FLS 6415: Class 7 Homework

October 5, 2017

Remember to answer all the questions in R markdown and produce a PDF. Email your completed homework (R markdown file and PDF) to jonnyphillips@gmail.com by midnight the night before class. Remember to refer to the example code from this week and the last couple of weeks for coding guidance.

To analyze a regression discontinuity we will conduct the same analysis as Titiunik (2011). However, we will not use her pre-prepared dataset - we will start the analysis by constructing our own. As with all analysis, 90% of the work is in preparing the dataset.

1. Titiunik measures the effect of being an incumbent mayor for the 2000-2004 period on the vote share of the incumbent in 2004. We can download this in a clean simple format from cepesp.io. Choose "Eleicoes por Cargo" and we want prefeito data at the municipal level for *parties* in the 2000 elections first. Then 'Consultar' and 'Adicionar Colunas' to add the COD_MUN_IBGE variable. Then export this dataset to 'CSV'. Then make the same selection for 2004 and download this as a separate CSV. Finally, load the data in R. Details for the description of each variable can be found on cepesp.io (see adicionar colunas).

2. First, prepare the 2000 dataset:
a. Filter only for the first round (primeiro turno), calculate the percentage vote share for each party in each municipal contest and calculate which position the party came in the municipal election (their `rank`).
b. Next, we need to calculate the winning margin of each party as defined on page 9 of Titiunik (2011). One way to do this is to `arrange` the data by *COD_MUN_IBGE* and by the *rank* of vote share that you just created (so the winning party is at the top of each municipality in your data.frame). Then calculate new columns for the highest and second-highest vote share in each municipality. *Hint:* Use `nth` within `mutate` to calculate the second-highest vote share.
c. Then create another column for *win_margin* using `ifelse` or `case_when` to calculate the gap between each party and the second-placed party (if they win), or between each party and the first-placed party (if they did not win). Check your *win_margin* values have the correct sign.
d. Finally, let's add an 'Incumbent' variable to code whether the party actually won the election and an 'Electorate' variable to measure the total number of votes in each municipality (we'll use this later).

3. Now prepare the 2004 data. Filter for the first round (primeiro turno), and calculate our outcome measure: the vote share of each party in each municipal contest.

4. We will first conduct the analysis for the PMDB. Filter the 2000 dataset so it only contains the voting results data for the PMDB. Then each row in the 2000 dataset needs to also have a value for the outcome in 2004 - we need to merge the 2004 dataset into the 2000 data using the keys (merging variables) for municipality and party. How many rows are now in your dataset?

There are 1961 in the dataset.

5. If we did not know about regression discontinuity, the observational regression we might run is of the outcome (vote share in 2004) on treatment (becoming an incumbent in 2000). Run and interpret this regression. Give specific examples of reasons why the estimated causal effect may be biased.

The observational regression suggests that incumbents have a 0.027% points higher vote share in subsequent elections. Since treatment is not randomly assigned, many confounding variables could affect both incumbency in 2000 and vote share in 2004, most obviously candidate competence which will make some candidates do better in both elections.

Table 1: Q4, Observational Regression of the Effect of Incumbency on 2004 Vote Share for PMDB

	Dependent variable:
	vote_pct_2004
Incumbent	0.027***
	(0.009)
Constant	0.426***
	(0.007)
Observations	1,157
R^2	0.008
Adjusted R^2	0.007
Residual Std. Error	0.150 (df = 1155)
F Statistic	8.883*** (df = 1; 1155)
Note:	*p<0.1; **p<0.05; ***p<0.01

