Longitudinal Data Analysis

Michael L. Berbaum
Institute for Health Research and Policy
University of Illinois at Chicago

Course description:

Longitudinal analysis is the study of short series of observations obtained from many respondents over time and is also referred to as panel analysis (of a cross-section of time series) or repeated measures or growth curve analysis (polynomials in time) or multilevel analysis (where one level is a sequence of observations from respondents). Longitudinal analysis is used for panel surveys, experiments and quasi-experiments in health and biomedicine, education and psychology, and the evaluation of prevention and treatment programs. This course treats the statistical basis and practical application of linear models for longitudinal normal data and generalized linear models for longitudinal binary, count, and ordinal data. The approach involves inclusion of random effects in linear models to reflect within-person cross-time correlation. Techniques for irregularly observed (unequally spaced) data will be covered. Other ICPSR courses focus on time series and structural equations approaches, including latent growth curve models, which are only briefly discussed in this course. The technical level will be at Track II, with interludes at Track III (matrix algebra, probability distributions). Examples and exercises will use both standard and special-purpose software. Participants should have a good understanding of linear regression or analysis of variance.

Required texts:


Recommended texts:


**Schedule of topics (approximate):**

**Day 1**
Introduction, course scope and organization, examples of longitudinal data, marginal and random-effects models for longitudinal data, models for longitudinal categorical data (binary, count, and ordinal responses), the “diamond of models” to be addressed. Read FLW, Ch. 1, “Longitudinal and Clustered Data,” and Ch. 2, “Longitudinal Data: Basic Concepts.”

**Day 2**
Linear Model (LM) and Linear Mixed Model (LMM). Read FLW, Ch. 3, “Overview of Linear Models for Longitudinal Data,” and scan Ch. 8, “Linear Mixed Effects Models.”

**Day 3**
Continue discussion of LM and LMM models; the SAS MIXED procedure. FLW is brief about MIXED: Sections 5.9, 6.6, and 7.8. For more detail read V&M, Ch. 8, “Fitting Linear Mixed Models with SAS.” Chapters 1-5 of Littell cover the mixed model approach to common experimental designs, with examples in SAS.

**Day 4**
Maximum likelihood (ML) estimation and restricted maximum likelihood (REML). Read FLW, Ch. 4, “Estimation and Statistical Inference.”

**Day 5**
Modelling the mean. Read FLW, Ch. 5, “Analyzing Response Profiles,” and Ch. 6, “Parametric Curves.”

**Day 6**
Continue discussion of modelling the mean. For another account of this subject read V&M, Ch. 6, “Inference for the Marginal Model.”

**Day 7**
Modelling the covariance. Read FLW, Ch. 7, “Modelling the Covariance.”

**Day 8**
LMM. Read FLW, Ch. 8, “Linear Mixed Effects Models.”

**Day 9**
Continue LMM.

**Day 10**
Model diagnostics. Read FLW, Ch. 9, “Residual Analysis and Diagnostics.” (For further discussion of diagnostics, see John Fox’s Sage monograph, *Regression diagnostics.*)

**Day 11**
Missing data. Read FLW, Ch. 14, “Missing Data and Dropout.” (V&M devote Chapters 14-21 to this important topic. See Ch. 14, “Exploring Incomplete Data,” Ch. 15, “Joint Modeling of Measurements and Missingness,” and Ch. 16, “Simple Missing Data Methods.”)
Day 12  Continue discussion of missing data.

Day 13  Power analysis for random effects models. Read FLW, Section 15.2, “Sample Size and Power,” and handouts provided in class. See also V&M, Ch. 23, “Design Considerations,”

Day 14  Generalized linear models (GLM) framework. Read Gill. Read FLW, Ch. 10, “Review of Generalized Linear Models”

Day 15  Generalized linear mixed model (GLMM). Read FLW, Ch. 12, “Generalized Linear Mixed Effects Models.”


Day 17  Continue discussion of GLMM; examples of binary and count responses. Overdispersion. (Receive handouts for ordinal responses.)

Day 18  Mixed models for ordinal responses. Handouts on the SAS NLMIXED procedure and Hedeker’s MIXOR program.)

Day 19  Multilevel or hierarchical models. Read FLW, Ch. 17, “Multilevel Models.”

**Supplemental readings:**

The texts are thoroughly referenced. In addition, additional references will be provided as class handouts, including a bibliography on experimental design and analysis of variance, and another on GLM and GLMM.

**Software:**

The primary software package for this course is SAS. Some examples will employ the lme/nlme libraries in the S family of packages (S/S-Plus/R). Stata has comparable capabilities for certain problems.