

EVENT HISTORY/HAZARDS/SURVIVAL ANALYSIS
SOCIOLOGY 313
VANDERBILT UNIVERSITY, SPRING 2004

Professor Michael E. Ezell

Time: Wednesdays 2:10– 4:25

Email: mike.ezell@vanderbilt.edu

Office Hours: Monday and Friday 1:00-2:00 & by appointment

Office: 312 Garland Hall

COURSE OBJECTIVES

This course provides a systematic exposition of the methods of event history/hazards/survival analysis, with particular emphasis on the regression models used to investigate the effects of covariates on the “timing” of an event. This class will involve both a practical, applied focus and a theoretical focus, as it is important to know not only how to obtain estimates of the various quantities of survival analysis and their resulting statistical properties, but it is important to know what the estimates substantively communicate to you about the event process.

The course begins with an introduction to both the statistical and data problems associated with failure time data. We will then transition to covering the descriptive and inferential statistical methods used for survival data. This will include covering such regression models as discrete time survival models, the semi-parametric Cox model (which is the workhorse of survival analysis), as well as various parametric models (e.g., lognormal models). The course will conclude with introductions to more advanced topics of survival analysis situations, especially what to do when an event can be repeated more than once and what to do in the situation where individuals are at risk of experiencing more than one type of event at any given time (competing risks).

PREREQUISITES

At a minimum, participants should have completed both an introductory statistics course and a course that covers the OLS regression model. Additionally, it is recommended that participants have had some exposure to the generalized linear model (e.g., logistic regression, probit regression). If you have not had exposure to such statistical treatments, you are advised to discuss this course with the professor. Students are assumed to have a thorough understanding of basic topics such as measures of central tendency and dispersion, sampling distributions, hypothesis testing, estimators and properties of estimators, and statistical inference.

In order to successfully complete this course, you have to be able to understand matrix algebra and be able to read matrix notation. All advanced statistical methods are based on matrix algebra, and there is no way around that fact. Participants in this class are assumed to have had some exposure to matrix algebra. If you have not had exposure to matrix algebra, you will have to learn to use matrix algebra in this class. Matrix algebra, at least what is necessary for this class, is really not that terribly complicated and, with a little effort on your part, anyone with basic algebra skills can master the depth of matrix algebra required to be successful in this class. There will be a special lab on the use of matrix algebra sometime during the first three weeks of this course to help you in this endeavor.

TEXT & READINGS

Required Textbook:

Hosmer, David W., and Stanley Lemeshow. 1998. *Applied Survival Analysis: Regression Modeling of Time to Event Data*. New York: John Wiley & Sons. [denoted as H&L below]

“Required Readings”

On the class schedule below, “Required Readings” denote the readings that you are expected to have read *prior* to the beginning of class for that day.

For the “Required Readings” that are not from the textbook, I will indicate where you are to obtain them from. Some of these readings will be placed in the copy room (in the Sociology department office) in the “Soc 313” folder at least one week prior to class and you may read them at your convenience.

“Recommended Readings”

In addition to the assigned readings, I have also listed other “Recommended Readings” that further discuss relevant topics. These are not required readings, but are expositions that are relevant for more in-depth or duplicative coverage of the material should you be interested in further information.

“Substantive Readings”

On relevant days, I will also be providing the citations for substantive (not methodological) articles that employ the use of the methods of survival analysis in answering a particular substantive question. Several of these readings will be mandatory readings for certain class days. Even if not mandatory, these substantive articles are highly recommended and will help you in the transition from the learning the method to employing the method in substantive application.

LECTURE NOTES

I will be providing you with reduced versions of my lecture notes in order to minimize the time you spend in class writing down formulas and to increase the amount of time you spend critically thinking about the material. These notes are in no way a substitute for the content of the required readings.

COURSE REQUIREMENTS

Your course grade will be determined on a grand total of 1,000 points.

	<u>%</u>	<u>Points</u>
Problem Sets (5 @ 12%):	60%	600
Final Paper:	40%	400

Problem Sets (60%)

Participants will complete four problem sets during the semester. The problem sets will involve questions focusing on both your substantive understanding of the topics and your ability to complete specified data analysis tasks.

You will need to turn in the both the “.do” Stata syntax file and the corresponding “.log” output file with each problem set. Make a copy of each of your problem sets prior to handing it in! It may not be returned before you need to refer to it (e.g., for the midterm).

Late problem sets will not be accepted unless you have received permission from the professor in advance of the assignment due date. The datasets for the problem sets will be available through the class website:

`http://people.vanderbilt.edu/~mike.ezell/courses/soc313.htm`

Also note that Stata is “web aware” and can directly load the data sets into Stata over the web using the command:

`use http://people.vanderbilt.edu/~mike.ezell/courses/soc313/data/filename`

Final Paper (40%)

Each student in this class will complete a paper employing the use of the methods of survival analysis on a topic they choose. The paper will be graded primarily on the quality and extensiveness of its application of the topics of this course. The topic of your paper and the data source will need to be submitted by February 18th. The paper is expected to be 15-30 pages in length, and will be written in an academic journal style.