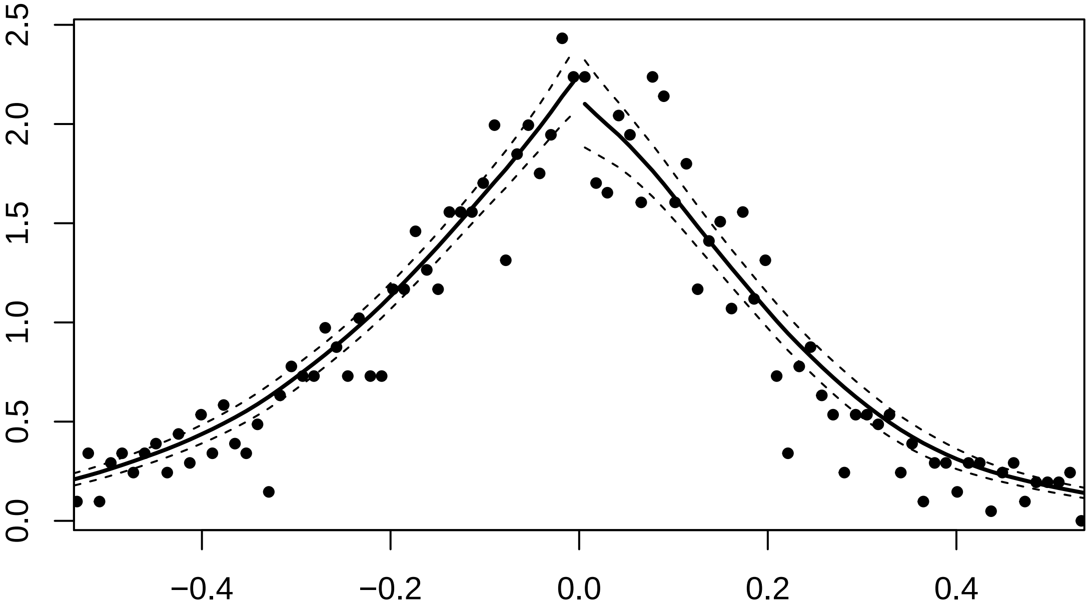
6. Before we implement the regression discontinuity methodology, we can check for balance. We don't have many variables in our dataset, but we can at least check if the size of the electorate (a good proxy for population) is the same for municipalities where the PMDB just lost compared to where they just won. First, assess balance for the 'Electorate' variable you created in Q2 within a 5% bandwidth either side of the cutoff of winning margin=0 for treatment and control. Then compare this difference with the balance on Incumbency in the full dataset. Interpret the results.

Incumbent	Electorate_3pct	Electorate_all
0	14878	19160
1	16688	10007

There is better balance in the dataset when it is narrowed to a smaller bandwidth either side of the threshold.

7. Before we implement the regression discontinuity methodology, we can check for sorting of units to either side of the threshold. Implement the McCrary density test and interpret both the graphical and statistical results. *Hint:* Use the `DCdensity` function with option verbose=TRUE.

