**SCHOOL OF POLITICS AND INTERNATIONAL STUDIES**

**PIED 5734M**

**Analysing Data in Political Science**

**MID-TERM ASSESSESSMENT**

**November 2023**

**Please refer to the Assessment Guide for direction on submitting your work.**

You should submit your assignment electronically on

You should upload your assignment onto Turnitin via Minerva **BEFORE 12:00 noon.**

Any assignment uploaded after 12:00 noon will be considered one day late and incur a five-mark penalty. Failure to conform to the word limit will also incur a penalty.

**1 x 1,000 word practical**

Your midterm project requires an application of the statistical theory and methods and STATA skills taught from weeks 1 through 5. Your goal here should be to demonstrate your understanding of how to: correctly and efficiently (1.) code a dataset and (2.) analyse data and interpret the results.

A few notes to keep in mind:

The 1000 word-count limit only applies to your interpretations and explanations in PART 2. It does not apply to your do-file, your log-file, or your tables.

You want to retain as many observations as possible for your analyses, so make sure you are not needlessly coding values as missing.

When creating your own tables for the STATA output, think about what information is crucial to present and include only what's necessary (i.e., don’t just copy and paste what’s presented in the STATA results window).

Finally, for Part 2, when we ask you to copy the do-file and the log-file into a Word document, please use the font that is used in STATA to ensure the results are readable. Part of this assessment is to determine your ability to present information in a clear and concise way – a messy log-file or do-file does not demonstrate this ability.

Create a new folder for this project. Copy the project datafile (“ESS9GB”) and the codebook (“ESS9\_appendix\_a7\_e03\_1”) into this folder.

Part 1 (50% of mark):

Create a do-file to do the following:

1. Create a text log-file to record all of the following commands and resultant output. The log-file should be closed at the end of your do-file.
2. Change the directory to the folder you have created for this project. Open the project datafile in STATA.
3. Retain only the following variables and drop everything else from the dataset: citizenship (**ctzcntr**), religious denomination (**rlgdnm**), gender (**gndr**), interest in politics (**polintr**), transparency in politics (**poltran**), satisfied with democracy (**stfdem**), left-right self-placement (**lrscale**), and vote choice in last national election (**prtvtcgb**).
4. Change the variable names to something that is more helpful in identifying the variables.
5. Where appropriate, recode missing values into useful values that can be used in your analyses. Be sure to label any recoded values if necessary.
6. Drop any observations that contain missing values.
7. Create one new dummy variable from any non-binary variable. In your do-file, provide a brief explanation of why changing the particular variable you choose into a dummy variable could be helpful in a future analysis. [Keep in mind that this is an exercise for assessment purposes. In non-assessment circumstances, reducing the information provided by a variable – e.g., changing a continuous variable to an ordinal variable – should generally be avoided unless you have a very good reason for doing so.]
8. Create a new variable label and new values labels for this new dummy variable.
9. Examine the means and standard deviations of all appropriate variables in the dataset. Examine the distribution of values for each categorical variable.
10. Save this revised dataset in the folder you created for this assessment.
11. Create a single histogram for any appropriate variable in your dataset. Save the histogram as a .png file.
12. Create a single boxplot for any appropriate variable in your dataset for each gender. Save the boxplot as a .png file.
13. Conduct a two-sample difference-in-means test using any appropriate variables in your dataset.
14. Conduct a Chi2 test on any two nominal variables in the dataset.
15. Calculate the appropriate correlations and their p-values for the following six sets of variables:
    1. satisfaction with democracy and left-right self-placement,
    2. satisfaction with democracy and transparency in politics,
    3. left-right self-placement and gender
    4. transparency in politics and interest in politics
    5. interest in politics and gender
    6. gender and the new dummy variable you created in STEP 7.

Part 2 (50% of mark):

Copy the do-file and the log-file into a Word document. Make sure to highlight all command lines in your log-file (and only in your log-file – DO NOT highlight the command lines in your do-file). Use **bolded** headers to make it clear which is your do-file and which is your log-file.

1. After the do-file and log-file, create another header called “**Analysis of Figures**”. Insert the histogram (from STEP 11) and the boxplot (from STEP 12) image files. Under each image, explain what the image is illustrating. Keep this brief but be thorough.
2. Create another header called “**Analysis of Tables**”. Create a set of tables that presents the useful/relevant information from STEPs 13 and 14. Under each table, interpret what the tables are telling us. In your interpretations, feel free to reference any relevant correlations if/where appropriate. As with the above, keep your interpretations brief but be thorough. [REMINDER: Do not simply copy and paste the tables from STATA. Create your own tables that contain only the important/relevant information you wish to discuss.]