

Homework Assignment 3

DPI 610

Assignment Details

This homework assignment is due *March 29* before midnight. Everyone must complete their own code and assignment, though you may discuss the homework with classmates.

To access the code and data for the assignment, go to RStudio.cloud. To turn in the assignment, upload two files to Canvas: (1) your code file (.Rmd), and (2) a PDF version (click Knit to create PDF).

Question 1: Ecological Inference

Legal cases on voting rights often require making estimates of the turnout rate of a minority group. In most states, however, it is not possible to know the exact turnout rate of any single population group due to the secret ballot. As a result, analysts often use methods of ecological inference to estimate turnout rates.

The data set `turnout` contains turnout data for North Carolina for the 2012, 2014 and 2016 general elections. Each row is a precinct-year combination. This question explores what we can learn about turnout among racial groups from the aggregate precinct data.

Q1(a): For one of the election years (2012, 2014 or 2016) in the data, plot the turnout rate (on the y-axis) against the share of black voters in the precinct. Include a regression line – from the linear regression of turnout share on black voter share – in the plot. Putting aside other factors, what is the general relationship, if any, between turnout rate and black population share in a precinct? Does this tell you anything about turnout among black voters? What pitfalls are there in drawing a conclusion based on the information in this plot?

#Insert answer here

Insert answer here.

Q1(b): What is the overall turnout rate (as a proportion of registered voters) in the state of North Carolina for each year in the data? What is the overall black population share for each year in the data?

#Hint: There are several ways to do this. One makes use of the function `weighted.mean()`. #Type `?weighted.mean()` to see the syntax.

#Insert answer here

Q1(c): For the 2012 election, suppose you want to learn what you can about turnout among black voters in precincts where black registrants comprised more than 75 percent of the registered voters in the precinct. One approach is to use the “method of bounds.” Use the function defined above, `bounds()`, to perform this calculation. Create a plot that illustrates how these bounds vary with the share of black voters in a precinct (so, include black population share on the x-axis and black voter turnout on the y-axis). How much variation is there in black turnout rate across precincts? Does the turnout rate appear to depend on the population share?

#Insert answer here

Q1(d): Estimate ecological regressions with the goal of determining an estimate for the black and non-black turnout rate among registrants in the state. What are your estimates for 2012, 2014 and 2016? In which year is there the smallest gap between estimates in turnout for black and non-black citizens?

#Insert answer here

*#Hint: To force R to omit the constant from a regression, include a "-1"
#(i.e., $\text{lm}(Y \sim -1 + X)$ would regress Y on X and omit the constant).*

Q1(e): The data.frame `turnout_by_race` includes actual, precinct-level turnout and registration data for NC voters in the 2012, 2014 and 2016 elections. That is, it includes information that allows us to know exactly the turnout rate among black and non-black registrants (rather than having to estimate it through ecological regression). Use this data to create a table that displays turnout as a percentage of registered voters in the state for the racial groups in the ecological regression above (i.e., black versus non-black). How much do your estimates from the ecological regression differ from the true turnout rates for black and non-black voters in North Carolina?

#Insert Answer Here

Q1(f): How would you explain differences between the ecological inference estimates and the results from the precinct-level turnout by race data? Is there evidence that key assumptions of ecological regression may not be met?

See the diagnostic plot, which displays residuals of the turnout rate by the black population share in each precinct, as one clue as to what assumptions from ecological regression may be violated. (Hint: Also see the Gelman et al. (2001) article discussed in class for a list of key assumptions that must be met for ecological regression to produce valid estimates).

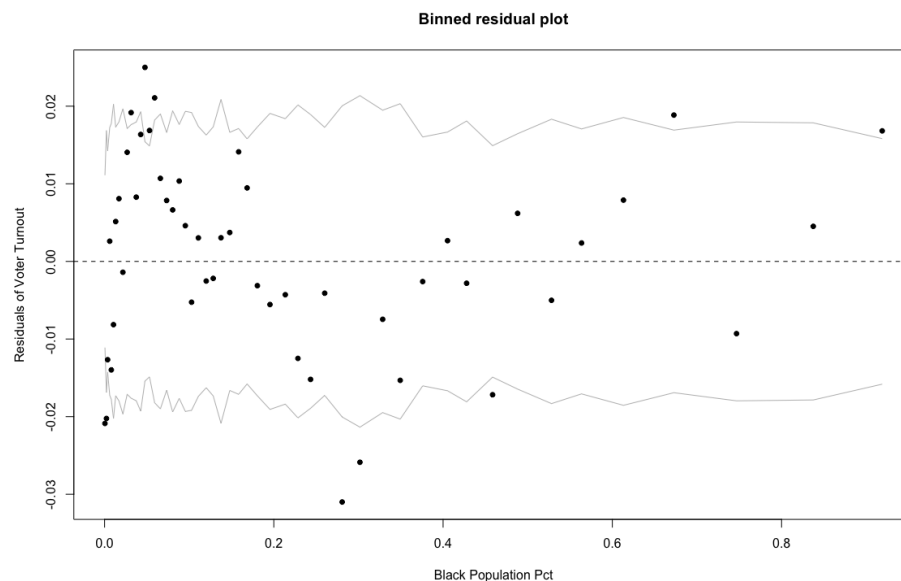


Figure 1: Residual Plot

Insert Answer Here

Question 2: Assessing Gerrymandering

A popular method for calculating seats-votes curves (allowing for estimates of how votes are translated into congressional seats in a state, for example) is uniform swing. This approach makes the assumption that swings in partisanship occur equally across all districts and, as a result, allow analysts to project the translation of votes into seats for counterfactual scenarios that have not occurred. This question explores some of those assumptions. To calculate uniform swing, use the function defined in the setup code chunk at the start of this document `uniform_swing()` as well as the data.frame `svdist`. (Note both of these were used in the in-class exercise on redistricting.)

Q2(a): Calculate uniform swing for North Carolina in 2016 and plot a seats-votes curve. What was the observed vote share and seat share (state-wide) for North Carolina *in 2018*? Plot this point on the graph as well. Is it in line with predictions from the 2016 seats-votes curve?

#Insert Answer Here

Q2(b): Now do the same for Ohio, California, Georgia, and Virginia. (Hint: You may want to write a function so that you can do this quickly/easily for multiple states).

#Insert Answer Here

Q2(c): For those states where the counterfactual predictions of the 2016 seats-votes curve appear different than the actual outcomes in the 2018 election, what do you think happened? What assumption or assumptionss underpinning uniform swing likely did not hold up?

Insert Answer Here.

Question 3 (Optional Bonus Question)

Create a plot that illustrates Seats-Votes curves for the US Congress (i.e., at the national level rather than for a particular state) for every election since 2002. What accounts for any observed differences in the seats-votes curve over time? What accounts for the similarities?

#Insert answer here

Insert answer here.