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

Antje Schwennicke, Indiana University
aschwenn@indiana.edu

Melinda Ward, University of Texas at Austin
mrw594@mail.la.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 23 to August 17, 2007
Lecture, 3:00 to 5:00; Lab and Office hours to be announced

This course is intended to give you an understanding of the theoretical basis, practical application, and interpretation of the main methods for analyzing counts and categorical outcomes 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, revised edition, by J. Scott Long and Jeremy Freese, 2003, 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, but the teaching assistants are also familiar with SPSS and 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 for the homework. 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 30  
**HW#2:** Monday, August 6  
**HW#3:** Monday, August 13  
**HW#4:** Friday, August 17

**Week 1**

**Monday, July 23:** Introduction; review of the binomial, normal, and poisson distributions  
(Long, Chapter 1; Powers and Xie, Chapter 1)

**Tuesday, July 24:** Review of the logic of estimation and testing with OLS regression  
(Long, Chapter 2; Powers and Xie, Chapter 2)

**Wednesday, July 25:** Categorical predictors in OLS regression: dummy variables, including the interpretation of interactions

**Thursday, July 26:** Measuring and testing for association in two-way and three-way contingency tables  
(Agresti, chapters 2 and 3)

**Friday, July 27:** Reasoning behind maximum likelihood estimation and tests; framework of generalized linear models  
(Powers and Xie, Appendix B.2.1; Agresti, chapter 4)

**Week 2**

**Monday, July 30:** Log linear models in two-way tables  
(Powers and Xie, Chapter 4; Agresti, Chapter 6)

**Tuesday, July 31:** Log linear models in three-way tables
**Wednesday, August 1:** 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, August 2:** Rate models, illustrated with fertility rates; age-period-cohort models
(Powers and Xie, Chapter 5)

**Friday, August 3:** Odds ratios, log odds, and logit regression with categorical predictors; link with log linear models

**Week 3**

**Monday, August 6:** Logit regression with interval-level predictors and interaction terms
(Long, Chapter 3; Powers and Xie, Chapter 3; Agresti, Chapter 5)

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

**Wednesday, August 8:** Logit regression: use of fitted values; applying the same model to different data sets

**Thursday, August 9:** Logit regression: applications and examples

**Friday, August 10:** Log probability models, illustrated with infant mortality rates and age-specific probabilities of dying

**Week 4**

**Monday, August 13:** Ordered logit regression
(Long, Chapter 5; Powers and Xie, Chapter 6; Agresti, Chapter 8.2)

**Tuesday, August 14:** Multinomial logit regression
(Long, Chapter 6; Powers and Xie, Chapter 7; Agresti, Chapter 8.1)

**Wednesday, August 15:** Conditional logit regression

**Thursday, August 16:** Related alternative models: probit, negative binomial

**Friday, August 17:** No class