

Regression Discontinuity Design

July 3, 2023

In this problem set, we will use *regression discontinuity design* (RDD) to estimate the effect of incumbency disadvantage in developing democracies.

Regression discontinuity design is usually considered as an identification strategy of “natural” experiment – using the conceptualization provided by Thad Dunning (UC Berkeley), natural experiment provides us with the opportunity of causal inference as the process of treatment assignment can be considered **as if** random and often the randomization is out of the control

Before you start, please download the dataset from Moodle and import it into **RStudio**. If you would like to submit this problem set, please complete the questions at the end.

1 Basic Setup

The intuition of regression discontinuity design (RDD) is as follows – suppose we can find a variable (**running variable**) that includes an exogenously determined threshold such that observations above the threshold will receive the treatment of interest.

Then we will need to justify that our observations in the study cannot pre-determine which value they will receive with respect to the running variable, which further means that when it is very close to the threshold, no observations can set themselves for sure above or below the threshold. Since only observations above the threshold will receive the treatment, we can infer that this setting will give us the scenario that **for a specified bandwidth around the threshold of the running variable, the assignment of the treatment status can be considered as random.**

For this particular, the idea goes as follows: In close elections, candidates do not know if they will win the elections for sure. As a result, which candidate will win the elections may be considered as a random event. Candidates who win this election will get the incumbency randomly and we can see if they get more or less votes in the next term to estimate the size of **incumbency advantage**.

For similar ideas, please see the following paper:

Lee, David S. 2008. “Randomized Experiments from Non-random Selection in U.S. House Elections.” *Journal of Econometrics* 142(2): 675-697.

Let’s talk about some possible examples of regression discontinuity in class, using the examples compiled by Dunning (2012).

2 Variables

For this exercise, we will replicate the following article:

Klašnja, Marko. 2015. “Corruption and the Incumbency Disadvantage: Theory and Evidence.” *Journal of Politics* 77(4): 928-942.

We will need the following variables:

- `vote_margin` – a continuous variable to indicate the vote margin of the winner and the first runner-up in the current election
- `win` – a binary indicator to see if the candidate wins the next election
- `educ_t` – a continuous variable for the education level in the municipal-year

2.1 Load Packages

Other than the packages we have used before, we will need the following packages, we will need `rdrobust`. For more information, please see <https://rdpackages.github.io/rdrobus> t/.

```
library(haven)
library(ggplot2)
library(stargazer)
library(tidyverse)
library(rdrobust)
```

Warning: package 'rdrobust' was built under R version 4.3.1

2.2 Read Data

We will first select the variables we need for this exercise – note that we will compare candidates/parties who/which won the majority of votes in close elections vs who/which lost.

For the purpose of comparison, let's create another dataset for really close elections (in which the margin of victory is within 0.1 percentage points).

Here we can use four functions to take a peak at the dataset.

```
names(dta_rdd) # list all columns (variables)
ls(dta_rdd) # list all columns (variables) alphabetically
summary(dta_rdd) # show summary statistics
head(dta_rdd) # show first 6 rows
```

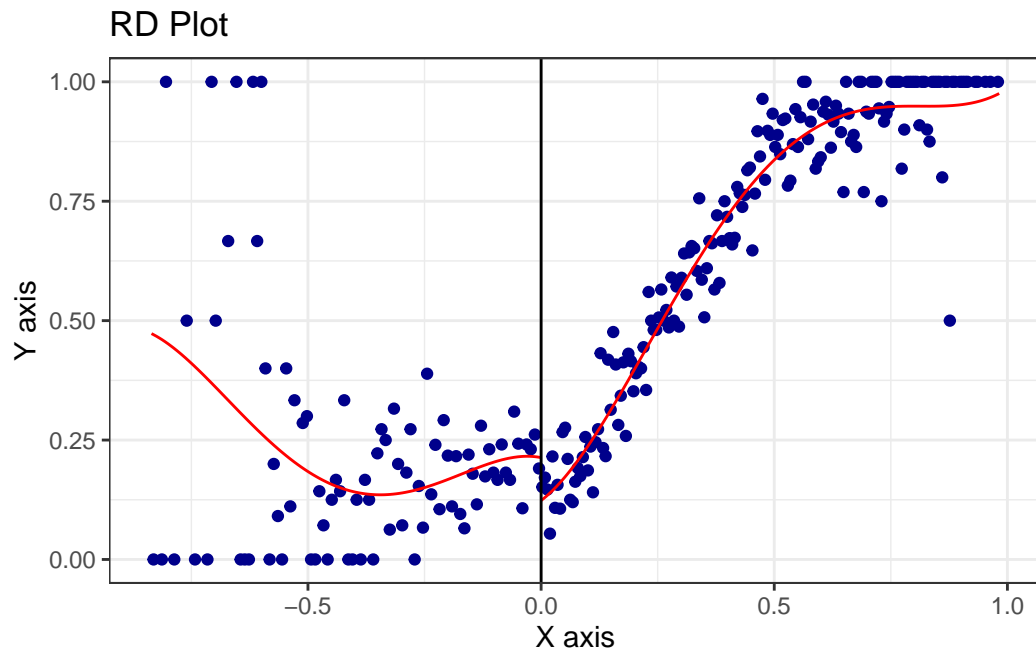
2.3 Disable Scientific Notation (optional)

We will also need to use `options` at the very beginning to disable print out our results in scientific notation.

