

**The University of Calgary
Faculty of Social Sciences
Department of Sociology**

Sociology 711.02
Modeling Categorical Dependent Variables
And Longitudinal Data
Winter, 2010
Course Outline

Instructor: Dr. Richard Wanner
Office: Social Sciences 1058
Office Hours: Tues., Wed. 1:30-2:30
Email: wanner@ucalgary.ca

Associate: Charlie Victorino
Office: MLT 116
Office Hours: TBA
Email: rdc@ucalgary.ca

Course Objectives

This course is designed to survey a variety of specific modeling techniques within the generalized linear model. Given the explosion of approaches to modeling in the social sciences over the past two decades, our survey can only be selective. The course will examine regression-based methods for categorical variables (various forms of logistic regression models), the occurrence of events over time (event history models), data arranged hierarchically at different levels of analysis (multilevel models), models for count dependent variables, and models appropriate for panel designs.

Given the wide scope of the course, our approach will be mainly applied and involves reading both technical work and a substantive application on each technique, as well as learning to use computer software to estimate each type of model. The emphasis will be on choosing an appropriate modeling strategy based on theoretical, design, and measurement considerations and the interpretation of parameter estimates and fit statistics.

Required Texts

Paul D. Allison, *Event History Analysis*. Thousand Oaks, California: Sage, 1984.

Douglas A. Luke, *Multilevel Modeling*. Thousand Oaks, California: Sage, 2004.

Charles W. Ostrom, Jr. *Time Series Analysis: Regression Techniques* (2nd Ed.).
Thousand Oaks, Calif.: Sage, 1990.

Fred C. Pampel, *Logistic Regression: A Primer*. Thousand Oaks, California: Sage,
2000.

In addition, a number of articles and book excerpts treating specialized topics will be distributed. To supplement the readings on statistical and methodological topics, for each unit we will also read and discuss a substantive article from a recent journal using the techniques taught in the course.

A Note on Software

The main statistical software package to be used in this course is Stata. It is rapidly becoming the standard for sociologists, economists, and other social scientists, as well as biomedical researchers, who are doing serious data analysis. It's main virtue is that, because it is programmable, much of it is designed by users themselves. This means that state-of-the-art statistical procedures are generally available first in Stata. While SPSS or SAS can do many of the procedures discussed in this course, they will not be supported in the lectures or labs. For full credit, all lab exercises, except for those using other software, must include Stata output.

Some Web Resources

The following are web pages that have substantial amounts of information about Stata and HLM. The UCLA page in particular is an important tool for learning to use Stata.

www.stata.com

www.ats.ucla.edu/stat/stata/

www.ssicentral.com/hlm/hlm.htm

Grading

Your mark for the course will be largely based upon a paper in which you will be asked to carry out a secondary analysis of some data in a substantive area of interest to you. The paper will comprise 80 percent of your final grade. In addition, each unit of the course will be accompanied by a computer exercise utilizing statistical software such as Stata, or HLM. The computer exercises will be worth 20 percent of your final mark. These exercises will be marked on a "pass-fail" basis, with a pass awarded if all questions are attempted. Unsubmitted or incomplete exercises will be awarded a "fail." Exercises are due on the day of the lab to which they apply. None will be accepted late. While you are encouraged to consult with the instructor, lab assistant, and other students in the course in carrying out the labs, *each student must submit his or her own work for marking*. Under this system, you will receive an 'A' for 20 percent of your mark by simply submitting the completed exercises on time. Eight percent will be deducted for each exercise not received on time.

TENTATIVE CALENDAR

Mar. 4 Models for Binary Dependent Variables*Technical Readings*

Fred C. Pampel, *Logistic Regression: A Primer*

*J. Scott Long, 1997. *Regression Models for Categorical and Limited Dependent Variables*, Chapter 3.

Substantive Example

A. Portes, W. J. Haller, and L.E. Guarnizo. 2002. "Transnational Entrepreneurs: An Alternative Form of Immigrant Economic Adaptation." *American Sociological Review* 67:278-98.

