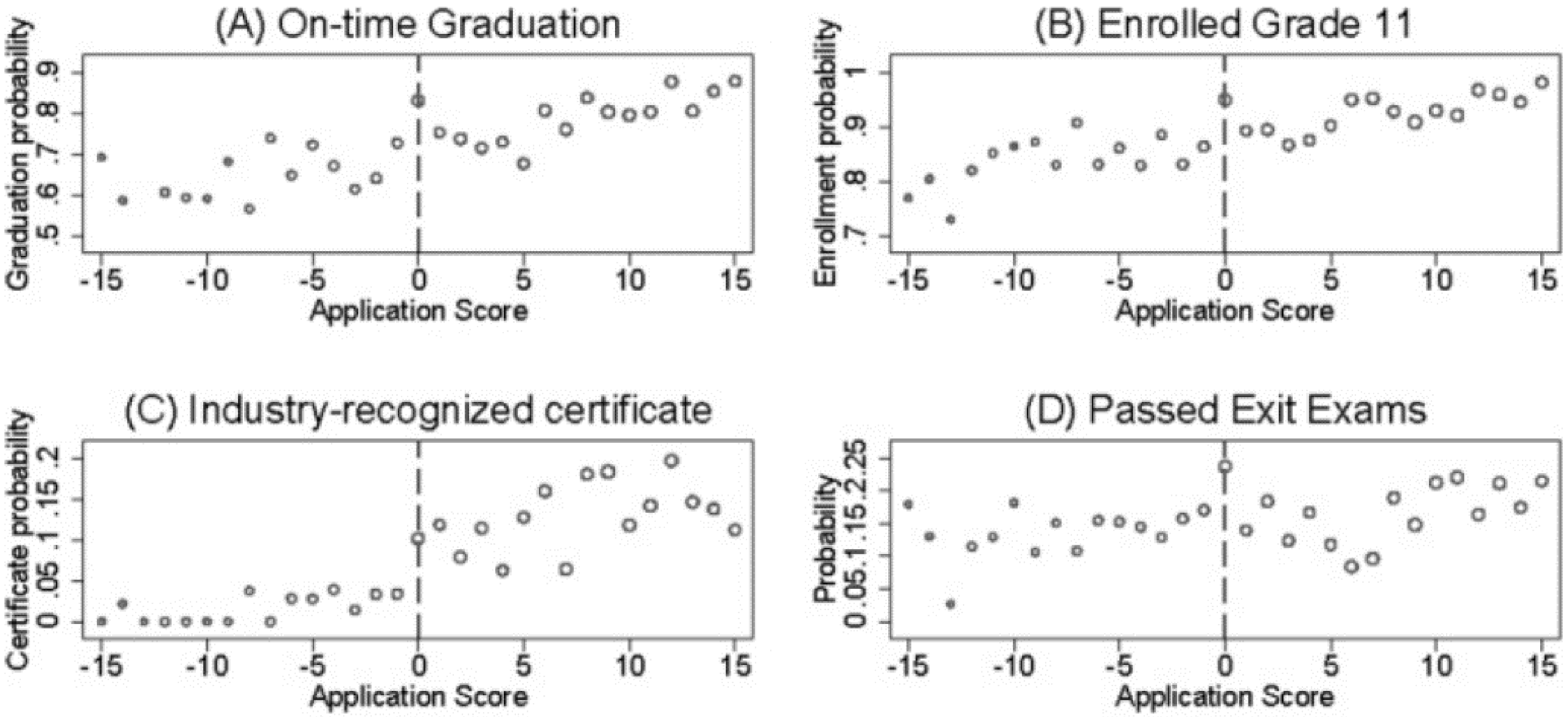
Smoothed Plots





Source: Jacob, R., Zhu, P., Somers, M., & Bloom, H. (2012). *A Practical Guide to Regression Discontinuity*. New York, NY: MDRC. Reprinted with permission of MDRC.

Smoothed Plots: CTE Example



Source: Dougherty, S. (2018). The Effect of Career and Technical Education on Human Capital Accumulation: Causal Evidence from Massachusetts. *Education Finance and Policy,* 13(2), 119–148.

Estimation: Two Strategies

- The goal of RDD is to use the jump at the cutoff score to estimate the impact of the program on the outcome.
- There are two approaches to doing this:
  - The **global** strategy uses all the data.
  - The **local** strategy uses only the observations close to the cutoff score.
- We will talk about both but focus more on the local strategy.

Global Strategy

- The global strategy approach to RDD uses all the data to estimate a regression model:

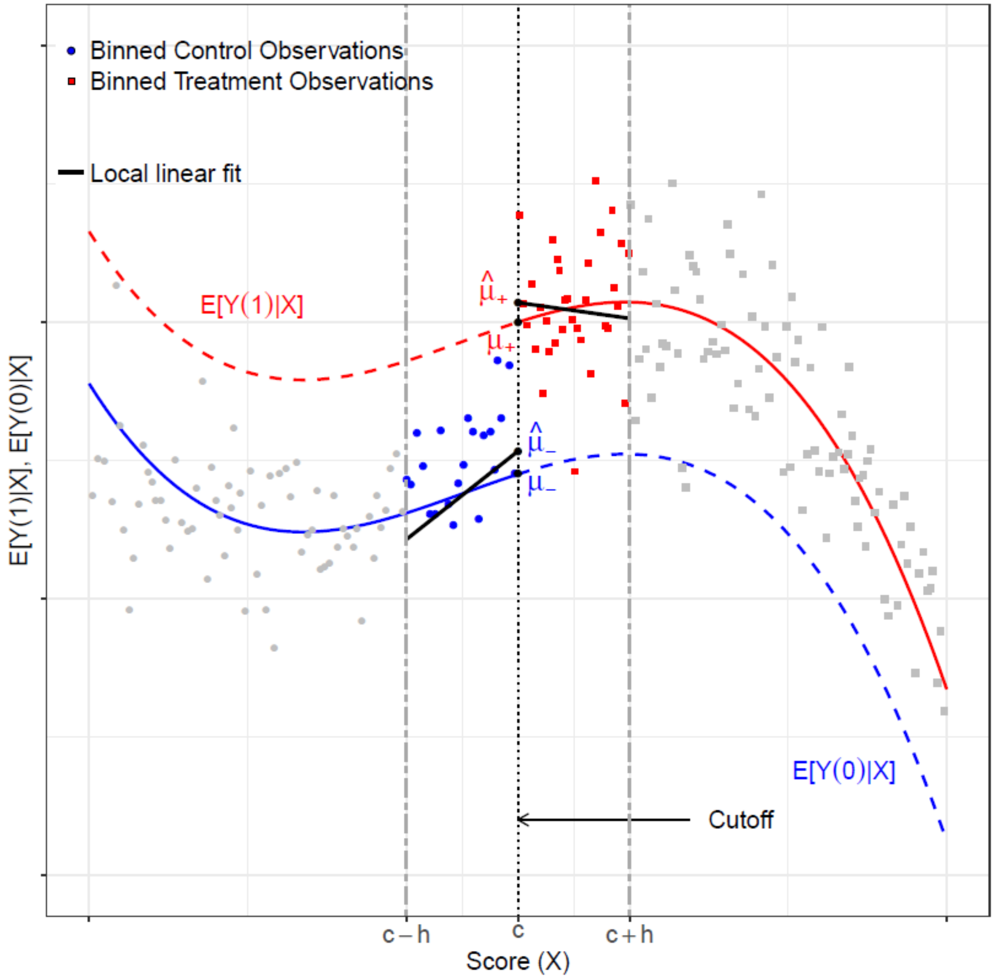
$$Y_i = \alpha_i + \beta_0 \cdot T_i + f(r_i) + \varepsilon_i$$

