

Sociology 952
MATHEMATICAL AND STATISTICAL APPLICATIONS IN SOCIOLOGY
Topic: Models for Categorical and Limited Dependent Variables

Syllabus, Spring, 2005

PREREQUISITE: Sociology 361 and 362, or consent of instructor.

TIME & PLACE: Wednesday, 3:30-5:25 p.m., 6116 Social Science.

INSTRUCTOR: John Allen Logan, 4458 Social Science, 262-0995, logan@ssc.wisc.edu.

Office hours: M 2:00-4:00, and by appointment.

SUBJECT MATTER

This course covers the analysis of data where the dependent variables are categorical or partly-categorical and partly continuous. Such dependent variables can be dichotomous, polytomous, counted, ordered, censored, and/or subject to selection. We will consider categorical observations that are either independent or else clustered, as in panel designs. We will also study discrete-choice models that provide behavioral interpretations of categorical responses. In addition we will look at general methods for handling missing data, which is logically similar to a categorical response. Please consult the list of topics below for details of the models to be covered. The course emphasizes understanding the logic of the methods, obtaining familiarity with software, and practical experience.

REQUIRED TEXTS (ordered from University Book Store):

Agresti, Alan. 2002. *Categorical Data Analysis*. 2nd ed. Wiley.

Train, Kenneth. 2003. *Discrete Choice Methods With Simulation*. Cambridge Univ. Press.

Breen, Richard. 1996. *Regression Models: Censored, Sample Selected, or Truncated Data*. Quantitative Applications in the Social Sciences 111. Thousand Oaks, CA: Sage.

Allison, Paul D. 2002. *Missing Data*. Quantitative Applications in the Social Sciences 136. Thousand Oaks, CA: Sage.

RESEARCH PAPER

The main course requirement is an empirical research paper that applies some of the methods covered in the seminar to a substantive problem in social science. The paper should include formulation of a research problem, linkage to substantive literature, analysis of suitable data, and discussion of findings. Purely methodological papers will not be accepted.

Your paper cannot use simple (binary) logistic regression as its only method. It should also carefully consider and deal with any missing data problems, using the advice and/or methods covered in this class.

Assigned reading: Students are **assigned** the following reading as a guide to producing a good statistical analysis and research paper: Leland Wilkinson, et al. 1999. "Statistical Methods in Psychology Journals: Guidelines and Explanations." *American Psychologist* 54: 594-604, which is available at this address: <http://www.apa.org/journals/amp/amp548594.html>. (Your term paper will be graded in part on whether you show signs of having read this.)

This paper is **strongly recommended** for advice on organizing your research data: Svend Juul, "Take Good Care of Your Data." www.biostat.au.dk/teaching/software/STATA/takecare.pdf.

Paper proposal: Students should submit a paper proposal that outlines the problem to be addressed, the data to be used, and the empirical analyses to be undertaken. **Due: March 16th** in class.

The instructor will respond to proposals with comments, if the proposals are turned in on time.

The final paper is **due on the last class day, May 4th**.

HOMEWORK

In addition to the paper, there will be homework assignments most weeks. Most assignments will involve use of a computer. Homework will be due in class the following week, unless announced otherwise. Grading of the homework will be simplified; rather than detailed marking of each homework, I will distribute an answer key.

CLASS WEBSITE

Students will be expected to make use of a web browser for downloading homework assignments, exercise data and other information from the class website, which can be found from this address: <http://www.ssc.wisc.edu/~logan>.

COMPUTING

SAS and Stata will be the main statistical analysis packages. These programs are available on the SSC Unix network. (Stata is also available on the SSC's Windows servers.) Class members are eligible for free accounts on these systems. (Stata can also be purchased for one's own PC or Mac at a very substantial discount, compared to the usual academic price. Contact the Social Science Computing Coop on the 4th floor of the Social Science Building.)

Some class time will be devoted to use of Stata, SAS and other programs for model estimation. No time, however, will be given to data extraction and management using general packages or dataset-specific extractors. Generally, students will be expected to use a good deal of initiative in figuring out how to operate software.

GRADING

The term paper will determine 75 percent of the final grade. Homework assignments and class participation will determine 25 percent. **THE GRADE "INCOMPLETE" WILL NOT BE GIVEN.**

SUPPLEMENTARY READING

The following works can be consulted for additional treatment of some of the course topics. Some readings may be assigned as **required reading** for particular topics during the semester.

Amemiya, Takeshi. 1985. *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

Arminger, Gerhard, Clifford C. Clogg, and Michael E. Sobel (eds.). 1995. *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. New York and London: Plenum.

Ben-Akiva, Moshe, and Steven R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.

Berk, Richard A. 1983. "An Introduction to Sample Selection Bias in Sociological Data." *American Sociological Review* 48:386-98.

Breen, Richard. 1994. "Individual Level Models for Mobility Tables and Other

CrossClassifications." *Sociological Methods and Research* 23(2):147-173.

Cameron, Stephen V., and James J. Heckman. 1998. "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males." *Journal of Political Economy* 106(22):262-333.

Davidson, Russell, and James G. MacKinnon. 2004. *Econometric Theory and Methods*. Oxford.

DeMaris, Alfred. 2002. "Explained Variance in Logistic Regression; A Monte Carlo Study of Proposed Measures," *Sociological Methods and Research* 31(1): 27-74.

Diggle, Peter J., Liang, Kung-Yee, and Scott L. Zeger. 1994. *Analysis of Longitudinal Data*. New York: Oxford University Press.

Greene, William H. 2002. *Econometric Analysis*. 5th Edition. New York: Macmillan.

