Impact Evaluation: Assignment 3

Mukhammed Zainidinov

Regression Discontinuity Design

Government Transfer and Poverty Reduction in Brazil

```
tr = read.csv('data/transfer.csv')
```

1. The fundamental assumption for this design is that there are discontinuities in outcome variables at the cutoff. In our example, we assume that there is:

- a sharp increase in literacy rate at the population cutoffs;
- a sharp increase in educational attainment at the population cutoffs;
- a sharp decline in poverty rate at the population cutoffs.

If the above-stated assumptions are violated, then the results of the RDD will be spurious and the internal validity of the design will be low. The advantage of RDD for this specific application is holding the above assumptions.

2. There are three separate population cutoffs: 10188, 13584, and 16980.

```
cut1 = 10188

cut2 = 13584

cut3 = 16980
```

We have to create three segments for the population. Based on *transfer.csv* data, the minimum value for *pop82* is 7504 and the maximum one is 23751 but we do not have to worry about these two bounds. We take the midpoint between two cutoffs to determine how close each municipality was to the corresponding cutoff. So, we have to identify two midpoints that divide the population into three segments:

$$mp_1 = 11886, \qquad mp_2 = 15282$$

```
mp1 = 11886

mp2 = 15282
```

The conditions for estimating the difference from corresponding difference are as follows:

$$diff = \begin{cases} pop82 - 10188 & \text{if} \quad pop82 \leq 11886 \\ pop82 - 13564 & \text{if} \quad 11886 < pop82 \leq 15282 \\ pop82 - 16980 & \text{if} \quad 15282 < pop82 \end{cases}$$

Then we create a new variable *closeness* for normalized percentage score running the following:

```
tr$closeness = ifelse(tr$pop82 <= mp1, (tr$pop82 - cut1)/cut1*100,
               ifelse((tr$pop82 > mp1) & (tr$pop82 <= mp2),
                      (tr$pop82 - cut2)/cut2*100,
                      (tr$pop82 - cut3)/cut3*100))
```

3. Keep only observations for whom *closeness* is within 3 points of the funding cutoff on either side:

```
library(tidyverse)
tr3 = tr %>%
    dplyr::filter(closeness >= -3 & closeness <= 3)
```

Create regression discontinuity data:

```
library(rddtools)
lit3 = rdd_data(literate91, closeness, data = tr3, cutpoint = 0)
edu3 = rdd_data(educ91, closeness, data = tr3, cutpoint = 0)
pov3 = rdd_data(poverty91, closeness, data = tr3, cutpoint = 0)
```

Run parametric regressions to estimate treatment effect:

```
lit3_para = rdd_reg_lm(lit3, order = 1)
edu3_para = rdd_reg_lm(edu3, order = 1)
pov3_para = rdd_reg_lm(pov3, order = 1)

library(stargazer)
stargazer(lit3_para, edu3_para, pov3_para,
          dep.var.labels = 'Outcome variable',
          column.labels = c('literacy', 'education', 'poverty'),
          title = 'Parametric regressions to estimate the treatment effect',
          type = "text", out = 'output/rdd_parametric.rtf')
```

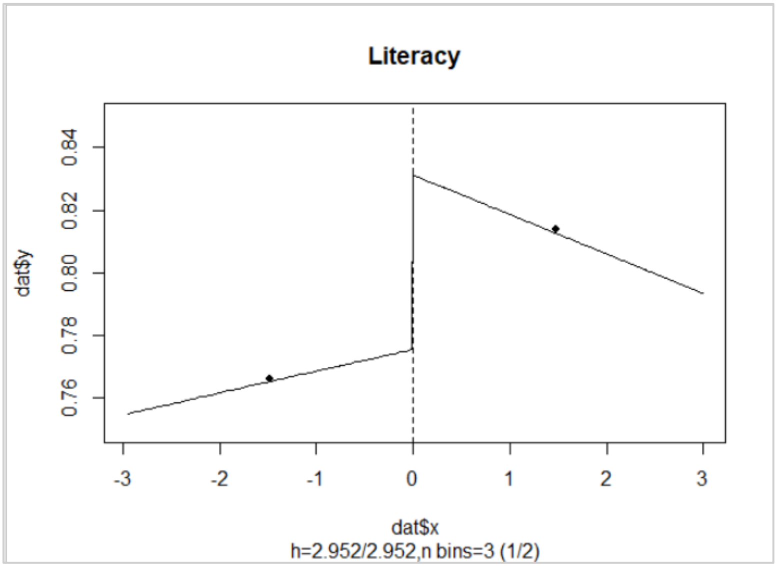
Parametric regressions to estimate the treatment effect

