Categorical Analysis

Thomas Pullum
University of Texas at Austin

Instructor:
Thomas Pullum, University of Texas at Austin
tom.pullum@mail.utexas.edu

Teaching Assistants:
Ryan Masters, University of Texas at Austin
rmasters@prc.utexas.edu

Hiro Yoshioka, University of Texas at Austin
hiro12@prc.utexas.edu

ICPSR Summer Program in Quantitative Methods of Social Research
Second Session: July 21 to August 15, 2008
Lecture, 3:00 to 5:00 p.m.; Lab and Office hours to be announced

This course is intended to give you an understanding of the theoretical basis, practical applications, and interpretation of the main methods for analyzing counts and categorical outcomes and predictors in sociology and related fields. It will focus on logit regression but will also include log linear models and poisson regression, and ordered and multinomial versions of logit regression. These methods will be compared with the more familiar methods of the general linear model.


Additional references are *Regression Models for Categorical and Limited Dependent Variables with Stata*, 2nd edition, by J. Scott Long and Jeremy Freese, 2006, Stata Press; and *An Introduction to Categorical Data Analysis*, by Alan Agresti, 1996, Wiley. These books will be on reserve.

It is expected that you are already familiar with multiple regression using ordinary least squares, and that you are familiar with at least one major statistical package. Examples will be given with Stata, and students who are not already familiar with that package will be encouraged to learn it. The Long and Freese text is especially recommended for students who want to use Stata. The teaching assistants are also familiar with SAS.
Students are encouraged to bring their own data sets to the workshop, and to do the homework assignments using those data, but other data sets will be available. There will be one assignment each week, and the grade will be based completely on the homework. The schedule of topics is approximate. Some topics may take more or less time than indicated.

Due dates for homework:

HW#1: Monday, July 28
HW#2: Monday, August 4
HW#3: Monday, August 11
HW#4: Friday, August 15

Week 1. Review of OLS regression and categorical predictor variables

Monday, July 21: Introduction; framework of generalized linear models; review of the normal distribution. (Long, Chapter 1; Powers and Xie, Chapter 1)

Tuesday, July 22: Review of the logic of estimation and testing with OLS regression. (Long, Chapter 2; Powers and Xie, Chapter 2)

Wednesday, July 23: Categorical predictors in OLS regression. Dummy variables, including the interpretation of coefficients.

Thursday, July 24: Interactions that include categorical covariates.

Friday, July 25: The poisson distribution; maximum likelihood estimation and tests. (Powers and Xie, Appendix B.2.1; Agresti, chapter 4)

Week 2. Log linear models and poisson regression

Monday, July 28: Log linear models in two-way tables. (Powers and Xie, Chapter 4; Agresti, Chapter 6)

Tuesday, July 29: Log linear models in three-way tables.

Wednesday, July 30: Poisson regression with categorical and interval-level predictors; structural zeroes; mover-stayer models and homogamy models. (Long, Chapter 8; Agresti, Chapter 4.3 and Chapter 9.2-5)

Thursday, July 31: Rate models; age-period-cohort models. (Powers and Xie, Chapter 5)

Friday, August 1: History of the logistic function. Odds, odds ratios, log odds.
Week 3. Logit regression

**Monday, August 4:** Logit regression with interval-level predictors and interaction terms. (Long, Chapter 3; Powers and Xie, Chapter 3; Agresti, Chapter 5)

**Tuesday, August 5:** Logit regression, testing and model selection. (Long, Chapter 4; Agresti, Chapter 7)

**Wednesday, August 6:** Logit regression, use of fitted values; applying the same model to different data sets.

**Thursday, August 7:** Logit regression, applications and examples.

**Friday, August 8:** Log probability models, illustrated with infant mortality rates and age-specific probabilities of dying.

Week 4. Extensions of logit regression

**Monday, August 12:** Multinomial logit regression. (Long, Chapter 6; Powers and Xie, Chapter 7; Agresti, Chapter 8.1)

**Tuesday, August 11:** Ordered logit regression. (Long, Chapter 5; Powers and Xie, Chapter 6; Agresti, Chapter 8.2)

**Wednesday, August 13:** Related alternative models: probit, negative binomial, tobit. (Long, Chapter 7)

**Thursday, August 14:** Models for matched pairs. (Agresti, Chapter 9)

**Friday, August 15:** Review.