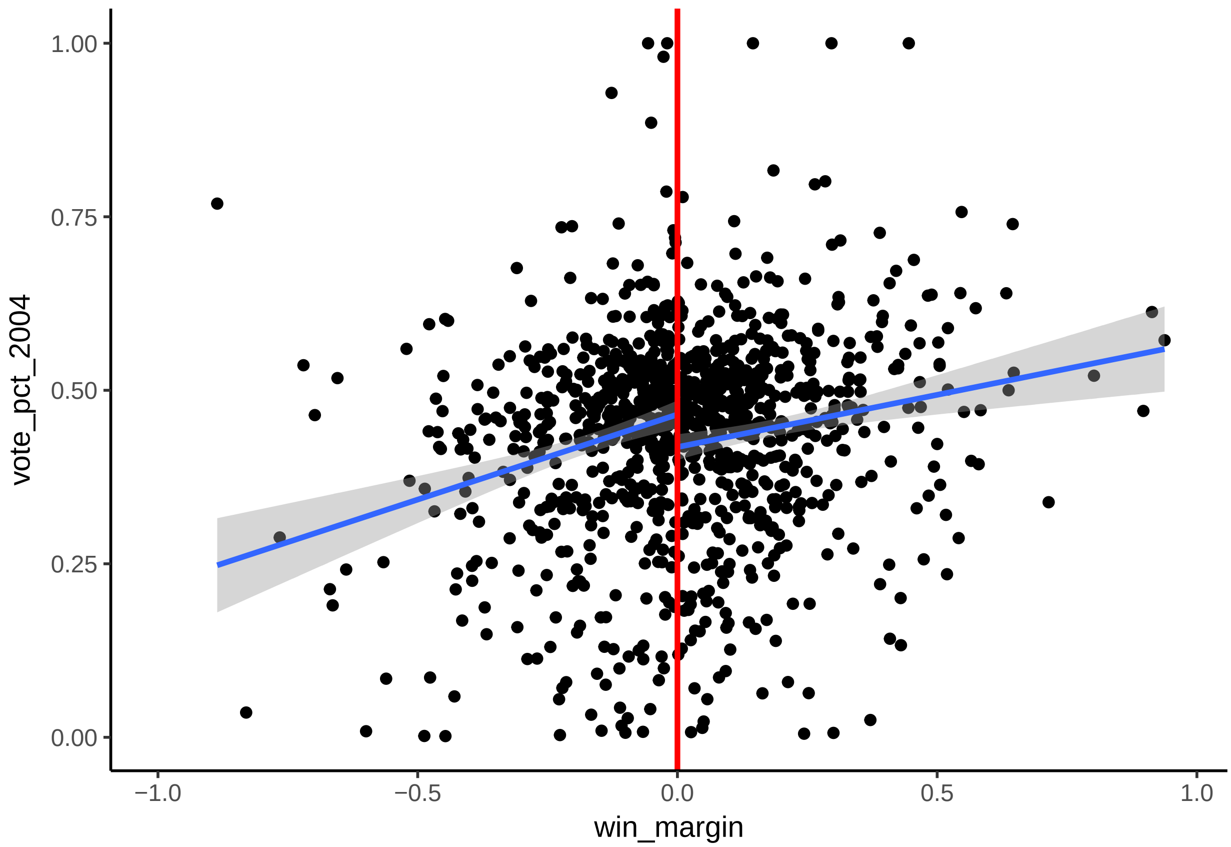
```
## Using calculated bin size:  0.012
## Using calculated bandwidth:  0.250
```

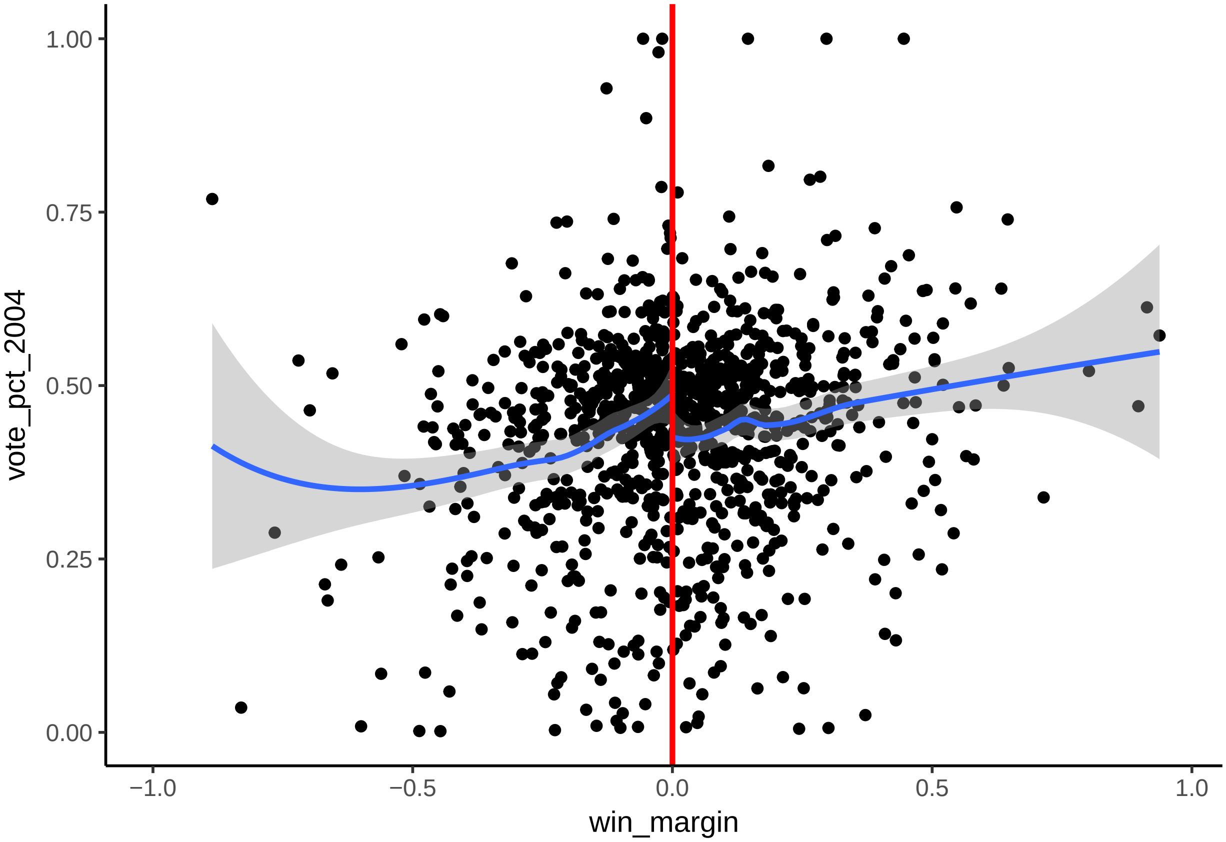


```
## Log difference in heights is  -0.056  with SE  0.101
##   this gives a z-stat of  -0.557
##   and a p value of  0.578
```

