Analyzing Vote Choice Data

Assignment 2 - Deadline: May 10, 2023

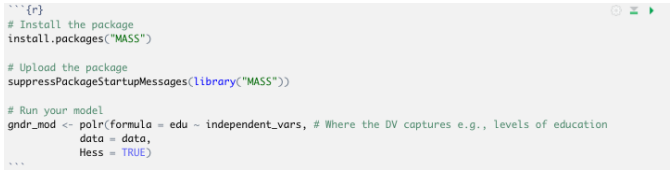
TA: Francesco Raffaelli (francesco.raffaelli@politics.ox.ac.uk)

1 R-PACKAGES

Before starting, let us review four `R-packages`, and their functions, that will be useful for this assignment - and in your research life!

- `MASS` - This package is pivotal if you want to run an ordered logistic regression, i.e, a model in which the dependent variable is made up by values representing categories that can be hierarchically ordered).

Figure 1. `MASS`



```r
# Install the package
install.packages("MASS")

# Upload the package
suppressPackageStartupMessages(library("MASS"))

# Run your model
gndr_mod <- polr(formula = edu ~ independent_vars, # Where the DV captures e.g., levels of education
                 data = data,
                 Hess = TRUE)
```

- `coefplot` - This package allows you to plot the coefficients of a model, picking the intervals of confidence you prefer, the colour, and the axis labels

Figure 2. `coefplot`

```r
# Install the package
install.packages("coefplot")

# Upload the package
suppressPackageStartupMessages(library("coefplot"))

# Draw the Plot
coefplot(model, title = "Title", xlab = "Label for the x-axis", ylab = "Label for the y-axis", color =
"colour", intercept = FALSE)

# Intercept allows you to add or remove the the coefficient of the model intercept
```

- `marginaleffects` - This package is new and it is frequently updated: watch out for changes, updates, and additional features[1]. Thanks to its function `plot cap`, it allows you to compute and plot adjusted predicted probabilities

- `patchwork` - This package allows you to patch different graphs together in a single imagine

[1] For example, for some versions the argument `effect` of the function `plot_cme` has been changed to `variables`

Figure 3. `marginaleffects, patchwork`

```
# In-class Example
```{r}
# Model
example_model <- glm(formula = extremist_party ~ distance + place_self,
                     data = data,
                     family = binomial(link = "logit"))

# Plot conditional adjusted predictions
example_preds <- example_model  %>%
   plot_cap(condition = c("distance", "place_self"))

# Plot conditional marginal effect
example_margeff <- example_model %>%
   plot_cme(effect = "place_self", condition = "distance")

# Put the two figures together
example_preds+example_margeff

```
```

2 The Paper

This assignment focuses on stacked data, marginal effects, interactions, and fixed effects with a focus on spatial models of voting. In order to do so, we will use as reference Catherine De Vries' *Sleeping Giant: Fact or Fairytale? How European Integration Affects National Elections* (2007). In this paper, the author argues that EU issue voting - that is, when attitudes towards European integration affect vote choice in national elections - is more likely to take place when *both* the level of partisan conflict over European integration *and* the salience of EU-related issues are high. For simplicity reasons, in this assignment we will focus only on one country: Italy

3 The Dataset

Dataset `ITANES1994-2001.dta` contains some variables from the Italian National Election Study (ITANES) for elections in 1994, 1996, and 2001. You need to upload it on `R` as you will use it for this assignment. In particular, the variables that we have selected for you are as follows:

- `year`: year in which the survey was conducted
- `prtystd-name`: standardized name of the party voted by the respondent that year
- `partystd`: standardized code of the party voted by the respondent that year
- `place-self`: respondent's right-left self-placement
- `place-votedprty`: respondent's right-left placement of the party voted that year
- `gndr`: respondent's gender (female = 2)
- `edu`: respondent's educational level (1 to 6, with 6 = university degree)
- `unemp-issue`: whether unemployment was the respondent's main issue of concern (yes = 1)

- `debate`: how often the respondent watches political debates in TV, as a proxy of respondent's interest for politics (1 to 3, with 3 = never)
- `relig`: how often the respondent goes to Church, as a proxy of respondent's religiosity (1 to 5, with 1 = never)

Upload the assignment on Canvas by the aforementioned deadline. Rename the pdf document obtained from the R-markdown as follows:

"AVCD-Assignment2-YOURLASTNAME"

4 Exercise 1

1. Let us do some data preparation:

   - Make sure that variable `gndr` is a dummy taking values 0/1. In this exploratory phase, get also a sense of the missing values in the dataset: if you want to measure extremism, you need to make sure that you know the left-right placement of parties and respondents.
   - Following De Vries (2007), create a dummy variable for parties and individuals with "extreme" righ-left (self-)placements (Hint: you should ask yourself how much an individual/party differ from the average individual/party)
   - Following De Vries (2007), build a measure of "issue salience" for unemployment (Hint: in order to measure issue salience, you should ask yourself what is the share of respondents that think that that issue is very relevant. However, do not forget that this dataset is longitudinal, meaning that there are more years)

2. Let us do some descriptives:

   - Plot the mean right-left self-placement of individuals over years. Has the public become more or less (left-) right-wing?
   - Can you actually say anything meaningful about this? Compare the mean and the median values for each year
   - Following De Vries (2007), measure the perceived distance betwen voters and their party of choice. Why would do you square it? And what is the main implication for results interpretation?

3. What pushes individuals to vote for parties with extreme right-left placements? Estimate the appropriate model and report it a nice, tidy table

   - Explain why you controlled for the covariates that you picked
   - What model did you use in the previous point? Why?
   - Plot the coefficients and comment their magnitude and level of statistical significance

5 Exercise 2

1. What is the effect of gender on interest for politics (measured as watching TV debate)? Does it change when individuals are very concerned by unemployment? If so, how?

   - Write down the formal models you have estimated, explain what the coefficients represent, and interpret the results

2. Represent graphically:

    - The predicted probabilities of gender on interest in politics when unemployment is the most important issue (or not)
    - The marginal effects of gender on interest in politics when unemployment is the most important issue (or not)
    - Put the two graphs in a singular imagine
    - What is the difference between marginal effects and predicted probabilities?
    - What happens to the marginal effects if the confidence intervals overlap?

3. Estimate a multinomial model: who do individuals with an extreme right-left self-placement vote for? Report it in a nice, tidy table

4. Conditional Logit:

    - What is a Conditional Logit Model, and does it differ from a Multinomial Logit Model?
    - What types of data should be used (or need to be used) for this type of model?
    - What is the best advantage of a Conditional Logit Model?