	Dependent variable:		
	Outcome variable		
	literacy	education	poverty
	(1)	(2)	(3)
D	0.056	0.584*	-0.060
	(0.036)	(0.315)	(0.051)
x	0.007	0.059	-0.034
	(0.016)	(0.136)	(0.022)
x_right	-0.020	-0.215	0.065**
	(0.021)	(0.185)	(0.030)
Constant	0.776***	4.407***	0.592***
	(0.026)	(0.228)	(0.037)
Observations	297	297	297
R2	0.025	0.030	0.035
Adjusted R2	0.015	0.020	0.025
Residual Std. Error (df = 293)	0.159	1.380	0.225
F Statistic (df = 3; 293)	2.473*	3.056**	3.491**
Note:	*p<0.1; **p<0.05; ***p<0.01		

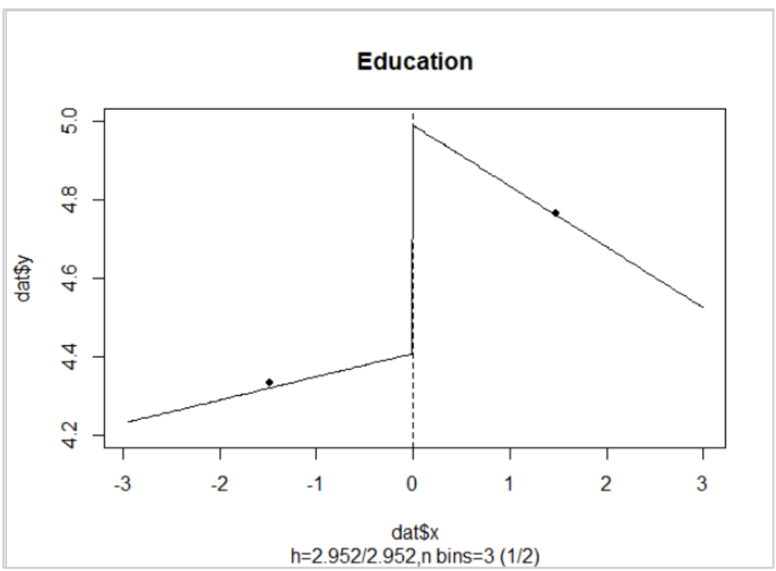
The regression results show that government transfer has no significant causal effect on literacy and poverty. However, the government transfer has a causal effect on educational attainment at a 10% significance level. We can say with a 90% confidence level that the average causal effect of government transfer on educational attainment is 0.584 units.

4. Plotting data points, fitted regression lines, and the population threshold

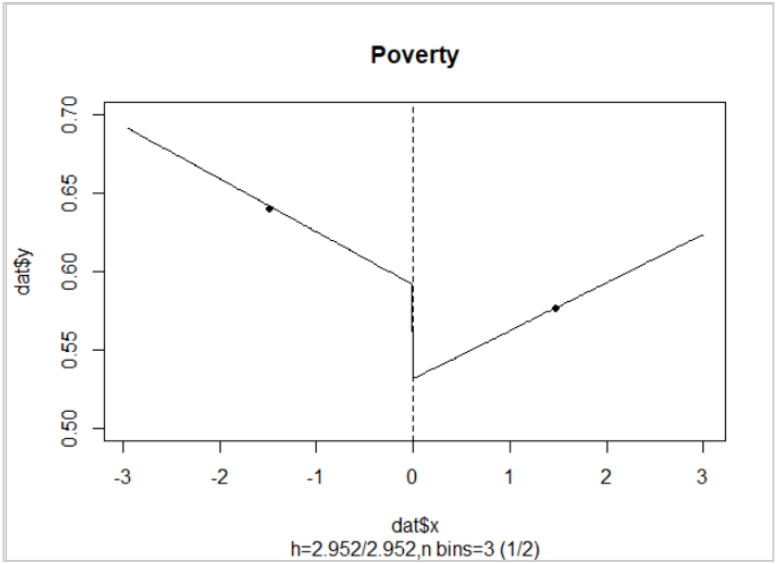
```
plot(lit3_para, ylim = c(0.75, 0.85)) + title('Literacy')
plot(edu3_para, ylim = c(4.2, 5)) + title('Education')
plot(pov3_para, ylim = c(0.5, 0.7)) + title('Poverty')
```



