

SC706: LONGITUDINAL DATA ANALYSIS

Instructor: Natasha Sarkisian
Email: natasha@sarkisian.net
Phone: (617) 755-3178
Office: McGuinn 417

Class time: Tuesdays 4-6:20 PM
Class location: O'Neill 245
Office hours: By appointment
Webpage: <http://www.sarkisian.net/sc706/>

COURSE DESCRIPTION*

This applied course is designed for graduate students with a prior background in statistics at the level of SC703: Multivariate Statistics (or its equivalent). This means that students should have considerable experience with ordinary-least-squares (OLS) regression: I assume you have an understanding of multiple OLS regression and an ability to conduct such analyses using some statistical software (e.g., SPSS, SAS, Stata, etc.). The major topics of the course will include event history analysis (also known as survival analysis) and panel data analysis (fixed and random effects, GEE models, mixed models, change models).

The goals of the course are to develop the skills necessary to identify an appropriate technique, estimate models, and interpret results for independent research and to critically evaluate contemporary social research using advanced quantitative methods. The course will be applied in the sense that we will focus on estimating models and interpreting the results, rather than on understanding in detail the mathematics behind the techniques. I hope that the course will provide you with a solid foundation in longitudinal data analysis, which is a type of advanced quantitative skill that is in high demand in many fields, both in and out of academia. For those of you in the Sociology Department, the course can also provide a foundation for the "Advanced Quantitative Methods" area examination.

We will be using Stata for all the analyses throughout the course. No previous Stata experience is necessary: I will provide an introduction to Stata in the beginning of the course and guide you throughout the course. We will use Stata on Citrix: see <http://apps.bc.edu> to install Citrix on your computer.

COURSE POLICIES

For each topic in the course, I will give a lecture focusing on the reasoning behind the technique, and provide a review of the syntax used to do analyses and the output generated by Stata. Throughout that process, you will get a chance to practice conducting the analyses and interpreting the results. We will also discuss and critically evaluate published research based on the various techniques. Make sure that you carefully read these examples of published research before class and be prepared to discuss them. The course is based on an interactive relationship between the instructor and students, as well as on collaboration among the students. You are strongly encouraged to ask questions and discuss the material in class. I also encourage collaboration among the students. Please feel free to help each other when running analyses for assignments. However, everyone must turn in their own report and statistical output.

I also would like to stress that you are always welcome to come and see me with any additional questions. Email is the best way to get in touch with me – I check my email very often. Email is the best way to get a quick question answered or to set up an appointment to discuss something at length. You are also welcome to call me either in my office or at home (any time between 9 AM and 10 PM); however, be prepared to leave your name and number if I am not available to pick up the phone. Also, please check our course website regularly: various course materials (assignments, handouts, etc.) will be posted there regularly. And make sure to check your email, too – from time to time I may send some announcements.

* This syllabus draws upon ideas presented in syllabi by a number of people, including Robert Kunovich, John Williamson, Joya Misra, and Doug Anderton.

Finally, a note on feedback. I would like to know how I could make this course experience as useful and interesting as possible. Therefore, every class in the end of class I will ask you to submit a sheet of paper with the date and at least one sentence of reaction to that class meeting, indicating what you learned, or something you liked or did not like, found interesting or controversial, found clear or too simplistic, or found confusing and in need of further (or better) explanation. You may also submit comments on the course in general.

REQUIRED MATERIALS

The following book will be available for purchase at the BC bookstore. It will be also placed on reserve at the library: Cleves, Mario Alberti, William Gould, and Roberto G. Gutierrez. 2004. *An Introduction to Survival Analysis Using Stata. Revised Edition*. College Station, TX: Stata Corporation. Other required readings will be available on electronic reserve in the library: see <http://www.bc.edu/reserves>

COURSE REQUIREMENTS AND GRADING

There will be two assignments in this course, each worth 40% of your grade. These assignments will involve selecting a research question and variables, running analyses, conducting diagnostics and applying remedies, and interpreting the results. You will create a running log to document all the steps and then edit it, keeping only the relevant steps, adding explanations, and interpreting the results.