The chart shows a close intersection of the two lines either side of the threshold, suggesting there is no bunching of extra points on one side or the other. In addition, the p-value on the McCrary test is 0.578, so we cannot reject the null hypothesis of no sorting.

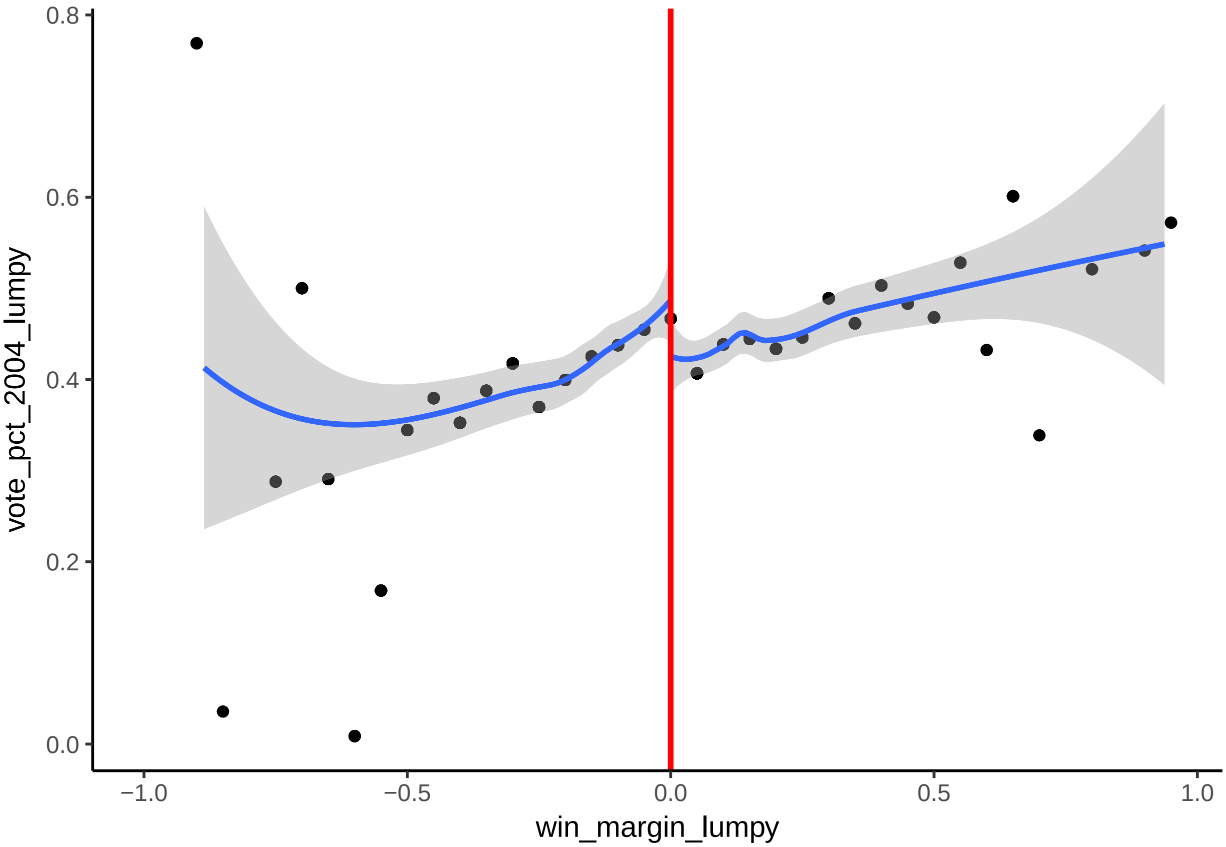
8. Before we run the analysis, let's construct a regression discontinuity plot to visually inspect the causal effect of incumbency at the threshold. Create two charts, one using linear (`lm`) and one using smooth local regression (`loess`) trend lines. Interpret what each chart tells you about the size and direction of the causal effect. *Hint:* Use two layers of geom_smooth to create separate 'regression' lines for each half of the data (above and below winning margin=0).


