Microeconometrics (Causal Inference)
Week 9 - Regression discontinuity

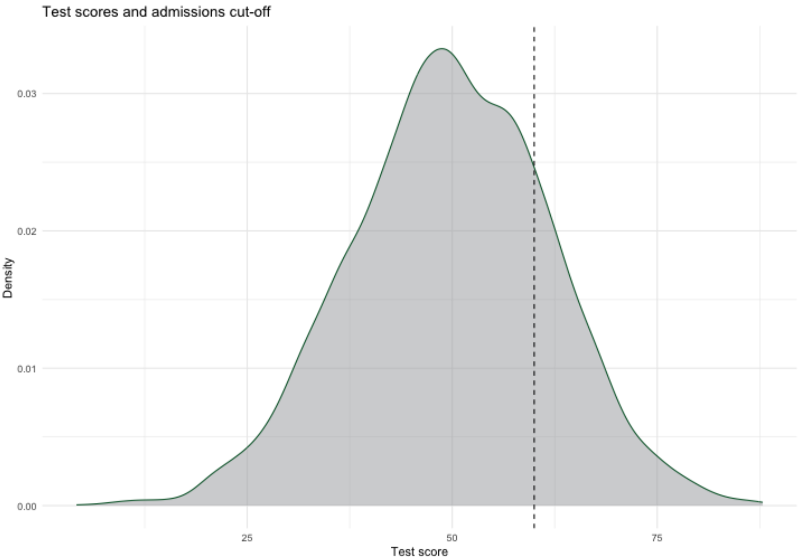
Joshua D. Merfeld
KDI School of Public Policy and Management

2023-11-13

What are we doing today?

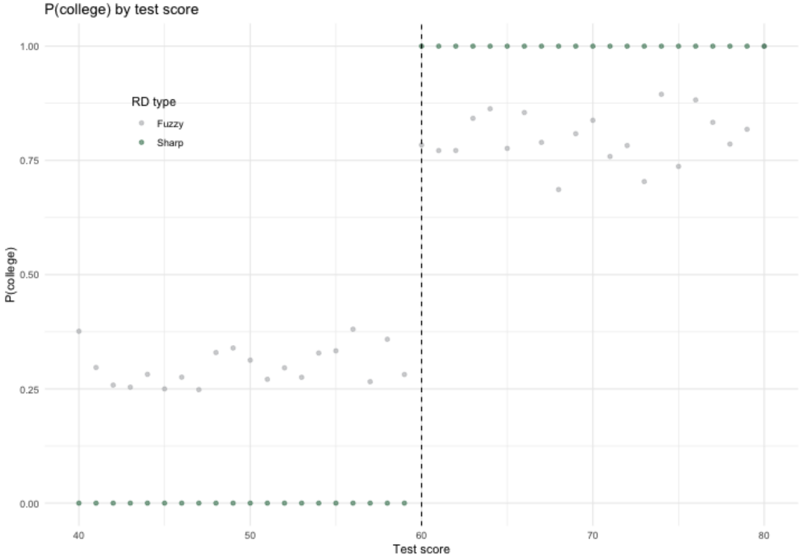
- Regression discontinuity
  - Requirements/assumptions
- Sharp and fuzzy RD
  - IVs and RDs

Motivation - standardized tests (fictitious data)



- In our example, you get into college if you score 60 or higher on a standardized test
- On average, "smarter" (in a broad sense) students will score higher on the test
- However, there is a lot of variation in scores among students with similar "smartness"
  - If one of us took the test multiple times, we'd probably get slightly different scores each time
  - We each have our own "distribution"
  - On a given day, how well (or not) we do is somewhat random

- Continuing with the example, imagine all of the students around the cut-off score of 60
- On average, students just below and just above the cut-off score are similar
  - They have similar "smartness"
  - They should also be similar on other variables!
- This is especially true if the test is a one-off test that you can't retake
  - Or if we don't know what the cut-off is
  - If we know the cut-off is 60 *and* we can take the test multiple times, what might we do?



