POLS5769 - Intermediate Quantitative Social Research
Central European University
Fall 2020
Gabor Simonovits

Assignment 2

Due date: November 16th, 2020, 11.59pm

General Instructions

Please, submit your assignment in Moodle as a *.Rmd* file and your R script. Remember that **if your *.Rmd* file doesn't compile, you will receive no points for this assigment!** To get full credit for this assignment:

- Your answers should be correct. You will get partial points if your code is mostly correct or if it shows that you have "put in the work". Incomplete or incoherent code won't receive any points. But, if you don't get the right answer we will take at least some points off.
- Your code should always produce all the results.
- Provide a correct written answer.

Inroduction

Surveys are frequently used to measure political attitudes. In this assignment you will use a large American survey called the Cooperative Congressional Election study (CCES). The CCES is fielded each year to a large number of American citizens and measures political attitudes as well as demographic variables. The CCES is fielded online and it uses a complicated algorithm to render it representative of the underlying population. For the purpose of this assignment you can assume that this dataset is indeed representative of American adults.
You will work with a subset of the survey questions that are included in this survey which are listed below in Table 1. Please also consult the survey questionnaire (also available on Moodle) to see how each of these variables are coded and what the corresponding scales are.

Table 1: CCES survey data

Variable	*Description*
`caseid`	case identifier
`birthyr`	birth year
`gender`	biological sex
`educ`	highest educational attainment
`race`	race/ethnicity
`state`	state of residence
`pid3`	partisanship (Democrat, Republican, Independent, Other)
`CC19343a`	White people in the U.S. have certain advantages because of the color of their skin.
`CC19343b`	Racial problems in the U.S. are rare, isolated situations.
`CC19343c`	When women lose to men in a fair competition, they complain about discrimination
`CC19343d`	Feminists are making entirely reasonable demands of men.
`CC19322a`	Overturn Trump's order to use 6 billion to pay for the construction of a wall.
`CC19322b`	Withhold federal funds from police departments that do not report illegal immigrants
`CC19322c`	Reduce legal immigration by 50 percent.
`CC19322d`	Grant legal status to all illegal immigrants who have held jobs and paid taxes for at least 3 years and did not commit crimes.
`CC19322e`	Increase the number of border patrols on the U.S.-Mexican border
`CC19308a`	Job approval of Trump (1-4 scale)

1 Exercise 1: Visualization

1.1. (15 points) Create an index of sexism, racism and immigration preferences using the corresponding survey items. Describe their distribution. To do so, identify which variables can be used to measure sexism, racism, and immigration, respectively. Then, calculate the index by averaging these variables. Be careful here! The scales need to be the same before you average. This means that if for one variable the maximum denotes a given attitude and for another, the minimum denotes that attitude, you need to reverse one of the scales scale.

1.2. (10 pts.) How does the distribution of sexism vary across men and women? Illustrate the difference by plotting the distributions for each category.

1.3. (10 pts.) How does the distribution of racism vary across racial/ethnic groups? How would you interpret this? Illustrate the difference by plotting the distributions for each category.

1.4. (10 pts.) How does the extent of anti-immigration sentiment vary across states? List the top 5 and bottom 5 states and comment on their geographical pattern!

1.5. (10 pts.) Calculate the average of sexism by age and gender and compute the gender difference in sexism for 4 quartiles of the sample by age. Visualize and interpret your findings!

1.6. (5 pts.) Use the ddply function to calculate the correlation between sexism and racism by education. Comment!

Exercise 2: Regression

In this section, you will use regression analysis to assess two prevalent explanation for the success of Donald Trump: sexism, racism and immigration preferences. In your regression models you will use `CC19308a` (job approval of Trump) as the outcome variable.

2.1. (10 pts.) Recode `CC19308a` by placing people saying "not sure" in a middle category. Show the results by summarizing the proportions in a table and barplot.

2.2. (10 pts.) Run a simple regression predicting approval of Trump by anti immigration preferences. Interpret your finding!

2.3. (10 pts.) Run a regression predicting approval of Trump by racism. Repeat this exercise separately for Whites, Blacks and Hispanics. How do the regression coefficients compare? Explain!

2.4. (10 pts.) Run a regression predicting approval of Trump by sexism. Then, run the same regression including two indicators for Republicans and Democrats (based on `CC19308a`). How do your results change? Explain!

Bonus exercise:

3.1. (5 pts) Calculate the average number of Hispanics in each state as well as the average of your measure of anti-immigration attitude using ddplyr.

3.2. (15 pts) Visualize and explain your findings!