

Discussion 8

Week of 03.23

Agenda

- Quiz Review
- Regression Discontinuity

Announcements

- Homework 5 due Friday, March 27 @ 9PM
- Quiz 6 due Tuesday, March 24 @ 9PM
- **REMINDER:** if a HW question asks you to “compute the difference” or something similar, please provide the stata output *and numerical answer* and not *just* the Stata code you used to *find* the numerical answer.

Quiz Review

Question 1

Suppose you are estimating the effect of education on earnings, but you believe education is not exogenous. Instead, you use distance to the nearest college as an instrument for education. What concerns might you have about this instrument?

- Distance to the nearest college may affect earnings through local labor market effects
- Distance to the nearest college may make students more likely to attend college
- Distance and education are uncorrelated
- Distance and education might be strongly correlated

Question 1

Suppose you are estimating the effect of education on earnings, but you believe education is not exogenous. Instead, you use distance to the nearest college as an instrument for education. What concerns might you have about this instrument?

- Distance to the nearest college may affect earnings through local labor market effects
- Distance to the nearest college may make students more likely to attend college
- Distance and education are uncorrelated
- Distance and education might be strongly correlated

Question 2

A researcher wants to estimate the effect of pretrial detention on conviction. In order to determine if the defendant is detained before the trial conviction the defendants get randomly assigned a judge for the judge to decide. This judge could be a lenient or a strict judge. This is a different judge that your "conviction" judge. The research uses this judge assignment as an instrument for whether the defendant is detained before trial. Provide the option that correctly identifies Y, D and Z.

- D=Conviction, Z=pre-trial detention, Y=judge leniency
- Y=defendant's guiltiness, Z=pretrial detention, D=judge leniency
- D=judge leniency, Y=conviction, Z=pretrial detention
- Y=conviction, D=pretrial detention, Z=judge leniency

Question 2

A researcher wants to estimate the effect of pretrial detention on conviction. In order to determine if the defendant is detained before the trial conviction the defendants get randomly assigned a judge for the judge to decide. This judge could be a lenient or a strict judge. This is a different judge that your "conviction" judge. The research uses this judge assignment as an instrument for whether the defendant is detained before trial. Provide the option that correctly identifies Y, D and Z.

- () D=Conviction, Z=pre-trial detention, Y=judge leniency
- () Y=defendant's guiltiness, Z=pretrial detention, D=judge leniency
- () D=judge leniency, Y=conviction, Z=pretrial detention
- (✓) Y=conviction, D=pretrial detention, Z=judge leniency

Question 3

Which of these equations identifies the reduced form?

$\text{Conviction} = \beta_0 + \beta_1 \text{PreTrialDetention} + \varepsilon$

$\text{JudgeLeniency} = \beta + \beta_1 \text{PreTrialDetention} + \varepsilon$

$\text{PreTrialDetention} = \beta_0 + \beta_1 \text{JudgeLeniency} + \varepsilon$

$\text{Conviction} = \beta_0 + \beta_1 \text{JudgeLeniency} + \varepsilon$

Question 3

Which of these equations identifies the reduced form?

Conviction = $\beta_0 + \beta_1 \text{PreTrialDetention} + \varepsilon$

JudgeLeniency = $\beta + \beta_1 \text{PreTrialDetention} + \varepsilon$

PreTrialDetention = $\beta_0 + \beta_1 \text{JudgeLeniency} + \varepsilon$

Conviction = $\beta_0 + \beta_1 \text{JudgeLeniency} + \varepsilon$

Regression Discontinuity

“Natura non facit saltus”

Nature does not make jumps.

Regression Discontinuity (RD): a quasi-experimental research design used to estimate causal effects when treatment assignment is based on whether an observed variable crosses a known cutoff or threshold.

- This method takes advantage of “rules” and “regulations” that are not embedded in nature.
- Thresholds → Treatments

Conditions for Regression Discontinuity

In order to apply RD, some conditions must be satisfied:

1. Whether or not you receive a “treatment” is determined by some score or ranking.
2. The score has to be continuous around the cutoff and has to be arbitrarily and exogenously determined.
3. We have outcomes above and below the cutoff.
4. People cannot game the system to receive the treatment.

In other words:

1. Units just below the threshold are a valid control for units just above the threshold.
2. There are no other jumps (“treatments”) at the discontinuity.

Regression Discontinuity Equation

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

T_i is a dummy/indicator for whether person i received the treatment

X_i is the assignment variable (running var.) measuring the score or ranking of person i

C is the cutoff value

$X_i - C$ is the assignment variable centered around C .

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

Adding an interaction term allows us to account for changes in *slope* from our “control” side to our “treatment” side.

RD Practice

We'll use this paper:

Morales, Camila, and Monica Mogollon. 2024. "The Effects of a Newcomer Program on the Academic Achievement of English Learners." *Journal of Policy Analysis and Management* 43 (3): 735–60. <https://doi.org/10.1002/pam.22601>.

Read through the introduction and try to identify the running variable, cutoff, and outcome(s).

RD Practice

Read the regression continuity design section on page 743. This is the regression the authors use:

$$A_{igc(t_0+a)} = \beta_0 + \beta_1 \text{LowWIDA}_{it_0} + \beta_2 f(WIDA_{it_0} - s) + \gamma_g + X_{it_0} \lambda + \varepsilon_{igc(t_0+a)}$$

How would you interpret β_0 , β_1 , and β_2 ?

Read through the results section (p. 746) and table 2 (p. 747). What were the key results of the study?