P(college) by test score
1.00
0.75
0.50
0.25
0.00
P(college)
RD type
Fuzzy
Sharp
40
50
60
70
80
Test score

Returns to college - RD example

Regression discontinuity assumptions

- RD only works in a very specific context: when there is a clear cut-off in some variable (called the running or forcing variable) that determines treatment
- The best-case scenario is something we already discussed:
  - People don't know the cut-off at the time
  - The cut-off is not something you can manipulate (for example if you can only take a test once)
- In these cases, we can assume that people just above and just below the cut-off are similar
  - Implication: they should be similar on variables unaffected by treatment
    - We can check this!
  - Implication: density on either side of the cut-off should be similar
    - We can check this!

Example: Bleemer and Mehta

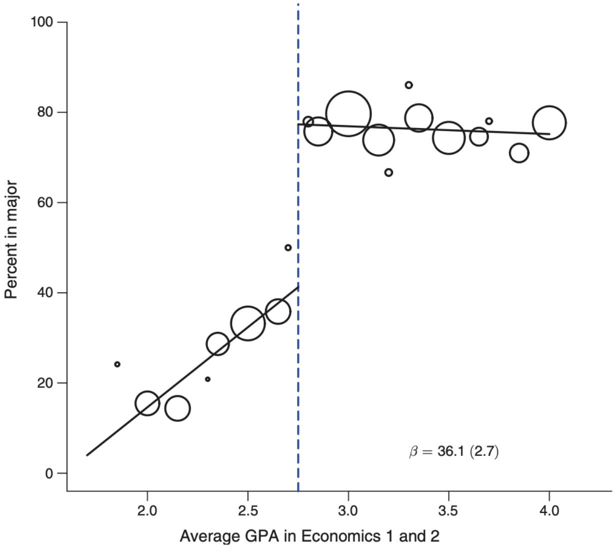
- Bleemer and Mehta (2022): Will studying economics make you rich? A regression discontinuity analysis of the returns to college major
  - *AEJ: Applied*
- Note: The data is confidential, so we can't replicate the results
  - We'll just go through the paper and discuss
- We'll replicate a common RD design later

Background for Bleemer and Mehta

- Data from UC Santa Cruz
  - Public university
- Starting in 2003, the econ department instituted a GPA restriction
  - Common for majors that are oversubscribed
  - Students with a GPA below 2.8 were not allowed to declare an econ major
    - (It's a little more complicated than that, but we'll just go with this for)
- Originally, grades in Economics 1 and 2 were counted
  - Added calculus in 2013

Data

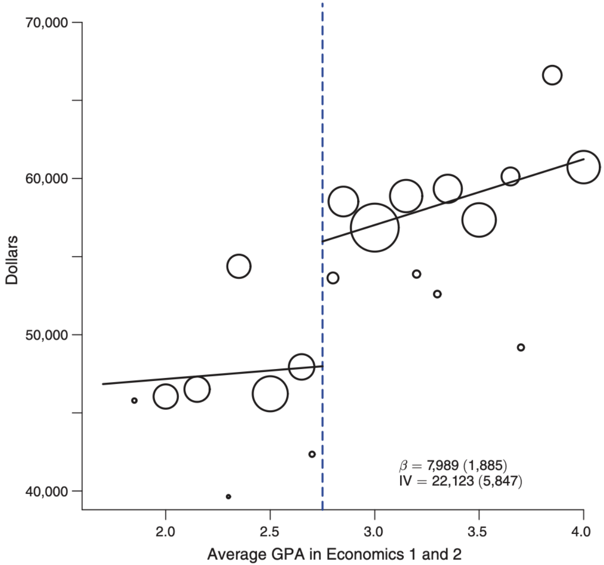
- They have information on individual students from their time in school
  - Information on econ GPA (EGPA) as well as other grades
  - Gender, ethnicity, cohort year, home address, residency status, high school, and SAT score
- They link the data to employment records from the California Employment Development Department
  - Annual wages and six-digit industry (NAICS) code
- You can probably tell by now why the data is confidential



100
80
60
40
20
0
Percent in major
2.0
2.5
3.0
3.5
4.0
Average GPA in Economics 1 and 2
$\beta = 36.1$ (2.7)