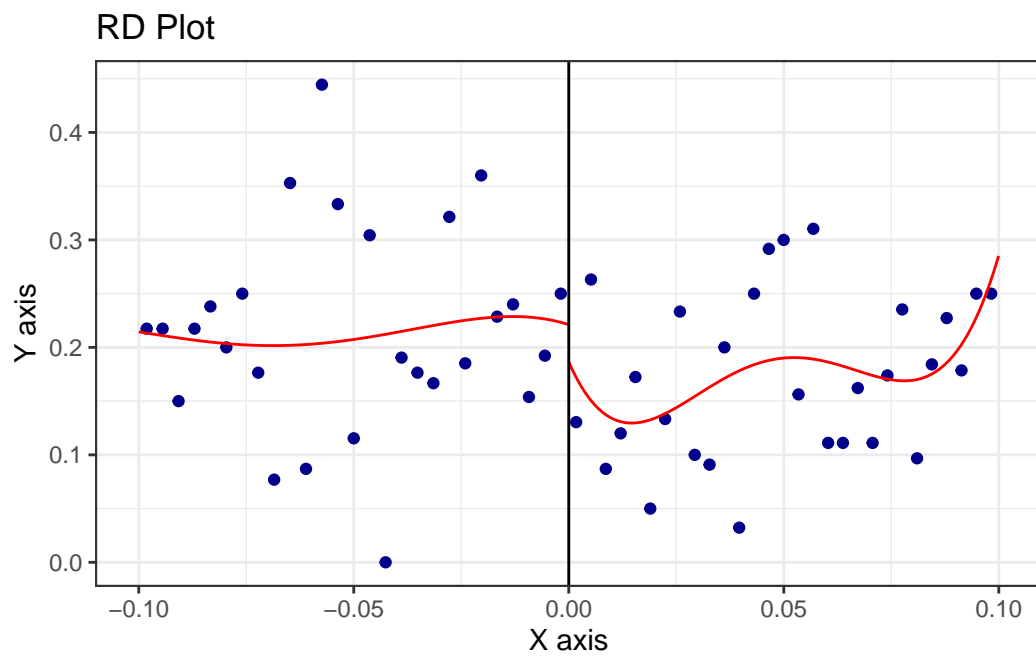
```
options(scipen=999)
```

3 Visualizing the Discontinuity

```
rdplot(dta_rdd$win, dta_rdd$vote_margin)
```



```
rdplot(dta_rdd_sel$win, dta_rdd_sel$vote_margin)
```



4 Balance Statistics

In a nutshell, we want to see if within the specified bandwidths our treated and control units are statistically identical with respect to the pre-treated covariates. Here we will use the education of the municipality to illustrate.

```
t.test(educ_t ~ treat, data=dta_rdd)
```

Welch Two Sample t-test

```
data: educ_t by treat
t = 2.3426, df = 2488.2, p-value = 0.01923
alternative hypothesis: true difference in means between group 0 and group 1 is not equal
95 percent confidence interval:
 0.0004588012 0.0051732725
sample estimates:
mean in group 0 mean in group 1
 0.1108894      0.1080733
```

```
t.test(educ_t ~ treat, data=dta_rdd_sel)
```

Welch Two Sample t-test

```
data: educ_t by treat
t = -0.40275, df = 1439.3, p-value = 0.6872
alternative hypothesis: true difference in means between group 0 and group 1 is not equal
95 percent confidence interval:
-0.005235257 0.003451695
sample estimates:
mean in group 0 mean in group 1
 0.1095820      0.1104738
```

5 Estimating the Effect

5.1 Use Manually Selected Bandwidths

```
rdd_simple_1 <- lm(win ~ treat, data=dta_rdd)
rdd_simple_2 <- lm(win ~ treat, data=dta_rdd_sel)

stargazer(list(rdd_simple_1, rdd_simple_2),
  omit.stat = c("f", "rsq", "ser"),
  covariate.labels = c("Treatment"),
  omit = c("as.factor"),
  type = "text",
  digits = 3,
  no.space = T,
  intercept.bottom = TRUE,
  star.cutoffs = c(0.1, 0.05, 0.01))
```

=====		
Dependent variable:		

	win	
	(1)	(2)

Treatment	0.366***	-0.039*
	(0.013)	(0.021)
Constant	0.193***	0.214***
	(0.012)	(0.016)

Observations	7,924	1,446
Adjusted R2	0.086	0.002
=====		
Note:	*p<0.1; **p<0.05; ***p<0.01	

5.2 Use Bandwidths Selected by the Algorithm

```
rdd_robust <- rdrobust(dta_rdd$win, dta_rdd$vote_margin, c=0)
summary(rdd_robust)
```

Sharp RD estimates using local polynomial regression.

```
Number of Obs.          7924
BW type                 mserd
Kernel                  Triangular
VCE method              NN
```

```
Number of Obs.          1586      6338
Eff. Number of Obs.     959       1399
Order est. (p)           1         1
Order bias (q)           2         2
BW est. (h)              0.159     0.159
BW bias (b)              0.264     0.264
rho (h/b)                0.603     0.603
Unique Obs.              1585     6307
```

Method	Coef.	Std. Err.	z	P> z	[95% C.I.]
Conventional	-0.102	0.037	-2.770	0.006	[-0.173 , -0.030]
Robust	-	-	-2.159	0.031	[-0.178 , -0.009]

Let's take a look at the bandwidths selected by the `rdrobust` package.

```
rdd_robust$bws
```

6 Extended Discussion

- Sharp v fuzzy regression discontinuity design
- Geographic discontinuity design and regression discontinuity in time