Assignment 3 (< 2,000 words)

Benjamin Jarvis & Richard Öhrvall

February 20, 2024

In this assignment, you will examine how immigrant attitudes relate to voting behavior in the contemporary UK. To do this, you will *individually* analyse data from the European Social Survey (see "Data" below). Beyond obtaining and loading the data, the assignment has two main parts. The "Analysis Tasks" section describes how you should prepare and analyze your data. Writing an R script to perform these tasks is necessary to complete the assignment, but is not sufficient to receive a passing grade. The bulk of your grade will be based on your written report, as described in the "Report" section. Read all of the prompts carefully before proceeding and consider the overall design of your report as you develop your analyses and responses.

Submission

The assignment is **due on Thursday, February 29, at 09.00**. You submit in Lisam, under assignments. The deadline is strict and non-negotiable, so please do not wait until the last minute to submit. If Lisam is not working for you, e-mail your files to Ben, `benjamin.jarvis@liu.se`, before the deadline.

You should submit 2 files in Lisam: one Quarto (or R Markdown) file with your code and text **and** one file that is the rendered pdf version of the same Quarto (or R Markdown) file. The pdf file should not include code or console output, just tables, graphs and your written answers. Remember that you can control the output from the code chunks by setting echo, messages, etc. to FALSE.

Your final, compiled report (not the R script) should not exceed 2,000 words, and should ideally be less than 1,500 words. Focus on clarity and concision.

Note that this is an individual assignment and LiU takes plagiarism very seriously. Plagiarism can include handing in an assignment that is, in whole or in part, identical to another person's work, not citing other peoples' texts and ideas properly, and more. Please consult LiU's guidelines at https://liu.se/en/article/plagiering-upphovsratt.

Data

You will work with voting data for the United Kingdom drawn from the 10th (2020) round of the European Social Survey (ESS). You can download data from the ESS website, but for this assignment you can find a Stata dataset with the data you may use in Lisam, under Course documents/Assignments/Assignment 3/ . You can download the dataset to your computer and use the `haven` package to read the data into R, see the code below (and the computer labs).

```
library(tidyverse)
library(haven)

ess2018_uk <- read_dta("path to where you stored the dataset/ESS10_gb.dta")
```

The ESS is a large-scale, general population survey conducted in many European countries every two years. There are many variables for each individual included in the survey. More information about the survey can be found on the ESS website https://www.europeansocialsurvey.org/. There you can find information about how data was collected in each country, response rates, variables, question wording, etc. The ESS has a complex sampling design, but you can ignore this design and sampling weights in your analysis.

Analysis

You will analyze how attitudes about immigration relate to party vote choice in the 2019 UK parliamentary election. Your main dependent variable will be a polychotomous (i.e., multiple category) outcome indicating which party respondents voted for. Your main explanatory variable is whether respondents think that immigration makes the country a better or worse place to live. You will use sex, age, and education as control variables. You should prepare your data and estimate models as follows:

1. Select UK respondents who are non-missing on the outcome and explanatory variables. Note that there are different variables relating to the respondents' view on immigration, so make sure to use the correct one, i.e., whether respondents thinks that immigration makes the country a worse or better place to live.
2. Change the party vote variable into a factor and recode it so that it has the values Conservative, Labour, Liberal Democrats, and others. Set Labour as the reference category.
3. Prepare a suitable table of descriptive statistics for the variables you use in your analysis.
4. Estimate two multinomial logistic regression models with party vote as the outcome variable:

- One where you include sex, age and educational attainment as the regressors. For education, dichotomize so that those with lower or higher tertiary education are coded 1 and all others are coded 0. You should decide (and justify) how to parameterize the age effect in your models.
- A second where you add respondents' views on immigration as an additional explanatory variable.

5. Produce one table containing estimates from both models. It should present the odds ratios with associated confidence intervals.
6. Generate appropriate model fit statistics and put them in the preceding table or in a separate table.
7. Produce a graph illustrating the predicted probabilities of different party choices as a function of whether immigration is perceived as being good for the country – this should be based on the second model and include separate panels for individuals with and without high education (hint: you could e.g. use `facet_wrap()` in `ggplot` or the `patchwork` package).

Report

You will be graded primarily based on your written report of your findings, **not** on the R code you wrote to obtain your findings. You must use complete sentences and paragraphs in your report. Your report should include and refer to figures and tables developed during your analysis. Figures and tables should be clearly and logically laid out, formatted, labelled, and footnoted. The quality and interpretability, not just the numerical accuracy, of your tables and figures will be a key component of your grade.

Your report should include the following sections:

- **Introduction**: briefly explain the study and your expectations regarding the main question. You don't have to do any extensive research to support your expectations, but please provide some rationale.
- **Data**: describe your data, e.g. when and how it was collected, nonresponse rate, etc., and how you defined your variables. This part should include your table of descriptive statistics for the variables in your models, and a brief discussion of notable patterns observed in this table.
- **Results**: discuss your key results from your models, including appropriate interpretations of odds ratios, p-values, model fit statistics, and predicted probabilities. Make reference to the figures/tables containing your regression results, fit statistics, and predicted probabilities.
- **Conclusions**: Summarise your main findings and compare with your expectations.