Lecture 18

Regression Discontinuity Design (RDD) III

- Problem Set 5 (last problem set) due Wednesday 04/29.

Overview

1. Basic regression discontinuity (RD) model
2. More flexible RD models
3. Windows and bins
4. Case study: universal pre-K
5. Limitations and diagnostics

Context: How we fight endogeneity (Omitted variable Bias)

1. Controlling for measurable factors
    - Multivariate Regression.
2. Create exogeneity: Experiments
3. Use exogeneity
    - Natural or Quasi-experiments (Difference in Difference)
    - Instrumental variables (Not covered)
4. Regression discontinuity

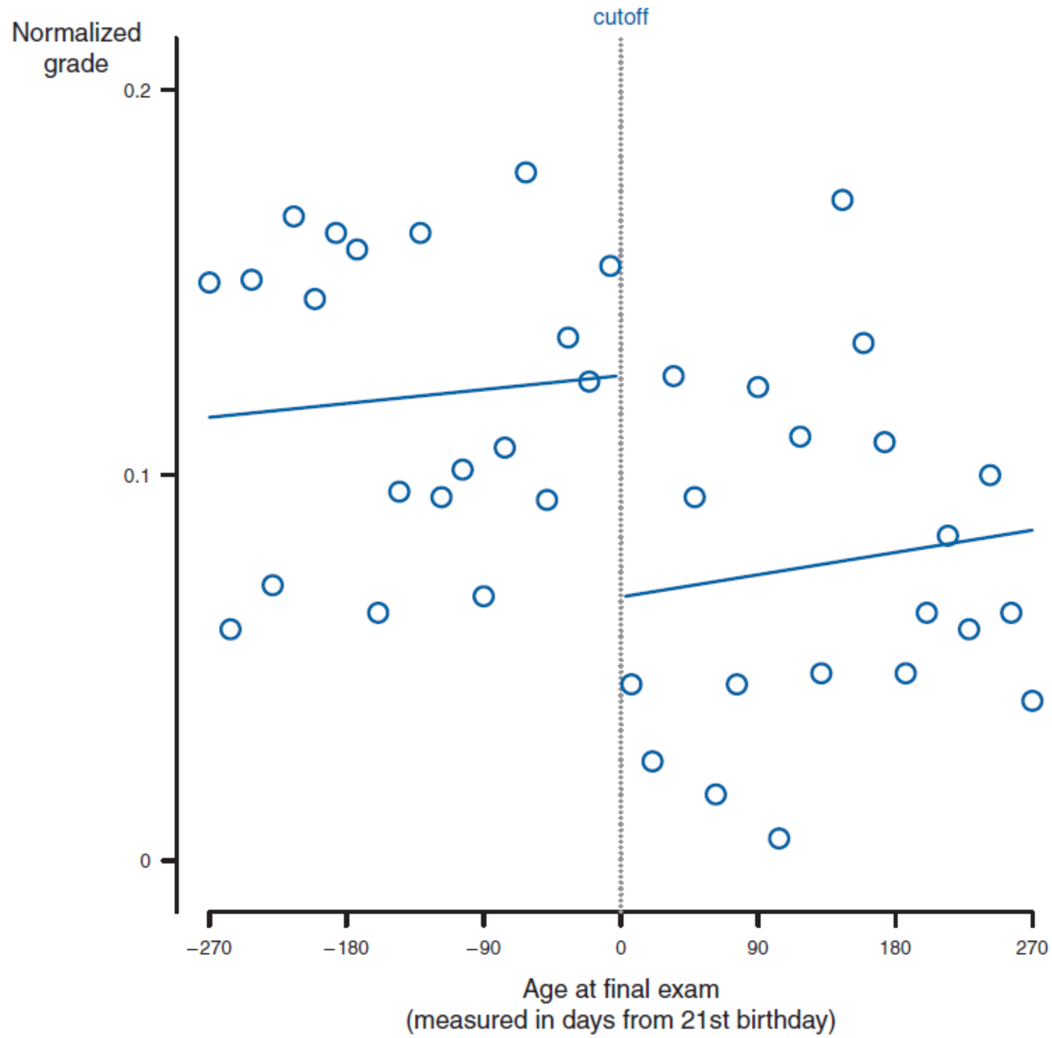


FIGURE 11.1: Drinking Age and Test Scores

Regression discontinuity model

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2(X_{1i} - C) + \epsilon_i$$

$$T_i = 1 \text{ if } X_{1i} \geq C$$

$$T_i = 0 \text{ if } X_{1i} < C$$

- T_i is the **treatment variable**. We seek to understand the effect of the treatment.
- X_{1i} is the "**assignment variable**" for which individuals get treatment if their assignment variable is greater than some **cutoff**.
- C is the value of the **cutoff**

Example: Effect of drinking on grades at school where drinking age is rigorously enforced

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (X_{1i} - C) + \varepsilon_i$$

