ANES Assignment

The American National Election Study is a scientifically rigorous landmark survey that has been conducted around every presidential election since 1948. For asignment, you will analyze the 2016 ANES. In 2016, the ANES was completed in both face-to-face format (~1,100 responses) and online (~3,000). For simplicity, we will focus on the face-to-face format only. Each respondent was interviewed twice: once a few weeks before the election, and once a few weeks after. You will have a raw dataset with both waves.

You will turn in two files:

1. An .R script that cleans the data and produces your analysis, according to the framework below. This script should produce all the code necessary to complete this assignment. This file should be **well-commented,** meaning that you’ve written notes in the file so that someone else can understand what your code is doing. Unlike previous assignments, I will not give you a template: you must make this script from scratch!
2. A Word document of your answers to each part of the assignment.

**Research question:** One of the central theories in political science is that people do what’s called *economic voting.* That is, they vote for the candidate they expect to most benefit their own economic situation. In particular, if a voter thinks that the economy is doing well, they will vote for the incumbent’s party (so, Clinton in 2016), and if a voter thinks the economy is not doing well, they will vote for an opposition party. Often, we proxy for this using perceptions of the general state of the economy. You may hear this in campaigns as “Are you better off now than you were 4 years ago?”

You will test a version of this theory using data from the ANES 2016 survey. There are several steps to this analysis, outlined below. You should write all code necessary to answer each question. It is completely acceptable (recommended, even) to use our in-class activities, assignments, etc. to help you complete the project. This assignment is equal parts coding and interpretation.

1. Consider the research question: whether a person’s belief about whether the economy is getting better or worse affects whether they voted for Clinton in 2016. Write 1-2 paragraphs that discuss why the nature of this relationship is not obvious. A good way to do this is to state one (or more) reasons why the relationship might be positive, but also one or more reasons why it might be negative (or zero). I’m not expecting you to know all the existing theories on this – I just want you to think logically about what might be going on in the real world.
2. Time for data! Write an R code to do the following:
   1. Clear the decks, programmatically import the raw ANES dataset, and load any necessary packages
   2. Make a new dataframe that only has the variables you need for this assignment.
   3. Make your data useable by creating and cleaning variables. To do this, look at the instructions in the final column of the codebook table at the end of this document.
3. Use the template below to create a well-formatted table of summary statistics showing, for each variable you created above, summary statistics including the variable’s minimum, maximum, mean, and median.

Table 1: Summary Statistics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable Name | Minimum | Maximum | Mean | Median |
| Cliton |  |  |  |  |
| Econ\_worse |  |  |  |  |
| Econ\_better |  |  |  |  |
| Female |  |  |  |  |
| Republican |  |  |  |  |
| College |  |  |  |  |
| Age |  |  |  |  |
| Religiosity |  |  |  |  |

1. Now let’s do some initial analysis.
   1. What is the correlation between beliefs about the economy and voting behavior? (i.e. the correlation coefficient?) Write a sentence explaining what it tells you.
   2. Run a bivariate regression to see how your economy variable predicts vote choice. Put this as Column 1 in your output table at the end of this document.
   3. Graph your result. Because a scatter plot doesn’t work well for this type of data, instead I want you to make a bar graph that shows the mean vote share for Clinton broken down by the econ\_worse variable. Make sure you format your graph nicely, including informative axis labeling.
2. Write a paragraph discussing the regression result. This discussion should include the direction of the overall results (positive or negative), and interpretation of the size of the effect. You should also discuss the predicted value of the DV for each possible value of the independent variable, and how those numbers compare to the graph you made in the previous question.

1. Look at the N of your regression, and compare it to the total number of observations in your dataset. How are they different? *Why* are they different, and is that a problem for your analysis?

1. Time to think about omitted variables and bias. For each of the following variables, discuss briefly whether *in theory* you think it’s a likely confounder. That it, talk about what would need to be true for it to be a confounder, and assess how likely it is that those conditions are met.
   1. Partisanship
   2. Age
   3. Gender
   4. Education
   5. Religiosity
2. You get distracted, and start wondering whether older people are more conservative. Make a graph that includes a scatterplot of age (x axis) and republican (y axis), and add the line of best fit. Make the graph pretty, too! What do you find?
3. Ok back to the matter at hand. Fill in the table below to show all the correlations between the economy variable, vote choice, and the 5 possible confounders.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Econ\_better | clinton | partisanship | age | gender | education | relig |
| Econ\_better |  |  |  |  |  |  |  |
| Clinton |  |  |  |  |  |  |  |
| Partisanship |  |  |  |  |  |  |  |
| Age |  |  |  |  |  |  |  |
| Gender |  |  |  |  |  |  |  |
| Education |  |  |  |  |  |  |  |
| Religiosity |  |  |  |  |  |  |  |

1. Based on this table, discuss which one of the potential confounders you think is most important to control for, and why.
2. Run a regression where you include the economy variable and just the confounder you think it most important. Make this Column 2 in the table below. Discuss how it changes your initial result, and how to interpret the new results (coefficients, intercept, etc).
3. Run a regression where you include the economy variable and all 5 potential confounders. Discuss the results, including 1) how they differ from the original bivariate regression and 2) how they differ from the regression that controls for only 1 confounder.
4. Up until now, we’ve been using the “econ\_worse” variable. Now rerun your main 3 specifications using the econ\_better variable, and fill in Table 2 below.
5. Discuss whether using econ\_better vs. econ\_worse changes the results. Why is this the case?
6. Write a 3 paragraph summary of your analysis. The first paragraph should summarize all of your results so far. The second paragraph should discuss how confident you are in the results: do you think this is a causal relationship? Why or why not? What are the limits of this approach? The third paragraph should discuss what you would do next if you were continuing this analysis. Are there other potential confounders you would include? Other ways you’d want to test this relationship?