1.00
0.75
0.50
0.25
0.00
vote_pct_2004
−1.0
−0.5
0.0
0.5
1.0
win_margin



Examining the distance between the two regression lines at the threshold suggests that when the PMDB just wins an election it gets a lower vote share in the previous election compared to if it just loses. This is also true where the regression lines are non-linear.

9. This graph is 'messy' because there are too many points on it to see clearly the pattern. To create a clearer regression discontinuity chart, first group observations with similar values of the running variable (win margin) together - round each value to the nearest 0.05. Next, for each of these rounded win margins, calculate the average value of the outcome (vote share in 2004). Then plot this average outcome on the y-axis against the rounded values of the win margin on the x-axis. Finally, add the same two non-linear regression lines you used in Q8 - but these lines should use the full original data, not the rounded data you plotted as points). *Hint:* To round to the nearest 0.1 we would use `round(x,1)`; to round to the nearest 0.05 we can use `round(x/0.05)*0.05`



10. For the first version of the regression discontinuity analysis, implement a simple difference-in-means test, comparing the average vote share received by the PMDB in 2004 for all contests in 2000 where the PMDB just won or lost by less than 3%. I.e. keep only observations with a win margin of +/-3%. Interpret these results and compare to the observational regression in Q5.

The difference in means between treated and control units within +/-3% of the threshold is -4.9% points. This suggests, contrary to the observational regression in Q5, that incumbency produces a *disadvantage*.

11. For the second version of the regression discontinuity analysis, use *all* the data and run the linear regression of the outcome (vote share in 2004) on treatment (Incumbency) and the running variable (win margin). Interpret this regression and compare it to your results in Q10.

The second methodology using all the data provides a statistically significant estimate of the effect of incumbency on 2004 vote share of -0.046% points. That's very similar to the simple difference in means estimate.

12. The regression in Q11 assumes that there is a linear relationship between the winning margin in 2000 and the party's vote share in 2004. However, the chart in Q6 suggests the relationship is non-linear. To get an accurate measure of the causal effect we need to accurately match the shape of the relationship between the running variable and the toucome. We can do this by adding quadratic (ˆ2) and cubic (ˆ3) terms of win margin as controls in the regression. Implement this non-linear regression discontinuity and interpret the results.

The non-linear methodology estimates a statistically-significant effect of incumbency on 2004 vote share of –0.059% points, slightly larger than the other estimtes.