Hoffman, Saul d. and Greg J. Duncan. 1988. "Multinomial and Conditional Logit Discrete-Choice Models in Demography." *Demography* 25: 415-427.

Hosmer, David W., and Stanley Lemeshow. 2000. *Applied Logistic Regression*. 2nd. Ed. New York: Wiley.

Hout, Michael, O.D. Duncan, and Michael E. Sobel. 1987. "Association and Heterogeneity: Structural Models of Similarities and Differences." *Sociological Methodology* 17: 145-184.

Hsiao, Cheng. 2003. *Analysis of Panel Data*, Second edition. Cambridge.

Kalbfleisch, John D., and Ross L. Prentice. 1980. *The Statistical Analysis of Failure Time Data*. New York: Wiley.

Lindsey, J.K. 2004. *Introduction to Applied Statistics*. 2nd ed. Oxford.

_____. 1996. *Parametric Statistical Inference*. Oxford.

Little, Roderick J.A., and Donald B. Rubin. 1987. *Statistical Analysis with Missing Data*. New York: Wiley.

Little, Roderick J.A., and Nathaniel Schenker. 1995. "Missing Data." Pp. 39-75 in *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, edited by Gerhard Arminger, Clifford C. Clogg, and Michael E. Sobel. New York and London: Plenum.

Logan, John Allen. 1996. "Opportunity and Choice in Socially Structured Labor Markets." *American Journal of Sociology* 102(1; July):114-160.

_____. 1996. "Rules of Access and Shifts in Demand: A Comparison of Log-Linear and Two-Sided Logit Models." *Social Science Research* 25 (June):174-199.

Long, J. Scott. 1997. *Regression Models for Categorical and Limited-Dependent Variables*. Thousand Oaks, CA: Sage.

Long, J. Scott, and Jeremy Freese. 2003. *Regression Models for Categorical Dependent Variables Using Stata*. Stata Press.

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Mare, Robert D. 1980. "Social Background and School Continuation Decisions." *Journal of the American Statistical Association* 75:295-305.

_____. 1990. "Five Decades of Educational Assortative Mating." *American Sociological Review* 56:15-32.

Powers, Daniel A., and Yu Xie. 2000. *Statistical Methods for Categorical Data Analysis*. Academic Press.

Pudney, Stephen. 1989. *Modelling Individual Choice*. Oxford: Basil Blackwell.

Raftery, Adrian E. 1995. "Bayesian Model Selection in Social Research." Pp. 111-163 in *Sociological Methodology 1995*, edited by Peter V. Marsden. Cambridge, MA: Basil Blackwell.

Schafer, Joseph L. 1997. *Analysis of Incomplete Multivariate Data*. New York, London: Chapman and Hall.

Simonoff, Jeffrey S. *Analyzing Categorical Data*. New York: Springer, 2003.

Stolzenberg, Ross M., and Daniel A. Relles. 1997. "Tools for Intuition about Sample Selection Bias and Its Correction." *American Sociological Review* 62: 494-507.

Vermunt, Jeroen K. 1997. *Log-Linear Models for Event Histories*. Thousand Oaks, CA: Sage.

Vermunt, J.K., Rodrigo, M.F., and Ato-Garcia, M. (2001) "modeling joint and marginal distributions in the analysis of categorical panel data." *Sociological Methods and Research* 30:170-196.

Winkelmann, Rainer. 2000. *Econometric Analysis of Count Data*. 3d edition, revised and enlarged. Berlin: Springer-Verlag.

Winship, Christopher, and Larry Radbill. 1994. "Sampling Weights and Regression Analysis." *Sociological Methods and Research* 23(2):230-257.

Woolridge, J.M.. (2001) *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.

Zorn, "Generalized Estimating Equation Models for Correlated Data: A Review With Applications," *American Journal of Political Science*, 45(2): April, 2001.

SCHEDULE AND READINGS (next page).

Both schedule and readings are subject to change. Many articles can be downloaded from www.jstor.org or other sites via the University Library's MadCat catalog. Other articles may be made available via the class website.

Date	Topic	Assigned Textbook Readings
1/19	Introduction, probability models and likelihoods.	(Review of material from Soc. 361.) After class, read Agresti, Ch. 1.
1/26	Generalized linear models and logistic regression.	Agresti 4.1-4.3, 4.5, 4.6, 5.1-5.5.
2/2	Logistic diagnostics; multinomial (including ordinal) logistic models.	Agresti 6.1, 6.2, 7.1-7.4.
2/9	Censored and sample-selection models.	Breen 1.1-3.4.
2/16	Censored/selection models: extensions, limitations, and alternatives.	Breen 4.1-5.5. (and additional reading)
2/23	Missing data: conventional and ML methods.	Allison, chapters 1 through 4.
3/2	Missing data: multiple imputation methods.	Allison, chapters 5 and 6.
3/9	Log-linear models.	Agresti, 8.1-8.5, 8.6.5, 8.6.7, 8.7.2-8.7.4.
3/16	Extensions of log-linear models. (<i>Paper proposals due.</i>)	Agresti 9.1-9.6, 9.8.
3/23	(Spring Break)	
3/30	Models for matched pairs.	Agresti 10.1-10.6.
4/6	Repeated categorical responses.	Agresti 11.1-11.3, 11.5.
4/13	Random effects models for categorical responses.	Agresti 12.1-12.6.
4/20	Discrete choice models.	Train 1.1, 1.2, 2.1-2.5, 2.8, 3.1-3.3, 3.6, 3.8-3.9.
4/27	Nested logit and probit models.	Train 4.1-4.4, 5.1-5.6.
5/4	Mixed logit.	Train 6.1-6.8.