In addition, for one of these assignments, you will write up the results like you would for a journal publication. This will include an Introduction that will provide a short substantive description of your theoretical argument, your research questions, and hypotheses (1 page max.). Second, your write-up will include a brief Data and Methods section (1-2 pages) describing the variables and the methodology. You will include any discussion of diagnostics and modifications in this section. Also, here you will include a table with descriptive statistics for your sample. Third, you will provide a 3-4 page description of the results including tables (in journal format) and graphs assisting in the interpretation of results. Finally, you will include a brief conclusion summarizing your findings (1 page max.). This journal-style write-up will be worth 20% of your grade. For all of these assignments, I will provide data, although you can use your own data if they are appropriate for the technique (see me in advance).

Each assignment will consist of two drafts, to be submitted electronically. If you turn in the first draft by the due date, I will comment on it, assign a temporary grade, and return it to you. At that point, we will also discuss the common problems and mistakes. You will then get a chance to submit a revised draft. If you are satisfied with your temporary grade, you do not need to revise the assignment – just let me know. This system will allow you to push yourself beyond your comfort level without worrying how it will affect your grade. For example, you might try to interpret some of the results not required for the assignment or you might decide to present the results in a more meaningful and perhaps less conventional way. The letter grades for the assignments will be determined as follows:

93-100	A
90-92	A-
87-89	B+
83-86	B
80-82	B-
60-79	C
0-59	F

COURSE OUTLINE

September 8. Introduction to Longitudinal Data Analysis and Stata.

Menard, Scott. 2002. Chapters 2, 5 from: *Longitudinal Research*. Beverly Hills, CA: Sage Publications.
RESERVE

September 15. Event History Analysis: Introduction to Discrete Time Models.

Willett, John B., and Judith D. Singer. 2004. Discrete-Time Survival Analysis. Chapter 11 from: David Kaplan (Ed.), *The Sage Handbook of Quantitative Methodology for the Social Sciences*. Thousand Oaks, CA: Sage Publications. RESERVE

September 22. Event History Analysis: Diagnostics for Discrete Time Models.

Singer, Judith D., and John B. Willett. 2003. Extending the Discrete-Time Hazard Model. Chapter 12, pp. 443-467, from: *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford University Press. RESERVE.

Gupta, Sanjiv, Pamela J. Smock, and Wendy D. Manning. 2004. Moving Out: Transition to Nonresidence among Resident Fathers in the United States, 1968-1997. *Journal of Marriage and Family*, 66, August, 627-638. RESERVE

September 29. Event History Analysis: Nonparametric Models.

Taris, Toon W. 2000. Chapter 6 from: *A Primer in Longitudinal Data Analysis*. Thousand Oaks, CA: Sage Publications. RESERVE

Cleves, Mario Alberti, William Gould, and Roberto G. Gutierrez. 2004. Chapter 4-8 from: *An Introduction to Survival Analysis Using Stata*. College Station, TX: Stata Corporation.

October 6. Event History Analysis: Semi-Parametric Models.

Cleves, Mario Alberti, William Gould, and Roberto G. Gutierrez. 2004. Chapters 9-11 from: *An Introduction to Survival Analysis Using Stata*. College Station, TX: Stata Corporation.

Teachman, Jay D., and Mark D. Hayward. 1993. Interpreting Hazard Rate Models. *Sociological Methods and Research*, 21, 3, 340-371. RESERVE

Hill, Twyla J. 2000. Legally Extending the Family: An Event History Analysis of Grandparent Visitation Rights Laws. *Journal of Family Issues*, 21, 246-261. RESERVE

October 13. Event History Analysis: Parametric Models.

Cleves, Mario Alberti, William Gould, and Roberto G. Gutierrez. 2004. Chapters 12-15 from: *An Introduction to Survival Analysis Using Stata*. College Station, TX: Stata Corporation.

Palloni, Alberto, and Elizabeth Arias. 2004. Paradox Lost: Explaining the Hispanic Adult Mortality Advantage. *Demography*, 41, 385-415. RESERVE

October 20. Event History Analysis: Competing Risk and Repeated Events Models.

