FLS 6415: Replication 5 - Discontinuities

April 2019

To be submitted (code + answers) by midnight, Wednesday 15th May.

First read the paper by Titiunik (2011) on the class website. However, we will not use her pre-prepared dataset - we will start the analysis by constructing our own. As with all regression discontinuities, 90% of the work is in preparing the dataset.

1. We need data from 2000 and 2004 Mayors. We can download this from cepespdata.io. Choose "Consultar resultados eleitorais" and we want prefeito data at the municipal level for *parties* in the 2000 elections first. Then 'Selecionar Colunas' to add the `COD_MUN_IBGE` variable. Then export this dataset to 'CSV'. Finally, make the same selection for 2004 and download this as a separate CSV. Details for the description of each variable can be found on cepesp.io (see adicionar colunas).

```r
d2000 <- read.csv("TSE_PREFEITO_MUNICIPIO_PARTIDO_2000.csv")
d2004 <- read.csv("TSE_PREFEITO_MUNICIPIO_PARTIDO_2004.csv")
```

2. First, prepare the 2000 dataset: a. Filter the data to include only the first round and to remove municipalities where only one party ran. b. Calculate the total number of votes in each municipality. c. Calculate the percentage vote share for each party in each municipal contest d. Calculate which position the party came in the municipal election (their `rank`). e. Filter the dataset to focus on only the first and second-placed parties. f. Make a binary variable that is equal to '1' for the winning party that becomes the incumbent. g. Add a column for the vote share of the winning party, and a second column for the vote share of the second-placed party. *(potentially tricky, email me if you're stuck)* h. Remove two annoying cases where the election result was tied (so we dont know who became the incumbent)! i. Calculate the winning margin of each party. - For the winning party, this is the vote share of the winner minus the vote share of the second place party. - For the second-place parties, this is the vote share of the second place party minus the vote share of the winning party.

```r
d2000 <- d2000 %>% filter(NUM_TURNO==1) %>%
  group_by(COD_MUN_IBGE) %>%
  mutate(Num_Parties=n(),
         Tot_votes_2000=sum(QTDE_VOTOS,na.rm=T),
         Pct_Votes_2000=QTDE_VOTOS/Tot_votes_2000,
         Rank=rank(-QTDE_VOTOS)) %>%
  filter(Rank<3 & Num_Parties!=1) %>%
  mutate(Incumbent=ifelse(Rank==1,1,0)) %>%
  arrange(COD_MUN_IBGE, Rank) %>%
  mutate(first_rank_vote_pct=max(Pct_Votes_2000),
         second_rank_vote_pct=nth(Pct_Votes_2000,2)) %>%
  filter(first_rank_vote_pct!=second_rank_vote_pct) %>%
  mutate(Win_Margin=case_when(Rank==1~first_rank_vote_pct-second_rank_vote_pct,
                              Rank==2~second_rank_vote_pct-first_rank_vote_pct))
```

3. Next, prepare the 2004 dataset: a. Filter for the first round. b. Calculate our outcome measure: the vote share (*not* the winning margin) of each party in each municipal contest. c. Select only the Municipality Code, Party and Vote Share variables

```r
d2004 <- d2004 %>% filter(NUM_TURNO==1) %>%
  group_by(COD_MUN_IBGE) %>%
  mutate(Tot_votes_2004=sum(QTDE_VOTOS,na.rm=T),
         Pct_Votes_2004=QTDE_VOTOS/Tot_votes_2004) %>%
```

```r
  select(COD_MUN_IBGE, SIGLA_PARTIDO, Pct_Votes_2004)
```

4. Join the two datasets (2000 and 2004 for all parties) based on the municipality (`COD_MUN_IBGE`) and party (`NUMERO_PARTIDO`) so that for every party that ran in both 2000 and 2004 we know what vote share they got in 2004. (What type of join do we want here? Left, Right, Inner?)

```r
d <- d2000 %>% inner_join(d2004, by=c("COD_MUN_IBGE", "SIGLA_PARTIDO"))
```

We want an inner join. This ensures that our dataset includes only parties that ran candidates in both the 2000 and 2004 elections, which is necessary for us to have the data to analyze the effect of interest.