There is a sharp increase in the literacy rate at cutoff, and then the literacy rate decreases as the number of the population goes further away from the cutoff.



There is a sharp increase in average years of education at cutoff, and then average years of education decreases as the number of the population goes further away from the cutoff.



There is a sharp decrease in the poverty rate at cutoff, and then the poverty rate increases as the number of the population goes further away from the cutoff. In general, all plots depict what we expected.

5. Estimating the treatment effect for various percentage points below and above the threshold

Estimate for 1 percentage point below and above the threshold:

```
tr1 = tr %>%
  dplyr::filter(closeness >= -1 & closeness <= 1)

lit1 = rdd_data(literate91, closeness, data = tr1, cutpoint = 0)
edu1 = rdd_data(educ91, closeness, data = tr1, cutpoint = 0)
pov1 = rdd_data(poverty91, closeness, data = tr1, cutpoint = 0)

lit1_para = rdd_reg_lm(lit1, order = 1)
edu1_para = rdd_reg_lm(edu1, order = 1)
pov1_para = rdd_reg_lm(pov1, order = 1)
```

Estimate for 2 percentage points below and above the threshold:

```
tr2 = tr %>%
  dplyr::filter(closeness >= -2 & closeness <= 2)
```

```
lit2 = rdd_data(literate91, closeness, data = tr2, cutpoint = 0)
edu2 = rdd_data(educ91, closeness, data = tr2, cutpoint = 0)
pov2 = rdd_data(poverty91, closeness, data = tr2, cutpoint = 0)

lit2_para = rdd_reg_lm(lit2, order = 1)
edu2_para = rdd_reg_lm(edu2, order = 1)
pov2_para = rdd_reg_lm(pov2, order = 1)
```

Estimate for 4 percentage points below and above the threshold:

```
tr4 = tr %>%
  dplyr::filter(closeness >= -4 & closeness <= 4)

lit4 = rdd_data(literate91, closeness, data = tr4, cutpoint = 0)
edu4 = rdd_data(educ91, closeness, data = tr4, cutpoint = 0)
pov4 = rdd_data(poverty91, closeness, data = tr4, cutpoint = 0)

lit4_para = rdd_reg_lm(lit4, order = 1)
edu4_para = rdd_reg_lm(edu4, order = 1)
pov4_para = rdd_reg_lm(pov4, order = 1)
```

Estimate for 5 percentage points below and above the threshold:

```
tr5 = tr %>%
  dplyr::filter(closeness >= -5 & closeness <= 5)

lit5 = rdd_data(literate91, closeness, data = tr5, cutpoint = 0)
edu5 = rdd_data(educ91, closeness, data = tr5, cutpoint = 0)
pov5 = rdd_data(poverty91, closeness, data = tr5, cutpoint = 0)

lit5_para = rdd_reg_lm(lit5, order = 1)
edu5_para = rdd_reg_lm(edu5, order = 1)
pov5_para = rdd_reg_lm(pov5, order = 1)
```