- Y_i is the outcome, T_i is an indicator for being in the treatment group, and r_i is the rating variable.
- The parameter β_0 is what we want to know—the effect of the program at the cutoff.
- We do not know the true $f(r_i)$, so in practice you may test different models (e.g., linear, quadratic, cubic).
- As sensitivity tests, you can re-run the models excluding some of the observations with the highest and lowest values of the rating variable.

Local Strategy

- The local strategy approach to RDD uses only observations close to the cutoff score to estimate the treatment effect.
- The range around the cutoff score that defines which observations are included is called the **bandwidth.**
- The process for estimating the impact of the program on the outcome is:
  - Estimate two separate regression models, one for below the cutoff and one for above it, using the model $Y_i = \alpha_i + f(r_i) + \varepsilon_i$.
  - Calculate two predicted values of the outcome at the cutoff, using each set of estimated regression coefficients.
  - Subtract the predicted value of the outcome using the below-cutoff regression coefficients from the predicted value using the above-cutoff regression coefficients.

Visualizing the Local Strategy



Source: Cattaneo, M., Idrobo, N., & Titiunik, R. (2020). A Practical Introduction to Regression Discontinuity Designs: Foundations (Elements in Quantitative and Computational Methods for the Social Sciences). Cambridge: Cambridge University Press. Reproduced with permission of the authors.

Choosing a Bandwidth

- To apply the local approach to RDD, you have to choose a bandwidth.
- There is a tradeoff you have to balance:
  - Choosing a smaller bandwidth always reduces bias (which is good).
  - But, a smaller bandwidth also reduces precision (which is not good).
- There are accepted empirical methods for choosing a bandwidth.

Validating RDD Estimates

- As part of an RDD analysis, you should assess whether critical assumptions for the validity of the design are satisfied.
- Some common approaches are:
  - Learn about how the rating variable is created and how the cutoff score was determined.
  - Analyze the probability of participating in the program.
  - Analyze variables that should not be affected by the program, such as **placebo outcomes.**
  - Look for evidence of manipulation in the density of the rating variable.

Other Sensitivity Tests

- There are other types of sensitivity tests you can use:
  - Checking for **placebo cutoffs**
    - Look for a jump in the outcome at other values of the cutoff.
  - Using the **donut hole approach** to test for manipulation of the rating variable
    - Drop observations near the cutoff.
  - Using different bandwidths
    - Use bandwidths slightly larger and smaller than the optimal bandwidth.