Allison, Paul D. 1985. Chapters 5 and 6 from: *Event History Analysis: Regression for Longitudinal Event Data*. Beverly Hills, CA: Sage Publications. RESERVE

Cleves, Mario Alberti, William Gould, and Roberto G. Gutierrez. 2004. Chapters 4-7 from: *An Introduction to Survival Analysis Using Stata*. College Station, TX: Stata Corporation.

Metraux, Stephen, and Dennis P. Culhane. 1999. Family Dynamics, Housing and Recurring Homelessness Among Women in New York City Homeless Shelters. *Journal of Family Issues*, 20(3): 371-396. RESERVE

De Graaf, Paul M. and Matthijs Kalmijn. 2003. Alternative Routes in the Remarriage Market: Competing-Risk Analyses of Union Formation after Divorce. *Social Forces*, 81(4):1459-1498. RESERVE

October 27. Panel Data Analysis: Fixed Effects Models.

Assignment 1 first draft due

Worrall, John L. 2008. An Introduction to Pooling Cross-Sectional and Time Series Data. Chapter 15 from *Handbook of Longitudinal Research: Design, Measurement, and Analysis* edited by Scott Menard. RESERVE

Allison, Paul D. 2009. Chapter 2 and Appendix 1 (portion related to Ch. 2), from: *Fixed Effects Regression Models*. Sage Publications.

Baum, Christopher. 2006. Chapter 9, pp.219-226, from: *An Introduction to Modern Econometrics Using Stata*. College Station, TX: Stata Press. RESERVE

November 3. Panel Data Analysis: Random Effects Models.

Rabe-Hesketh, Sophia, and Anders Skrondal. 2005. Chapter 2 from: *Multilevel and Longitudinal Modeling Using Stata*. College Station, TX: Stata Press. RESERVE

Baum, Christopher. 2006. Chapter 9, pp.226-232, from: *An Introduction to Modern Econometrics Using Stata*. College Station, TX: Stata Press. RESERVE

November 10. Panel Data Analysis: GEE models.

Twisk, Jos W. R. 2003. Chapters 4, 5 from: *Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide*. Cambridge University Press. RESERVE

Waldfoegel, Jane. 1997. The Effect of Children on Women's Wages. *American Sociological Review*, 62, 209-217. RESERVE

November 17. Panel Data Analysis: Mixed Effects Models.

Rabe-Hesketh, Sophia, and Anders Skrondal. 2005. Chapter 3 from: *Multilevel and Longitudinal Modeling Using Stata*. College Station, TX: Stata Press. RESERVE

Hedecker, Donald. 2004. An Introduction to Growth Modeling. Chapter 12 from: David Kaplan (Ed.), *The Sage Handbook of Quantitative Methodology for the Social Sciences*. Thousand Oaks, CA: Sage Publications. RESERVE

Farkas, George, and Kurt Beron. 2004. The Detailed Age Trajectory of Oral Vocabulary Knowledge: Differences By Class And Race. *Social Science Research*, 33, 464-497. RESERVE

November 24. Panel Data Analysis: Dynamic Models.

Assignment 2 first draft due

Halaby, Charles N. 2004. Panel Models in Sociological Research: Theory into Practice. *Annual Review of Sociology*, 30, 507-544. RESERVE

Baum, Christopher. 2006. Chapter 9, pp.232-236, from: *An Introduction to Modern Econometrics Using Stata*. College Station, TX: Stata Press. RESERVE

Finkel, Steven E. 2008. Linear Panel Analysis. Chapter 29 from *Handbook of Longitudinal Research: Design, Measurement, and Analysis* edited by Scott Menard. RESERVE

December 1. Panel Data Analysis: Change Models.

Taris, Toon W. 2000. Chapters 4 in: *A Primer in Longitudinal Data Analysis*. Thousand Oaks, CA: Sage Publications. RESERVE

Johnson, David. 2005. "Two-Wave Panel Analysis: Comparing Statistical Methods for Studying the Effects of Transitions." *Journal of Marriage and Family* 67(4):1061-75. RESERVE

December 8. Missing Data in Longitudinal Research.

Final drafts of Assignments 1 and 2 due December 15

Twisk, Jos, and Wieke de Vente. 2002. Attrition in Longitudinal Studies: How to Deal with Missing Data. *Journal of Clinical Epidemiology* 55:329-37. RESERVE