

Homework 3

Submission Instructions (please read carefully)

Your homework submissions will look a little different this time. There are no longer boxes and blank columns where you will be prompted to write code or insert plots. Instead, you will submit a PDF that you will start from a blank Word document. This should serve as a transition from the format of the previous homework assignments to the format that your Final Project will take. As before, you should submit your HW3 using the **Gradescope** tab on Canvas.

- Be sure to number your questions appropriately and to indicate when you are answering Part I and Part II.
- All code should be copy and pasted on the last page (or pages) of your homework.
 - When copy-pasting code into Word, be sure to format the code using **monospaced font**. A font is monospaced if all characters have the same width. Examples of monospaced fonts include `Lucida Console`, `Consolas`, and `Courier New`.
- When asked to run a regression, please include a screenshot of the regression output in R. Do not copy-paste the code in the middle of your answers: leave all code to the end of the document.
- When asked to produce a plot, include your plot in the document. Do not copy-paste the code in the middle of your answers: leave all code to the end of the document.
- Please do not write your name.
- Include the names of students you have worked with, if any.

There is a template included on the assignments page, your assignment may look like it.

Part I: Regression Analysis in R

Because the dataset is in .dta (Stata) format, base R does not have a function that can read it in your computer. You will need to install a package. Some common options are *foreign*, *have*, and *rio*. See this link for instructions on how to use the *rio* package: <https://cran.r-project.org/web/packages/rio/vignettes/rio.html>

Read the dataset into R and store it as an object called `ukdata`. For example, if you choose the *rio* package, you can run the following code:

```
install.packages("rio")
library(rio)
ukdata <- import("insert filepath")
```

This dataset contains data for part of the analysis in the paper [Global Competition and Brexit](#) by Colantone and Stanig (2018). We recommend reading the abstract and skimming the paper for an insight on the type of analysis you will do in Part 1.

Before moving onto the questions, **be sure to read pages 1-4 of the codebook (up until “Replication DB Individual Wave8.dta”)**. You are encouraged to explore the data by making plots and tables before answering the questions.

As Colantone and Stanig did, you will use regression to analyze results of the [Brexit Referendum](#) at the regional level. The variable *leave_share* shows the percentage of voters in each region of the United Kingdom that voted for the UK to leave the European Union (see page 3 of the codebook).

With *leave_share* as your dependent variable:

1. Choose **one variable** in the dataset that, in your view, would be a good predictor of *leave_share*. (Do not choose a variable with the “_int” suffix. These variables are used to measure interaction effects, a topic not covered in this course.)
 - a. Justify your choice: why should we expect this variable to have a (positive or negative) correlation with *leave_share*? (It is okay if your variable turns out not to be correlated with *leave_share*. You will be graded based on the plausibility of your justification.)
 - b. State a hypothesis and a null hypothesis based on your choice of variable and your justification.
 - c. Produce a scatter plot with *leave_share* on the y axis and the variable of your choice on the x axis. Be sure to add a regression line. Do you see the correlation you expected?
 - d. Run a simple regression with *leave_share* as the dependent variable and the variable of your choice as the independent variable.
 - e. What is estimated effect of the variable of your choice on *leave_share*? In other words: according to the regression results, for an increase of 1 in the variable of your choice, *leave_share* will vary by how much?
 - f. Is this result statistically significant at the .05 level?

Part II: Final Project

Using R, begin analyzing the dataset of your choice for your Final Project. You are allowed to make changes to your research question and/or dataset from HW2 to HW3, and you will be allowed to make changes to your Final Project from HW3, but being consistent will help you do better at your Final Project. We encourage you to read the Final Project Description one more time before moving onto the questions in Part II.

1. Restate your research question. It can be the same from HW2, or a new research question. Feel free to revise the material in this course about what constitutes a good, empirical research question.
2. Formulate a theory to guide your investigation into your research question. Remember: a social science theory is “a reasoned and precise speculation about the answer to a research question.”¹
3. Formulate a hypothesis that (1) follows logically from your theory and (2) can be tested using the dataset of your choice.
4. What is your dependent variable? And your independent variable?
5. Can you think of a possible confounding variable in your dataset?
6. Create plots to visualize your projects main variables of interest. Be sure to add titles and axis labels:
 - a. Based on the type of your dependent variable, create a plot that illustrates the distribution of observations across all possible values of the variable.
 - b. Based on the type of your independent variable, create a plot that illustrates the distribution of observations across all possible values of the variable.
 - c. Based on the types of your dependent and independent variables, create a plot that illustrates the relationship between them.
7. Write a paragraph interpreting the plots you created for Question 6. How are your variables distributed? What is the relationship between your dependent and independent variables?

¹ Verba S, King G, Keohane RO. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press; 1994. p. 19.