
Categorical Data Analysis

Course #: 034:218

Fall Term 2006

T, TH 8:05-9:20 am

Main Library, Arcade Classroom (first floor)

Instructor: Mary Noonan
Office: W121 Seashore Hall
Phone: 335-2490
E-mail: mary-noonan-1@uiowa.edu
Office Hours: T, TH 11:00-12:30 pm (or by appt)

Course Description

This course deals with the most basic regression models for categorical outcomes. These include binary logit and probit, ordinal logit and probit, multinomial logit, and count models. The course focuses on (1) conceptual development of different methods, (2) practical issues in model estimation, (3) interpretation of results, and (4) using standard software to estimate and interpret the models. The emphasis will be on practical applications rather than statistical theories. Students will learn to use Stata, a statistical software package, to analyze data.

The course assumes a good working knowledge of the linear regression model for continuous variables (for Sociology students 34:214 and 34:216) and basic algebra.

Class Meetings

The class consists of two 75-minute meetings per week.

For most class meetings, I will present a lecture, interspersed with discussion and questions. You should feel free to discuss in class any problems you are having with the material. Outlines of my lecture notes will be available on the course web page the day before the class meeting (by 3 pm). Please print a copy to bring to class.

During our “computer lab” sessions I will teach you how to use Stata. Each week, I will give a short demonstration of the new commands, and then you will have time to practice them with an in-class exercise. Bring a disk/CD/memory stick to the lab session so that you can save your work. Alternatively you can e-mail your Stata programs to yourself as an e-mail attachment.

Textbook

The text for this course is: Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage. The text is available at University Book Store, located at the Iowa Memorial Union.

This supplementary book covers how to fit and interpret regression models for categorical data with Stata: Long, J. Scott and Freese, Jeremy. 2001. *Regression Models for Categorical and Limited Dependent Variables with Stata*. Stata Press: College Station, TX.

This book is not required for the class (I have a copy that you are welcome to borrow). If you would like to purchase it, you can buy it through Stata (<http://www.stata.com> → click on “Bookstore”).

Occasionally, I will ask you to read journal articles that use the method/models we are currently discussing. These readings will be posted on the course web page as pdf documents.

Web Page

The course web page is <http://www.uiowa.edu/~c034218>. Lecture outlines, assignments, data sets, codebooks, additional readings, and Stata tips are all located on the course web page. You do not need a password to access the web page.

Requirements

It is extremely important that you keep up with the course material. The study of statistics is cumulative; later material builds on earlier material. You will get the most out of the class if you complete the reading assignments before the day in which the material is presented (see course schedule). If you are having problems understanding the readings, then you should see me as soon as possible.

Your final grade is based on three in-class exams and eight problem sets. The weight of each will be as follows:

First Exam	15%
Second Exam	20%
Final Exam	25%
<u>Problem Sets</u>	<u>40%</u>
Total	100%

Letter grades are then assigned as follows: 97-100% (A+), 93-96% (A), 90-92% (A-), etc.

The first exam will cover material from the beginning of the course. The second exam will cover material since the first exam. The final exam is cumulative, although material covered since the second exam will be given more emphasis. All exams are quasi-open book. You will be allowed to consult 1 note card (single-side; 4 X 6 inches) and a calculator during exams.

In each problem set, you will use Stata to analyze a data set that I provide for you. The problem sets will consist of obtaining correct output and manipulating and interpreting the output. At the beginning of the semester, I will hand out a detailed set of formatting specifications and a sample completed assignment so that the expectations for the assignments are clear. Stata 8.0 is available in nearly all of the ITCs on campus. It is also available in the Sociology Grad computer lab (W36). If you are interested in purchasing Stata software for your personal computer, let me know and I can help you order it (cost is approximately \$100).

I will hand out assignments two weeks before they are due. Late problem sets will only be accepted under special circumstances. If you find yourself in a situation in which you need to turn in a problem set late, submit your request to me by e-mail with a specific proposal for when the homework will be completed. My goal is to be reasonable, but fair. No “incompletes” will be given for the course.

You are encouraged to discuss your work with other students, but you must complete assignments on your own. Do not save your work on the Sociology shared drive. If you find that you are having trouble doing the assignments, you should see me as soon as possible. If you want to ask me about a computing

problem, bring a hard copy and electronic copy of the Stata do file and log file. Make sure that your log file contains the error message that you received.

