

SOC386L/SSC385: Longitudinal Data Analysis

Unique # 46530 59255

Spring 2010

MW 3:30–5:30 PM, RAS 313A

Instructor: Dan Powers, *e-mail* dpowers@mail.la.utexas.edu

Office Hours: MW 1–3 502 Burdine (232–6335), and by appointment.

Course Description

This is a course in statistical methods for longitudinal data analysis. We will cover two main areas: multiple regression models for data collected on the same subjects over time (repeated measures/panel data), and methods for modeling event occurrences over time.

Topics

Multilevel/Hierarchical Models for Change. The first half of the course introduces multilevel models for change (i.e., growth curve models), which are appropriate for the analysis of change in a continuous dependent variable over time. No familiarity with this subject is assumed or required; the major aspects of estimating and interpreting multilevel/hierarchical models will be reviewed in the context of growth curve modeling. We will also review latent growth models from the structural equation modeling (SEM) perspective. Growth curve models for categorical outcomes (counts) and nonlinear growth curve models will also be discussed.

Event History Analysis. The second half of this course deals with event history analysis (i.e., survival analysis, hazard rate models, etc.), which is a technique for modeling the transition from one status (or state) to another. Examples include life course transitions like marriage, birth, divorce, entry and exit from the labor force, etc. We will focus on discrete time and continuous time models that make few assumptions regarding time dependence of the hazard, such as the piecewise constant exponential and the Cox proportional hazard model.) We will focus mainly on single transition models. Multi-state and competing risks methods will be discussed in somewhat less detail toward the end of the course. We will also consider frailty and shared frailty models as well as discrete and continuous-time multilevel hazard models.

Background.

Students in this course should have prior exposure to statistical methods at the graduate level. This normally includes a course in basic statistics and a course in linear regression, or a course that combines the two. It should be noted that this class serves a wide range of students; it is an elective course for students in the MS in statistics program, as well as for students in the social sciences. Some students will have taken courses in mathematical statistics and will have an understanding of matrix algebra and calculus. Other students may have more familiarity with the advanced statistical models in their substantive research area, but might lack the formal statistical/mathematical training. Given this heterogeneity in background, the course material presented here aims to be useful for students at *all* levels. Those without formal training can acquire a working understanding of the necessary concepts using the resources listed on the syllabus. Most of the material in the exercises and handouts is based on *applied*, as opposed to theoretical (i.e., mathematical statistical) problems. In keeping with the applied nature of this course, we will provide examples drawn mainly from sociological and demographic research.

Readings

The readings be selected from required text. Handouts containing supplementary reading material will be posted to Blackboard (Bb).

- Required Text:
 - J. D. Singer, and J. B. Willett (2003) *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press.
- Recommended Textbooks on Specific Subjects:
 - H. Blossfeld and G. Rohwer (1995) *Techniques of Event History Modeling*. Rahwah, NJ: Erlbaum.
 - K. Bollen (1989) *Structural Equations with Latent Variables*, New York: Wiley.
 - A. S. Bryk and S. W. Raudenbush (2002) *Hierarchical Linear Models: Applications and Data Analysis Methods (2nd Edition)*, Thousand Oaks: Sage.
 - D. R. Cox and D. Oakes (1984) *Analysis of Survival Data*. London: Chapman Hall.
 - Gelman A., J.B. Carlin, H.S. Stern, D.B. Rubin (2004) *Bayesian Data Analysis*, 2nd Edition. New York: Chapman and Hall.
 - Gelman A., and J. Hill (2005) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge.
 - J. Kalbfleisch and R. L. Prentice (1980) *The Statistical Analysis of Failure Time Data*. New York: Wiley.
 - I. G. G. Kreft and J. de Leeuw (1998) *Introducing Multilevel Modeling*, Thousand Oaks: Sage.
 - J. S. Long (1997) *Regression Models for Categorical and Limited Dependent Variables*, Beverly Hills: Sage.
 - K. Namboodiri and C. M. Suchindran (1987) *Life Table Techniques and Their Applications*. New York: Wiley.
 - D. Powers and Y. Xie (2008) *Statistical Methods for Categorical Data Analysis*. 2nd Edition. London: Emerald.
 - N. B. Tuma and M. Hannan (1984) *Social Dynamics: models and methods*. London: Academic Press.