$$Grades_i = \beta_0 + \beta_1 Legal_i + \beta_2 (Age_i - 21) + \varepsilon_i$$

$$Y = \alpha + \delta Above + \gamma Score + \varepsilon$$

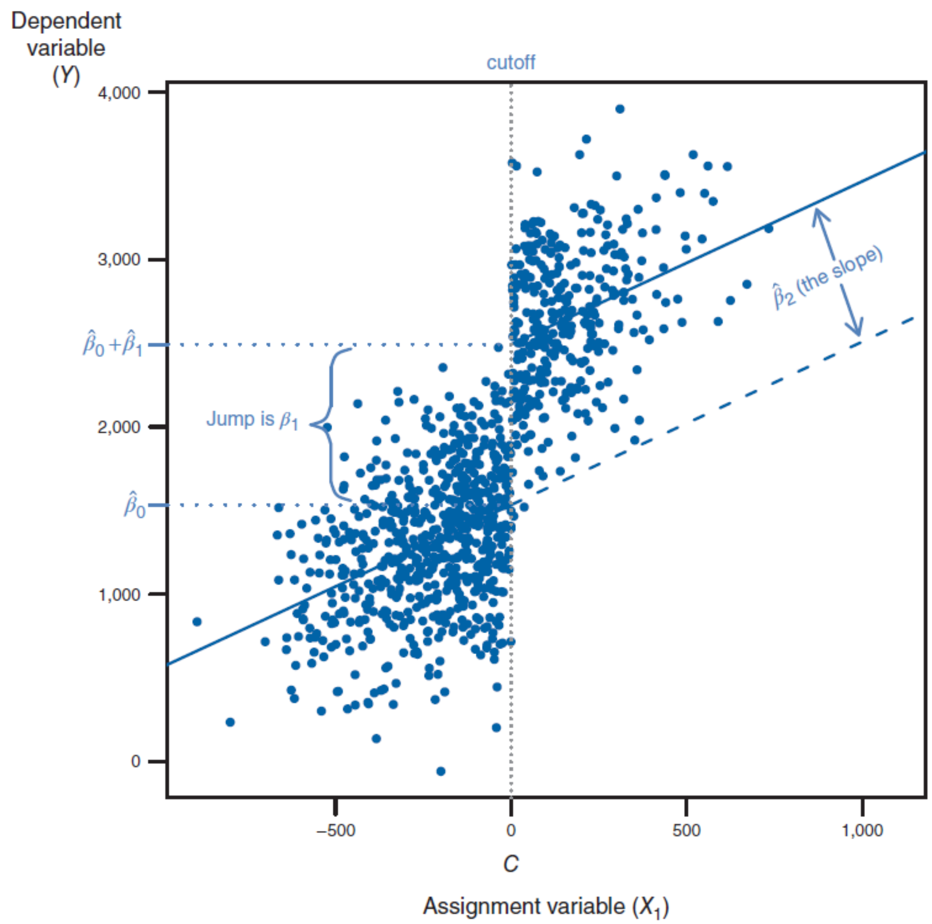
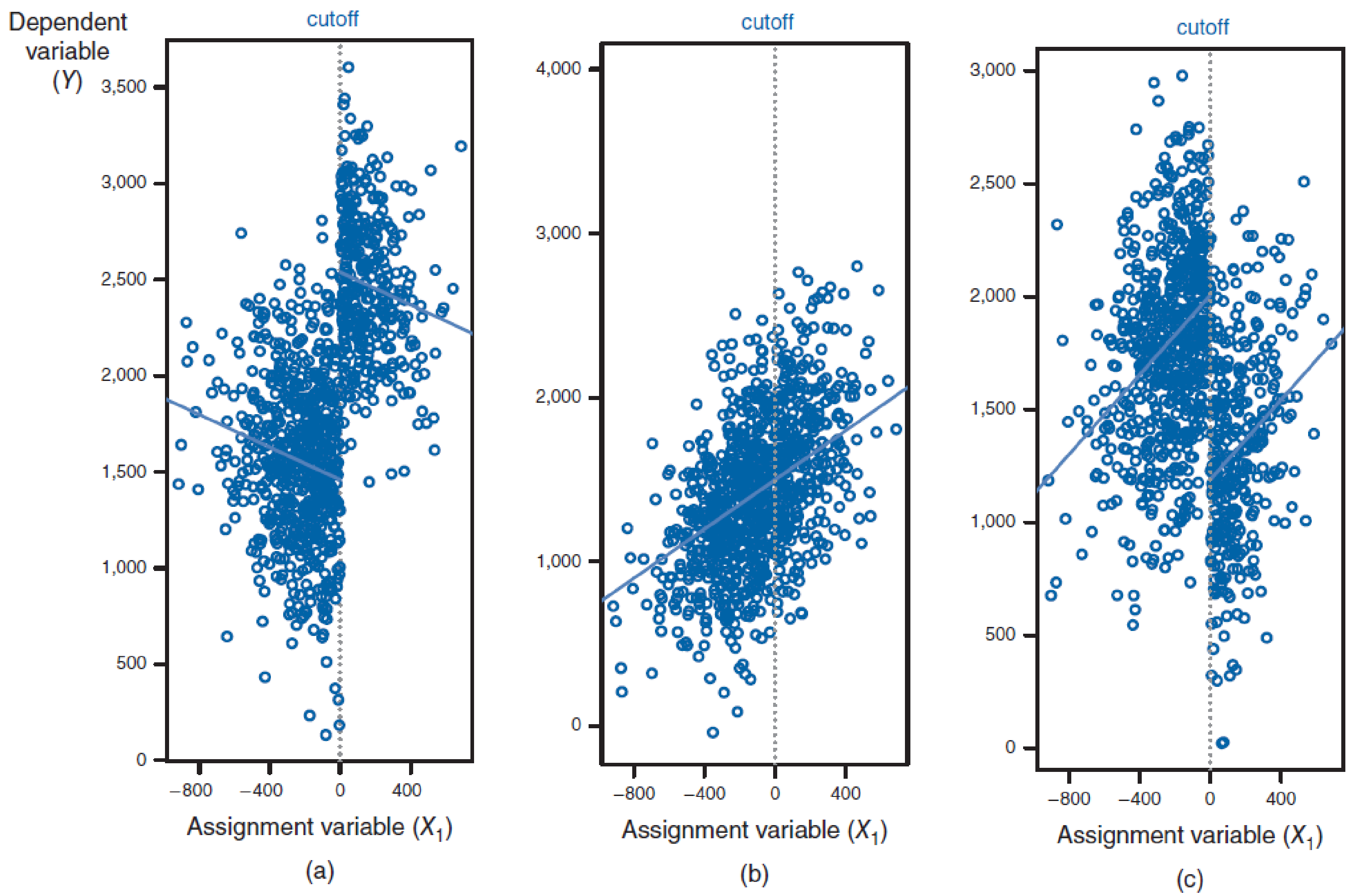


FIGURE 11.2: Basic RD Model, $Y_i = \beta_0 + \beta_1 T_i + \beta_2(X_{1i} - C)$

The jump ("bump") in dependent variable at point of discontinuity is the estimated effect of the treatment.

FIGURE 11.3: Possible Results with Basic RD Model

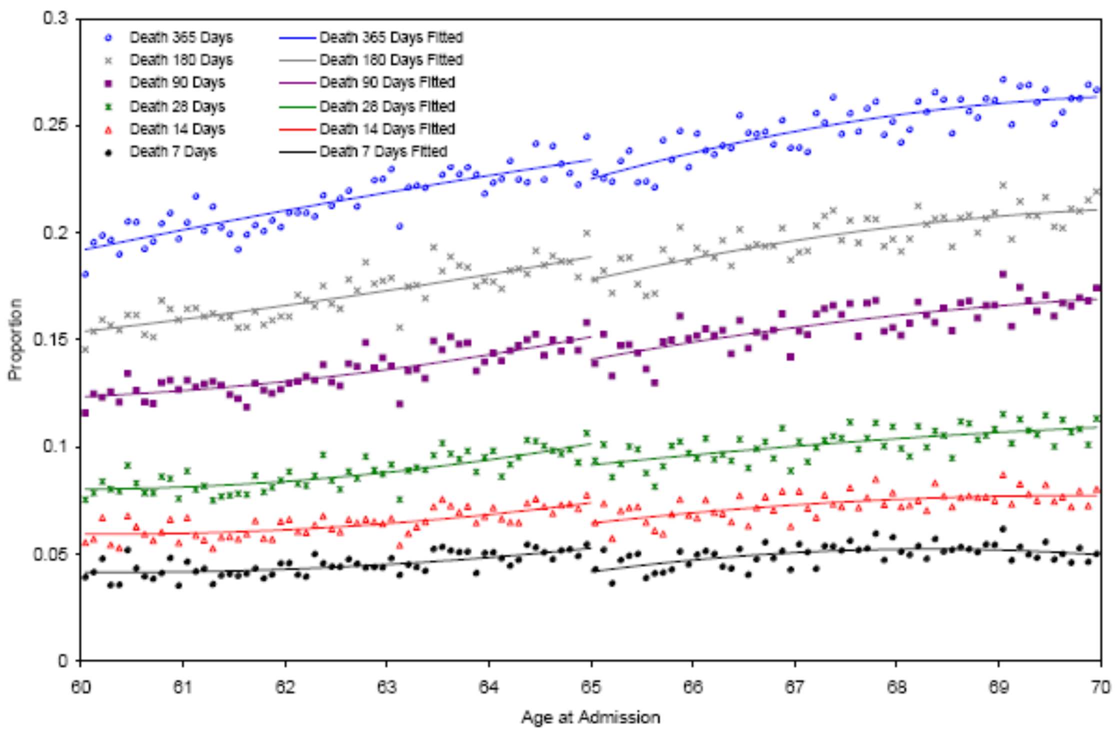


Why does RD work?