Estimated treatment effects for literacy rate, education, and poverty rate by percentage points below and above the threshold:

```
stargazer(lit1_para, lit2_para, lit3_para, lit4_para, lit5_para,
          dep.var.labels = 'literate91',
          column.labels = c('-1% <= X <= 1%', '-2% <= X <= 2%',
                            '-3% <= X <= 3%',
                            '-4% <= X <= 4%', '-5% <= X <= 5%'),
          title = 'Parametric regressions to estimate treatment effect on
                    literacy rate',
          type = "text", out = 'output/rdd_lit_para.rtf')


stargazer(edu1_para, edu2_para, edu3_para, edu4_para, edu5_para,
          dep.var.labels = 'educ91',
          column.labels = c('-1% <= X <= 1%', '-2% <= X <= 2%',
                            '-3% <= X <= 3%',
                            '-4% <= X <= 4%', '-5% <= X <= 5%'),
          title = 'Parametric regressions to estimate treatment effect on
                    years of education',
          type = "text", out = 'output/rdd_edu_para.rtf')


stargazer(pov1_para, pov2_para, pov3_para, pov4_para, pov5_para,
          dep.var.labels = 'poverty91',
          column.labels = c('-1% <= X <= 1%', '-2% <= X <= 2%',
                            '-3% <= X <= 3%',
                            '-4% <= X <= 4%', '-5% <= X <= 5%'),
          title = 'Parametric regressions to estimate treatment effect on the
                   poverty rate',
          type = "text", out = 'output/rdd_pov_para.rtf')
```

Parametric regressions to estimate treatment effect on literacy rate

	Dependent variable: literate91				
	-1% <= X <= 1%	-2% <= X <= 2%	-3% <= X <= 3%	-4% <= X <= 4%	-5% <= X <= 5%
	(1)	(2)	(3)	(4)	(5)
D	0.037	0.054	0.056	0.075**	0.065**
	(0.061)	(0.045)	(0.036)	(0.033)	(0.029)
x	0.007	0.014	0.007	-0.005	-0.011
	(0.071)	(0.029)	(0.016)	(0.010)	(0.007)
x_right	0.015	-0.031	-0.020	-0.011	0.005
	(0.102)	(0.039)	(0.021)	(0.014)	(0.010)
Constant	0.779***	0.781***	0.776***	0.762***	0.756***
	(0.045)	(0.033)	(0.026)	(0.023)	(0.020)
Observations	107	202	297	391	479
R2	0.026	0.029	0.025	0.018	0.012
Adjusted R2	-0.002	0.014	0.015	0.010	0.005
Residual Std. Error	0.162 (df = 103)	0.161 (df = 198)	0.159 (df = 293)	0.162 (df = 387)	0.162 (df = 475)
F Statistic	0.934 (df = 3; 103)	1.969 (df = 3; 198)	2.473* (df = 3; 293)	2.322* (df = 3; 387)	1.880 (df = 3; 475)
Note:					*p<0.1; **p<0.05; ***p<0.01

The effect of the treatment on literacy rate is statistically significant only when we take 4 and 5 percentage points below and above the threshold. The estimated average effect of government transfer on literacy rate is 0.075 units when we take 4 percentage points below and above the threshold, 0.065 when we take 5 percentage points below and above the threshold.

Parametric regressions to estimate treatment effect on years of education

	Dependent variable:				
	educ91				
	-1% <= X <= 1%	-2% <= X <= 2%	-3% <= X <= 3%	-4% <= X <= 4%	-5% <= X <= 5%
	(1)	(2)	(3)	(4)	(5)
D	0.215	0.406	0.584*	0.712**	0.690***
	(0.538)	(0.391)	(0.315)	(0.281)	(0.252)
x	0.366	0.203	0.059	-0.051	-0.116*
	(0.625)	(0.250)	(0.136)	(0.090)	(0.063)
x_right	-0.182	-0.252	-0.215	-0.095	0.032
	(0.898)	(0.339)	(0.185)	(0.125)	(0.090)
Constant	4.595***	4.503***	4.407***	4.290***	4.209***
	(0.395)	(0.286)	(0.228)	(0.201)	(0.179)
Observations	107	202	297	391	479
R2	0.035	0.042	0.030	0.021	0.016
Adjusted R2	0.007	0.028	0.020	0.014	0.010
Residual Std. Error	1.429 (df = 103)	1.413 (df = 198)	1.380 (df = 293)	1.403 (df = 387)	1.414 (df = 475)
F Statistic	1.252 (df = 3; 103)	2.897** (df = 3; 198)	3.056** (df = 3; 293)	2.832** (df = 3; 387)	2.549* (df = 3; 475)
Note:					*p<0.1; **p<0.05; ***p<0.01

The effect of the treatment on educational attainment is statistically significant at a 5% significance level only when we take 4 and 5 percentage points below and above the threshold. The estimated average effect of government transfer on literacy rate is 0.712 units when we take 4 percentage points below and above the threshold, 0.690 when we take 5 percentage points below and above the threshold.

