Final Project: POLI 281 (Fall 2020)

Optional Feedback: **November 11th, 11:59**
Paper & Analysis Due: **November 22nd, 11:59**

1 The Assignment

In the field of political science, there is substantial research on the infrastructure of elections – this research shows that everything from ballot design to the kind of primary election system a state adopts has an influence in determining election outcomes. Policy experts have most recently turned to long wait times as a key factor in affecting who does and does not vote in U.S. elections. In 2018, everything from broken scanners to insufficient polling locations brought about grueling waits for voters – causing some to stand in line for upwards of 6 hours. After this debacle, questions arose as to whether certain Americans – namely people of color – waited, on average, in longer lines than white Americans. From this question, we produce a null hypothesis and alternative hypotheses:

H_0 *: There is no difference in the voting wait times of white Americans and non-white Americans*

H_{a1}*: Black Americans had longer voting wait times that white Americans*

H_{a2}*: Hispanic Americans had longer voting wait times that white Americans*

H_{a3}*: Asian Americans had longer voting wait times that white Americans*

H_{a4}*: Americans of "other" races/ethnicities had longer voting wait times that white Americans*

For your final project, you will be using real data from the 2018 Congressional Cooperative Elections Study to investigate racial disparities in Election Day wait times for voting. This assignment will be composed of two parts: (1) a quantitative analysis, and (2) a written essay.

2 Quantitative Analysis

Using your knowledge of linear regression and coding skills in R, your task is to conduct a robust quantitative analysis of voter wait times in the 2018 U.S. midterm election. In this exercise, the main variable being investigated is cces$wait, which takes on six values = 1 (not at all); 2 (less than 10 minutes); 3 (10-30 minutes); 4 (31 minutes-1 hour); 5 (more than 1 hour); and 6 (don't know). This analysis will be broken into several parts. Each component is described in detail below. Data and a codebook will be provided, please use only these materials to complete the quantitative portion of your assignment:

2.1 Provide Summary Statistics

Using the variable of interest, calculate summary statistics for wait times. Report these in your essay.

2.2 How Long Did Americans Wait By Region?

For each of region in the data set (South, West, Midwest, and Northeast), calculate the percentage of people who reported waiting more than ten minutes. Plot these proportions using a bar plot in ggplot2, what trends do you notice? Place this plot in your essay and discuss the trends you observe.

Helpful tips:

- Using the margins command in prop.table might be helpful here!
- Remember, ggplot2 is only compatible with type data.frame, no matrices allowed!

2.3 How Long Did Americans Wait By Party?

Using the same approach as above, create a plot in ggplot2 that assesses wait times by party, what trends do you notice? Place this plot in your essay and discuss the trends you observe.

2.4 Preparing Your Data For Analysis

We're almost ready to embark on our main analysis, but first we'll have to clean up the data. Our main independent variable will include: (1) race, (2) faminc, (3) income_county, (4) density, and (5) vote2016. Please refer to the bulleted list below for all variable cleaning instructions. In addition to these instructions, please take whatever steps you think are additionally necessary to prepare these variables for your linear regression analysis:

- Race – what race/ethnicity did the respondent self-identify?
    - Simply the variable "race" such that the categories include "Black", "White", "Hispanic", "Asian" and "Other".
    - Is there anything else that should be done to this variable? Discuss in your essay.

- Family Income – how wealthy is the respondent?
    - Set all individuals who earn less than $30,000/year to "Low Income"; all individuals who earn between $30,000/year and $150,000/year to "Medium Income"; and all other individuals with reported incomes to "High Income."
    - Is there anything else that should be done to this variable? Discuss in your essay.

- Income County – how prosperous is the county where the respondent lives?
    - Check the distribution of this variable? Is there anything else that should be done to this variable? Discuss in your essay.

- Density – how rural or urban is the county where the respondent lives?
    - This variable is not currently in the data set, you have to create it using county_pop and land_area. This variable should reflect density in the *thousands.* So, for instance, if a county had a density of 2,000 the variable should have a numerical value of 2.

- Vote Choice – did the respondent vote for Trump or Clinton in 2016?
    - Set all respondents who did not vote for Clinton or Trump to NA
    - Is there anything else that should be done to this variable? Discuss in your essay.