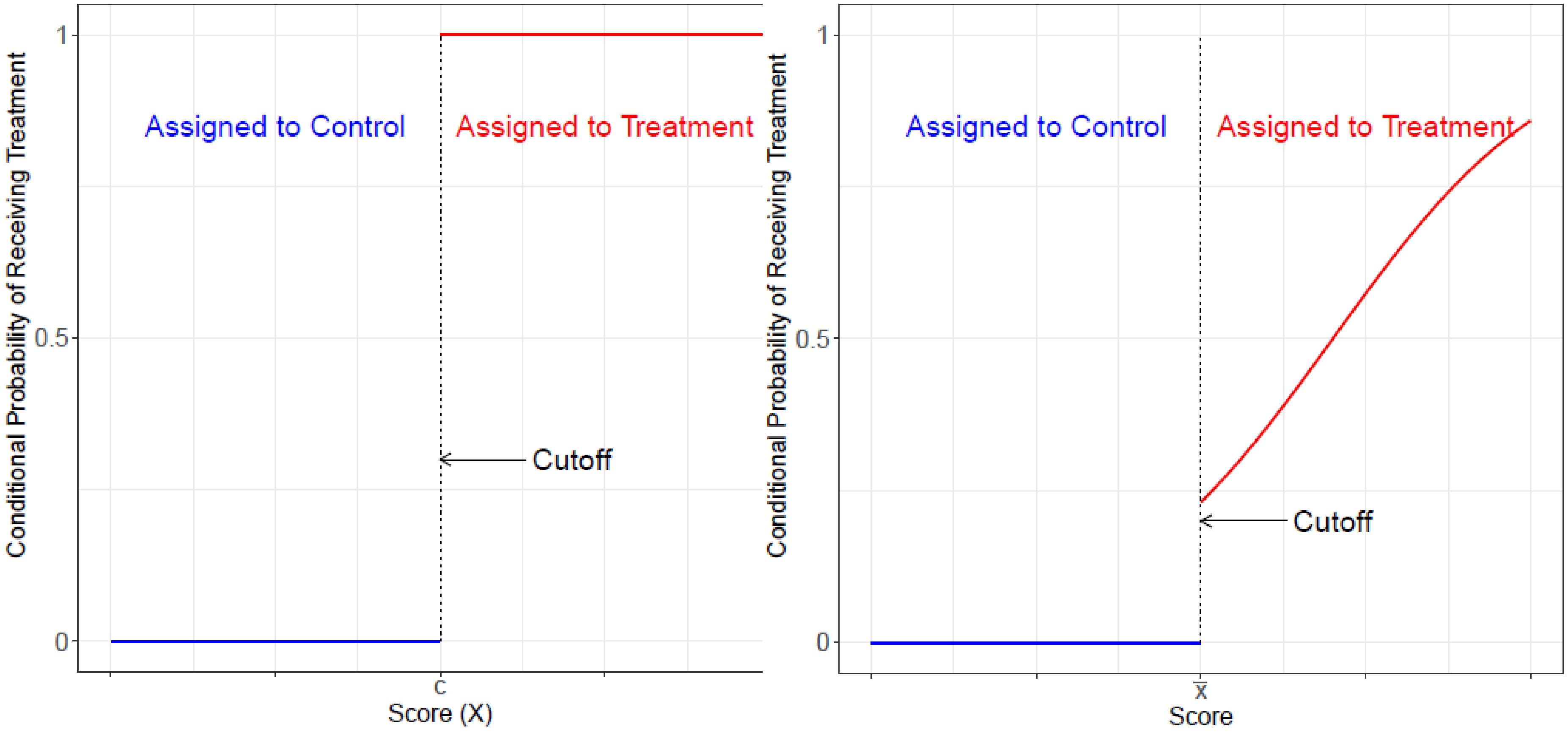
Knowledge Check #2

Validation Tests

Fuzzy RDD

- Up to this point, we have assumed that the cutoff completely determines whether an individual is in the program.
- But program participation may not be completely determined by the cutoff score.
  - Some people with scores above the cutoff may not enroll (**no-shows**).
  - Others with scores below the cutoff may find a way to enroll (**crossovers**).
- Non-compliance with assignment is a well-known problem in experimental settings.
- In an RDD, we call an application with non-compliance a **fuzzy RDD.**

Visualizing Sharp Versus Fuzzy RDDs



(a) Sharp RD (b) Fuzzy RD (One-Sided)

Source: Cattaneo, Matias D., Idrobo, Nicolás, and Titiunik, Rocío. Forthcoming. A Practical Introduction to Regression Discontinuity Designs: Extensions. Cambridge Elements: Quantitative and Computational Methods for the Social Sciences. Cambridge University Press. Reproduced with permission of the authors.

Estimation in a Fuzzy RDD Context

- Estimation in a fuzzy RDD context is not the same as the process in a sharp RDD setting.
  - Data on either side of cutoff include a mix of participants and non-participants.
- You can just ignore non-compliance and estimate the intent-to-treat (ITT) effect.
  - The estimate tells you the effect of the program on those eligible for it.
- A more common approach is to estimate the local average treatment effect (LATE).
  - The estimate tells you the effect of the program on compliers.

Estimation in a Fuzzy RDD Context (cont.)

- The LATE is estimated equal to the jump in the outcome at the cutoff divided by the jump in the probability of participating at the cutoff:

$$\text{LATE} = \widehat{\beta_{\text{outcome}}} / \widehat{\beta_{\text{assignment}}}$$

- In practice, to estimate the LATE, researchers use an **instrumental variables** (IV) approach to estimate two equations:

$$T_i = \alpha_1 + \gamma_0 \cdot D_i + f_1(r_i) + \epsilon_i$$
$$Y_i = \alpha + \beta_0 \cdot T_i + f_2(r_i) + \mu_i$$

- T_i is an indicator for treatment, and D_i is an indicator for treatment assignment.
- The two-equation model is estimated using two-stage least squares (2SLS).
- You can test different functional forms for $f_1(r_i)$ and $f_2(r_i)$.

Other Topics in RDD

- **Precision**
  - RDDs require much larger sample sizes than experiments (~ 2x to 4x).
  - The need for a larger sample to compensate for the reduced precision of an RDD is called the **design effect.**
    - Larger for non-parametric methods than parametric methods
    - Larger for fuzzy RDD than sharp

- **Extensions**
  - Multiple cutoffs
  - Multiple scores
  - Clustered RDDs
  - Combinations with experiments
  - Adding a nonequivalent group

WWC Standards for RDDs

- WWC standards for RDDs are based on five key aspects of the design:
  1. Integrity of the rating variable
  2. Sample attrition
  3. Continuity
  4. Bandwidth/functional form
  5. Fuzzy RDD
- To be rated **Meets WWC RDD Standards Without Reservations**, the study has to completely satisfy all five individual standards.
- To be rated **Meets WWC RDD Standards With Reservations**, the study has to at least partially satisfy 1, 4, and 5 plus either 2 or 3.

Regression Discontinuity Designs: Practice

Overview

- The goal is to let you practice carrying out RD analysis using real data.
  - First, I will demonstrate some examples.
  - Second, you will have time to work in small groups through a series of activities.
- The first part is meant to show you how to implement some of the techniques we discussed.
  - All commands are from **Stata** but have **R** equivalents.
- After your independent work time, you will meet with trainers in virtual office hours.

Example Analysis Scenario

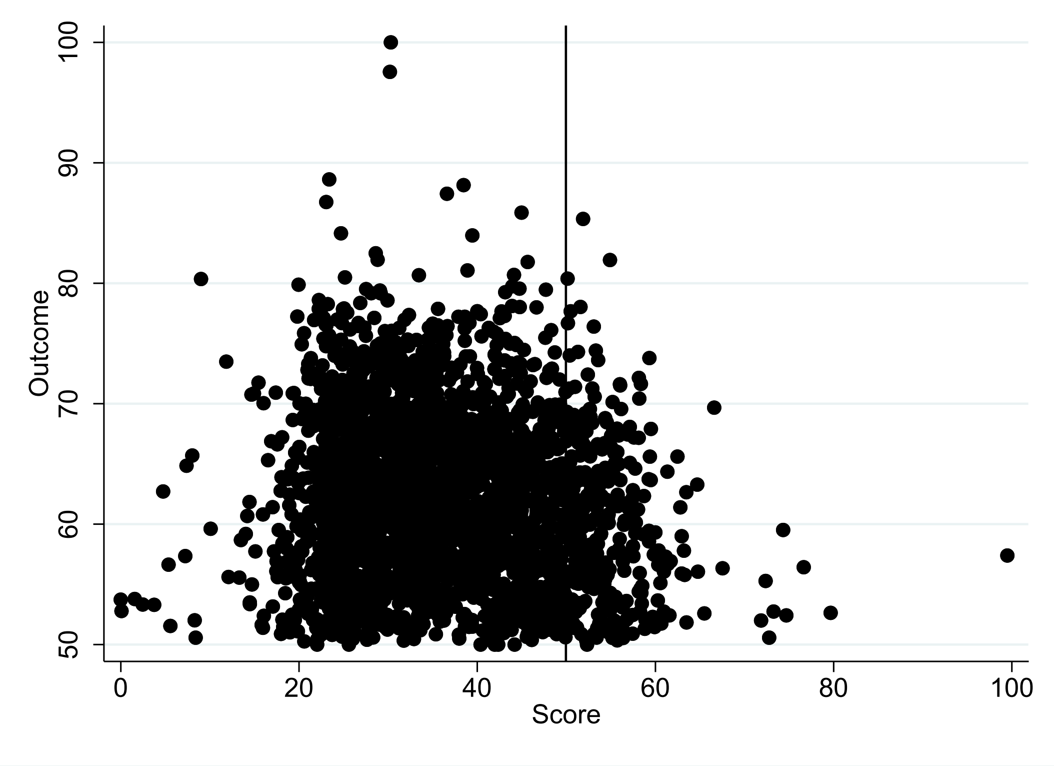
- Suppose that a state features a number of CTE-focused high schools.
  - Students have to apply for admission.
  - Admission is based on several factors that are combined into a single score for each applicant, with a threshold for acceptance.
- Assume all students who are accepted to the CTE-focused schools actually enroll.
- We can use **RD** to estimate the impact of attending a CTE-focused high school on an outcome of interest.