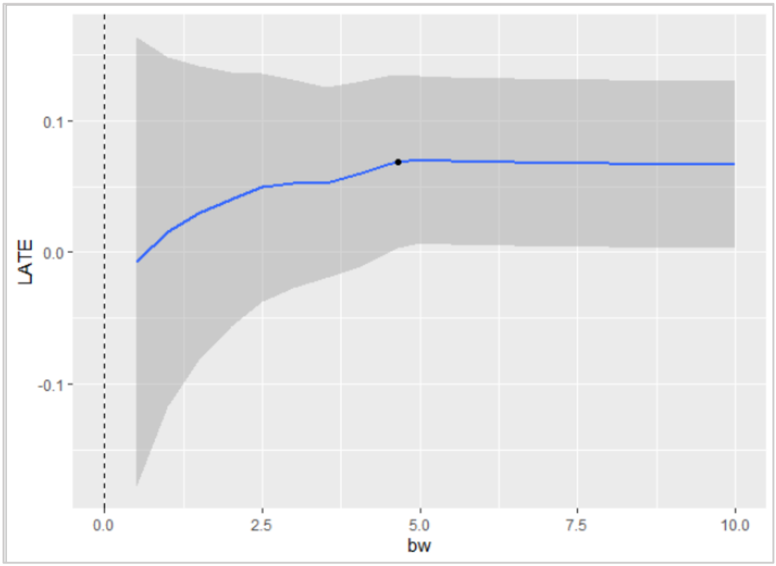
Parametric regressions to estimate treatment effect on the poverty rate

	Dependent variable:				
	poverty91				
	-1% <= X <= 1%	-2% <= X <= 2%	-3% <= X <= 3%	-4% <= X <= 4%	-5% <= X <= 5%
	(1)	(2)	(3)	(4)	(5)
D	-0.040	-0.046	-0.060	-0.081*	-0.083**
	(0.091)	(0.064)	(0.051)	(0.044)	(0.040)
x	-0.018	-0.040	-0.034	-0.008	0.007
	(0.106)	(0.041)	(0.022)	(0.014)	(0.010)
x_right	-0.011	0.057	0.065**	0.030	0.004
	(0.152)	(0.056)	(0.030)	(0.020)	(0.014)
Constant	0.599***	0.587***	0.592***	0.619***	0.639***
	(0.067)	(0.047)	(0.037)	(0.032)	(0.028)
Observations	107	202	297	391	479
R2	0.019	0.028	0.035	0.021	0.011
Adjusted R2	-0.010	0.013	0.025	0.014	0.004
Residual Std. Error	0.241 (df = 103)	0.233 (df = 198)	0.225 (df = 293)	0.221 (df = 387)	0.224 (df = 475)
F Statistic	0.653 (df = 3; 103)	1.892 (df = 3; 198)	3.491** (df = 3; 293)	2.798** (df = 3; 387)	1.717 (df = 3; 475)
Note:					*p<0.1; **p<0.05; ***p<0.01

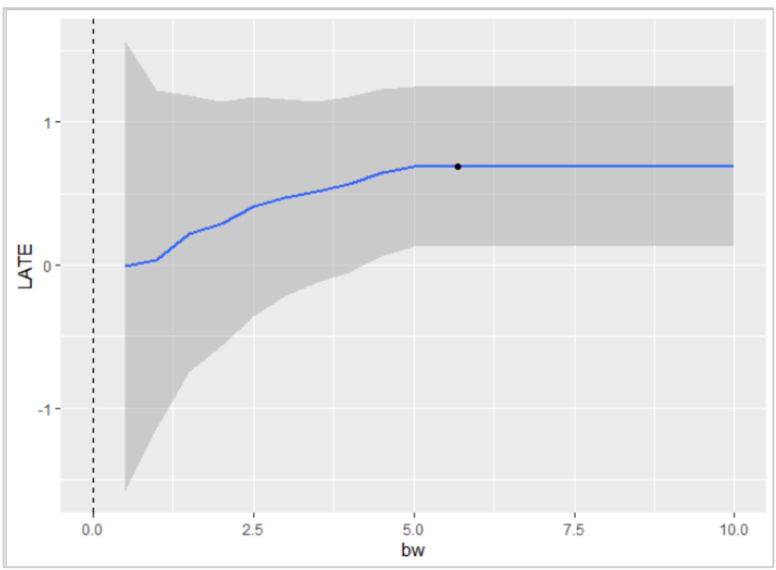
The effect of the treatment on poverty is statistically significant at a 5% significance level only when we 5 percentage points below and above the threshold. The estimated average effect of government transfer on the poverty rate is -0.083 units.

