POLS 7012 Final Exam 2022

Joe Ornstein

Due December 14, 2022 @ 5pm

Pereira & Fernandez-Vazquez (2022) investigate whether electing women to public office reduces corruption on Spanish municipal councils, taking advantage of a population-based cutoff that Spain used to enforce gender quotas. You can find a repository of the study's data and code here. For the final exam, we will replicate and extend some of their findings. Please write and submit a PDF report rendered from an R script (optionally: a Quarto document) that conducts the following analyses:[1]

1. Plot a histogram of the Spanish city population distribution in 2007. What is th mean, median, and standard deviation of the distribution? Why are the mean and median so different?
2. What about the *logarithm* of population? What does that distribution look like?
3. Are larger cities likely to have a larger share of women on the council? Plot the relationship. Estimate a linear model and interpret the slope of the relationship.
4. What percent of municipal councils had a corruption scandal during the 2007-2011 period? Compare the average share of women on councils facing corruption scandals versus those not facing corruption scandals.
5. Estimate the *difference-in-means* – the difference in women's share between the cities with corruption scandals and those without. Compute and interpret a 95% confidence interval and p-value associated with that difference-in-means.
6. Can we interpret that relationship you just estimated as causal? Why or why not? Draw a DAG illustrating some potential alternative explanations.
7. Estimate a *conditional* difference-in-means for cities that don't have any banks. Interpret the results.
8. Using the bandwidth reported in Table 2, just keep the cities slightly above and below the quota threshold in the 2007 election. Why is this subset of cities more appropriate for estimating the effect of women councilmembers on corruption than using the entire dataset?
9. What is the average women's share of elected municipal councilmembers for the cities just above the cutoff? Just below? Plot the difference between the groups. Is the difference statistically significant?

[1] Unlike the problem sets, this exam must be completed individually, without help from others.

10. It can often be useful to conduct a *placebo test*. Do the same thing you just did with the 5,000 population cutoff, but for a 4,000 person cutoff. There's no law implementing a quota at that point, so we shouldn't expect to find a significant difference. What do you find?
11. What about corruption? Is the rate of corruption scandals higher for cities just to the left or right of the threshold?
12. What about the *change* in corruption scandals between the 2003-2007 period and the 2007-2011 period? Is that higher for cities just barely above or below the threshold?
13. Re-estimate that difference-in-means using the local linear regression approach from the `rdrobust` package. What is the associated p-value and 95% confidence interval? Interpret the results.
14. **Optional Bonus Fun:** Create a chart that visualizes what's going on in the local linear regression analysis. Keep data points only within the bandwidth, with population on the x-axis, outcome on the y-axis, a vertical line at the threshold, and two linear models on either side. What's going on here?