

ICPSR Workshop on Categorical Data Analysis Using Stata:  
Models for Binary, Ordinal, Nominal, and Count Outcomes

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

*Teaching Assistant:* Shawna Rohrman, Indiana University  
rohrman@indiana.edu

ICPSR Summer Program Workshop

June 15 - June 19, 2009

This workshop deals with the basic regression models for categorical dependent variables. While advances in software have made it simple to estimate these models, post-estimation interpretation of these models is difficult due to the nonlinearities of the models. The workshop begins by considering the general objectives for interpreting the results of any regression type model and then considers why achieving these objectives is more difficult when models are nonlinear. Basic concepts and notation are introduced by reviewing the linear regression model. Within this familiar context, the method of maximum likelihood estimation is presented. These ideas are used to develop the logit and probit models for binary outcomes. A variety of practical methods for interpreting nonlinear models are presented. The models and methods of 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 component of the course is showing how Stata can be used to estimate and interpret the models using special commands for post-estimation interpretation. The use of Excel to interpret the results from other statistical packages is illustrated in lab. 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. 2009. *Lecture and Lab Notes for Categorical Data Analysis*. These notes contain copies of the overheads for the lectures and materials used in 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. They will be available for purchase at ICSPR.

Long, J. Scott and Jeremy Freese. 2005. *Regression Models for Categorical Dependent Variables Using Stata*. 2nd Edition. College Station, TX: Stata Press. (LF2)

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage. Hereafter RM4. (RM4)

Long, J. Scott. 2008. *The Workflow of Data Analysis Using Using Stata*. College Station, TX: Stata Press. (LF2)

### *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.

### *Workshop Outline*

The content of the workshop will vary 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. *Readings*: LF2-Chapter 3; RM4-Chapter 2. Day 1.
3. *Maximum Likelihood Estimation*. *Readings*: LF2-Chapter 3; RM4-Chapter 2. Day 1.
4. *Binary Outcomes*: The linear probability model, logit and probit. *Readings*: LF2-Chapter 4; RM4-Chapter 3. Days 1 and 2.
5. *Hypothesis Testing and Goodness of Fit*: Common tests for all of the models estimated by ML; various measures of goodness of fit. *Readings*: LF2-Chapters 3, 4; RM4-Chapter 4. Day 3.
6. *Ordinal Outcomes*: Extensions of the logit and probit model for ordinal outcomes. *Readings*: LF2-Chapter 5; RM4-Chapter 5. 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 4.
8. *Count Outcomes*: Poisson regression and negative binomial regression. *Readings*: LF2-Chapter 8; RM4-Chapter 8. Day 5.