Table 1: Results for econ\_worse

|  |  |  |  |
| --- | --- | --- | --- |
|  | Col 1: Bivariate Model | Col 2: One control | Col 3: All controls |
| Econ\_worse | Coefficient |  |  |
|  | (standard error) |  |  |
| Female |  |  |  |
|  |  |  |  |
| Republican |  |  |  |
|  |  |  |  |
| College degree |  |  |  |
|  |  |  |  |
| Age |  |  |  |
|  |  |  |  |
| Religiosity |  |  |  |
|  |  |  |  |
| Constant |  |  |  |
|  |  |  |  |
| N |  |  |  |

Table 2: Results for econ\_better

|  |  |  |  |
| --- | --- | --- | --- |
|  | Col 1: Bivariate Model | Col 2: One control | Col 3: All controls |
| Econ\_better |  |  |  |
|  |  |  |  |
| Female |  |  |  |
|  |  |  |  |
| Republican |  |  |  |
|  |  |  |  |
| College degree |  |  |  |
|  |  |  |  |
| Age |  |  |  |
|  |  |  |  |
| Religiosity |  |  |  |
|  |  |  |  |
| Constant |  |  |  |
|  |  |  |  |
| N |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| ANES variable | Description | Coding | Cleaning instructions |
| V162034a | Who did you vote for? Hillary Clinton, Donald Trump, Gary Johnson, Jill Stein, or someone else?  **Fun fact:** actual question randomized whether Clinton or Trump’s name came first. | 1 = Hillary Clinton,  2 = Donald Trump,  3 = Gary Johnson,  4 = Jill Stein,  5, 7, & 9 = other  Negative numbers indicate no data for that question. | Make a variable called “clinton” that’s equal to 1 if the person voted for Clinton, or 0 if they voted for someone else. All negative values of V162034a should be recoded as NA. |
| V161140 | Now thinking about the economy in the country as a whole, would you say that over the past year the nation’s economy has gotten better, stayed about the same, or gotten worse? | 1 = Gotten better  2 = Stayed about the same  3 = Gotten worse  -8 = don’t know  -9 = refused | Make 2 binary variables. The first, “econ\_better” should =1 if respondent thinks economy got better, and =0 if they thought it got worse or is about the same.  The second, econ\_worse, should = 1 if they think the economy got worse, and 0 if they thought it got better or stay the same. Both should be NA if V161140 is negative. |
| V161002 | Gender | 1=male, 2=female (ANES does not include other gender options) | Make a variable called “female” that’s 1 for female respondents and 0 for male respondents. |
| V161158x | Party ID | 1 = Strong Democrat  2 = Not very strong Dem  3 = Independent-Dem  4 = Independent 5 = Independent-Rep 6 = Not very strong Rep  7 = Strong Republican -8 = Don’t know -9 = Refused | Make a variable called Republican that is equal to V161158x, except any negative values are recoded as NA. |
| V161270 | What is the highest level of school you have completed or the highest degree you have received? | 1. Less than 1st grade  2. 1st, 2nd, 3rd or 4th grade  3. 5th or 6th grade  4. 7th or 8th grade  5. 9th grade  6. 10th grade  7. 11th grade  8. 12th grade no diploma  9. High school graduate- high school diploma or equivalent  10. Some college but no degree  11. Associate degree in college – occupational / vocational program  12. Associate degree in college – academic program  13. Bachelor’s degree  14. Master’s degree  15. Professional school degree (MD, DDS, DVM, LLB, JD)  16. Doctorate degree ( 90. Other specify given as: high school graduate 95. Other SPECIFY  -9 = refused | Make a variable called “college” that =1 if the person has a BA/BS degree or higher, and a 0 otherwise. This variable should be equal to NA if respondent said “don’t know” or “other”   NOTE: In real analysis I’d look at the people who said “other” and figure out how to code them. For simplicity I’m not going to ask you to do this here, just treat them as missing. |
|  | Age | Age in years.  - 9 = Refused to answer | Make a variable called “age” that is the respondent’s age in years. Set to NA if refused to answer. |
| V161244 | Lots of things come up that keep people from attending religious services even if they want to. Thinking about your life these days, do you ever attend religious services, apart from occasional weddings, baptisms or funerals? | 1 = yes  2 = no  -8 = Don’t know  -9 = Refused | Make a variable called “religious” that is 1 if they report going to religious services, and 0 if they do not. Code negative values as NA. |