CATEGORICAL DATA ANALYSIS

Spring, 2009
Tuesday, 6:15-9:00 p.m. (Eastern)
5:15-8:00 p.m. (Central)
Instructor: Allan McCutcheon
Course Web Site: www.jpsm.umd.edu/surv699a

Instructor's Office:

Allan McCutcheon
UNL-Gallup Research Center
University of Nebraska-Lincoln
402-472-7793 (O)
402-472-7764 (F)
AMcCutcheon1@UNL.Edu

A. Overview of the Course

This course is organized into three related parts. First, we will briefly explore some of the standard methods for categorical data analysis developed prior to about 1960. In these early sessions, we focus on assessing association in 2- and 3-variable models in which all variables are categorical. The second part of the class focuses on models for categorical response variables. Following a brief review of standard Gaussian-based data-analytic methods—especially regression models with continuous response variables—we consider the logistic regression and loglinear model, which assume binomial, or Poisson, rather than normal, distributions for the response variable. We will also briefly consider some alternatives such as probit models. The third part of the course will extend the binomial response model to consider multinomial logit and loglinear models for ordered and unordered response variables, and for models with and without measurement error. As we will see, these models share much in common with the Gaussian-based data-analytic methods of structural equation modeling (SEMs). Time permitting, we will briefly discuss some of recent advances in categorical data (e.g., hierarchical/multilevel modeling, mixed Markov models). Course discussion will focus on model estimation, identification, evaluation and interpretation.

B. Grading and Course Assignments

Grades will be based on homework assignments, a midterm, and a final. Together the homework assignments will count as 40% of the grade; the midterm and final will each count 30%. Both the midterm and final will be open-book exams; the final will include material from the entire course, but will concentrate on material presented following the midterm. Participation in class discussions will be evaluated informally; contributions to class discussions should demonstrate familiarity with the readings. A final exam will be given during the regular class time on April 25.

C. Office Hours

The instructor will be available to meet with each student on a regular basis. Please call (402) 458-2035 or send an e-mail to set up an appointment.

D. Course Readings

There are four required texts.

- Agresti, Alan (2007) *An Introduction to Categorical Data Analysis, Second Edition.* New York: Wiley.
  Dayton, C. Mitchell (1999) *Latent Class Scaling Analysis*. Sage.
  Hagenaars, Jacques A. (1993) *Loglinear Models with Latent Variables* Sage.
  McCutcheon, Allan L. (1987) *Latent Class Analysis*. Sage.

And two recommended texts.

Agresti, Alan (2002) *Categorical Data Analysis*. New York: Wiley.
Hagenaars, Jacques A. and Allan L. McCutcheon (2002) *Applied Latent Class Analysis*. New York: Cambridge University Press.

E. A Note on Software

This is not a course on the use of a particular brand of software, though you must have available, and must know how to use, one of the "major" statistical software packages (e.g., SPSS, SAS, STATA) for estimating several of the models considered in the course; these programs can not yet estimate the measurement error models considered in the final segment of this course. Example syntax for these programs will be provided in the class, though students are encouraged to develop proficiency in the use of one or more of these software packages. For those using the SAS program, a useful guidebook for categorical data analysis is *Categorical Data Analysis Using the SAS System, Second Edition,* **by Maura E. Stokes, Charles S. Davis and Gary G. Koch (2000, Wiley). For those using the STATA program, a useful guidebook for categorical data analysis is** *Regression Models for Categorical Dependent Variables Using STATA, Revised Edition,* **by J. Scott Long and Jeremy Freese (2003, Stata Press). Neither of these books is required for the course.**

A widely available freeware program called LEM will be used for estimation of the models with measurement error. LEM can also be used to estimate most of the models presented in this course.

F. Lecture Topics, Readings, and Schedule

JANUARY 27: INTRODUCTION TO CATEGORICAL DATA

Agresti, Chapter 1.

FEBRUARY 3: CONTINGENCY TABLES

Agresti, Chapter 2.

ASSIGNMENT 1 HANDED OUT.

FEBRUARY 10: GENERALIZED LINEAR MODELS; POISSON VS. GAUSSIAN-BASED REGRESSION MODELS

Agresti, Chapter 3.

ASSIGNMENT 1 DUE.

FEBRUARY 17: LOGISTIC REGRESSION

Agresti, Chapters 4-5.

ASSIGNMENT 2 HANDED OUT.

FEBRUARY 24: LOGLINEAR MODELS FOR CONTINGENCY TABLES

Agresti, Chapter 7.
[Hagenaars, Chapters 1-2.]

ASSIGNMENT 2 DUE.

MARCH 3: MULTINOMIAL LOGIT MODELS; REVIEW

Agresti, Chapter 7.

MARCH 10: MIDTERM EXAM

MARCH 17: SPRING BREAK

MARCH 24: MODELS FOR MATCHED PAIRS

Agresti, Chapter 8.

MARCH 31: RANDOM EFFECTS: GENERALIZED LINEAR MIXED MODELS

Agresti, Chapter 10.

APRIL 7: LATENT CLASS ANALYSIS: MEASUREMENT ERROR MODELS FOR CATEGORICAL DATA

McCutcheon, *Latent Class Analysis*, Chapters 1-4
[Hagenaars and McCutcheon, *Applied Latent Class Analysis*. Chapters 2-3.]

ASSIGNMENT 3 HANDED OUT.

APRIL 14: LATENT CLASS SCALE MODELS.

Dayton, *Latent Class Scaling Models,* Chapters 1-5.
[Hagenaars and McCutcheon, *Applied Latent Class Analysis*. Chapter 5.]

ASSIGNMENT 3 DUE.

APRIL 21: CAUSAL MODELS WITH LATENT CATEGORICAL VARIABLES.

McCutcheon, *Latent Class Analysis*, Chapter 5.
Dayton, *Latent Class Scaling Models*, Chapter 6.
Hagenaars, Loglinear Models with Latent Variables, Chapters 3-5.
[Hagenaars and McCutcheon, *Applied Latent Class Analysis*. Chapter 9.]

ASSIGNMENT 4 HANDED OUT.

APRIL 28: ADVANCED TOPICS, CATEGORICAL MODELS OF CHANGE.

Hagenaars, Loglinear Models with Latent Variables, Chapter 6.
[Hagenaars and McCutcheon, *Applied Latent Class Analysis*. Chapter 10-11.]

ASSIGNMENT 4 DUE.

MAY 5: FINAL EXAM