Applied Quantitative Analysis II

Spring 2020

Lab assignment II

Due before Feb 20th 6:30 pm

Please upload your answer on NYU Class.

Basic questions:

1. Download and open stata file GSS_lab_2.dta[1]. And finish the following tasks
   - Describe the dataset concisely, be sure to mention the number of observations, name of variables, and years covered.
   - Pick a variable to be your outcome variable; and at least 2 regressors of your choice, and you should justify your choice. A line of explanation/motivation is enough.
   - Keep the data if year >=2010. Provide descriptive statistics for these variables you chose. You can provide summary table, cross-tabs, etc. What can you see from these numbers?
   - Run an OLS regression on your variables, and interpret your model, including the coefficients and standard errors.
   - Draw a graph to illustrate the underlying relationship you observed in the regression results.[2]

2. Use the same dataset as in question 1. Finish the following tasks
   - Pick a binary variable indicating 1 for being extremely liberal or liberal, 0 for extremely conservative or conserative. Tabulate this variable, what proportion do you see?
   - Snapshot save the data, then collapse the to the mean of this binary variable, by sex and year

[1] If you do not use stata, you can find a function similar to read_stata/dta () in either Python or R, or export the data to a csv file

[2] Refer to the codes we posted on NYU class if you get stuck. It's generally easier if you have a story to tell and a graph is therefore made to help.

- Browse the data, what do you see? Drop the observations with missing value.
- Make a plot of yearly trend for the outcome with one line for female and another for male. Refine the graph, save it and describe what you observe.
- Restore the data. Run a logistic model with the outcome on sex, age, age square, education, and year dummy. Be sure to drop educ >= 98 and age >=98. Interpret the model.
- Run the same model as above but with OLS. Interpret the model.
- Summarize your findings. What can you say about the political polarizations in the US?

3. Briefly there are multiple ways of interpreting a logit\probit model. Review this article, feel free to ask questions, but not required.
   https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-how-do-i-interpret-odds-ratios-in-logistic-regression/