This is a fuzzy regression discontinuity

- There appears to be a clear jump at the cut-off
- However, The jump is not from 0 to 1
    - The department actually had some discretion in who they let in below 2.8


70,000
60,000
50,000
40,000
Dollars
2.0
2.5
3.0
3.5
4.0
Average GPA in Economics 1 and 2
β = 7,989 (1,885)
IV = 22,123 (5,847)

Estimating RD empirically

- Graphs are nice, but we want to estimate the effect of majoring in economics on earnings
- Simplest specification:

$$y_{it} = \alpha_0 + \alpha_1 EGPA + \alpha_2 \mathbb{I}(EGPA \geq 2.8) + \alpha_3 \mathbb{I}(EGPA \geq 2.8) \times EGPA + \epsilon_{it} \quad (1)$$

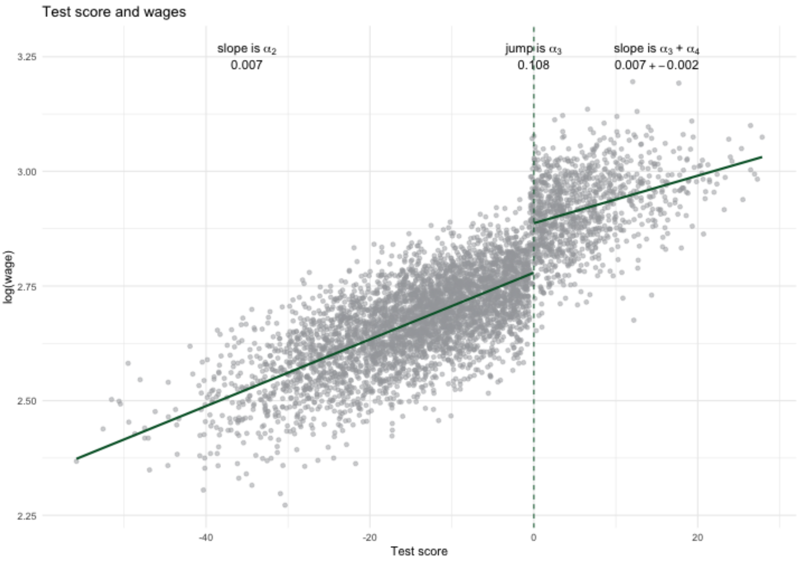
- $EGPA$ is the student's econ GPA
- $\mathbb{I}(EGPA \geq 2.8)$ is an indicator for whether the student had a GPA high enough to declare an econ major
- We are allowing the effect of EGPA to be different for students above and below the cut-off
- Usually first check the intermediate outcome (econ major) and then final outcome (wages)
- NOTE: Common to recenter the running variable to zero at the cut-off

With our fictitious data - test score and wages

```
df <- as_tibble(cbind(scores = scores, wages = wages))
# Recenter running
df$scores <- df$scores - 60
df$abovecut <- ifelse(df$scores>=0, 1, 0)
(reg1 <- feols(log(wages) ~ scores + abovecut*scores,
              data = df,
              vcov = "HC1"))
```

```
## OLS estimation, Dep. Var.: log(wages)
## Observations: 5,000
## Standard-errors: Heteroskedasticity-robust
##                  Estimate Std. Error    t value   Pr(>|t|)
## (Intercept)      2.779060   0.002114 1314.53641  < 2.2e-16 ***
## scores           0.007271   0.000128   56.85130  < 2.2e-16 ***
## abovecut         0.107807   0.004295   25.10096  < 2.2e-16 ***
## scores:abovecut -0.002090   0.000434   -4.81940 1.4826e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.072836   Adj. R2: 0.718956
```

