Mini-project 3: Election 2016

S520

Upload your draft through the Assignments tab on Canvas by 11:59 pm, Sunday 16th April.

What factors helped to predict whether 2012 voters for Barack Obama voted for Donald Trump in 2016?

To answer this, we'll use the 2016 Cooperative Congressional Election Study, a very large survey of a nationally representative sample of 64,600 adults. The investigators asked questions to the sample both before and after the election, although not all the pre-election respondents replied to the post-election survey. The data is available in various formats at:

`http://cces.gov.harvard.edu/data`

Make sure you get the 2016 data and not some other year. The .RData file is the easiest to get into R.

Questions

- **Specific question:** In U.S. politics, three factors that are very important in predicting voting are gender, education, and race. Fit and display a model that gives the probability of an 2012 Obama voter switching to Trump in 2016 using gender, education, and race as predictors. Justify the form of your model (e.g. whether or not you included interactions.)
- **Open question:** Add ONE variable to your model that improves prediction. Fit and display this new model. Convince us that the variable does improve prediction (and isn't just adding noise), and explain what insight this variable gives you into what might have driven Obama voters to Trump. (Note: Only variables that were measured before the election can be used here.)

About the data

Some variables you'll need to use:

- **commonweight_post**: The weights for people who took the *post-election* survey.
- **tookpost**: Whether the respondent took the post-election survey. Limit your study to those for whom this is "Yes."
- **gender**: Male or Female.

- **educ**: Education (a factor with six levels.)
- **race**: A factor with eight levels.
- **CC16_326**: The respondent's vote in the 2012 Presidential election. Limit your study to those who voted for Barack Obama.
- **CC16_410a**: The respondent's vote in the 2016 Presidential election. "NA" could mean they didn't vote or that they didn't take the post-election survey. *Do not* limit your study to those who voted for Donald Trump; otherwise you won't be able to give probabilities.

The full documentation of the variables and question wording is on the CCES website.

Directions and tips

Work in groups of three or four. One person per group should submit a set of answers (making sure that everybody's name is on the submission.) The submission should include a PDF, your code, and the additional data you use besides the unemployment CSV file. You can use any statistical or graphical technique you want except find P-values.

The maximum length of your write-up, including graphs, is four pages. Your graphs should be of readable size. Your write-up should aim to convince someone with reasonable knowledge of statistics of your main points.

Prediction is not the main goal of this analysis, so you do not have to make the best possible predictions. Rather, the goal is to see how the data helps you understand what happened in the 2016 Presidential election.

Try not to start political fights with other members of your group, at least until your final project is submitted.

Grading

You'll be graded on four criteria:

- **Answer to the specific question.**
- **Answer to the open question.**
- **Presentation.** This include clarity and correctness of the graphs (i.e. we should understand what you're drawing) and clarity and correctness of the writing. Spelling and grammar will be a small but non-zero proportion of the grade.
- **Reproducibility.** Jake will attempt to reproduce all your graphs and numerical results from the materials you provide. If he can, you get all the points for this; if he can't, you get no points.

We will grade your initial submission and tell you how you can improve it, then you will have a week to resubmit it.