

POLS6382 Quantitative Methods III: Maximum Likelihood Estimation

University of Houston
November 12, 2025

Homework Assignment 5

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 5 is due by noon on Friday (November 21).
- We will review Homework Assignment 5 during class on November 26.

1 The Politics of Drug Approval

In this exercise, you'll use a replication data file "CarpenterFdaData" in the R package "simPH" to conduct event history analysis. Carpenter originally compiled the data for his 2002 study of the determinants of FDA drug approval (bureaucratic delay/or efficiency). Luke Keele (2010) used this data example to illustrate the estimation and diagnostic analysis when using a Cox model. For more detailed information about the dataset, read:

- Carpenter, Daniel P. 2002. "Groups, the Media, Agency Waiting Costs, and FDA Drug Approval." *American Journal of Political Science* 46:490-505.
- Keele, Luke. 2010. "Proportionally Difficult: Test for Nonproportional Hazards in Cox Models." *Political Analysis* 18:189-205.

A subsample of the original Carpenter dataset is included as an empirical example in the package `simPH`. Use the datafile "CarpenterFdaData" from package `simPH` to complete this exercise. The data file contains 408 drugs that were subject to FDA review. Variable `acttime` measures the duration of drug review (in months). Variable `censor` is the event-occurrence indicator.

- 1a. Load the data file from the R package "simPH". Estimate a Cox regression model for the time until approval, including the following covariates as the explanatory variables: `mandiz01`(a drug for disease mainly affect men), `femdz01`(drug for disease mainly affect women), `peddz01`(drug for disease mainly affect children), `deathr1`(death rate per 1,000), `lethal`(lethal condition), `hhosleng`(average hospitalization length), `hosp01` (population of hospitalization associated with the targeted disease), `stafcdcr` (FDA drug review staff, full-time employee), and `wpnoavg3` (Washington Post disease stories).
- 1b. Discuss the results from the Cox model that you estimated.

- 1c. Use a proper method to visualize the hazard ratio across the full range of the variable `stafcder`. Discuss how the hazard ratio changes as the value of `stafcder` varies across its full range.
- 1d. Test for the proportional hazard assumption. Do you find any evidence that this assumption is violated? If so, specify an alternative model and report what you find.
- 1e. Enjoying your time searching for the “secret” of FDA drug approval!

2 Review and Recap

In this “Review and Recap” section, you will answer a few essay questions about Maximum Likelihood Estimation. For each question, provide a list of key bullet points to reflect your understanding of the important methodological concepts covered in this class. Do not write more than one single-space page for each question. Aim at providing 4 to 5 concise bullet points for each question.

- 2a. Briefly explain the method of maximum likelihood estimation. How is this approach connected and/or different from the ordinary least squares method? Describe the key steps of building an MLE model (you may choose one particular model to illustrate the process).
- 2b. Suppose that you want to develop a proper statistical model to describe and explain a binary outcome. What are potential model choices? What particular characteristics of the dependent variable do you need to consider when specifying your statistical model? How would you compare and assess different model specifications? Once you choose the most suitable model, how would you substantively interpret the statistical results?
- 2c. Describe what an ordered logit or ordered probit model is. In what empirical context is an ordered-logit/probit model an appropriate specification? What are key assumptions of this model, and what if these assumptions are violated?
- 2d. Political scientists are often interested in studying phenomena that can be quantitatively measured as counts. Discuss why the ordinary least squares (OLS) approach is not a proper method for modeling count data. What are more appropriate model specifications? What are the key assumptions to be evaluated when you choose a particular count model?
- 2e. What is event history analysis (also known as duration/survival analysis)? What are the major types of models suitable for survival analysis? Briefly discuss the differences between semi-parametric and full-parametric event history models.