With our fictitious data - test score (recentered) and wages



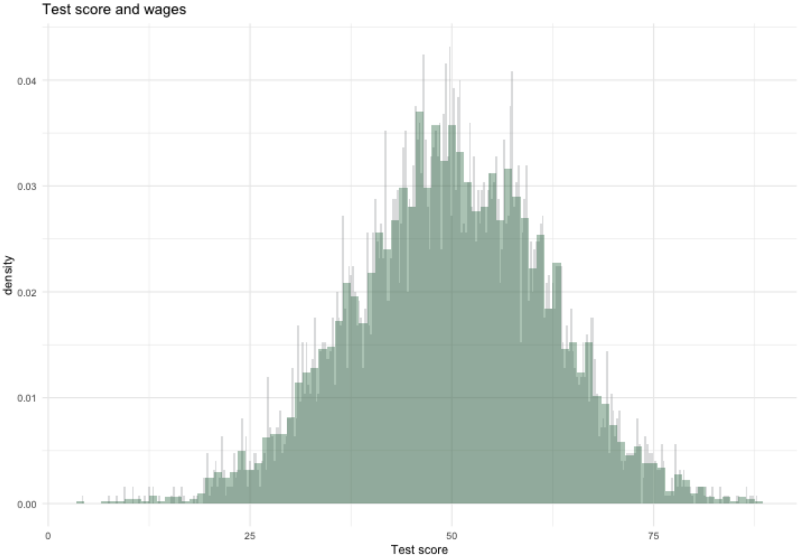
But there's a problem

- There's an issue with fitting a regression like this
- RD is really only valid *around the cut-off*
  - But when we fit a regression like this, we're using all of the data
  - This includes points far from the cut-off
- So in practice nowadays, it's more common to use a local linear regression

- This is an example of *non-parametric estimation*
- You're actually all familiar with this, even if you didn't realize it
  - Density estimates as commonly implemented are a non-parametric estimator
- Consider a histogram:

$$\hat{f}(x) = \frac{\sum_i \mathbb{I}(x_i \in \text{interval } k)}{n} \tag{2}$$

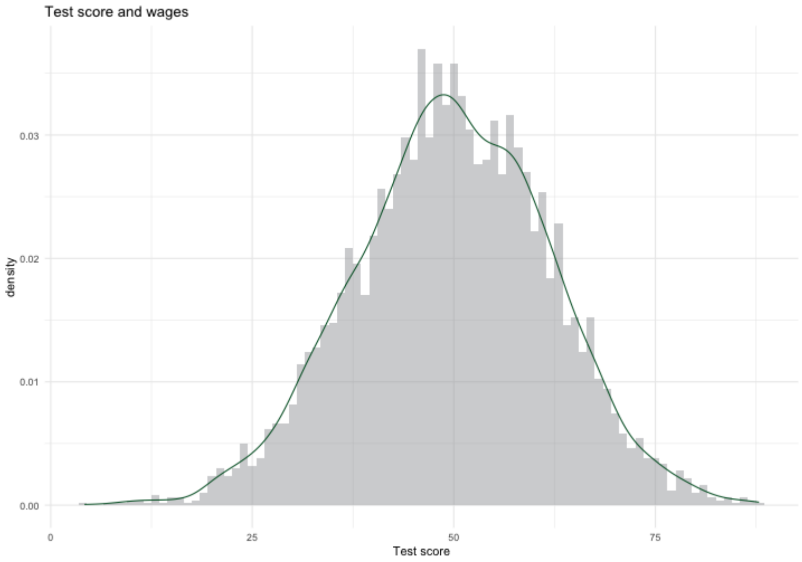
Histograms with different bin widths (0.25 and 1)



Bin width clearly matters for how the density looks

- The size of each bin affects how the density looks
- We can manually choose the bin width
  - It's really somewhat arbitrary
- There's a trade-off between bias and variance
  - The larger the width, the more the bias but the less the variance
- We can call the width of the bin the *bandwidth*
  - Now let's see how this works with non-parametric estimators