5. If we did not know about regression discontinuity we might run the observational OLS regression of 2004 vote share on incumbency in 2000. For the next set of questions we will focus only on the PMDB. Subset the data so it includes only the PMDB, run and interpret this regression.

```r
d_PMDB <- d %>% filter(SIGLA_PARTIDO=="PMDB")

d_PMDB %>% lm(Pct_Votes_2004~Incumbent, data=.) %>%
  stargazer(header=F, keep.stat=c("n"), title="Q5")
```

Table 1: Q5

	Dependent variable:
	Pct_Votes_2004
Incumbent	0.017**
	(0.007)
Constant	0.438***
	(0.005)
Observations	1,415
Note:	*p<0.1; **p<0.05; ***p<0.01

PMDB Incumbency in 2000 is associated with a 1.7% point increase in the 2004 vote share for the PMDB, which is significant at the $p = 0.05$ level.

6. Before implementing any regression discontinuities, let's check for balance around the discontinuity. Within a +/-1% winning margin in 2000 check the balance of the total number of voters in treated and control municipalities in 2000 (we created this variable in Q2). Compare this to the balance for a winning margin of +/-3%.

```r
d_PMDB %>% filter(Win_Margin<0.01 & Win_Margin>-0.01) %>%
  t.test(Tot_votes_2000~Incumbent, data=.)
```

```
##
##  Welch Two Sample t-test
##
## data:  Tot_votes_2000 by Incumbent
## t = -0.1679, df = 75.947, p-value = 0.8671
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6157.341  5199.912
## sample estimates:
```

```
## mean in group 0 mean in group 1
##        9143.030        9621.745
```

```r
d_PMDB %>% filter(Win_Margin<0.03 & Win_Margin>-0.03) %>%
  t.test(Tot_votes_2000~Incumbent, data=.)
```

```
##
##  Welch Two Sample t-test
##
## data:  Tot_votes_2000 by Incumbent
## t = -0.54597, df = 151.9, p-value = 0.5859
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -22502.61  12758.49
## sample estimates:
## mean in group 0 mean in group 1
##        12789.56        17661.62
```

There is statistically good balance for both bandwidths. However, there is substantively much better balance (a much smaller absolute difference) for the smaller +/-1% bandwidth.

7. Next, check for sorting and manipulation of the threshold with the McCrary density test using the `rddensity` function. Interpret the results and produce a density plot using the `rdplotdensity`.

The results fail to reject the null hypothesis of no sorting. There are a very similar number of units just either side of the threshold, and the distribution is simmilar to what we would expect if there was no sorting.

```r
library(rddensity)
density_test <- rddensity(d_PMDB$Win_Margin)

density_test %>% summary()
```

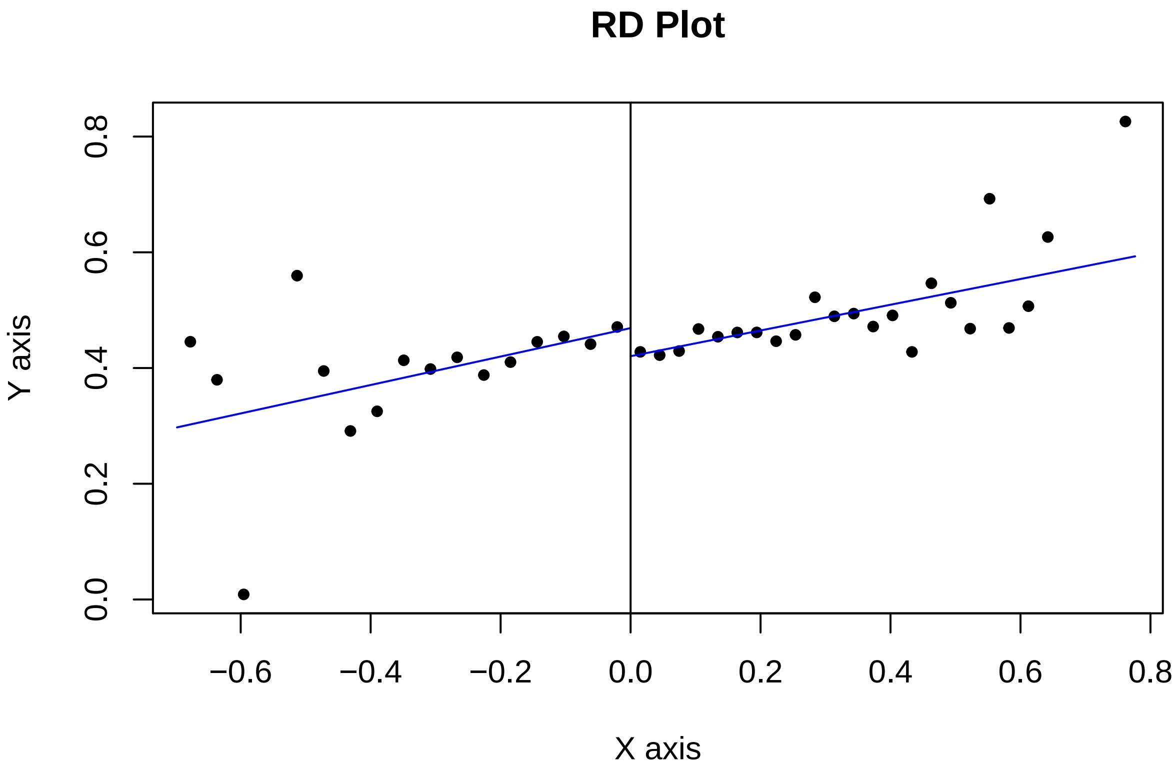
```
##
## RD Manipulation Test using local polynomial density estimation.
##
## Number of obs =       1415
## Model =               unrestricted
## Kernel =              triangular
## BW method =           comb
## VCE method =          jackknife
##
## Cutoff c = 0          Left of c           Right of c
## Number of obs         649                 766
## Eff. Number of obs    507                 600
## Order est. (p)        2                   2
## Order bias (q)        3                   3
## BW est. (h)           0.196               0.231
##
## Method                T                   P > |T|
## Robust                -0.716              0.474
```

```r
#density_test %>% rdplotdensity(d_PMDB$Win_Margin)
```

