POL S 411/511
Assignment 4
Total Marks: 22

Using **Module 9 of the European Social Survey**, write a do-file that:

Part 1: Theory (9pt)

1. Contains the frontmatter for the do-file
   a. Contains a comment at the top of the file with your name, the course, date and instructor (each on separate lines)
   b. Sets the working directory and opens the dataset

2. Contains in comments (2pt):
   a. A strong, falsifiable research question (which identifies your dependent and independent variables)
   b. The null hypothesis
   c. An alternative hypothesis that **explains why you think it is theoretically-justifiable**
   d. Ensure that your dependent variable is either continuous or ordinal with at least 5 meaningful categories (NOT including "don't know", "other", etc).

3. Summarizes your two variables
   a. In comments, identify the mean, median, minimum, maximum, standard deviation and skewness (1pt)

4. Identify (in comments) at least 4 control variables that could account for alternative explanations of your dependent variable. Explain why you selected your control variables (i.e., the theoretical justification) and what you expect to find in a multiple regression between your IV, DV and the controls (5pt)
   a. At least one variable must be continuous (or an ordinal approximation of a continuous variable)
   b. At least one (other) variable must be ordinal.
   c. At least one variable must be categorical.
   d. You must have at least one interaction between two **binary** variables (don't forget to justify why you chose this)

Part 2: Regression (6pt)

5. Create a multiple regression investigating the relationship between your dependent variable, independent variable and controls. Don't forget to include the interaction. (2pt)

6. Interpret your regression, variable by variable. Make sure to include: (4pt)
   a. Whether each variable is statistically significant
   b. What the substantive interpretation of each coefficient is
   c. What the interpretation of the interaction term is

Part 3: Diagnostics (8pt)

7. Create a kernel density plot (`help kdensity`) for your residuals and superimpose a normal curve overtop. Describe what you find in comments. Save the figure in memory by giving it a name, and export it to an image file or PDF (2pt).

8. Create a regression vs fitted plot (`help rvfplot`) for your variable. Describe what you find in comments. Save the figure in memory by giving it a name, and export it to an image file or PDF (2pt)

9. Perform a Breusch-Pagan test for heteroskedasticity (`help estat hettest`). Describe what you find in comments (2pt).

10. Calculate the variance inflation factors for your independent variables (`help vif`). Describe what you find in comments (2pt).

Once you are done this assignment, submit your .log file and figures from questions 7-10 through eClass.

Make sure to submit a .log file containing the comments, code and output from above through eClass, **as well as the figures (.png or .jpeg) from questions 7-10**. Make sure to submit a .log file, not a .smcl file.

To create a log file:
1. Start your log through the Stata menu (File → Log → Begin), making sure to select .log and not .smcl
2. Run your do file.
3. End your log through the Stata menu (File → Log → Close).
4. Open your .log file to verify the file has saved correctly with all of your output.
5. Make sure you have also saved your figures (as .png or .jpeg files) to submit through eClass.