Example Analysis Variables and Approach

- The data we will use are all hypothetical.
  - Derived from Meyersson (2014) but manipulated
- The key variables in the data set are:
  - **Y**, the outcome of interest (Grade 12 test score)
  - **X**, the rating variable (application score; assume a cutoff of 50)
  - **Z1**, an example student characteristic
- This analysis is a **sharp RDD** because the treatment (attending a CTE-focused high school) is completely determined by the application score cutoff.
  - We assumed that all accepted students will enroll.

Step 1: Plot the Raw Data

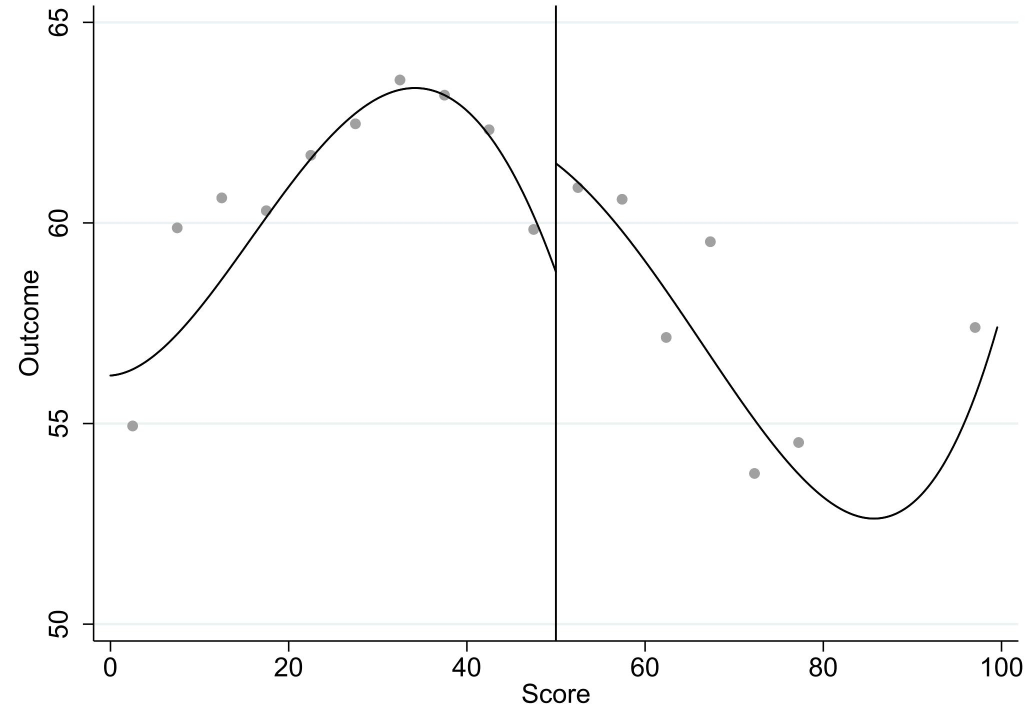
```
twoway (scatter Y X,
mcolor(black) xline(50,
lcolor(black))),
graphregion(color(white))
ytitle(Outcome) xtitle(Score)
```



Step 2: Plot the Data Using Bins (20)

Number of bins = **20**

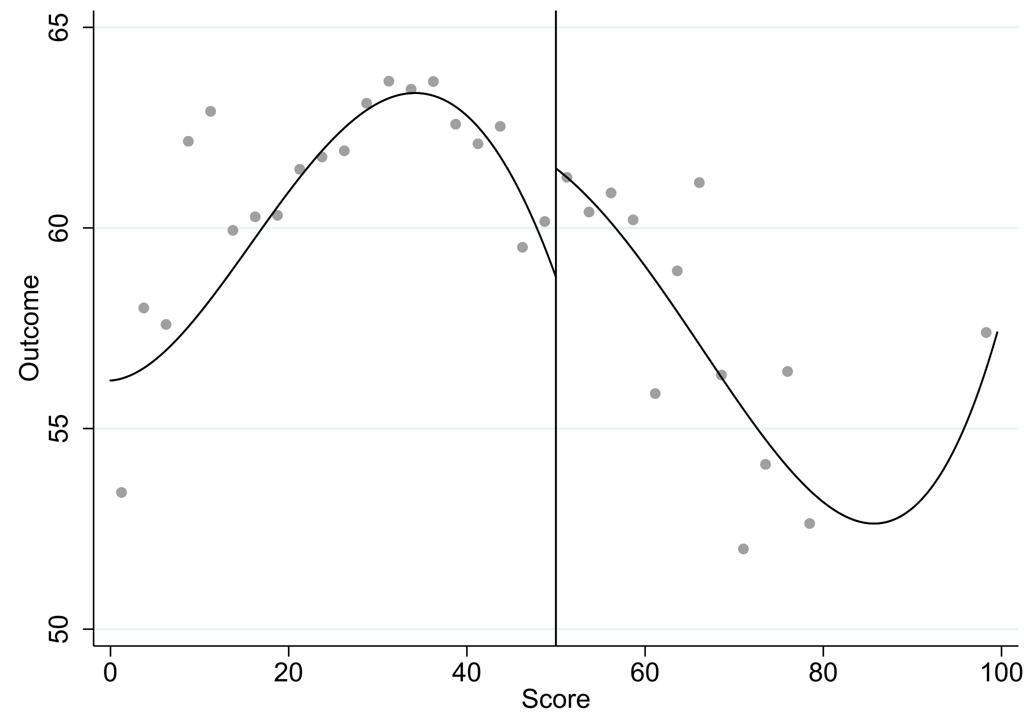
```
rdplot Y X, c(50) nbins(10 10)
binselect(esmv)
graph_options(
graphregion(color(white))
xtitle(Score) ytitle(Outcome)
legend(off))
```