For sensitivity analysis, I decided to take estimates of 5 percentage points below and above the threshold because statistical significance of treatment effect on each outcome variable at 5% significance level.

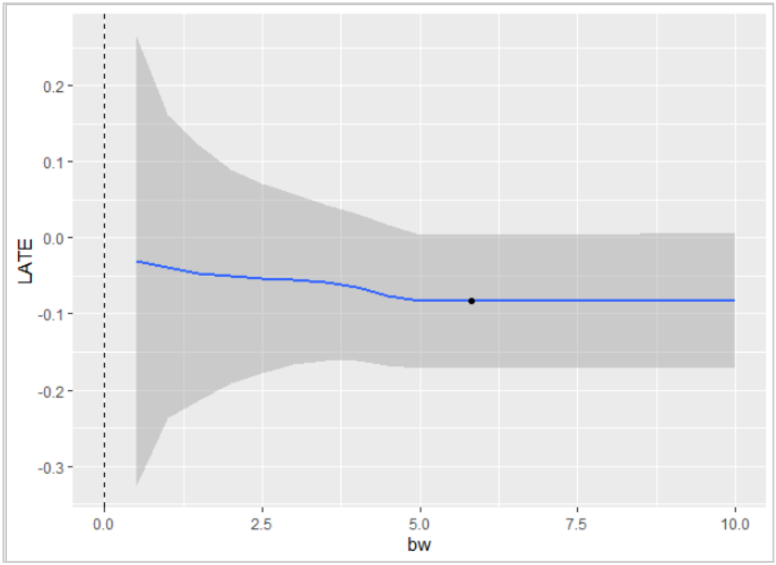
```
bw_lit5 = rdd_bw_ik(lit5)

lit5_nonpara = rdd_reg_np(rdd_object = lit5, bw = bw_lit5)

plotSensi(lit5_nonpara, from = 0.5, to = 10, by = 0.5)
```




```
bw_edu5 = rdd_bw_ik(edu5)

edu5_nonpara = rdd_reg_np(rdd_object = edu5, bw = bw_edu5)

plotSensi(edu5_nonpara, from = 0.5, to = 10, by = 0.5)
```



```
bw_pov5 = rdd_bw_ik(pov5)
pov5_nonpara = rdd_reg_np(rdd_object = pov5, bw = bw_pov5)
plotSensi(pov5_nonpara, from = 0.5, to = 10, by = 0.5)
```



6. Estimate RDD before the population-based government transfers began.

```
edu3_before = rdd_data(educ80, closeness, data = tr3, cutpoint = 0)
pov3_before = rdd_data(poverty80, closeness, data = tr3, cutpoint = 0)

edu3_before_para = rdd_reg_lm(edu3_before, order = 1)
pov3_before_para = rdd_reg_lm(pov3_before, order = 1)

stargazer(edu3_before_para, pov3_before_para,
          dep.var.labels = 'Outcome variable',
          column.labels = c('education', 'poverty'),
          title ='Parametric regressions to estimate the treatment effect',
          type = "text", out = 'output/rdd_parametric_before.rtf')
```