8. Before we run the analysis, let's construct a regression discontinuity plot to visually inspect the causal effect of incumbency at the threshold. Using a pre-packaged command like `rdplot` from the `rdrobust` package, create a regression discontinuity plot for the effect of incumbency in 2000 on vote share in 2004 for the PMDB. Use linear regression lines. Interpret the results.

Controlling for the running variable (win margin) there is a negative jump in 2004 vote share at the threshold of winning/losing. This suggests a negative incumbency effect.

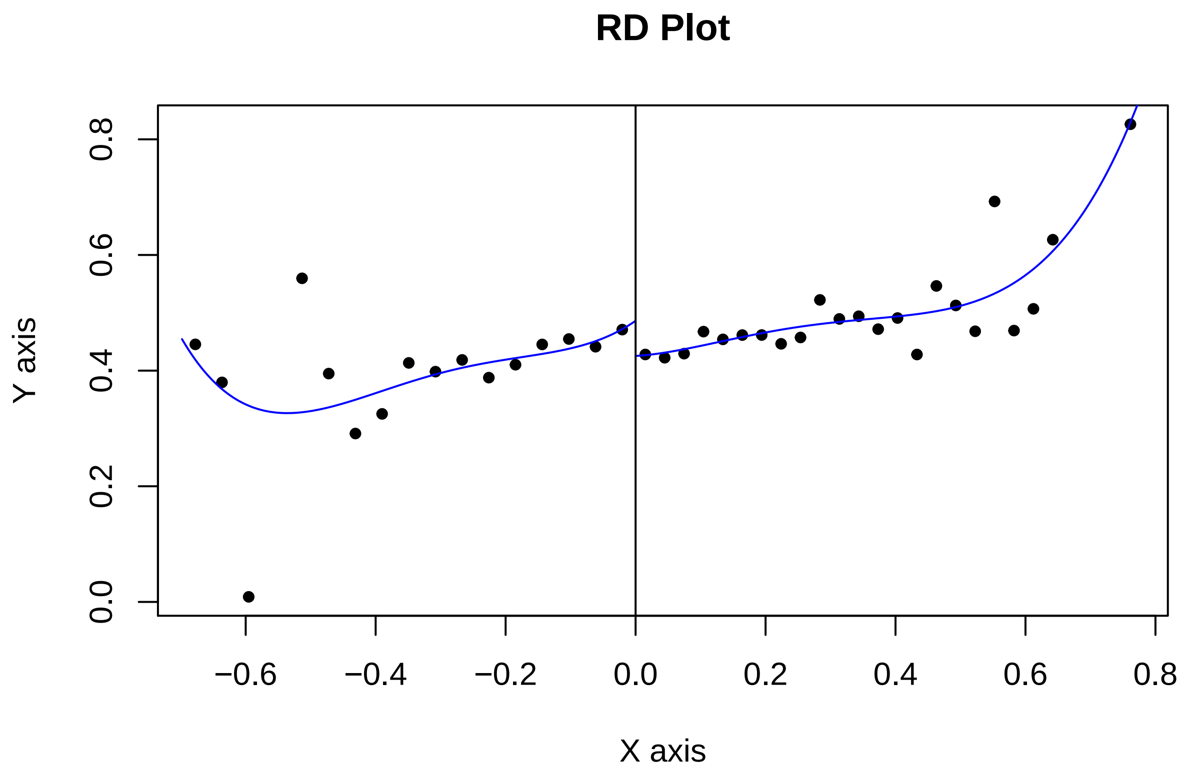
```
library(rdrobust)
rdplot(d_PMDB$Pct_Votes_2004, d_PMDB$Win_Margin, p=1)
```