Lab Exercise and Reference

Estimating and interpreting logistic regression using Stata `logistic` and `logit`
Stata Reference Manual Vol. 2 pp. 220-259 (entries for `logistic` and `logit`)

Mar. 11 Models for Nominal, Ordinal, and Count Dependent Variables*Technical Readings*

J. Scott Long, 1997. *Regression Models for Categorical and Limited Dependent Variables*, Chapters 5, 6 and 8.

Substantive Examples

A. Rhea and L.B. Otto. 2001. "Mothers' Influences on Adolescents' Educational Outcome Beliefs." *Journal of Adolescent Research* 16:491-510.

Nogle, June Marie. 1994. "Internal Migration for Recent Immigrants to Canada." *International Migration Review* 28: 31-48.

Lab Exercise and Reference

Estimating and interpreting logistic regression using Stata `ologit` and `mlogit`
Stata Reference Manual Vol. 2 pp. 358-373 and 451-459 (entries for `mlogit` and `ologit`)

Mar. 18 Event History (Survival) Analysis:*Technical Readings*

Allison, *Event History Analysis*.

*Paul Allison 2004. "Event History Analysis." Pp. 369-385 in M. Hardy and A. Bryman (eds.) *Handbook of Data Analysis*. Thousand Oaks, Calif.: Sage.

*M. A. Cleves, W.W. Gould, and R.G. Gutierrez. 2002. *An Introduction to Survival Analysis Using Stata*. College Station, Texas: Stata Press.

Substantive Example

J. Z. Zhao, F. Rajulton, and Z.R. Ravenera. 1995. "Leaving Parental Homes in Canada: Effects of Family Structure, Gender, and Culture." *Canadian Journal of Sociology* 20: 31-50.

Lab Exercise and Reference

Estimating and interpreting an event history model using Stata `stcox`
Stata Reference Manual Vol. 2 pp. 273-283 (entry for `ltable`); Vol. 3 pp. 266-306 (entry for `stcox`)

Mar. 25; April 1 Time-Series and Panel Regression: Fixed- and Random-Effects Models

Technical Readings

C.W. Ostrom, Jr. 1990. *Times Series Analysis: Regression Techniques* (Second Edition). Newbury Park: Sage, pp. 5-57.

Trond Petersen. 2004. "Analyzing Panel Data: Fixed- and Random-Effects Models." Pp. 331- 345 in M. Hardy and A. Bryman (eds.) *Handbook of Data Analysis*. Thousand Oaks, Calif.: Sage.

*W.H. Greene. 2000. *Econometric Analysis*. (4th ed.). Upper Saddle River, NJ: Prentice-Hall, pp. 557-584.

Substantive Examples

Michelle J. Buddig and Paula England. 2001. "The Wage Penalty for Motherhood." *American Sociological Review* 66:204-225.

Bruce Arai. 1997. "The Road Not Taken: The Transition from Unemployment to Self-Employment in Canada, 1961-1994." *Canadian Journal of Sociology* 22: 365-382.

Lab Exercise and Reference

Estimating and interpreting random- and fixed-effects models using Stata xtreg
Stata Reference Manual, Vol. 4 pp. 285-290, 434-455 (entries for xt and xtreg)

April 8; April 15 Multilevel (Mixed) Models

Technical Readings

Luke, *Multilevel Modeling*

*S.W. Raudenbush and A.S. Bryk. 2002. *Hierarchical Linear Models* (2nd Edition). Newbury Park: Sage, pp. 3-37; 68-95.

Substantive Example

Lincoln Quillian. 1995. "Prejudice as a Response to Perceived Group Threat: Population Composition and Anti-Immigrant and Racial Prejudice in Europe." *American Sociological Review* 60: 586-611.

Lab Exercise

Estimating and interpreting a multilevel model using HLM
HLM Manual, pp. 14-37