Step 2: Plot the Data Using Bins (40)

Number of bins = **40**

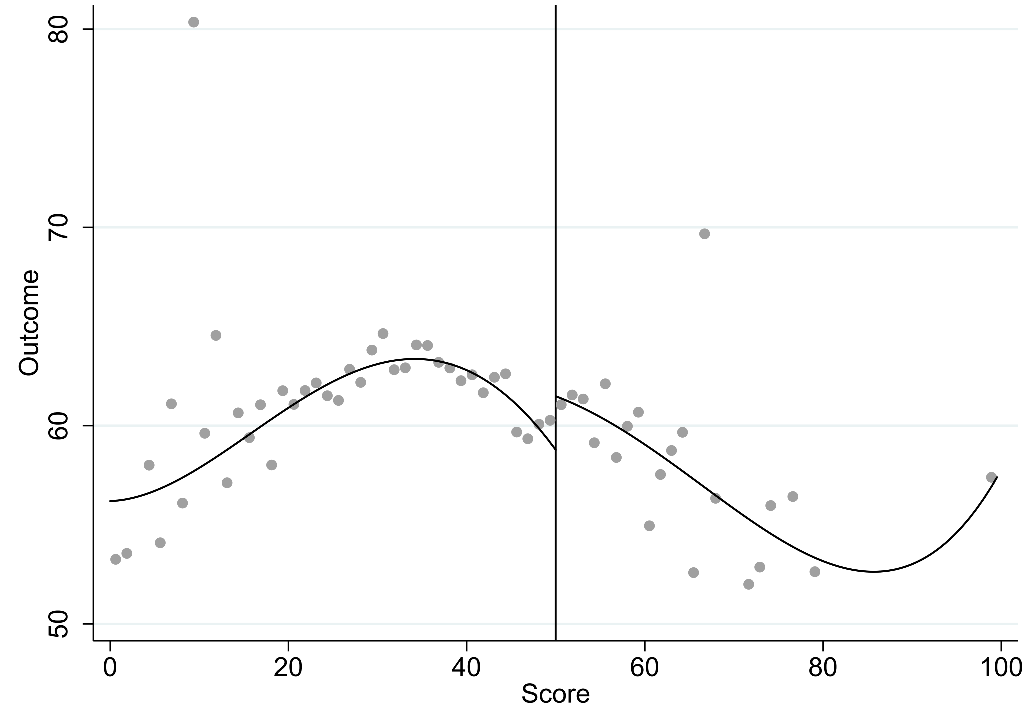
```
rdplot Y X, c(50) nbins(20 20)
binselect(esmv)
graph_options(
graphregion(color(white))
xtitle(Score) ytitle(Outcome)
legend(off))
```



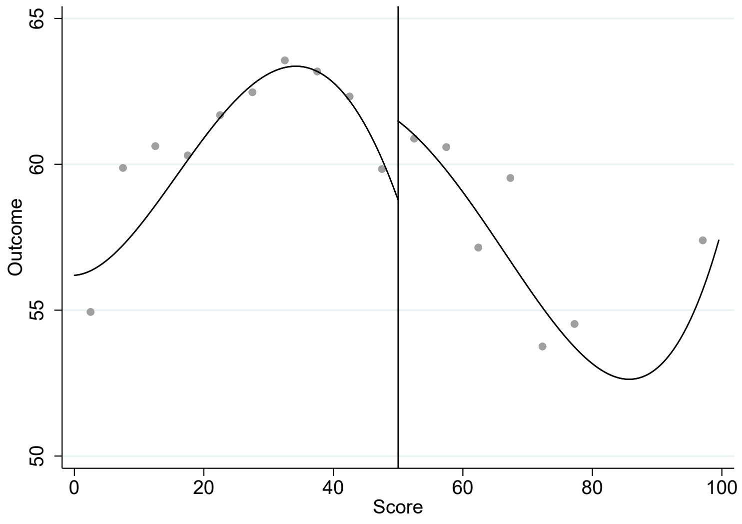
Step 2: Plot the Data Using Bins (80)

Number of bins = **80**

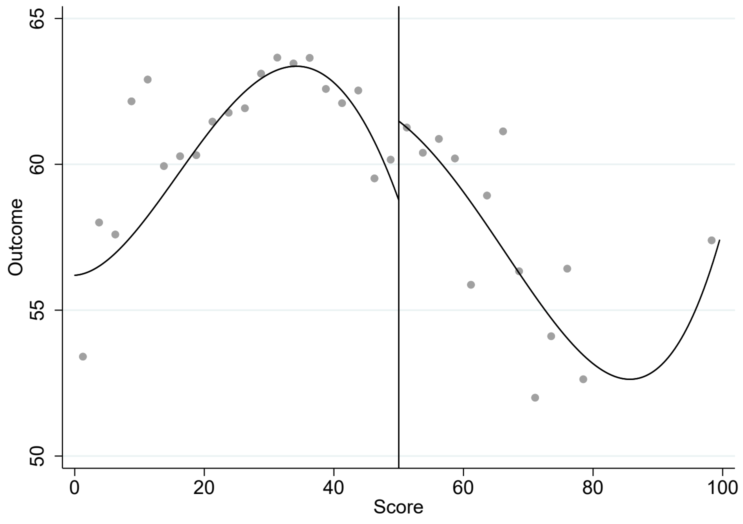
```
rdplot Y X, c(50) nbins(40 40)
binselect(esmv)
graph_options(
graphregion(color(white))
xtitle(Score) ytitle(Outcome)
legend(off))
```