Histograms with different bin widths, adding non-parametric



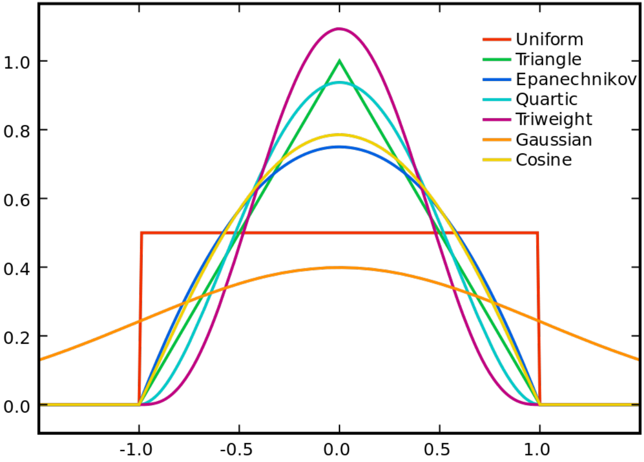
- Define the density estimator as:

$$\hat{f}(x) = \frac{1}{nh} \sum_i K\left(\frac{x - x_i}{h}\right), \tag{3}$$

where K is a kernel function and h is the bandwidth.

- The kernel function decides how to weight observations within the bandwidth
- Kernels often weight observations closer to x more heavily
  - Uniform, traingular, and Epanechnikov are most common
- The intuition: take different values of x and calculate the (weighted) average of the observations within the bandwidth using a given kernel

Kernel examples (Wikipedia)



A note on kernels

- The specific kernel *usually* doesn't make a big difference
- If it does, you probably have a bigger problem
  - You're probably not in a good situation for RD
  - Your results are too sensitive

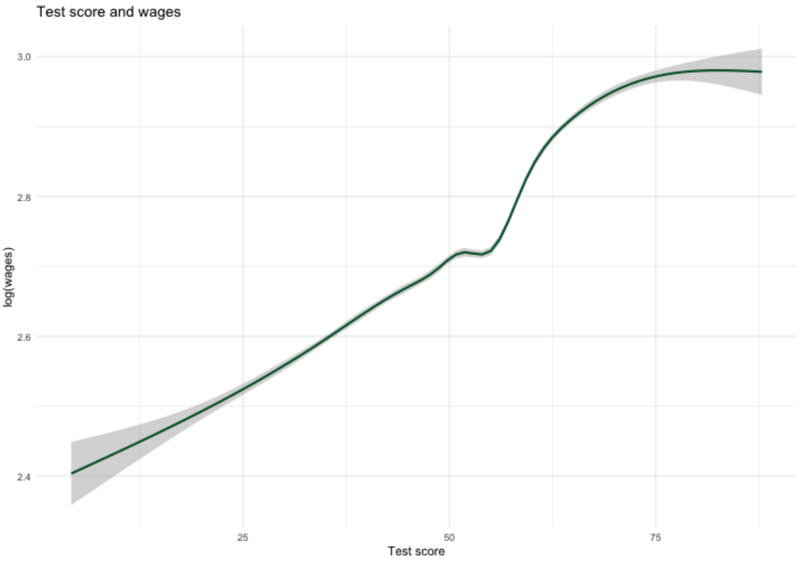
Non-parametric regression

- Let's stick to the simple example of estimating the effect of EGPA on wages
  - Just the two variables, nothing more

- Consider a general non-parametric estimator, where $K_h()$ is a kernel weight and h is the bandwidth:

$$\min_{\alpha,\beta} \sum_{i|x_i \in [x-h, x+h]} (y_i - \alpha - \beta(x_i - x))^2 K_h(x - x_i) \tag{4}$$

Non-parametric regression in R using `geom_smooth`



- What we are essentially going to do with RD is estimate the previous equation separately for $x < x_0$ and $x \geq x_0$, where x_0 is the cut-off
  - We are going to only look right around the cut-off!
- In other words, we are going to estimate:

$$\min_{\alpha_l,\beta_l} \sum_{i|c-h<x_i<c} (y_i - \alpha - \beta(x_i - c))^2 K_h(c - x_i) \tag{5}$$

$$\min_{\alpha_r,\beta_r} \sum_{i|c<x_i<c+h} (y_i - \alpha - \beta(x_i - c))^2 K_h(c - x_i) \tag{6}$$

- The RD estimate will be $\hat{\alpha}_r - \hat{\alpha}_l$

Example with data from Cunningham

- Cunningham has provided data from Lee, Moretti, and Butler (2004)
  - Do voters affect or elect policies? Evidence from the US House
  - *Quarterly Journal of Economics*