Method of Evaluation

This is an applied course. We will learn by applying the techniques learned in class to specific pedagogical examples. No previous programming experience is required. I am available to answer programming and statistics questions via email. We will use several statistical packages in this course depending on the problem. These include Stata, SAS, and R. When possible, example programs will be provided in each package. Stata is the most user friendly of the three, and can fit most of the models discussed in this course. Stata is available in the Liberal Arts Computer Instruction Laboratory (LACIL) in room 120 Burdine. A valid UTEID is required to log in. R is public domain statistical programming environment. R, as well as useful extension *packages*, can be obtained free by following the links at www.r-project.org. Students are encouraged to acquire a working knowledge of at least one of the statistical packages mentioned above early on in the course. Data and programming examples will posted to Bb.

Assignments. There will be 5 assignments covering the course material. The assignments will contain short problems and questions on conceptual issues and will account for 80% of the grade. These must be turned in on or before the dates indicated. Homework will be based on data sets from the readings and other sources that I will make available to you. A final assignment worth 20% of the grade will be due on the final exam date. *NOTE: Students in the MS in mathematical statistics program will be expected to complete the extra-credit problems to satisfy the accreditation requirements of this program.*

Topics and Readings

All readings will be from Singer and Willett. Supplementary reading material will be posted to Bb.

1. Introduction to Longitudinal Data (week 1). Ch. 1. Topics: frameworks for investigating change over time, distinguishing between types of questions about change
2. An Introduction to Longitudinal Data on Change (week 2). Ch. 2-3. Topics: working with longitudinal data, individual trajectories, improving OLS models.
3. Multilevel Models for Change (week 3). Ch. 4. Topics: individual change, interindividual differences in change, multilevel (hierarchical) models, fixed and random effects.
4. Treatments of Time, Time-varying Covariates, and Centering (week 4). Ch. 5. Topics: handling various aspects of time.
5. Nonlinearity and Discontinuous Change (week 5). Ch. 6. Topics: spline functions, transformations, nonlinear growth.
6. Error Covariance Structures of Multilevel Models (week 6). Ch. 7. understanding various components of error in multilevel models.
7. Structural Equation Models and Latent Growth Curve Models (week 7). Ch. 8. Topics: latent variables, covariance structure models, growth curve models as structural equation models, parallel change/growth processes.
8. Events and Event Occurrence Data (week 8). Ch. 9-10. Topics: introduction to the nature of event occurrence (event history) data, censoring, describing event data using life table techniques.
9. Discrete-Time Hazard Models and Extensions (week 9). Ch. 11-12. Topics: life-tables with covariates, discrete-time models, split episode data, time-varying covariates, time-varying effects, model testing/evaluation, multi-state and competing risk models.
10. Continuous-Time Hazard Models (week 10). Ch. 13. Topics: distributional aspects of time (T), survivor function, hazard function, the role of exposure, grouped data methods, life table and nonparametric techniques for describing continuous time data, the exponential model.
11. Piecewise Constant Hazard Rate Models (week 11). (Handouts) Topics: extending the exponential model using grouped and individual-level data, proportional hazards models, proportional hazards models with nonproportional effects.
12. The Cox Proportional Hazards Model (week 12). Ch. 14. Topics: development of the Cox PH model, partial likelihood estimation, fitting the Cox model, consequences of Cox's method, inference about effects and interpretation of results.
13. Extending the Cox Model (week 13) Ch. 16. Topics: time-varying covariates, assessing the proportionality assumption with statistical tests and residual plots, accounting for nonproportionality with stratification or via "interactions" with time, model diagnostics and tests.
14. Unobserved Heterogeneity in Hazard Rate Models (week 14-15). Topics: individual-level and shared frailty models, multilevel models for event occurrence, incorporating additional sample information using clustered data (pair data, sibling data, and other contexts).