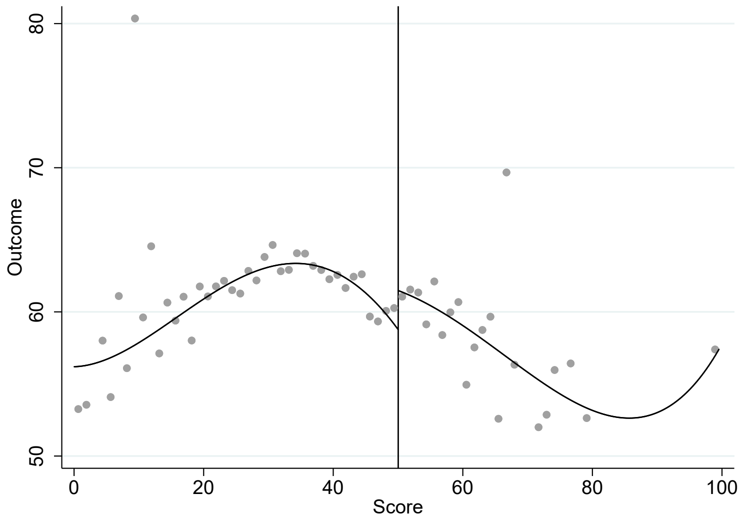
Step 2: Plot the Data Using Bins (Summary)



20 Bins



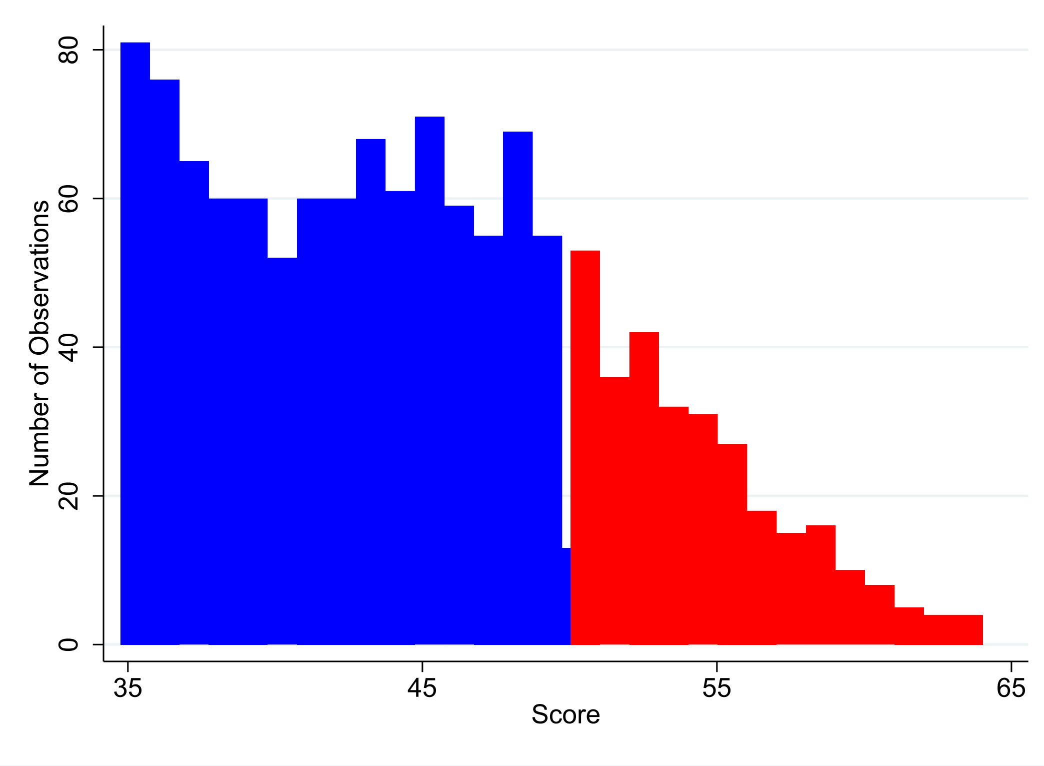
40 Bins



80 Bins

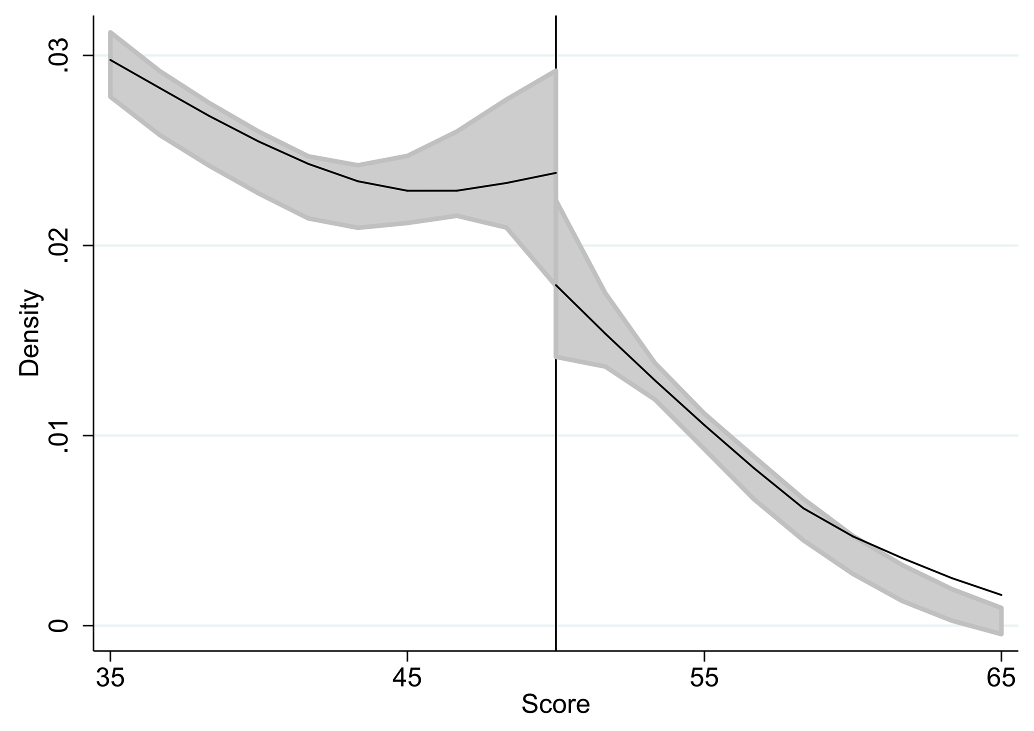
Step 3: Check for Manipulation (Histogram)

```
qui rddensity X, c(50)

local bandwidth_left = e(h_l)

local bandwidth_right = e(h_r)

twoway (histogram X if X >= 50 -
`bandwidth_left' & X < 50, freq
width(1) color(blue)) (histogram X
if X >= 50 & X <= 50 +
bandwidth_right', freq width(1)
color(red)), xlabel(35(10)65)
graphregion(color(white))
xtitle(Score) ytitle(Number of
Observations) legend(off)
```