STATISTICAL SOFTWARE

The problem sets all involve the use of a computer and it is recommended that they be completed with the statistical software package Stata. It is **strongly advised** that you do not use another package other than Stata for several reasons. First, and foremost, all in-class demonstrations and examples will be done using this package. Second, the survival analysis routines of Stata are extensive, complete, wide-ranging, and built-in. Third, every data analysis task you will be asked to complete can be successfully completed using Stata; I cannot make the same guarantee for any other program. Fourth, I have extensive exposure to Stata and I will be available during my office hours to help you with any Stata problems you encounter with the problem sets or your project. Let me state at the outset though that you cannot complete this class using the program SPSS due to its extremely limited coverage of the methods of survival analysis. If you wish to use another package, please let me know ASAP so that we can discuss it.

COURSE OUTLINE

Below is a tentative outline of the topics to be discussed in class each day and the dates that each of the problem sets will be handed out. This schedule is more or less firm, but if I have inaccurately projected the time required for each topic of discussion, I reserve the right to make modifications to the topics to be discussed and/or the dates of assignments/exam. The adjustments will be made on an as needed basis.

January 14 Introduction to Course/Stata

January 21 Introduction to Survival Analysis

Required Reading:

H&L—Chapter 1

Recommended Readings:

- Andersen, Per Kragh, and Niels Keiding. 1998. "Survival Analysis: An Overview." Pp. 4452-61 in *Encyclopedia of Biostatistics*, edited by P. Armitage and T. Colton. Chichester, NY: John Wiley & Sons.
- Chung, Ching-Fan, Peter Schmidt, and Ann D. Witte. 1991. "Survival Analysis: A Survey." *Journal of Quantitative Criminology* 7(March):59-98.
- Kiefer, Nicholas. 1988. "Economic Duration Data and Hazard Functions." *Journal of Economic Literature*. 26:646-79.
- Singer, Judith D., and John B. Willett. 1991. "Modeling the Days of Our Lives: Using Survival Analysis When Designing and Analyzing Longitudinal Studies of Duration and the Timing of Events." *Psychological Bulletin*: 1110: 268-290.

January 28 Basic Quantities of Survival Analysis & Descriptive/nonparametric Analysis

Required Reading:

H&L—Chapter 2, Appendix 2

Substantive Application Reading:

- Joo, Hee-Jong, Sheldon Eklund-Olson, and William R. Kelly. 1995. "Recidivism Among Paroled Property Offenders Released During a Period of Prison Reform." *Criminology* 33: 389-409.

February 4 Descriptive Analysis (cont.) & Using Stata for Survival Analysis

Required Reading:

Various segments of the Stata Manual

February 11 Discrete Time Models

- Assignment 1 Due

Required Reading:

Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42(October): 1260-88.

Recommended Readings:

Allison, P. 1982. "Discrete Time Methods for the Analysis of Event Histories," in Leinhardt S (ed) *Sociological Methodology 1982*, Jossey-Bass, San Francisco, 61-97. (Available through JStor)

Singer, Judith D., and John B. Willett. 1993. "It's About Time: Using Discrete-Time Survival Analysis to Study Duration and Timing of Events." *Journal of Education Statistics* 18:255-195.

Substantive Application Readings:

Berry, Frances Stokes and William D. Berry. 1990. "State Lottery Adoptions As Policy Innovations: An Event History Analysis." *American Political Science Review* 84(June): 395-415.

Grattet, Ryken, Valerie Jenness and Theodore R. Curry. 1998. "The Homogenization And Differentiation of Hate Crime Law in the United States, 1978-1995: Innovation and Diffusion in the Criminalization of Bigotry." *American Sociological Review* 63: 286-307.

Long, J. Scott, Paul D. Allison and Robert McGinnis. 1993. "Rank Advancement in Academic Careers: Sex Differences and the Effects of Productivity" *American Sociological Review* 58: 703-722.

February 18 Cox Proportional Hazards Models I: Theory and Estimation

- Paper Topics are due

Required Reading:

H&L—Chapter 3

February 25 Cox Proportional Hazards Models II: Interpretation

- Assignment 2 Due

Required Reading:

H&L—Chapter 4

Recommended Readings:

Teachman, Jay D., and Mark D. Hayward. 1993. "Interpreting Hazard Rate Models." *Sociological Methods and Research* 21:340-371.

Substantive Application Readings:

Albonetti, Celesta A., and John R. Hepburn. 1997. "Probation Revocation: A Proportional Hazards Model of the Conditioning Effects of Social Disadvantage." *Social Problems* 44: 124-138.

L. Wu. 1996. "Effects of Family Instability, Income, and Income Instability on the Risk of a Premarital Birth." *American Sociological Review* 61: 386-406.

