

Models for Categorical Outcomes Using Stata: Specification, Estimation, and Interpretation

Instructor: Scott Long, Indiana University
jslong@indiana.edu
www.indiana.edu/~jslsoc/teaching_CDAicpsr.htm

Teaching Assistant: Thomas VanHeuvelen
tvanheuv@indiana.edu

ICPSR Summer Program Workshop

June 16 - June 20, 2014

This workshop deals with the most fundamental regression models for binary, ordinal, nominal and count outcomes. While advances in software make it simple to estimate these models, post-estimation interpretation is challenging due to the nonlinearity of the models. The workshop begins by reviewing the general objectives for interpreting regression models and then considers why achieving these objectives is more difficult when models are nonlinear. Basic concepts and notation are introduced in a quick review of the linear regression model. In this context the ideas of maximum likelihood estimation and identification are introduced. These ideas are used to develop the binary logit and probit models. For these models numerous methods of interpretation are presented, including the use average marginal effects (AMEs) and marginal effects at the mean (MEMs). Wald and LR tests are illustrated along with information on how to analyze complex samples. The binary model is extended to explore “nonlinear nonlinear” models and methods for comparing groups. The multinomial logit model is developed by extending ideas from the binary model. The ordinal logit and probit models are then presented along with several less common models. Restrictions imposed by ordinality are explored and ways to choose between ordinal and nominal models are considered. Finally, models for count data, including Poisson regression, negative binomial regression, and zero modified models are presented. A major focus of the class is showing how to use Stata for sophisticated interpretation of nonlinear models. The course assumes familiarity with the linear regression model.

Tentative Schedule

8:30-12:30	Lecture and lab
12:30-1:30	Break
1:30-5:30	Lab and lecture

Texts

Long, J. Scott. 2014. Lecture and Lab Notes for Categorical Data Analysis. These notes contain copies of the overheads used in lectures and materials for the computing lab. The notes you receive may differ slightly from the overheads shown in class.

Computing

StataCorp is providing a temporary license to install Stata on your laptop. **Attendees are encouraged to bring a laptop.** If you do not have one, let the instructor know.

The workshop uses Stata for estimating and interpreting regression models for categorical outcomes. While Stata includes commands for estimating these models, we will use a set of ado files that Jeremy Freese and I have written to make it easier to interpret these models. If you are not familiar with Stata, don't worry. The lab guide will show you everything you need.

Recommended Texts

The lecture and lab notes have the critical material for your work during the workshop. After the workshop, the following books are recommended.

1. Long, J. Scott and Jeremy Freese. 2005. Regression Models for Categorical Dependent Variables Using Stata. 2nd Edition. College Station, TX: Stata Press. (LF2) Do not buy before the first day of class; if you have a copy, bring it!
2. Long, J. Scott. 1997. Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks, CA: Sage. (RM4) This book is more technical.
3. Long, J. Scott. 2008. The Workflow of Data Analysis Using Stata. College Station, TX: Stata Press. This book deals with general issues of efficient and replicable data analysis using any statistical method.

Workshop Outline

The content varies depending on the background of class members.

1. Overview: Types of variables and why the standard regression model may be inappropriate. Readings: LF2-Chapters 1-3; RM4-Chapter 1. Day 1.
2. Continuous Outcomes: The basic assumptions of the regression model; the idea of identification, and an introduction to maximum likelihood estimation. Readings: LF2-Chapter 3; RM4-Chapter 2. Day 1.
3. Binary Outcomes: The linear probability model, logit and probit. Readings: LF2-Chapter 4; RM4-Chapter 3. Days 1 and 2.
4. Testing and Complex Samples: Common tests for models estimated by ML are reviewed. Methods of estimation and testing for complex sampling designs are discussed. Readings: LF2-Chapters 3, 4; RM4-Chapter 4. Days 2 and 3.
5. Internal and External Assessment of Fit: Examining individual cases to determine how well they fit the model. Scalar measures of fit include R^2 -type measures and AIC/BIC measures. Readings: LF2-Chapters 3; RM4-Chapter 4. Day 3.
6. Complications to Models for Categorical Outcomes: Nonlinearities on the RHS of the model. Group comparisons with binary outcomes. Day 3.
7. Nominal Outcomes: Extensions of the logit and probit model for nominal outcomes. Readings: LF2-Chapter 6 and parts of Chapter 7; RM4-Chapter 6. Day 3 and 4.
7. Ordinal Outcomes: Extensions of the logit and probit model for ordinal outcomes. Readings: LF2-Chapter 5; RM4-Chapter 5. Day 4.
9. Count Outcomes: Poisson regression and negative binomial regression. Readings: LF2-Chapter 8; RM4-Chapter 8. Day 5.