Categorical Data Analysis: A Second Course  
ICPSR Summer Program in Quantitative Methods, July 2010  
Instructor: Jeremy Freese, Northwestern University (jfreese@northwestern.edu)  
TA: J. Alex Kevern, Northwestern University (kevernja@u.northwestern.edu)

OVERVIEW

There is no accepted canon for the contents of a first categorical data analysis course, which creates some problems for thinking about what to cover in a second. However, during an introduction to regression models for categorical outcomes, the instruction will often make references to material that would be covered or would be covered in more depth if only the instructor had more time. This course is intended to cover a variety of these topics. To get to them, we will review material presented in a standard categorical data analysis course as we go along, but in a considerably accelerated fashion.

PREREQUISITES

The course presumes that you have had a graduate-level regression analysis course and that you have had either a separate course in regression models for categorical outcomes or that your regression course spent a good deal of time on categorical outcomes.

SOFTWARE AND DATA

All demonstrations in the course will be conducted using Stata. For a short course, Stata offers a good balance of sophistication in its available routines, elegance in its syntax, and straightforwardness for those who are not already familiar with it.

The course will make use of add-on routines that are not part of official Stata but need to be downloaded and added.

Students are encouraged to use their own data to work on examples throughout the course. We will bring a couple of publicly available datasets as well, in addition to their own.

TOPICS

1. Models for Continuous Outcomes

   Selective review of linear regression model and assumptions [regress]  
   Special virtues of log-outcome regression models  
   Models for censored and truncated outcomes [tobit, intreg, cnreg, heckman]

2. Binary response models, specification and estimation
Standard binary response models [logit, probit]
Exact logistic regression [exlogistic]
Alternative link functions for binary outcomes [cloglog, scobit]
Post-estimation predictions [predict, prvalue, margins]

3. Model assessment and selection

Analogies to R-square [fitstat]
Information criterion measures [estat ic]
Sensitivity and specificity [lroc, lsens]
Link function tests [linktest, estat gof]
Testing alternative specifications [fracpoly]
Problem of interaction effects in binary response models [hetprob, inteff]

4. Correspondence analysis for multiclass classification data

Visual representation of trichotomous outcomes [spineplot, triplot]
Canonical correlation as a method of optimal scaling [canon]
Two-way correspondence analysis [ca, cabiplot]
Multiway correspondence analysis [mca, mcaplot]

5. Introduction to Event Outcomes

Introduction to Survival and Event History Analysis
Configuring data for survival analysis [stset]
Parametric Models for Event Outcomes [streg, stcurve]

6. Cox model for event outcomes

Specification and estimation of the Cox model [stcox]
The problem of ties for Cox models
Cumulative hazard, survivor, and hazard functions from Cox parameters
Testing the proportional hazards assumption

7. Additional issues in estimating models for event outcomes

Time-varying covariates
Competing risks
Frailty models
Split-population models
8. Ranked outcomes

Treating multivariate response data as ranks
Configuring data for analysis as ranked
Rank-ordered models [rologit, roprobit]

9. Models of event counts

Poisson / negative binomial regression [poisson, nbreg]
Models for truncated counts [ztp, ztnb]
Mixed process models for counts [zip, zinb]

10. Models for tabular counts

Modeling contingency tables as a Poisson process [poisson, glm]
Local independence and interval scoring
Model comparison and models for square tables
Simple three way tables

11. Conditional and mixed logit models

Joint binary outcomes [biprobit]
Multinomial logit model [mlogit]
Fixed-effect and conditional logistic regression model [clogit]
Models with nested choices [nlogit]
Random-intercepts logistic regression [xtlogit]
Generalized estimating equations [xtgee]
Rasch models

12. Ordered categorical outcomes

Standard models for ordered outcomes [ologit, oprobit]
Generalized ordered logistic regression [gologit2]
Stereotype ordered logistic regression [slogit]
Heterogeneous ordered logit model [oglm]
Random-coefficients and random-intercepts model for ordered data [gllamm]

BIBLIOGRAPHY

The readings below are among those most central to the preparation of materials for this course. We will discuss recommended readings over the course of our four days in class.

Baum, Christopher F. 2006. An Introduction to Modern Econometrics Using Stata. College Station, TX: Stata Press.


Some additional reading on specific topics:

*Group comparisons and interaction effects in binary response models*


**Rank-ordered logistic regression model [rologit]**


**Generalized ordered logistic regression [gologit2]**