- Common explanation:
  - “the group of participants just to the "left" of any cut-off are equal in expectations on all dimension, other than their exposure to the treatment, to those just to the "right" of the cutoff" (Murnane and Willett 2011, 152)
- More general: The assignment variable X soaks up correlation between error and treatment
  - As long as there is no discontinuity in relationship between error (omitted variables) and the outcome at the discontinuity, then we can attribute any bump in the dependent variable as the effect of the treatment.

Example: Medicare

Figure 9: Died Either in or Out of Hospital



Example: incumbency advantage

$$DemPercent_i = \beta_0 + \beta_1 Incumbent_i + \varepsilon_i$$

- Is there reason to suspect endogeneity? Why or why not?

Incumbency effects: vote share

From Lee & Lemieux (2010, 310).

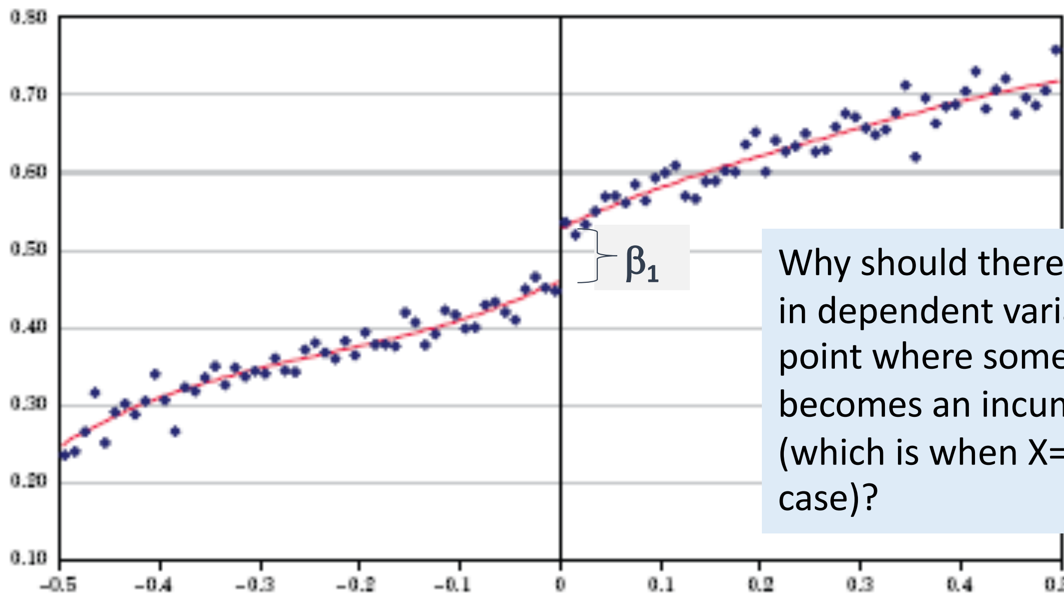


Figure 7. Share of Vote in Next Election, Bandwidth of 0.01 (100 bins)

Why should there be a bump in dependent variable right at point where someone becomes an incumbent (which is when X=0 in this case)?

$$VoteShare_{t+1}^{*} = \beta_0 + \beta_1 DemIncumbent_{t+1} + \beta_2(Vote\ Share_t - 0.5) + \varepsilon_{t+1}$$

Incumbency effects: probability of winning

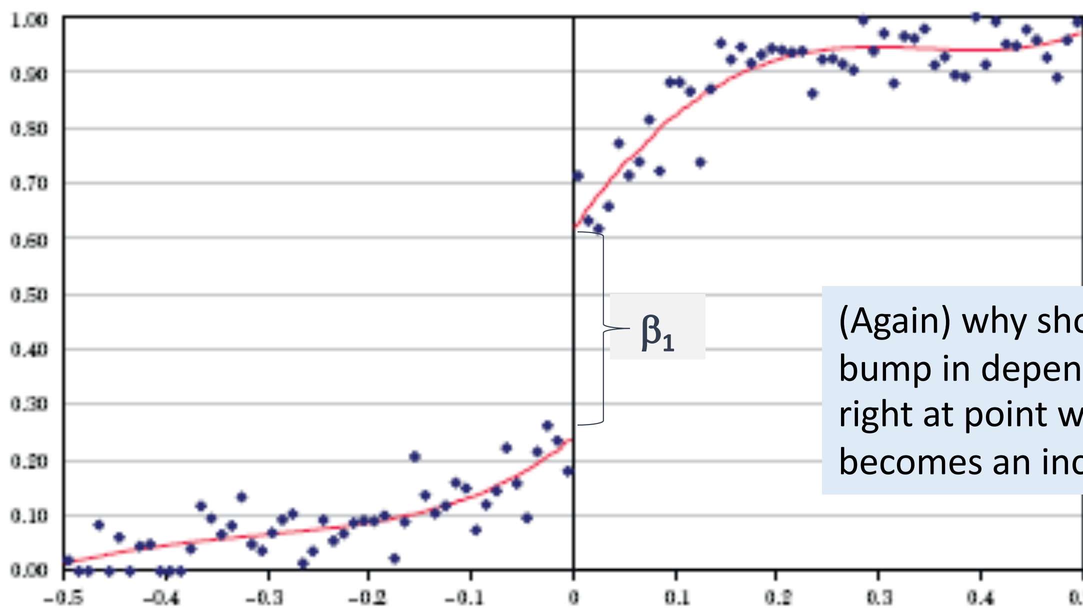


Figure 10. Winning the Next Election, Bandwidth of 0.01 (100 bins)

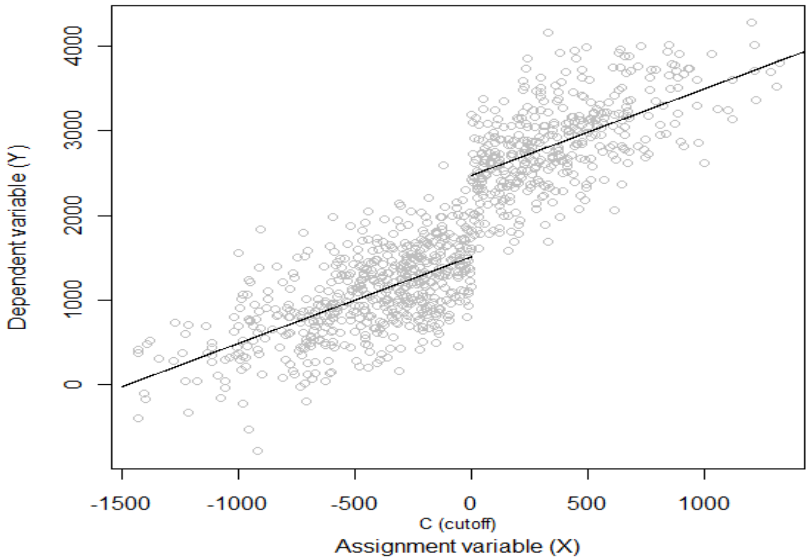
(Again) why should there be a bump in dependent variable right at point where someone becomes an incumbent?