13. For the third version of the regression discontinuity analysis, we apply the regression

Table 3: Q11, Second, linear method: Regression Discontinuity Estimate of the Effect of Incumbency on 2004 Vote Share for PMDB

	Dependent variable:
	vote_pct_2004
win_margin	0.194***
	(0.031)
Incumbent	−0.046***
	(0.014)
Constant	0.457***
	(0.008)
Observations	1,026
R^2	0.042
Adjusted R^2	0.040
Residual Std. Error	0.148 (df = 1023)
F Statistic	22.236*** (df = 2; 1023)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4: Q12, Second, non-linear method: Regression Discontinuity Estimate of the Effect of Incumbency on 2004 Vote Share for PMDB

	Dependent variable:
	vote_pct_2004
win_margin	0.252***
	(0.049)
Incumbent	−0.059***
	(0.016)
I(win_marginˆ2)	−0.035
	(0.049)
I(win_marginˆ3)	−0.158
	(0.109)
Constant	0.465***
	(0.009)
Observations	1,026
R^2	0.044
Adjusted R^2	0.041
Residual Std. Error	0.148 (df = 1021)
F Statistic	11.865*** (df = 4; 1021)
Note:	*p<0.1; **p<0.05; ***p<0.01

method only to data within a small window (bandwidth) of the threshold. Subset the data to +/-5% of the threshold and apply the same regression as in Q11 to this smaller dataset. Interpret the results and compare to your results in questions 10,11 and 12.

Table 5: Q13, Third method: Regression Discontinuity Estimate of the Effect of Incumbency on 2004 Vote Share for PMDB

	Dependent variable:
	vote_pct_2004
win_margin	−0.192
	(0.669)
Incumbent	−0.039
	(0.038)
Constant	0.465***
	(0.022)
Observations	230
R^2	0.028
Adjusted R^2	0.019
Residual Std. Error	0.146 (df = 227)
F Statistic	3.237** (df = 2; 227)
Note:	*p<0.1; **p<0.05; ***p<0.01

The third methodology is not statistically significant but provides a comparable point estimate of a -0.039% points lower vote share in 2004 due to incumbency. The lack of significance here reflects the smaller amount of data available when we look at a smaller bandwidth.

14. An alternative way to implement the regression discontinuity analysis in Q13 is to use the package `RDestimate`. Report and interpret the results for each of the three automatically-selected bandwidths. *Hint:* You can access the output from the `RDestimate` object using `$est` for the coefficient on treatment and `$p` for the p-value.

Table 6: Q14: RDestimate Regression Discontinuity Estimate of the Effect of Incumbency on 2004 Vote Share for PMDB

	rdestimate_bw	rdestimate_est	rdestimate_p
LATE	0.142	-0.060	0.019
Half-BW	0.071	-0.041	0.243
Double-BW	0.283	-0.061	0.002

The RDestimate results suggest that the effect of incumbency is consistently between -4% and -6% points, with the 14% points and 28% points bandwidths proving statistically significant.

15. Use a for loop to implement the regression discontinuity method in Q11 (linear regression on all the data) for all three parties: the PMDB, PSDB and PFL. Summarise the results in a single table and compare the results between parties.

The incumbency disadvantage is only evident for the PMDB, and not for the PSDB or PFL.

16. 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

Table 7: Q15: Regression Discontinuity Estimates of the Effect of Incumbency on 2004 Vote Share, by Party

	Dependent variable:		
	vote_pct_2004		
	PMDB	PSDB	PFL
	(1)	(2)	(3)
win_margin	0.194***	0.157***	0.194***
	(0.031)	(0.034)	(0.033)
Incumbent	−0.046***	−0.007	−0.009
	(0.014)	(0.017)	(0.017)
Constant	0.457***	0.431***	0.436***
	(0.008)	(0.010)	(0.010)
Observations	1,026	744	669
R^2	0.042	0.053	0.082
Adjusted R^2	0.040	0.050	0.079
Residual Std. Error	0.148 (df = 1023)	0.153 (df = 741)	0.154 (df = 666)
F Statistic	22.236*** (df = 2; 1023)	20.543*** (df = 2; 741)	29.681*** (df = 2; 666)
Note:			*p<0.1; **p<0.05; ***p<0.01

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?

You could not give him much advice at all. Our estimates do not tell us anything about the incumbency (dis)advantage for Mayors with large winning margins, they are only informative for mayors who won or lose by very small margins. I.e. they are Local Average Treatment Effects (LATE).