- On github: `lmb-data.dta`

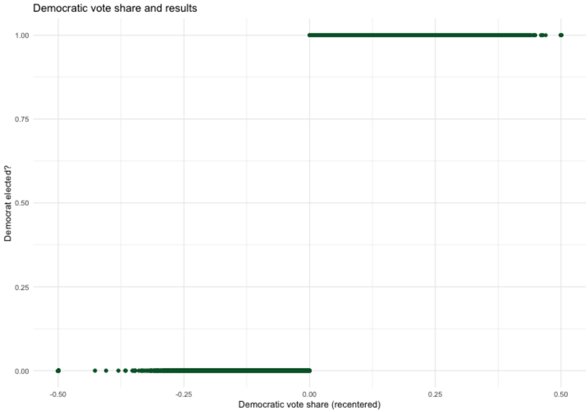
Example with data from Cunningham

```
library(haven)
df <- read_dta("lmb-data.dta")

# recenter vote (straightforward in US context)
df$demvoteshare <- df$demvoteshare - 0.5
```

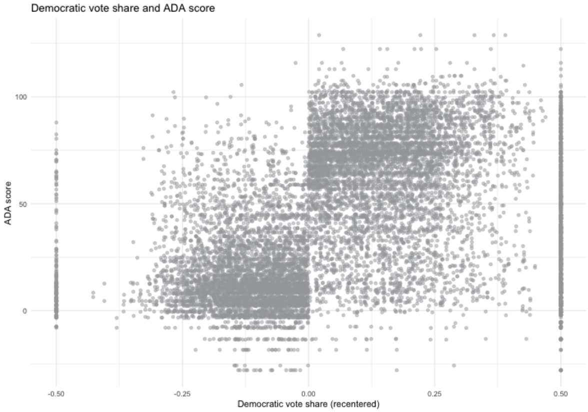
Example with data from Cunningham

```r
ggplot(data = df) +
  geom_point(aes(x = demvoteshare, y = democrat), color = kdisgreen) +
  theme_minimal() +
  labs(x = "Democratic vote share (recentered)", y = "Democrat elected?", title = "Democratic vote share and results")
```



- Let's look at the ADA score, which measures how "liberal" a representative is

Example with data from Cunningham



Estimating the simple linear RD

```r
df$abovecutoff <- ifelse(df$demvoteshare>=0, 1, 0)

(reg1 <- feols(realada ~ demvoteshare + abovecutoff + abovecutoff*demvoteshare,
                data = df,
                cluster = "state"))
```

```
## OLS estimation, Dep. Var.: realada
## Observations: 13,577
## Standard-errors: Clustered (state)
##                            Estimate Std. Error   t value   Pr(>|t|)
## (Intercept)                16.81598    1.59483 10.544034 3.3679e-14 ***
## demvoteshare               -5.68279    7.32303 -0.776016 4.4147e-01
## abovecutoff                55.43136    3.29304 16.832906  < 2.2e-16 ***
## demvoteshare:abovecutoff -55.15188   15.55239 -3.546201 8.7155e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 24.5   Adj. R2: 0.434324
```

Improving RD estimates using `rdrobust` in R

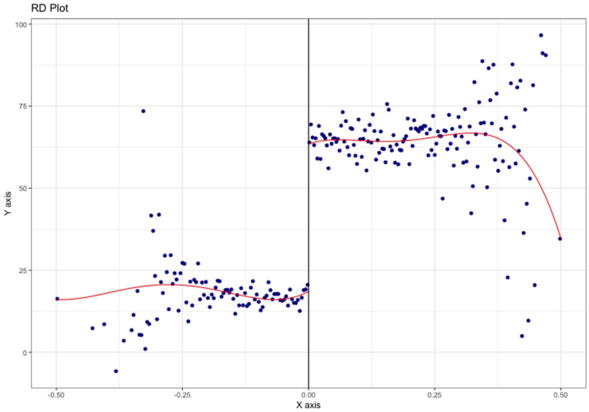
```r
library(rdrobust)
rd1 <- rdrobust(df$realada, df$demvoteshare, c = 0, cluster = df$state)
summary(rd1)
```

```
## Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                 13577
## BW type                        mserd
## Kernel                    Triangular
## VCE method                        NN
##
## Number of Obs.                  5480         8097
## Eff. Number of Obs.             2690         2506
## Order est. (p)                     1            1
## Order bias  (q)                    2            2
## BW est. (h)                    0.113        0.113
## BW bias (b)                    0.156        0.156
## rho (h/b)                      0.726        0.726
## Unique Obs.                     2770         3351
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    46.870     2.079    22.547     0.000    [42.796 , 50.945]
##         Robust         -         -    20.604     0.000    [42.092 , 50.942]
## =============================================================================
```