$$\text{ProbDemWins}_{t+1}^* = \beta_0 + \beta_1 \text{DemIncumbent}_t + \beta_2(\text{Vote Share}_t - 0.5) + \varepsilon_{t+1}$$

RD and policy evaluation

- Random experiment often infeasible as randomization of benefits is not fair or sensible to program leaders
- R-D can be viable alternative that is attractive to policymakers

Basic version of RD - revisited

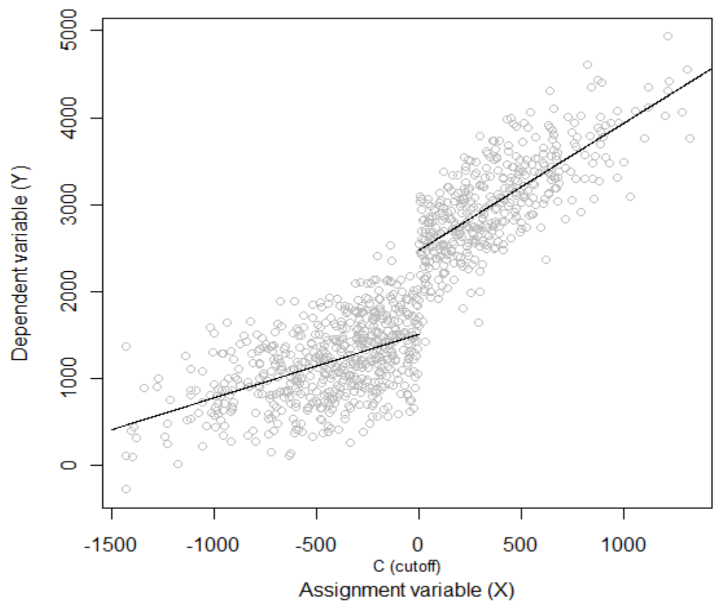


- In basic version of RD, the relationship between X and Y is linear and the same on both sides of treatment in this model
- ***but this need not be the case in data***

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (X_i - C) + \varepsilon_i$$

$$T_i = 1 \text{ if } \mathrm{X_i} \geq C$$

$$T_i = 0 \text{ if } \mathrm{X_i} < C$$



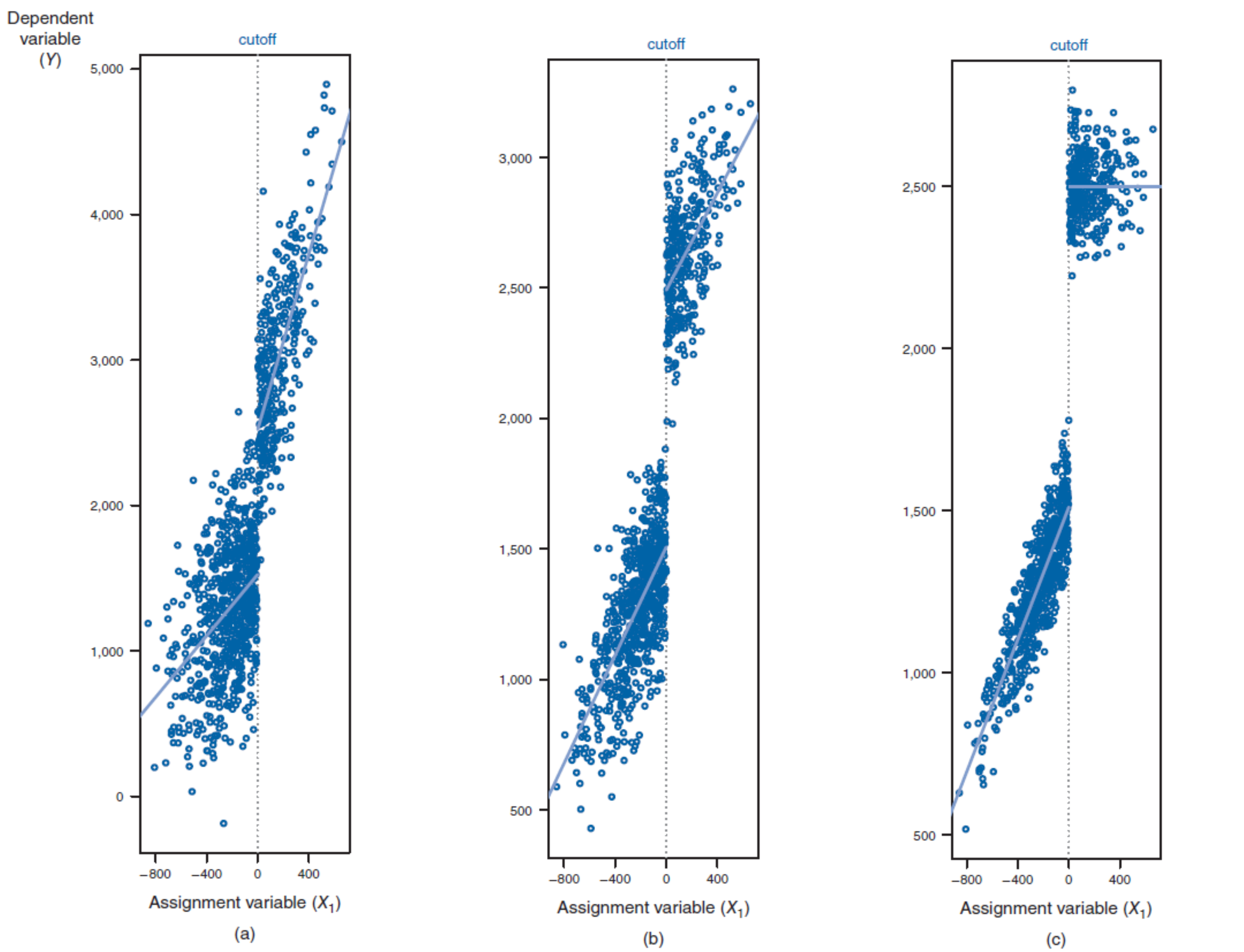
Slopes differ on sides of cutoff. This is most common form of RD

