

# Regression Models for Categorical Outcomes: Specification, Estimation, and Interpretation

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

## ICPSR Summer Program Workshop

May 23 – May 26, 2017

This workshop deals with the most useful regression models for categorical outcomes. While estimating these models is simple, interpretation is challenging due to the nonlinearity of the models. Innovative methods of interpretation using predictions and marginal effects are considered. The class begins with the binary model which is extended to show how to interpret models with interactions and nonlinearities introduced using polynomials (e.g., age-squared). The multinomial logit model is developed by extending ideas from the binary model to nominal outcomes. The ordinal logit and probit models are then presented. Limitations of ordinal models are explored and methods for choosing between ordinal and nominal models are considered. Finally, models for count outcomes, including Poisson regression, negative binomial regression, and zero inflated models are presented. The models presented in this class are essential prerequisites for modelling categorical outcomes using panel analysis or hierarchical models.

The workshop assumes that participants have a background in linear regression. Data analysis for this class will be carried out using Stata, but prior experience with Stata is not required.

### Tentative Schedule

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

### Texts

Long, J. Scott. 2017. Lecture and Lab Notes for Categorical Data Analysis. These notes contain the overheads from lectures and materials for the computing lab. The notes you receive may differ slightly from the overheads shown in class.

Long, J. Scott and Freese, Jeremy. 2014. *Regression Models for Categorical Dependent Variables Using Stata*. Third Edition. Stata Press (**LF3**). During the class, the lab guide and lecture notes might provide all of the information that you need.

## Computing

**Attendees are expected to have a laptop.** Stata is used to fit and interpret models. If you do not have a copy of Stata, StataCorp has provided a temporary license to install Stata on your laptop. We also use commands that Jeremy Freese and I wrote to make it easier to interpret these models. To install these, in Stata run `—search spost13_ado—` and follow the instructions. If you are new to Stata I encourage you to watch some of the videos at <https://www.youtube.com/user/statacorp>.

## Recommended Texts

1. Long, J. Scott. 1997. Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks, CA: Sage. **(RM4)** This book provides a more technical discussion.
2. 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

1. Overview: Types of variables and why the standard regression model may be inappropriate. Readings: LF3 Chapters 1-4. Day 1.
2. Continuous Outcomes: A quick review of the basic assumptions of the regression model; the idea of identification, and an introduction to ML estimation. Readings: LF3 Chapters 1-4. Day 1.
3. Binary Outcomes: The linear probability model, logit and probit. Readings: LF3 Chapters 5, 6. 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 briefly discussed. Readings: LF3 Chapters 3, 4; RM4 Chapter 4. Day 2.
5. Internal and External Assessment of Fit: Methods for examining individual cases to determine how well they fit the model are briefly discussed. Methods for using AIC and BIC to select models are discussed.  $R^2$ -type measures are briefly discussed. Readings: LF3 Chapters 3, 5. Day 3.
6. Complications to Models for Categorical Outcomes: Nonlinearities on the RHS of the model. Group comparisons with binary outcomes are considered if time permits. Day 3.
7. Nominal Outcomes: Extensions of the logit and probit model for nominal outcomes. Readings: LF3 Chapter 8; Day 3.
7. Ordinal Outcomes: Extensions of the logit and probit model for ordinal outcomes. Readings: LF3 Chapter 7. Day 3 and 4.
9. Count Outcomes: Poisson regression and negative binomial regression. Readings: LF3 Chapter 9. Day 4.