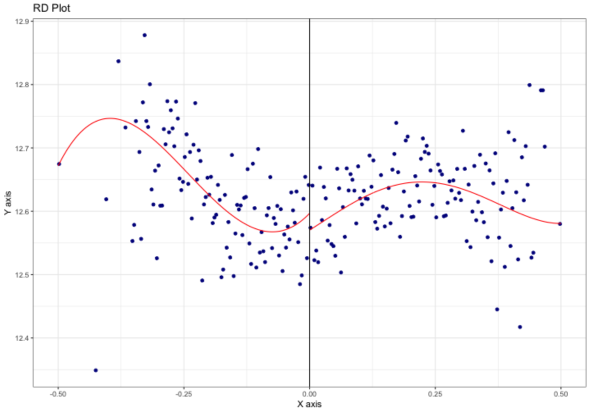
Plotting the estimates

```r
rdplot(df$realada, df$demvoteshare, c = 0)
```

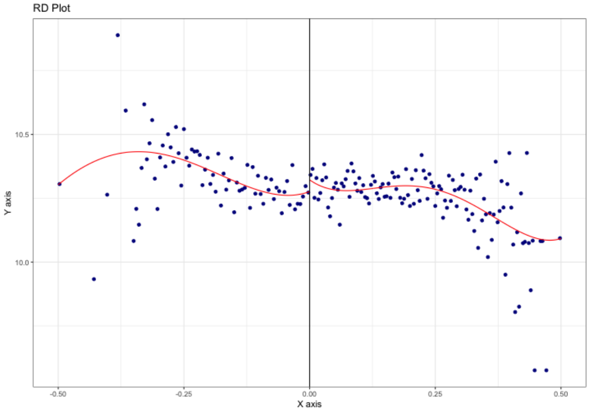
```
## [1] "Mass points detected in the running variable."
```



```
## [1] "Mass points detected in the running variable."
```

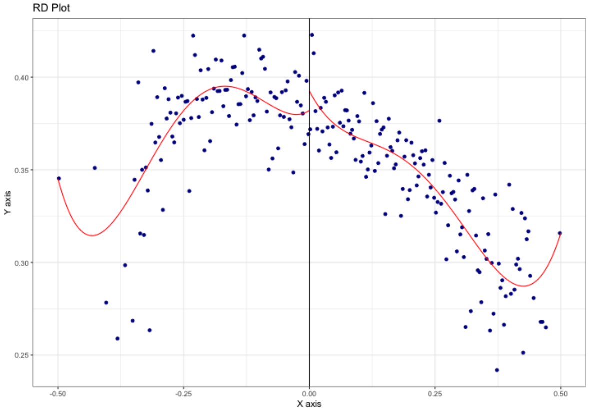


```
## [1] "Mass points detected in the running variable."
```



Checking variables unrelated to treatment - percent HS

```
## [1] "Mass points detected in the running variable."
```