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (X_i - C) + \beta_3 (X_i - C) T_i + \varepsilon_i$$

$$T_i = 1 \text{ if } \mathrm{X_i} \geq C$$

$$T_i = 0 \text{ if } \mathrm{X_i} < C$$

FIGURE 11.4: Possible Results with Differing Slopes RD Model



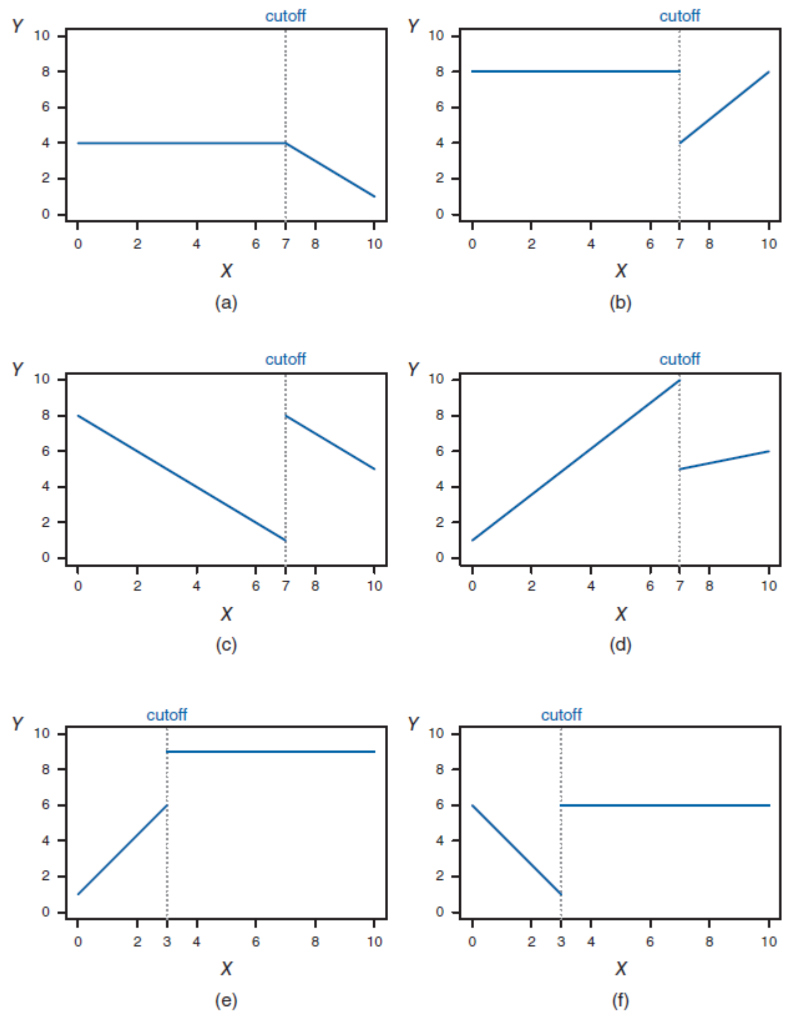


FIGURE 11.6: Various Fitted Lines for RD Model of Form $Y_i = \beta_0 + \beta_1 T_i + \beta_2(X_{1i} - C) + \beta_3(X_{1i} - C)T_i$ (for Review Question)

Higher order polynomial terms

(for reference)

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (X_i - C) + \beta_3 (X_i - C)^2 + \beta_4 (X_i - C)^3 + \beta_5 (X_i - C) T_i + \beta_6 (X_i - C)^2 T_i + \beta_7 (X_i - C)^3 T_i + \varepsilon_i$$

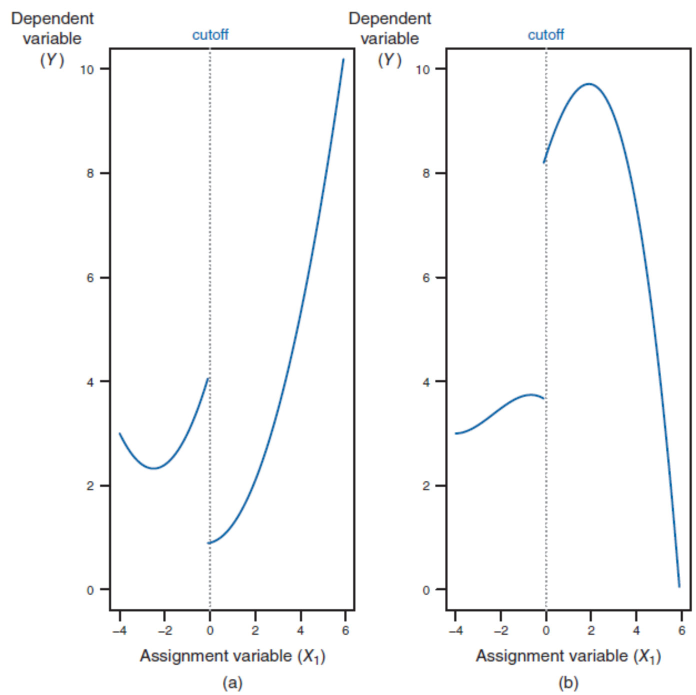


FIGURE 11.5: Fitted Lines for Examples of Polynomial RD Models

Window size

- The window is the range of X to which we limit our analysis.
- In theory, we'd like the window to be "epsilon" on each side of cutoff.
- In practice, almost always need fairly big window in order to get sufficient sample size.
- Reducing window size is great way to deal with non-linearities.
  - With smaller window, linear model should do better and better

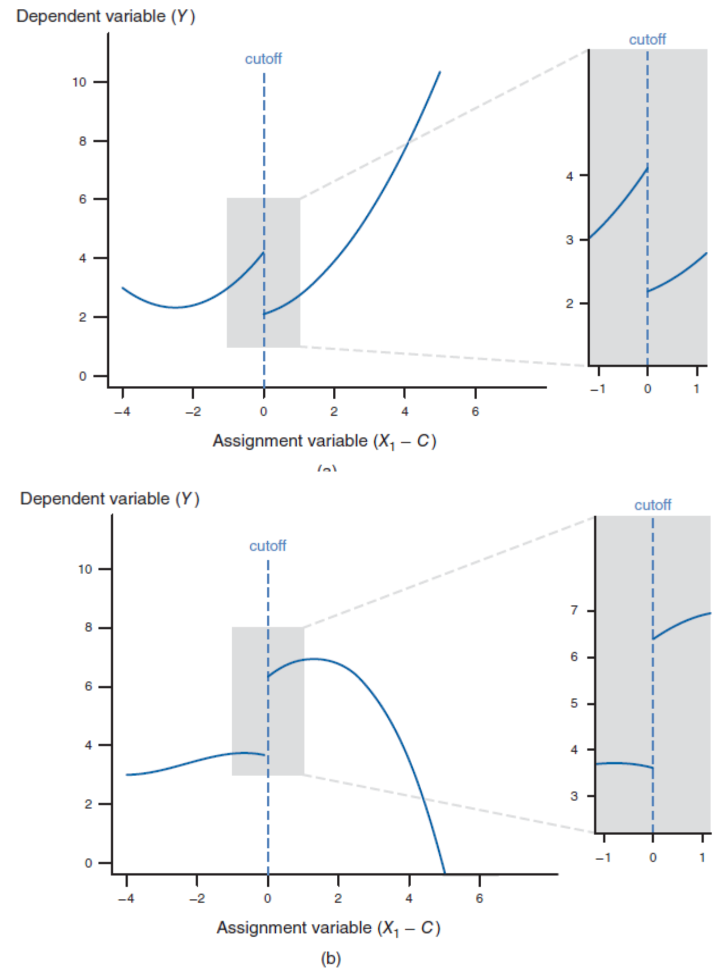


FIGURE 11.7: Smaller Windows for Fitted Lines for Polynomial RD Model in Figure 11.5

Graphs and RD

- Binned graphs
  - On right: “binned graph” where average value of Y for bins of X is plotted
- Virtues of graphs
  - If there is an effect, will almost always be able to “see” it with simple graph
  - Diagnostics – assess if there are non-linearities or discontinuities at other points.