Bienen, Henry and Nicolas van de Walle. 1992. "A Proportional Hazard Model of Leadership Duration." *Journal of Politics* 54: 685-717.

Box-Steffensmeier, Janet M., Laura W. Arnold and Christopher J. W. Zorn. 1997. "The Strategic Timing of Position Taking in Congress: A Study of the North American Free Trade Agreement." *American Political Science Review* 91(June): 324-338.

Fagan, Jeffrey, and Martin Guggenheim. 1996. "Preventive Detention and the Judicial Prediction of Dangerousness for Juveniles: A Natural Experiment." *The Journal of Criminal Law & Criminology* 86: 415-448.

March 3 Cox Proportional Hazards Models III: General Extensions

Required Readings:

H&L—Chapter 5, Chapter 7

Eide, Geir Eide, Ernst Omenaas, and Amund Gulsvik. 1996. "The Semi-Proportional Hazards Model Revisited: Practical Reparameterizations." *Statistics in Medicine* 15: 1771-1777.

March 10 Spring Break

March 17 Cox Proportional Hazards Models IV: Model Fit Issues

- **Assignment 3 Due**

Required Reading:

H&L—Chapter 6

March 24 Parametric Models I

Required Reading:

H&L—Chapter 8

Substantive Application Reading:

Hepburn, John R. , and Celesta A. Albonetti. 1994. “Recidivism Among Drug Offenders: A Survival Analysis of the Effects of Offender Characteristics, Type of Offense, and Two Types of Intervention.” *Journal of Quantitative Criminology* 10: 159-179

March 31 Parametric Models II

- **Assignment 4 Due**

Required Reading:

Bennett, D. Scott. 1999. “Parametric Models, Duration Dependence, and Time-Varying Data Revisited.” *American Journal of Political Science* 43:256-270.

April 7 Advanced Topics I: Repeatable Events—Variance-Corrected Hazards Models

▪ **Assignment 5 Due**

Required Reading:

H&L—Chapter 9, Section 9.1, 9.2

Ezell, Michael E., Kenneth C. Land, and Lawrence E. Cohen. 2003. “Modeling Multiple Failure Time Data: A Survey of Variance-Corrected Proportional Hazards Models with Empirical Applications to Arrest Data.” *Sociological Methodology*.

Lindsey, J.K. 1998. “Counts and Times to Events.” *Statistics in Medicine* 17:1745-1751

Recommended Readings:

Allison, Paul D. 1996. “Fixed Effects Partial Likelihood for Repeated Events.” *Sociological Methods and Research* 25: 207-222.

Willett, John B. and Judith D. Singer. 1995. “It’s Déjà Vu All Over Again: Using Multiple-Spell Discrete-Time Survival Analysis.” *Journal of Educational and Behavioral Statistics* 20: 41-67

Substantive Application Reading:

Taylor, Donald H., Gerda G. Fillenbaum, and Michael E. Ezell. 2002. "The Accuracy of Medicare Claims Data in Identifying Alzheimer's Disease." *The Journal of Clinical Epidemiology* 55: 929-937.

April 14 Advanced Topics II: Frailty Models

Required Reading:

H&L—Chapter 9, Section 9.3

Aalen, Odd O. 1998. “Frailty Models.” Pp. 59-74 in *Statistical Analysis of Medical Data: New Developments*, edited by Brian S. Everitt and Graham Dunn. New York: Arnold.

Recommended Readings:

Land, Kenneth C., Daniel S. Nagin, and Patricia McCall. 2001. “Discrete-Time Hazard Regression Models with Hidden Heterogeneity: The Semiparametric Mixed Poisson Regression Approach.” *Sociological Methods and Research* 29: 342-373.

Substantive Application Reading:

Kessing, Lars Vedel, Elizabeth Wreford Olsen, and Per Kragh Andersen. 1999. “Recurrence in Affective Disorder: Analyses with Frailty Models.” *American Journal of Epidemiology* 149: 404-411.

April 21 Advanced Topics III: Competing Risks Situations

Required Reading:

Allison, Paul D. 1995. "Competing Risks." Chapter 6 in *Survival Analysis Using the SAS System* by Paul D. Allison. Cary, NC: SAS Institute.

Substantive Application Reading:

Diermeier, Daniel, and Randy T. Stevenson. 1999. "Cabinet Survival and Competing Risks." *American Journal of Political Science*.

Recommended Readings:

Hachen, David S. 1988. "The Competing Risks Model: A Method for Analyzing Processes with Multiple Types of Events." *Sociological Methods and Research* 17:21-54.

Tai, Bee-Choo, David Machin, Ian White, and Val Gebski. 2001. "Competing Risks Analysis of Patients with Osteosarcoma: A Comparison of Four Different Approaches." *Statistics in Medicine* 20:661-684.

(Note: The Tai reading is a more advanced survey of competing risks approaches)