Other Considerations

- (1) Academic accommodation. If you have a disability that may require some modification of seating or class requirements, please notify me after class or during my office hours so that appropriate arrangements can be made.
- (2) Procedure for student complaints. If anyone is faced with an issue regarding this class that they feel they cannot resolve by speaking with me, please contact Michael Lovaglia, the Department of Sociology Executive Officer.

Additional Resources

Agresti, Alan. 1990. *Categorical Data Analysis*. New York: John Wiley & Sons, Inc. (*Deals mainly with logit and loglinear models for contingency table data, but there is also some treatment of logistic regression models*)

Cameron, A. Colin and Pravin K. Trivedi. 1998. *Regression Analysis of Count Data*. Oxford: Oxford University Press. (*Good resource if you will do a lot of work with count data*)

DeMaris, Alfred. 1992. *Logit Modeling: Practical Applications*. Sage Quantitative Applications in the Social Sciences Series, No. 106. Thousand Oaks, CA: Sage Publications.

Kleinbaum, David G. and Mitchel Klein. 2002. *Logistic Regression: A Self-Learning Text* (2nd edition). New York: Springer-Verlag.

Hosmer, David W. and Stanley Lemeshow. 2000. *Applied Logistic Regression* (2nd edition). New York: John Wiley & Sons, Inc. (*Used mostly by biostatisticians and health researchers – use Stata examples throughout*)

Pampel, Fred C. 2000. *Logistic Regression: A Primer*. Sage Quantitative Applications in the Social Sciences Series, No. 132. Thousand Oaks, CA: Sage Publications. (*Green booklet – concise overview of main issues*)

Powers, Daniel A. and Yu Xie. 2000. *Statistical Methods for Categorical Data Analysis*. San Diego: Academic Press. (*Includes discussion of loglinear models and event history models*)

Tabachnick, Barbara G. and Linda S. Fidell. 2001. *Using Multivariate Statistics* (4th edition). New York: Harper Collins. (*Good overall resource for review of multiple regression for linear models; includes short sections on logistic regression and survival analysis*)

Schedule of Topics, Readings, and Assignments

The schedule below is tentative. We will move at a pace that represents the ability and interests of the class. Some topics may be expanded, while others may be shortened. All changes in the schedule will be announced in class. The latest rendition of the syllabus will always be made available on the course web page.

- **Overview and Stata Basics**

Dates: Aug 22, 24, and 29

Reading: Long, Chapter 1

**Assignment 1

- **Continuous Outcomes**

Dates: Aug 31 and Sept 5

Reading: Long, Chapter 2.1-2.5

**Assignment 2

- **Maximum Likelihood Estimation**

Dates: Sept 7

Reading: Long, Chapter 2.6-2.7

- **Binary Outcomes**

Specification and Estimation

Reading: Long, Chapter 3.1-3.6

Dates: Sept 12, 14, and 19

Interpretation

Reading: Long, Chapter 3.7-3.9

Dates: Sept 21, 26, 28, Oct 3 and 5

**Assignment 3

**Exam 1 – Oct 10

- **Hypothesis Testing**

Dates: Oct 12 and 17

Reading: Long, Chapter 4.1

**Assignment 4

Goodness of Fit

Dates: Oct 19 and 24

Reading: Long, Chapter 4.2-4.4

**Assignment 5

- **Ordinal Outcomes**

Specification and Estimation

Dates: Oct 26 and 31

Reading: Long, Chapter 5.1-5.3

Interpretation

Dates: Nov 2, 7, 9, and 14

Reading: Long, Chapter 5.4-5.7

**Assignment 6

**Exam 2 – Nov 16

*** No Class Nov 21 and 23 due to Thanksgiving Recess ***

- **Nominal Outcomes**

Specification and Estimation

Dates: Nov 28

Reading: Long, Chapter 6.1-6.5

Interpretation

Dates: Nov 30 and Dec 5

Reading: Long, Chapter 6.6-6.10

**Assignment 7

**Exam 3 – Dec 7

Note: Many of the examples used in this course were made available through the generosity of Scott Long and are being used with his permission.