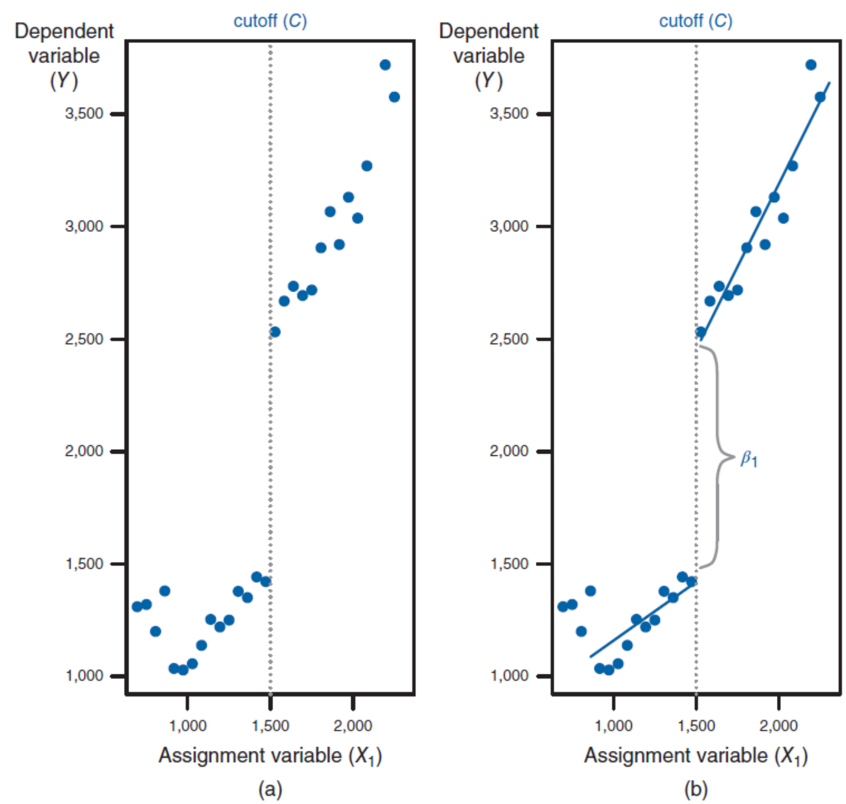


FIGURE 11.8: Bin Plots for RD Model

CASE STUDY

Universal Prekindergarten

RD model for universal pre-k

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2(X_{1i} - C) + \varepsilon_i$$

$$\text{Test score}_i = \beta_0 + \beta_1 \text{Pre-K}_i + \beta_2(\text{Age}_i - \text{PreK Cutoff}) + \varepsilon_i$$

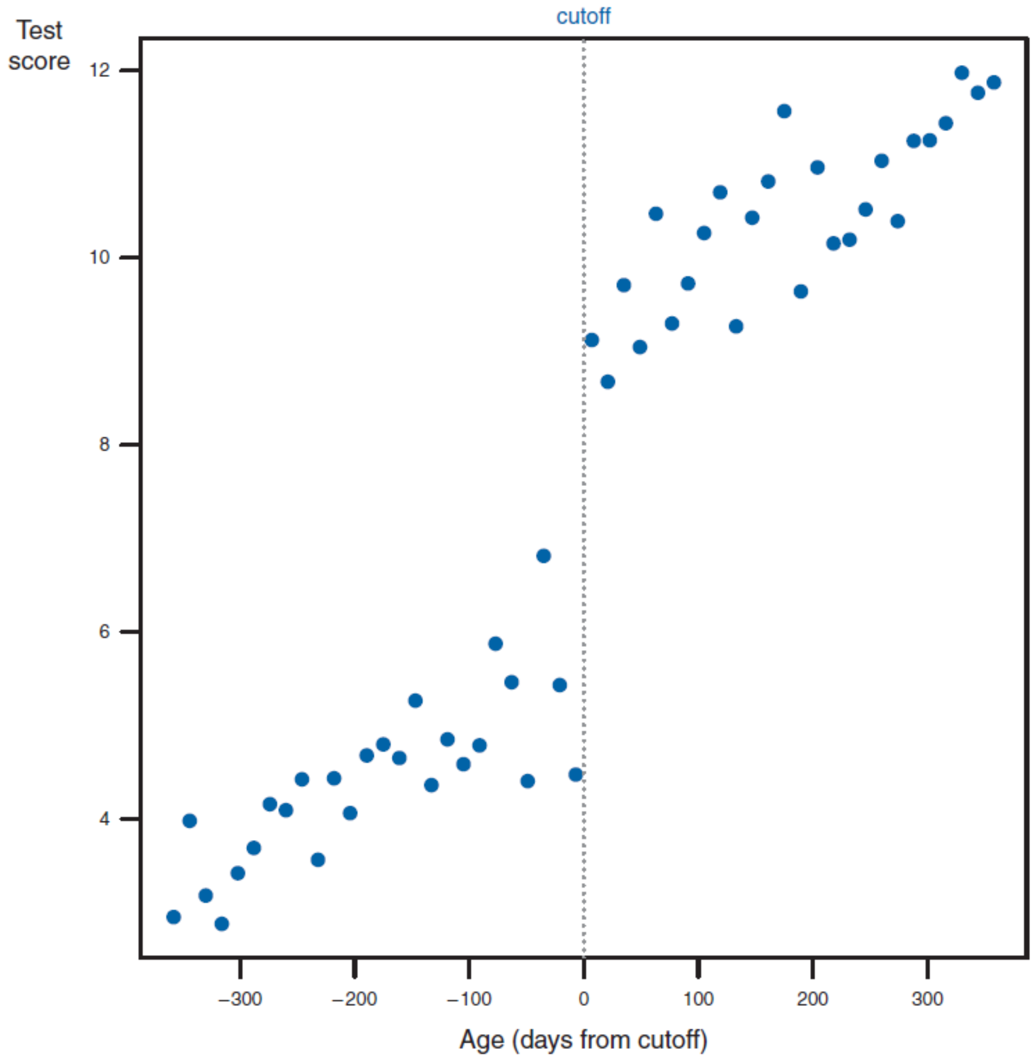


FIGURE 11.9: Binned Graph of Test Scores and Pre-K Attendance

TABLE 11.1 RD Analysis of Prekindergarten

	Basic	Varying slopes
Pre-K	3.492*	3.479*
	(0.339)	(0.340)
	[$t = 10.31$]	[$t = 10.23$]
Age – C	0.007*	0.007*
	(0.001)	(0.001)
	[$t = 8.64$]	[$t = 6.07$]
Pre-K × (Age – C)		0.001
		(0.002)
		[$t = 0.42$]
Constant	5.692*	5.637*
	(0.183)	(0.226)
	[$t = 31.07$]	[$t = 24.97$]
N	2,785	2,785
R^2	0.323	0.323

Standard errors in parentheses.

