Problem Set 4

POLS 8810

Due 27 March 2025

This problem set will require the same ANES data file (and codebook) available used last week. In case you did not save the link, it is available here.

In Problem Set 3, you looked at the conditional relationship between the income of respondents and their gender on their feelings towards the Republican Party. Imagine that after this, you were talking about it with the colleague who previously criticized your work to show them your new results. That colleague then expressed concern over whether you had an issue with heteroskedasticity. You decide that you should investigate this possibility and, if heteroskedasticity is present, reestimate your model from PS3 to account for this. (Note: for purposes of this problem set, even if you discover that you have no heteroskedasticity problem, you will still need to correct for it as if you did.)

Thus, *your primary goals for this assignment are 1) to test for the presence of heteroskedasticity, 2) to reestimate the linear regression model from Problem Set 3 correcting for heteroskedasticity, and 3) justify why your choice of approach for correcting for the presence of heteroskedasticity is the best among the possible options you could have chosen from.* In addition to clearly and correctly specifying ONE model that accounts for the presence of heteroskedasticity, you will need to present the results from your model in a clear way using table(s) and/or graph(s) as appropriate, and correctly interpret those results in a substantively meaningful way.

However, prior to conducting any analysis, it is a good idea to get a sense of what your data looks like. There are a variety of ways to do this using descriptive statistics and/or graphs to look at summaries of individual variables and relationships between variables. Doing this at the beginning of any analysis can save time and trouble later and help to ensure reliability of your estimates. *While you should always do this as a standard step in the process any time you are working with data, is not necessary to include any of this in your write up (unless you feel that something you found at this stage is relevant to the primary analysis/analyses).*

For the primary analysis, you will need to estimate an appropriately specified regression model that takes into account any relevant information learned from your preliminary examinations of the data you conducted. You will also need to ensure that all Gauss-Markov assumptions are met or, if not, why this is not a problem and/or how you are dealing with the problem(s).

Your write-up should contain any information the reader would need to replicate your data manipulations, recoding of variables, variable transformations, model estimation, etc. However, please pay close attention to how you write this up. Your goal is to emulate the way that the data, methods, and results sections of a publishable research paper are written. The handouts I have provide should help with this, but the best way to learn to write professionally is to read things that have been published and emulate their style. (*Note that repetition of the errors in how we discuss our variables and their coding highlighted as Problem 1 in the "Problem Set 1: Most Common Problems" handout (i.e. the use of variables coded in a way that is not suitable for a regression analysis, will negatively impact your grade on this assignment.*)

The primary grade will be based on an evaluation of the student's understanding of how to test for the presence of heteroskedasticity, how to appropriately estimate a regression model that accounts for heteroskedasticity, and how to present and interpret the results. However, quality/clarity of writing, quality of tables and figures, demonstrated understanding and description of the variables used, and other considerations matter as well. The write up should be approximately 5 pages, with 7 pages as an absolute maximum. **Assignments that fail to follow these direction or that cut and paste raw output from the software will be returned ungraded.**

Finally, while I *strongly* encourage you to work together to develop an understanding of data, coding, model estimation and interpretation, and the collaborative nature of research generally, the final write up must be your own. Moreover, it must be typed and written in the manner of the data, methods, and results sections of an article manuscript complete with any and all professional looking tables and graphs appropriate for the assignment.

Your Problem Set 4 write up is to be submitted *both* via iCollege under `/Assessments/Assignments/Problem Set 4` (in order to allow for automated plagiarism detection) and in physical hard copy following the instructions on the syllabus (for grading). Additionally, you should prepare a replication file with the code you use that should be submitted via email at the same time the problem set is submitted. If the replication code does not fully replicate all results presented in the write-up, you will be notified and your problem set will not be graded until this is corrected. If this is not resolved within one week of your receiving that notification, you will receive a grade of 0 for the problem set.