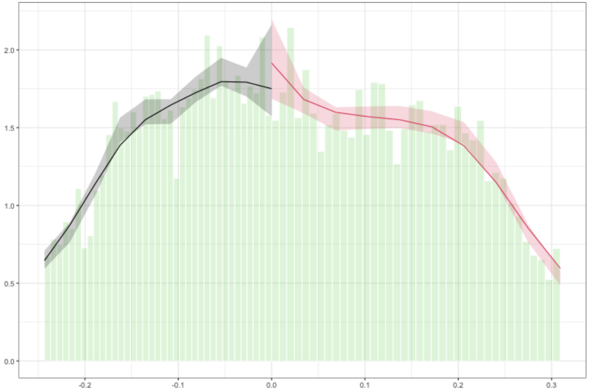


One more thing: checking the density around the cut-off

```r
library(rddensity)
density <- rddensity(df$demvoteshare, c = 0)
summary(density)
```

```
##
## Manipulation testing using local polynomial density estimation.
##
## Number of obs =       13577
## Model =               unrestricted
## Kernel =              triangular
## BW method =           estimated
## VCE method =          jackknife
##
## c = 0                 Left of c           Right of c
## Number of obs         5480                8097
## Eff. Number of obs    1994                2250
## Order est. (p)        2                   2
## Order bias (q)        3                   3
## BW est. (h)           0.081               0.103
##
## Method                T                   P > |T|
## Robust                0.3628              0.7168
##
##
## P-values of binomial tests (H0: p=0.5).
##
## Window Length / 2          <c     >=c    P>|T|
## 0.001                      20      23    0.7608
## 0.002                      47      39    0.4505
## 0.003                      65      58    0.5887
## 0.004                      86      82    0.8170
## 0.005                     101      97    0.8318
```

One more thing: checking the density around the cut-off




```
## $Estl
## Call: lpdensity
##
## Sample size                                      5480
## Polynomial order for point estimation    (p=)    2
## Order of derivative estimated            (v=)    1
## Polynomial order for confidence interval (q=)    3
## Kernel function                                  triangular
## Scaling factor                                   0.40357984678845
```

Fuzzy regression discontinuity

- This example is sharp; all Democrats who received the most votes were elected
- Let's go back to our studying economics example
  - Are we really interested in the effect of EGPA on wages?
  - No. We want to know the effect of majoring in economics on wages
- Well, if people just around the cut-off really are similar, then being just above the cut-off is a valid IV!

- Leaving out the forcing variable for simplicity:

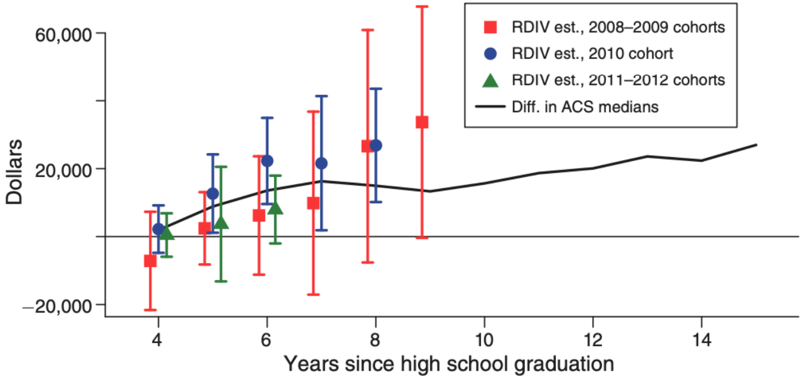
$$econ = \alpha_0 + \alpha_1 \mathbb{I}(EGPA \geq 2.8) + \varepsilon \tag{7}$$
$$wages = \beta_0 + \beta_1 econ + \upsilon \tag{8}$$

- We're going to do this non-parametrically, though.
- Following Hansen and what we learned last week with IVs:

$$\hat{\theta} = \frac{\hat{m}_{c+} - \hat{m}_{c-}}{\hat{p}_{c+} - \hat{p}_{c-}} \tag{9}$$

- We're going to scale the reduced form by the first stage!

Returns to majoring in economics



Notice the large standard errors

- This is a common problem with fuzzy RD using local polynomial regressions
- More generally, non-parametric estimators are *very* data hungry
  - The more controls you add, the worse it gets
  - "The curse of dimensionality"
  - With RD, we are also estimating at the boundary (edge) of the data, which involves similar issues
- By making parametric assumptions, we can get more precise estimates
  - But the estimates may be more biased
  - Trade off!

- We won't go into details here, but regression kink is another similar method
- This is about changes in *slopes*, not intercepts
- For more details, see Card et al. (2015 - *Econometrica*) and CI