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	Full sample		Grades 3 through 5		Grades 6 through 7	
	Baseline	Preferred	Baseline	Preferred	Baseline	Preferred
<i>Panel A: ELA scores</i>						
Program eligible	0.06 (0.09)	0.05 (0.09)	0.39*** (0.07)	0.20** (0.08)	-0.21* (0.12)	-0.17 (0.11)
Observations	4,469	4,469	2,390	2,390	2,001	2,001
Left window	0.62	0.91	0.52	0.66	0.93	0.97
Right window	0.27	0.25	0.31	0.28	0.34	0.30
<i>Panel B: Math scores</i>						
Program eligible	-0.07 (0.11)	-0.09 (0.09)	0.18** (0.09)	0.03 (0.12)	-0.23 (0.15)	-0.25*** (0.07)
Observations	4,609	4,609	2,509	2,509	2,022	2,022
Left window	0.49	1.00	0.57	0.86	0.80	0.74
Right window	0.33	0.33	0.31	0.30	0.33	0.34
Grade FEs	yes	yes	yes	yes	yes	yes
Covariates	no	yes	no	yes	no	yes

In Stata:

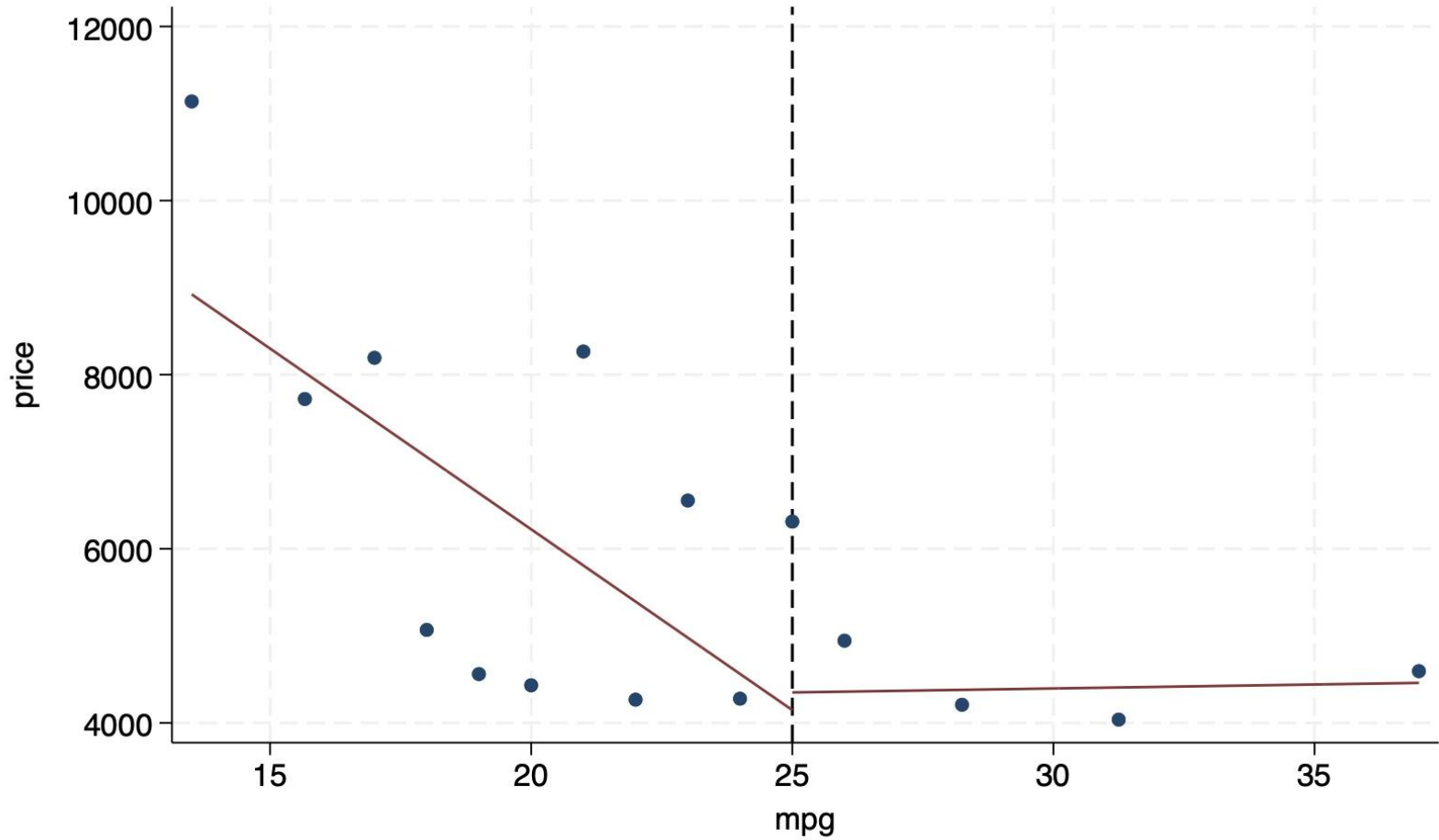
Open `auto.dta` (`sysuse auto.dta`)

The `binscatter` command allows us to make a scatter plot with data in “bins”. Try it out: `binscatter price mpg`

To make an RD graph with `binscatter`, use the RD option:

```
binscatter price mpg, rd(25) line(lfit) xline(25)
```

The `rd(25)` option gives us a cutoff of 25, `line(lfit)` gives us lines of best fit on either side, and `xline(25)` adds a vertical line at the cutoff.



Questions?