Parametric regressions to estimate the treatment effect

	Dependent variable: Outcome variable	
	education	poverty
	(1)	(2)
D	0.192	-0.027
	(0.219)	(0.050)
x	0.086	-0.042*
	(0.095)	(0.022)
x_right	-0.124	0.066**
	(0.129)	(0.029)
Constant	1.885***	0.570***
	(0.158)	(0.036)
Observations	297	297
R2	0.022	0.031
Adjusted R2	0.012	0.021
Residual Std. Error (df = 293)	0.958	0.218
F Statistic (df = 3; 293)	2.167*	3.131**
Note:	*p<0.1; **p<0.05; ***p<0.01	

The estimated results are not statistically significant. If we look at the results in question 3, the treatment effect on education is statistically significant at a 90% confidence level, but poverty is not statistically significant. So, the results in question 3 suggest the validity of treatment effect only on educational attainment.

Difference in Differences

Banks in Business

7. Import data

```
library(readr)

banks = read.csv('data/banks.csv')
```

8. Average number of banks in each year in districts 6 and 8

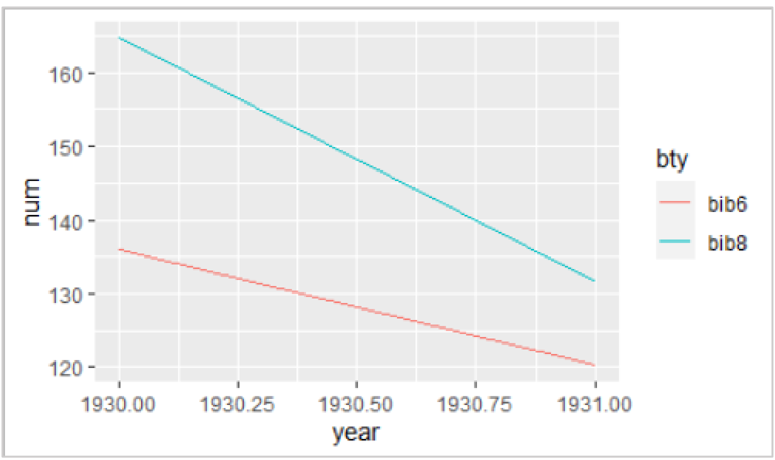
```
banks_mean = banks %>%
  group_by(year) %>%
  summarize_at(vars(bib6, bib8, bio6, bio8), mean)
```

9. Gather the banks in districts 6 and 8

```
banks_gather = gather(banks_mean, 'bty', 'num', -year)
```

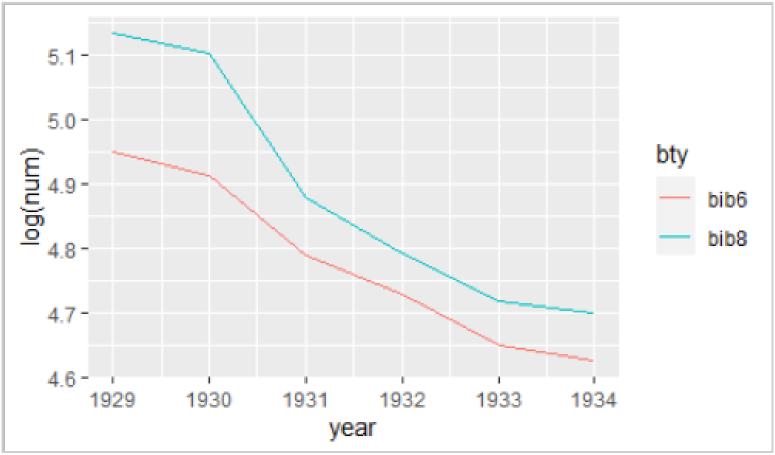
10. Filter for the years 1930 and 1931 and plot the banks in business

```
banks_30_31 = banks_gather%>%
  filter(year == '1930' | year == '1931')

ggplot(banks_30_31) + geom_line(aes(x = year, y = num, color = bty))
```



11. Explore the parallel trends assumption visually

```
ggplot(banks_gather, aes(x=year, y=log(num), colour=bty)) + geom_line()
```



From the graph, we see a parallel trend for the period 1929 – 1930. Under parallel trend assumption, if we assume that without treatment, the 6th and 8th districts would have the same trend, then there is an improvement in the 6th district. If there were no easy lending to troubled banks, the number of banks in business in the 6th district would be lower.