* *indicates significance at $p < 0.05$, two-tailed.*

11.4 Limitations and Diagnostics

RD limitations

- Recall, the key assumption of RD is that there is no "bump" (discontinuity) in the error term at the discontinuity
- This can fail when
  a) Individuals have control over assignment variable
     - Example: suppose a test score above "100" allows you to get a scholarship and that students are allowed to re-take test. More motivated students (where motivation is in error term) may re-take the test until they get a 100.
     - Example: financial aid officer may score financial need knowing applicant and score needed to get aid

  b) People may anticipate discontinuity
     - Example: suppose people save up visits to doctor until after they are 65

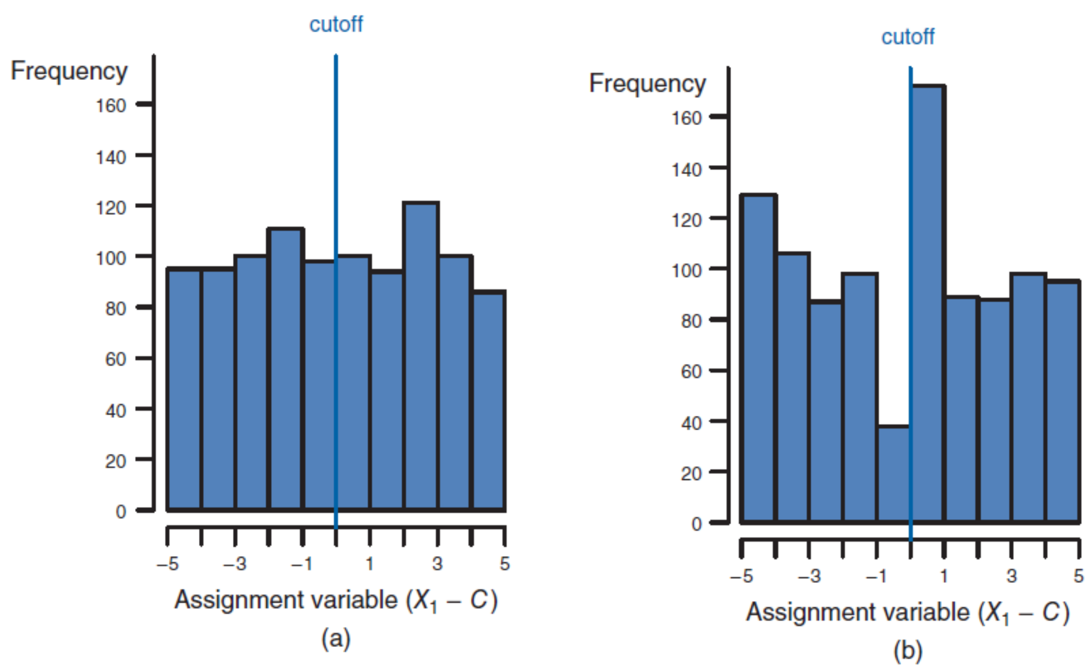
RD limitations - generalizability

- Effect of treatment on the subpopulation with $X_i = c$
  - Sometimes this is referred to as “local average treatment effect” (LATE)

RD Diagnostics - 1

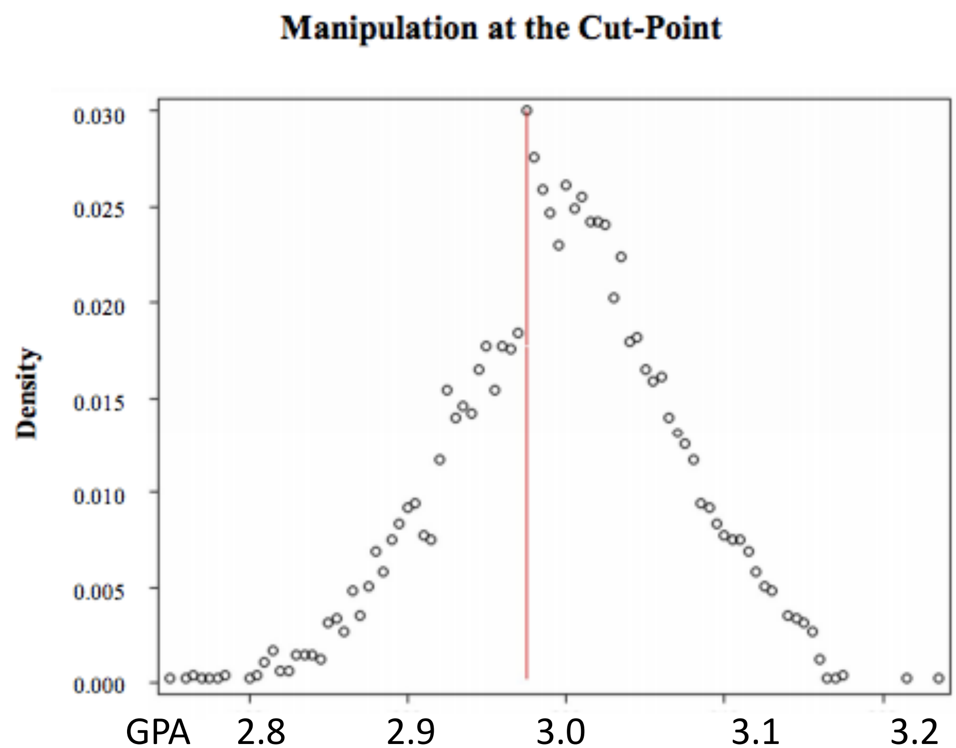
1) There should be no bump in assignment variable. Use histogram to assess.

FIGURE 11.10: Histograms of Assignment Variable for RD Analysis



Histogram test

- Draw a histogram of the running variable, showing distribution of observations (X-axis: GPA; Y-axis: frequency (density))
- If there is no manipulation of the running variable for the GPA example, what would the histogram of students look like?
  - If there is no manipulation, the histogram should be fairly smooth



- Can you draw a histogram when there IS manipulation?

RD Diagnostics - 2

2) At discontinuity, there should be no discontinuity in other covariates. For some "other" covariate W:

$$W_i = \gamma_0 + \gamma_1 T_i + \gamma_2(X_{1i} - C) + \varepsilon_i$$

γ_1 should equal zero.