9. Create a second regression discontinuity plot with fourth-order polynomial regression lines.

The regression lines now follow more closely the data, but there remains a negative jump at the threshold.

```
rdplot(d_PMDB$Pct_Votes_2004, d_PMDB$Win_Margin, p=4)
```



10. We will now implement four alternative specifications of the same regression discontinuity. For the first version of the analysis, implement a simple difference-in-means test comparing the average vote share received by the PMDB in 2004 within a bandwidth of +/-3% winning margin in 2000. Interpret these results and compare to the observational regression in Q5.

Incumbency now has a negative effect of approximately -5.5% points, and this is highly statistically significant.

```
d_PMDB %>% filter(Win_Margin<0.03 & Win_Margin>-0.03) %>%
  t.test(Pct_Votes_2004~Incumbent, data=.)
```

```
##
##  Welch Two Sample t-test
##
## data:  Pct_Votes_2004 by Incumbent
## t = 3.0003, df = 229.08, p-value = 0.002995
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.01859042 0.08972182
## sample estimates:
## mean in group 0 mean in group 1
##       0.4820535       0.4278974
```

11. For the second version, implement the full-data regression discontinuity analysis. Interpret this regression and compare it to your results in Q10.

The negative incumbency effect is now -4.8% points and similarly statistically significant.

```
d_PMDB %>% lm(Pct_Votes_2004~Incumbent + Win_Margin, data=.) %>%
  stargazer(header=F, keep.stat=c("n"), title="Q11")
```

Table 2: Q11

	Dependent variable:
	Pct_Votes_2004
Incumbent	-0.048^{***}
	(0.011)
Win_Margin	0.231^{***}
	(0.028)
Constant	0.467^{***}
	(0.006)
Observations	1,415
Note:	*p<0.1; **p<0.05; ***p<0.01

12. For the third version, implement the limited-bandwidth regression discontinuity analysis for a bandwidth of +/-3%. Interpret this regression and compare it to your results in Q10 and Q11.

The incumbency effect remains negative but is now only -2.9% points and is no longer statistically significant.

```
d_PMDB %>% filter(Win_Margin<0.03 & Win_Margin>-0.03) %>%
  lm(Pct_Votes_2004~Incumbent + Win_Margin, data=.) %>%
  stargazer(header=F, keep.stat=c("n"), title="Q12")
```

Table 3: Q12

	Dependent variable:
	Pct_Votes_2004
Incumbent	-0.029
	(0.037)
Win_Margin	-0.836
	(1.073)
Constant	0.468^{***}
	(0.022)
Observations	236
Note:	*p<0.1; **p<0.05; ***p<0.01

13. Fourth, let's implement the optimal-bandwidth linear regression discontinuity using the `rdrobust` command. What bandwidth was selected? How do the results compare to the other methodologies?

The optimal bandwidth chosen is +/-13.5% points, with an estimated effect of -5.6% points, which is highly statistically significant.

```r
rdrobust(d_PMDB$Pct_Votes_2004, d_PMDB$Win_Margin) %>% summary()
```

```
## Call: rdrobust
##
## Number of Obs.                 1415
## BW type                       mserd
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                  649          766
## Eff. Number of Obs.             415          417
## Order est. (p)                    1            1
## Order bias  (p)                   2            2
## BW est. (h)                   0.135        0.135
## BW bias (b)                   0.216        0.216
## rho (h/b)                     0.624        0.624
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    -0.056     0.020    -2.790     0.005    [-0.096 , -0.017]
##         Robust         -         -    -2.432     0.015    [-0.106 , -0.011]
## =============================================================================
```

14. Now let's try to adjust the functional form used to estimate the effect of the rnuning variable. Implement the optimal-bandwidth regression discontinuity but with a second-order polynomial (quadratic) trend. Also try a third-order polynomial (cubic) trend and assess the sensitivity of the results.