Step 3: Check for Manipulation (Density)

```
rddensity X, c(50) plot
plot_range(35 65)
graph_options(xlabel(35(10)65)
graphregion(color(white))
xtitle(Score) ytitle(Density)
legend(off))
```



Step 3: Check for Manipulation (Binomial Test)

```
sum T if abs(X-C) <= 1

bitesti r(N) r(mean) 0.5
```

```
. sum $T if abs($X-$C) <= 1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
           T |        100         .53    .5016136          0          1

. bitesti r(N) r(mean) 0.5

        N   Observed k   Expected k   Assumed p   Observed p
------------------------------------------------------------
      100         53           50       0.50000      0.53000

  Pr(k >= 53)            = 0.308650  (one-sided test)
  Pr(k <= 53)            = 0.757941  (one-sided test)
  Pr(k <= 47 or k >= 53) = 0.617299  (two-sided test)
```

Step 4: Estimate the Treatment Effect With an Ad-Hoc Bandwidth

Regression approach

```
qui reg Y X if X < 50 & X >= 40
matrix coef_left = e(b)
local intercept_left = coef_left[1, 2] +
50*coef_left[1,1]
qui reg Y X if X >= 50 & X <= 60
matrix coef_right = e(b)
local intercept_right = coef_right[1, 2] +
50*coef_right[1,1]
local difference = `intercept_right' -
`intercept_left'
display "The RD estimator is `difference'"
```

```
. qui reg $Y $X if $X < 50 & $X >= 40

. matrix coef_left = e(b)

. local intercept_left = coef_left[1, 2] + 50*coef_left[1,1]

. qui reg $Y $X if $X >= 50 & $X <= 60

. matrix coef_right = e(b)

. local intercept_right = coef_right[1, 2] + 50*coef_right[1,1]

. local difference = `intercept_right' - `intercept_left'

. display "The RD estimator is `difference'"
The RD estimator is 2.151067285696037
```

Step 4: Estimate the Treatment Effect With an Ad-Hoc Bandwidth

Local linear model, uniform kernel

```
rdrobust Y X, c(50) kernel(uniform) p(1) h(10)
```

```
. qui reg $Y $X if $X < 50 & $X >= 40

. matrix coef_left = e(b)

. local intercept_left = coef_left[1, 2] + 50*coef_left[1,1]

. qui reg $Y $X if $X >= 50 & $X <= 60

. matrix coef_right = e(b)

. local intercept_right = coef_right[1, 2] + 50*coef_right[1,1]

. local difference = `intercept_right' - `intercept_left'

. display "The RD estimator is `difference'"
The RD estimator is 2.151067285696037
```

```
. rdrobust $Y $X, c(50) kernel(uniform) p(1) h(10)

Sharp RD estimates using local polynomial regression.

      Cutoff c = 50 | Left of c  Right of c            Number of obs =        2629
--------------------+----------------------            BW type       =      Manual
      Number of obs |      2314         315            Kernel        =     Uniform
Eff. Number of obs  |       608         280            VCE method    =          NN
     Order est. (p) |         1           1
     Order bias (q) |         2           2
        BW est. (h) |    10.000      10.000
        BW bias (b) |    10.000      10.000
          rho (h/b) |     1.000       1.000

Outcome: Y2. Running variable: X2.
--------------------------------------------------------------------------------
             Method |   Coef.    Std. Err.     z     P>|z|     [95% Conf. Interval]
--------------------+-----------------------------------------------------------
       Conventional |  2.1511     .90724    2.3710   0.018     .372916      3.92922
             Robust |    -          -       1.6364   0.102    -.427825      4.75542
--------------------------------------------------------------------------------
```

Step 4: Estimate the Treatment Effect With an Ad-Hoc Bandwidth

Local polynomial (quadratic) model, uniform kernel

```
rdrobust Y X, c(50) kernel(uniform) p(2) h(10)
```

```
. rdrobust $Y $X, c(50) kernel(uniform) p(2) h(10)

Sharp RD estimates using local polynomial regression.

    Cutoff c = 50 | Left of c  Right of c            Number of obs =       2629
-------------------+----------------------            BW type       =     Manual
     Number of obs |      2314        315            Kernel        =    Uniform
Eff. Number of obs |       608        280            VCE method    =         NN
     Order est. (p) |         2          2
     Order bias (q) |         3          3
        BW est. (h) |    10.000     10.000
        BW bias (b) |    10.000     10.000
          rho (h/b) |     1.000      1.000

Outcome: Y2. Running variable: X2.
--------------------------------------------------------------------------------
            Method |   Coef.    Std. Err.    z     P>|z|    [95% Conf. Interval]
-------------------+------------------------------------------------------------
      Conventional |  2.1638     1.3223   1.6364   0.102   -.427825     4.75542
            Robust |     -          -     0.9279   0.353   -1.81448     5.07729
--------------------------------------------------------------------------------
```

Step 5: Estimate the Treatment Effect With the Optimal Bandwidth

Local polynomial (quadratic) model, uniform kernel, symmetric bandwidth

```
rdrobust Y X, c(50) kernel(uniform) p(2) bwselect(mserd)
```

```
. rdrobust $Y $X, c(50) kernel(uniform) p(2) bwselect(mserd)

Sharp RD estimates using local polynomial regression.

      Cutoff c = 50 | Left of c  Right of c            Number of obs =       2629
-------------------+----------------------            BW type       =      mserd
      Number of obs |      2314         315            Kernel        =    Uniform
Eff. Number of obs |       584         277            VCE method    =         NN
     Order est. (p) |         2           2
     Order bias (q) |         3           3
        BW est. (h) |     9.517       9.517
       BW bias (b) |    16.973      16.973
         rho (h/b) |     0.561       0.561

Outcome: Y2. Running variable: X2.
--------------------------------------------------------------------------------
            Method |   Coef.    Std. Err.    z     P>|z|    [95% Conf. Interval]
-------------------+------------------------------------------------------------
      Conventional |  2.6134     1.3397   1.9508   0.051   -.012343     5.23916
            Robust |     -          -     1.6310   0.103   -.479662     5.23579
--------------------------------------------------------------------------------
```