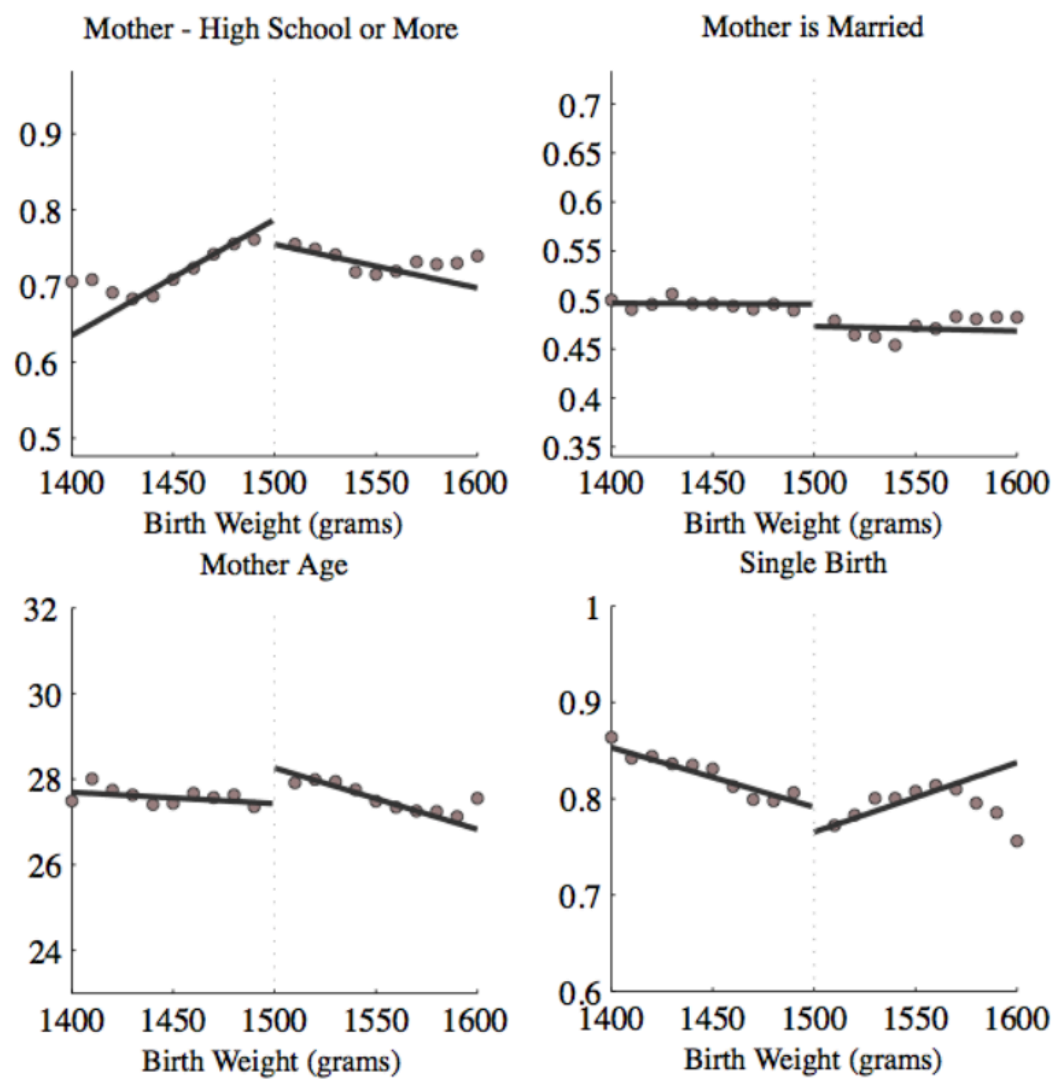
Covariate smoothness test-example

- Based on the covariates test, do you worry about manipulation?

- The idea is similar with the balanced table in OLS

  1. If covariates are smooth, we imply unobservables are smooth
  2. So the difference in Y is not caused by differences in observables and unobservables. **It's entirely due to the treatment status**

Figure 6: Baseline Covariates Around 1500 grams - Chile



Example: Alcohol and grades

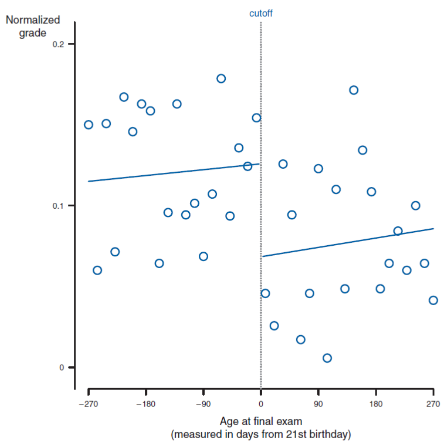


FIGURE 11.1: Drinking Age and Test Scores

TABLE 11.2 RD Analysis of Drinking Age and Test Scores

	Varying slopes	Varying slopes with control variables	Quadratic with control variables
Discontinuity at 21	−0.092*	−0.114*	−0.106*
	(0.03)	(0.02)	(0.03)
	[t = 30.67]	[t = 57.00]	[t = 35.33]
N	38,782	38,782	38,782

Standard errors in parentheses.

All three specifications control for age, allowing the slope to vary on either side of the cutoff. The second and third specifications control for semester, SAT scores, and other demographics factors.

* *indicates significance at $p < 0.05$, two-tailed.*

From Carrell, Hoekstra, and West (2010).

Diagnostics for alcohol and grades RD model

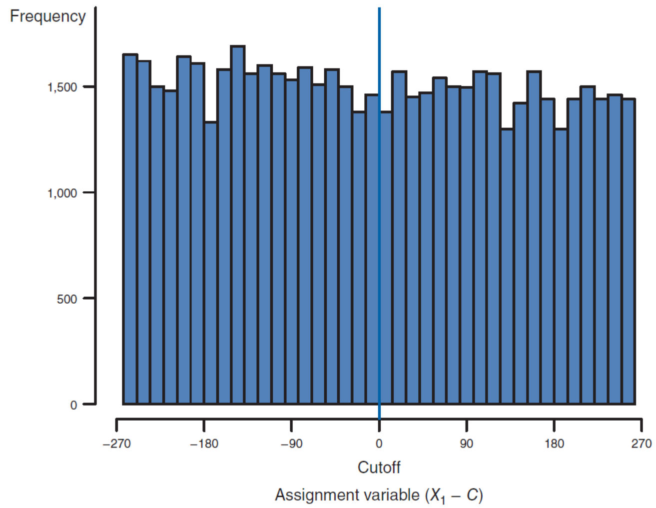


TABLE 11.3 RD Diagnostics for Drinking Age and Test Scores

	SAT math	SAT verbal	Physical fitness
Discontinuity at 21	2.371	1.932	0.025
	(2.81)	(2.79)	(0.04)
	[$t = 0.84$]	[$t = 0.69$]	[$t = 0.63$]
N	38,782	38,782	38,782

Standard errors in parentheses.

All three specifications control for age, allowing the slope to vary on either side of the cutoff.

RD steps

1. Assess appropriateness of RD
    - Sharp RD (assignment variable perfectly predicts treatment)
    - Qualitatively assess whether agents have control over assignment variable
    - Conduct balancing tests
2. Plot data – typically with binned graph
3. Estimate linear model using different window sizes

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (X_i - C) + \beta_3 (X_i - C) T_i + \varepsilon_i$$

4. Estimate non-linear models using different window sizes

Summary: Regression discontinuity

- RD uses X to soak up endogeneity
- Easy to summarize graphically with data
- Relatively little room for researcher discretion
- Applicable to many real-world evaluation contexts

Review for exams

- For RD implementation, understand the following formula,

$$Y = \alpha + \delta Above + \gamma S + \eta(S * Above) + \varepsilon$$

OR

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (X_i - C) + \beta_3 (X_i - C) T_i + \varepsilon_i$$

being able to explain the key parameters (α and β)

- Understand the first stage and reduced form of fuzzy RD
- Understand Why RD might fail