

Generalized Linear Model Lab

For today we'll be using the Danish country data from the European Social Survey Round 6 (2012–2013), which consists of a nationally representative survey of Danish adults asked questions on a variety of topics as part of a multi-country survey of European residents. The data, in Stata (.dta) format, are available to you on Blackboard in the **Content > ESS** folder, along with a complete codebook describing the questions and response options in the data, and a do-file to get you started with some simple variable recoding. *Note: Variables in the codebook are identified by an alphanumeric code, but these codes are not used as variable names in the Stata dataset. As a result, it is helpful to use **lookfor** along with particular keywords to find and identify relevant variables.*

The goal today is to look at and interpret a variety of Generalized Linear Models for ordered, multinomial, and count outcomes.

Ordered Outcomes

1. Open the .do file from Blackboard to contain the following analyses.
2. Load the ESS data file.
3. Take a look at Questions B29–B31. These questions address attitudes toward levels of immigration by different types of groups. Look at the current coding of these variables and settle on a reasonable coding that orders the alternatives. One possibility is included in the do file.
4. Generate a few covariates that will allow you to build a simple regression model. Some possibilities are given in the do file. Feel free to add others, change the recodings, or exclude some of the covariates.
5. Estimate an ordered logistic regression model for one of the outcomes using the `ologit` command.
6. Use `margins` to estimate predicted probabilities from the model at various levels of the covariates. For example, what do the predicted probabilities look like at varying levels of left-right political ideology?
7. Try estimating the probabilities for interesting, representative cases from the data using the `at()` option to `margins`.
8. Use `margins` with the `dydx` option to estimate marginal effects at various levels of the covariates.
9. Use `marginsplot` to plot predicted probabilities and marginal effects.
10. Compare these substantive interpretations to those from an ordered probit regression (`oprobit`) and an OLS regression.

Multinomial Outcomes

11. One prominent example of an unordered categorical variable in the ESS is party choice in the last election. Use `tab prtvtcdk` to examine the variable.
12. Use `clonevar partychoice = prtvtcdk` to create a new “party choice” variable and then use `recode partychoice 10/99=. to code all non-substantive answers as missing values.`
13. Use `mlogit` to estimate a multinomial logistic regression model predicting party choice as a function of your chosen covariates.
14. By default, `mlogit` used the lowest category as the baseline in the model, so all coefficients represent the effect of a given variable on the shift to a particular party relative to the baseline outcome category (i.e., party). Reestimate the model by specifying a different party using option `baseoutcome`. For example, set Venstre as the baseline using: `baseoutcome(7)`.
15. Define a new variable `sdchoice1` that represents voting for the Social Democrats as 1, Venstre as 0, and all other choices as missing.
16. Estimate a binary logistic model using `sdchoice1` as the outcome. How do the results compare to those from the `mlogit` estimates?
17. Define another new variable `sdchoice2` that represents voting for the Social Democrats as 1, and all other choices as 0.
18. Estimate a binary logistic model using `sdchoice2` as the outcome. How do the results compare to those from the `mlogit` and previous `logit` estimates?
19. Try expressing these results as predicted probabilities using `margins` and plotting the results.
20. To see that ordered logit and binary logit are identical, re-estimate your models of `sdchoice1` and `sdchoice2` using the `ologit` command instead of `logit`. Compare the results.

Count Outcomes

21. The ESS does not include a lot of count variables, but one — `njbbspv` — that is available indicates how many employees the respondent supervises in their job. Eliminate the missing values from the variable and code the “not applicable” category as 0 (i.e., that the individual supervises zero employees).
22. Use `poisson` to estimate a poisson model that regresses number of employees on any covariates.
23. Interpret the results using `margins`.