Step 5: Estimate the Treatment Effect With the Optimal Bandwidth

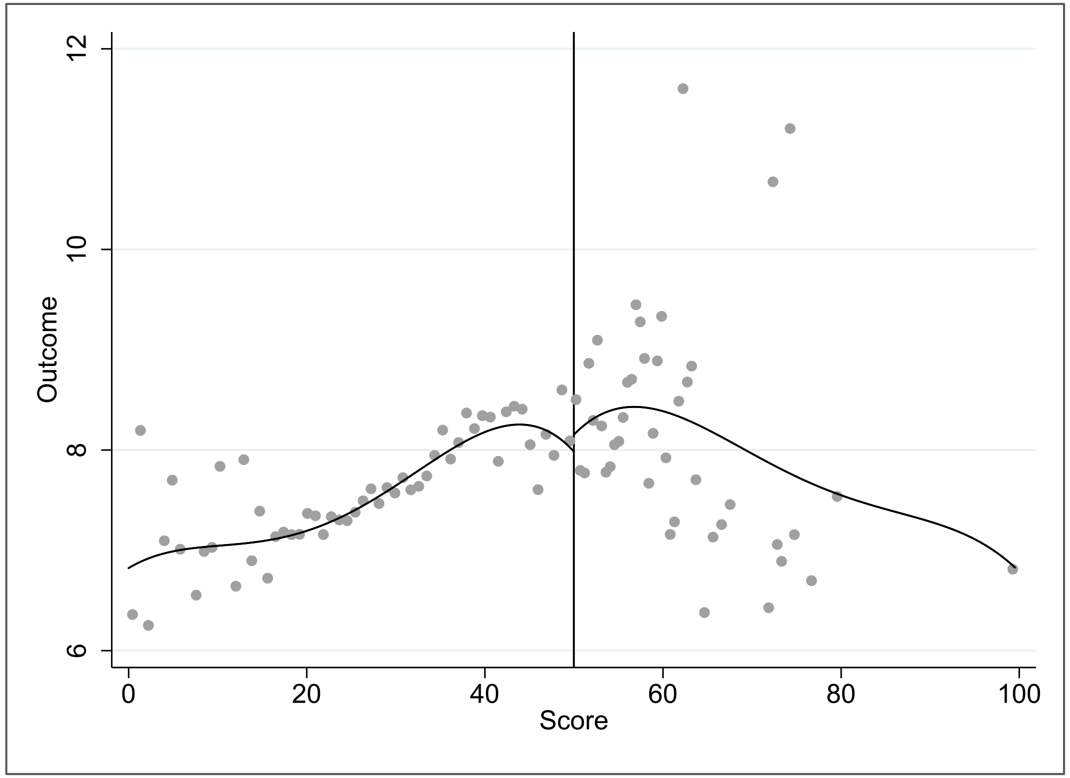
Local polynomial (quadratic) model, uniform kernel, asymmetric bandwidth

```
rdrobust Y X, c(50) kernel(uniform) p(2) bwselect(msetwo)
```

```
. rdrobust $Y $X, c(50) kernel(uniform) p(2) bwselect(msetwo)

Sharp RD estimates using local polynomial regression.

      Cutoff c = 50 | Left of c  Right of c            Number of obs =       2629
----------------------+----------------------            BW type       =     msetwo
      Number of obs |      2314         315            Kernel        =    Uniform
 Eff. Number of obs |       686         281            VCE method    =         NN
     Order est. (p) |         2           2
     Order bias (q) |         3           3
        BW est. (h) |    11.313      10.075
       BW bias (b) |    16.849      21.517
         rho (h/b) |     0.671       0.468

Outcome: Y2. Running variable: X2.
--------------------------------------------------------------------------------
            Method |   Coef.    Std. Err.    z     P>|z|    [95% Conf. Interval]
-------------------+------------------------------------------------------------
      Conventional |  2.2409     1.2822   1.7478   0.080   -.272028     4.75392
            Robust |     -          -     1.4485   0.147   -.724483     4.82787
--------------------------------------------------------------------------------
```

Step 6: Falsification Test Using Z1 (Plot)

```
rdplot Z1 X, c(50)
graph_options(legend(off)
graphregion(color(white))
ytitle(Outcome) xtitle(Score))
```



Step 6: Falsification Test Using Z1 (Estimate)

```
rdrobust Z1 X, c(50)
```

```
. rdrobust $Z1 $X, c(50)

Sharp RD estimates using local polynomial regression.

      Cutoff c = 50 | Left of c  Right of c            Number of obs =       2629
--------------------+----------------------            BW type       =      mserd
      Number of obs |      2314         315            Kernel        = Triangular
 Eff. Number of obs |       400         233            VCE method    =         NN
     Order est. (p) |         1           1
     Order bias (q) |         2           2
        BW est. (h) |     6.660       6.660
        BW bias (b) |    10.684      10.684
          rho (h/b) |     0.623       0.623

Outcome: lpop1994. Running variable: X2.
--------------------------------------------------------------------------------
             Method |   Coef.    Std. Err.    z     P>|z|    [95% Conf. Interval]
--------------------+-----------------------------------------------------------
       Conventional |  .01245      .2777   0.0448   0.964   -.531827      .556733
             Robust |     -          -     0.0011   0.999   -.644098      .644843
--------------------------------------------------------------------------------
```

Regression Discontinuity Designs: Small Group Work

Overview

- We'll send you an RDD Activity Guide just after we end this morning.
  - The guide presents a series of tasks for your group to complete, using the data sets we've provided.
- You have the rest of the afternoon to work on the tasks.
- Each group will have some time to meet with our trainers to talk about your progress and ask questions.
  - The agenda tells you when each group will meet and with whom.
- At 4:30, we will reconvene as a group to recap today's session.

References

Cattaneo, M., Idrobo, N., & Titiunik, R. (2020). A Practical Introduction to Regression Discontinuity Designs: Foundations (Elements in Quantitative and Computational Methods for the Social Sciences). Cambridge: Cambridge University Press.

Cattaneo, Matias D., Idrobo, Nicolás, and Titiunik, Rocío. Forthcoming. A Practical Introduction to Regression Discontinuity Designs: Extensions. Cambridge Elements: Quantitative and Computational Methods for the Social Sciences. Cambridge University Press.

Dougherty, S. (2018). The Effect of Career and Technical Education on Human Capital Accumulation: Causal Evidence from Massachusetts. *Education Finance and Policy*, 13(2), 119–148.

Jacob, R., Zhu, P., Somers, M., & Bloom, H. (2012). *A Practical Guide to Regression Discontinuity*. New York, NY: MDRC.

CTE Research Network Lead

The work of the CTE Research Network Lead is supported by the Institute of Education Sciences (IES), U.S. Department of Education, through grant R305N180005 to the American Institutes for Research. The opinions expressed are those of the authors and do not represent views of IES or the U.S. Department of Education.

CTEResearchNetwork.org

CTEResearchNetwork@air.org