POLS6382 Quantitative Methods III: Maximum Likelihood Estimation

University of Houston
October 8, 2025

Homework Assignment 3

Instructions

- Answer the following questions and submit your answers and R code in one document. Your document should be prepared using LaTeX or R-Markdown.

- Submit your report using the submission link on Canvas.

- Homework Assignment 3 is due on Friday (October 17), by noon.

- We will review Homework Assignment 3 during class on October 22.

1   Analyzing Oridnal Data: Preliminaries

The topic of the day is abortion, the premier hot-button issue in contemporary American politics. Dozens of studies have examined differences in individual opinions about abortion. We will examine demographic and political influences on opinions about abortion in the United States. The data (which are a bit dated, but none the worse for wear) come from the 1988 American National Election Study (ANES), and consist of 1,377 respondents to a national probability sample survey conducted that year; the data contain a numeric ID variable to identify respondents. The ANES abortion question reads: "There has been some discussion about abortion during recent years. Which of the opinions on this page best agrees with your view?" The relevant categories are:

- 1= By law, abortion should never be permitted.

- 2= The law should permit abortion only in case of rape, incest or when the woman's life is in danger.

- 3= The law should permit abortion for reasons other than rape, incest, or danger to the woman's life, but only after the need for the abortion has been clearly established.

- 4= By law, a woman should always be able to obtain an abortion as a matter of personal choice.

Using data, `Exercise3Part1.dta`, you will examine the influence of several variables on opinions about abortion. Variable `conservatism` is the respondent's self-reported political ideology, ranging on a seven-point scale from "extremely liberal" (coded one) to "extremely conservative" (coded seven). Variable `black` is a dichotomous variable, coded as one if the respondent is African-American, and zero otherwise; similarly, `female` is a dichotomous variable for the respondent's gender (0=male, 1=female). Finally, `college` is a rough measure of educational attainment; it is coded as one if the respondent attended college (at all), and zero otherwise (i.e., the respondent had a high-school education or less).

- 1a. Present a descriptive analysis of the dependent variable and the variables you plan to include as the key independent variable(s).

- 1b. Estimate an ordered logit model to show how the chosen independent variable(s) affect public opinion regarding abortion. Specify your model however you choose, discuss your results, and speculate more on the possible substantive reasons behind your findings.

- 1c. Use predicted probabilities to assess the substantive importance of the variables in your model. Frame your answer in terms of practical significance: What impact does a change in one or more of the independent variables have on the level of support for abortion? (Feel free to use whatever approach(es) you think are appropriate here).

- 1d. Examine the "fit" of your model, both statistically and substantively. How good a job does your model do of explaining and predicting the ANES respondents' abortion attitudes?

- 1e. Assess the "parallel regression assumption" or the proportional odds assumption, i.e., the idea that the explanatory variables' effects are constant across the ordered categories. Use whatever approach(es) you think might be useful, and discuss your test result.

- 1f. Find and download the ANES 2024 Pre-election wave and replicate the empirical model you estimated for question #1b. Discuss your sample size, the descriptive statistics of the key variables of your choice, and report whether findings based on the ANES 1988 still hold when using the much more recent ANES data in 2024. What do you find, similar or different statistical results? Have fun!

2 Analyzing Multinomial Response Data

In this exercise, you'll use data file `Exercise3Part2.dta` to estimate and interpret a multinomial logit model of voting in the 1996 U.S. Presidential election. Specifically, you will examine the relative impact of ideology and political party affiliation on voting in the 1996 Presidential election. The data are from a 1997 NES post-election pilot study and consist of a sample of respondents contacted as a follow-up to their participation in the 1996 NES (N = 351). `prezvote` is a variable indicating who the respondent voted for in the 1996 election: 1 for Bill Clinton, 2 for Robert Dole, and 3 for H. Ross Perot. `respideo` is a variable indicating the respondent's political conservatism, ranging from 1 (very liberal) to 7 (very conservative). `democrat` and `GOP` are dichotomous variables indicating self-identification as Democrats or Republicans, respectively (with independents serving as an omitted baseline category).

- 2a. After examining summary statistics, estimate a multinomial logit model of voting in the 1996 Presidential race.

- 2b. Interpret these findings in statistical terms. Are the results in the expected direction? Discuss their statistical significance. Also, how do changes in ideology and political party affect (a) the probabilities of a respondent voting for each candidate and (b) the relative odds of voting for each of the candidates?

- 2c. Generate and examine the predicted probabilities of voting for each candidate across the range of values for each of your independent variables using tables or graphs of the probabilities. Interpret these results in substantive terms: What effects do *conservatism* and *partisanship*, have on voting here?

- 2d. Finally, briefly discuss in substantive terms what your statistical conclusions suggest for the 1996 presidential election results.

- 2e. Rerun the model, change to a different baseline category, and discuss your results. How (if at all) does the interpretation change? In your opinion, does re-specifying the baseline category tell you anything you didn't know before or make anything clearer? [Hint: you can use function `relevel()` to specify a reference category.]

- 2f. Briefly explain what the IIA assumption is. Test for whether the data/model conforms to the IIA assumption. Use whatever tests you can/ are aware of, and discuss your findings on this point in statistical and substantive terms. [Hint: Check out package `mlogit`. The function `hmftest()` can be used to test for the IIA assumption.]

3 Your Own Research: Developing Preliminary Analysis

In these two weeks, you will build on your proposed research ideas in HW1 and HW2 and focus more on developing your project data and preliminary analysis.

3a Update and revise your project codebook if needed. If not, skip 3a.

3b Produce and present a table of descriptive statistics of variables. You may choose a peer-reviewed journal in your field and format the table according to the journal's guidelines.

3c Visualize your dependent variable and independent variables, using proper visualization methods. Format your figures based on the guidelines of the journal of your choice.

3d Present preliminary statistical models for testing your hypotheses. Discuss what you find.