The bandwidth increases to +/-20% points, and the effect is now -5.9% or -6.2% points, which is reasonably stable.

```r
rdrobust(d_PMDB$Pct_Votes_2004, d_PMDB$Win_Margin, p=2) %>% summary()
```

```
## Call: rdrobust
##
## Number of Obs.                 1415
## BW type                       mserd
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                  649          766
## Eff. Number of Obs.             516          560
## Order est. (p)                    2            2
## Order bias  (p)                   3            3
## BW est. (h)                   0.202        0.202
## BW bias (b)                   0.288        0.288
## rho (h/b)                     0.702        0.702
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    -0.059     0.025    -2.370     0.018    [-0.107 , -0.010]
##         Robust         -         -    -2.122     0.034    [-0.115 , -0.005]
## =============================================================================
```

```r
rdrobust(d_PMDB$Pct_Votes_2004, d_PMDB$Win_Margin, p=3) %>% summary()
```

```
## Call: rdrobust
##
## Number of Obs.                 1415
## BW type                       mserd
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                 649          766
## Eff. Number of Obs.            520          568
## Order est. (p)                   3            3
## Order bias  (p)                  4            4
## BW est. (h)                  0.206        0.206
## BW bias (b)                  0.265        0.265
## rho (h/b)                    0.775        0.775
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    -0.062     0.034    -1.847     0.065    [-0.129 , 0.004]
##         Robust         -         -    -1.690     0.091    [-0.138 , 0.010]
## =============================================================================
```

15. The Mayor of a small municipality calls you for political advice. He wants to know what vote share his party (the PMDB) is likely to receive in the next election. He is very confident because at the last election he won easily with a winning margin of 30%. Based on the evidence you have recorded above from the regression discontinuities, how would you advise the Mayor about his likely performance in the next election?

Our estimated negative incumbency effect only applies to candidates who won by a small margin, which this Mayor did not, so we cannot say anything causal about the impact incumbency will have on his 2004 performance. However, the running variable suggests that those who did better in 2000 tend to do well in 2004 as well. Specifically, the regression in Q11 suugests he will receive $0.467 - 0.048 + 0.231 * 0.3 = 0.488 = 48.8\%$ points in 2014. But remember this is not a causal effect of incumbency, only a prediction.

16. Choose your preferred specification and implement the regression discontinuity for the other two parties: the PFL and the PSDB. How similar are your results to those in Titiunik (2011) for the +/-3% window?

```r
d_PFL <- d %>% filter(SIGLA_PARTIDO=="PFL")
rdrobust(d_PFL$Pct_Votes_2004, d_PFL$Win_Margin) %>% summary()
```

```
## Call: rdrobust
##
## Number of Obs.                  882
## BW type                       mserd
## Kernel                   Triangular
## VCE method                       NN
##
## Number of Obs.                 393          489
## Eff. Number of Obs.            195          206
## Order est. (p)                   1            1
## Order bias  (p)                  2            2
## BW est. (h)                  0.099        0.099
## BW bias (b)                  0.174        0.174
```

```
## rho (h/b)                   0.571      0.571
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    -0.069     0.027    -2.526     0.012    [-0.122 , -0.015]
##         Robust         -         -    -2.447     0.014    [-0.139 , -0.015]
## =============================================================================
```

```
d_PSDB <- d %>% filter(SIGLA_PARTIDO=="PSDB")
rdrobust(d_PSDB$Pct_Votes_2004, d_PSDB$Win_Margin) %>% summary()
```

```
## Call: rdrobust
##
## Number of Obs.                 928
## BW type                      mserd
## Kernel                  Triangular
## VCE method                      NN
##
## Number of Obs.                 371        557
## Eff. Number of Obs.            137        150
## Order est. (p)                   1          1
## Order bias  (p)                  2          2
## BW est. (h)                  0.076      0.076
## BW bias (b)                  0.142      0.142
## rho (h/b)                    0.535      0.535
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional     0.075     0.036     2.056     0.040     [0.003 , 0.146]
##         Robust         -         -     2.213     0.027     [0.010 , 0.170]
## =============================================================================
```