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

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

Teaching Assistant: J.D. Wolfe, Indiana University
jdwolfe@indiana.edu
mypage.iu.edu/~jdwolfe/

ICPSR Summer Program Workshop
Institute for the Study of Societal Issues (ISSI)
University of California-Berkeley

July 11 - July 15, 2011

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 difficult due to the nonlinearity of the models. The workshop begins by considering the general objectives for interpreting results from any regression model 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, during which 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 examined. Standard statistical tests and the use of complex samples are considered. Methods of estimation and interpretation for binary outcomes are extended to ordinal outcomes using the ordinal logit and probit models. The multinomial logit model for nominal outcomes is then discussed. Finally, a series of 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
12:30-1:30 Break
1:30-5:30 Computer lab and lecture

Texts

Long, J. Scott. 2011. Lecture and Lab Notes for Categorical Data Analysis. These notes contain copies of the overheads used in lectures and materials for the computing lab. Be sure to bring these notes to the lectures and labs. The notes you receive may differ slightly from the overheads shown in class.

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.


**Computing**
The workshop focuses on using 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.

**Workshop Outline**

The content varies depending on the background of class members. You will get the most out of the lectures if you try to read the material before the class in which it is discussed.

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 R²-type measures and AIC/BIC measures. *Readings*: LF2-Chapters 3; RM4-Chapter 4. Day 3.