2.5 Running The Regression & Reporting The Results

Run two separate models. In the first, investigate a bivariate relationship between race and wait time. In the second model, use all the variable described above to investigate the relationship between race and wait times. Use the results you see to make a nicely-formatted regression table, where there is one column for model 1, and a second column for model 2. Your table must include the coefficients the

number of observations, and the r-squared statistic. You can include the standard errors if you wish, but this is not required. Please ensure you bold or place a star next to those coefficients that are statistically significant at an alpha level of 0.05.

Additionally, provide the mathematical equation from your model using all available data from your regression output (alpha value, coefficients, etc.)

In the body of your essay, provide an interpretation of all coefficients in model 2. This can be done in the form of bullet points in you desire (however, please use full sentences). Please also interpret the r2 value. Once this is complete, answer the following questions below:

- Did we find support for our alternative hypotheses? Discuss and provide a statement concerning whether we reject or fail to reject our null hypothesis (for each alternative hypothesis).
- Did poorer respondents spend more time in line than those wealthier respondents?
- Did respondents in poorer counties spend more time in line than those from wealthier counties?
- Did respondents in more urban areas spend more time in line than those from more rural areas?
- Did Democrats spend more time in line than Republicans?
- Why are the coefficients in model 1 and model 2 different? Which model provided a more robust investigation of our question of interest? How do you know?

3 Written Essay

The subheadings above should provide a structure for your written essay. This essay should include a short introduction, stating the question of interest and hypotheses being investigated. In the body of the essay, discuss your descriptive data, including all plots and statistics produced. Then, go on to discuss your data preparation. What steps did you take? Why did you take these steps? Why is it important that these steps be taken before a linear regression model is fit?

In the final section, provide your regression table output, discuss your results, and address all questions outlined above. In addition, please provide a thoughtful and ***thorough*** discussion of the policy implications of these results. What are these results' implications on the democratic process? What recommendations might you make to experts? This section will a count for a ***third*** of your final grade on this essay! Finally, think back to your model 2 and provide some concluding thoughts. What improvements could have been made? What could you have done differently?

3.1 Finding (Academic) Sources

If you would like to include help academic sources for your paper (this is not a requirement, but could be useful in your policy discussion) there are a number of good search engines to help. The three that I suggest that you use are:

Google Scholar (https://scholar.google.com/)

1. JSTOR (http://www.jstor.org.libproxy.lib.unc.edu/action/showAdvancedSearch)
2. UNC Library (http://library.unc.edu/)

As with all essays, any and all outside information used should be included in a bibliography/references section. It does not matter which citation style you use, so long as you are consistent. Plagiarism will be penalized to the fullest extent possible through the UNC Honor System.

4 CODEBOOK

Variable	Details
Caseid	A unique number for each respondent
Race	Respondent's answer to the question "What racial or ethnic group best describes you?" 1 White 2 Black 3 Hispanic 4 Asian 5 Native American 6 Mixed 7 Other 8 Middle Eastern
Vote2016	Respondent's answer to the question, "In the election for U.S. President, who did you vote for?" 1 Donald Trump 2 Hillary Clinton 3 Someone else 4 I did not cast a vote for president 5 I don't recall
faminc	Respondent's self-reported income, coded as follows: 1 Less than $10,000 2 $10,000 - $19,999 3 $20,000 - $29,999 4 $30,000 - $39,000 5 $40,000 - $49,999 6 $50,000 - $59,999 7 $60,000 - $69,999 8 $70,000 - $79,999 9 $80,000 - $99,999 10 $100,000 - $119,999 11 $120,000 - $149,999 12 $150,000 - $199,999 13 $200,000 - $249,999 14 $250,000 - $349,999 15 $350,000 - $499,999 16 $500,000 or more 97 Prefer not to say

wait	Respondent's answer to the question, "Approximately, how long did you have to wait in line to vote?" 1 Not at all 2 Less than 10 minutes 3 10-30 minutes 4 31 minutes – 1 hour 5 More than 1 hour 6 Don't know
land_area	The size of the respondent's county, in square miles
county_name	The name of the respondent's county of residence
state county_pop	The name of the state where the respondent lives (or DC) The population of the respondent's home county
income_county	The average income in the respondent's county, in thousands of dollars.
region	What region of the country the respondent lives in: MW: Midwest NE: Northeast S: South W: West