Graduate Group in Epidemiology & Biostatistics

BSTA 656: Longitudinal Data Analysis
Fall 2005 Syllabus

1.0 course unit; Prerequisites: BSTA6210, BSTA6300, BSTA6510

Lecturers

- Rui Feng, Ph.D., Blockley 209, ruifeng@upenn.edu
- Jing Huang, Ph.D., Blockley 625, jing14@upenn.edu

- Office Hours: By appointment

Teaching Assistant

- Binghao Yan, Binghao.Yan@Pennmedicine.upenn.edu

- Office Hours: TBD

Class Time and Place

- Tuesdays and Thursdays, 1:45–3:15PM
- Anat-Chem 202

Textbooks

- **Required:** (DHLZ) Diggle PJ, Heagerty P, Liang K-Y, Zeger SL. *Analysis of Longitudinal Data, 2nd Edition*. Oxford University Press, 2002.
- **Recommended:** (VM) Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. Springer Series in Statistics, 2000.
- **Recommended:** (MV) Molenberghs G, Verbeke G. *Models for Discrete Longitudinal Data*. Springer Series in Statistics, 2006.

Computing

- Analyses will be conducted in `R` (`http://r-project.org`).
- Packages: `gee`, `geepack`, `nlme`, `lme4`.

Objectives

This course will cover extensions of linear and generalized linear models to the analysis of longitudinal data, with an emphasis on parametric and semi-parametric regression-based estimation methods. At the conclusion of the course, students should:

- *Methodology:* Understand the theoretical framework for longitudinal data analysis and the statistical properties of longitudinal data analysis methods, including mixed-effects models and estimating equations;
- *Applications:* Apply appropriate techniques to summarize and generate inference from longitudinal data, verify the validity of the approach, and correctly interpret the results;
- *Research:* Be familiar with current research topics in longitudinal data methodology and applications.

Topics

1. Introductory topics (DHLZ 1, 2, 3, 4)
    - Design considerations
    - Exploring longitudinal data using graphics and summaries
    - General linear models for longitudinal data: Two-stage least squares, weighted least squares, restricted maximum likelihood, robust standard error estimation
2. Linear mixed-effects models (VM)
    - Estimation, inference, prediction, and model diagnostics
    - Extension to non-linear mixed-effects models
3. Marginal models (DHLZ 7, 8)
    - Quasi-likelihood
    - Estimating equations for continuous outcomes
    - Generalized estimating equations for discrete (count, binary) outcomes
4. Generalized linear mixed-effects models (DHLZ 7, 9; MV)
    - Linear mixed-effects models for continuous outcomes
    - Generalized linear mixed-effects models for discrete outcomes
    - Bayesian implementation
5. Advanced topics (DHLZ 11, 12, 13)
    - Time-dependent exposures, recurrent marked point process data
    - Missing data: Semi-parametric and likelihood-based methods
    - Transition models and marginalized models
    - Machine learning models

Guest Seminar Series

Guest seminars on active research problems will be integrated. Material may be included in exams.

Evaluation

- 40% Coursework: 8 Weekly Assignments
- 20% Take-Home Midterm Exam
- 25% Take-Home Final Exam
- 15% Participation

Policies

- Regular attendance is expected. Missing lectures may affect your understanding of the material.
- Late homework or exam submissions without prior approval will not be accepted.
- The use of artificial intelligence (AI) or other automated writing tools is limited to correcting grammar and typos; it is not allowed to solve problems or write content for assignment or exam questions.

Tentative Course Outline

Date	Topic	Coursework Assigned	Coursework Due
Section 1	*Introductory topics* (taught by RF)		
Aug 26	Introduction and Basics of R		
Aug 28	Design Considerations	HW1	
Sep 2	Exploratory Data Analysis		
Sep 4	General Linear Models for Longitudinal Data	HW2	HW1
Sep 9	Weighted Least Squares		
Sep 11	Robust Standard Error Estimation, Restricted Maximum Likelihood		
Section 2	*Linear mixed-effects models* (taught by RF)		
Sep 16	Motivation and Applications of Linear Mixed-Effects Models	HW3	HW2
Sep 18	Piecewise Regression		
Sep 23	Estimation and Inference	HW4	HW3
Sep 25	Prediction and Model Diagnostics		
Sep 30	Nonlinear Mixed-Effects Models		HW4
Oct 2	Latent Class Mixed-Effects Models	Midterm	
Section 3	*Marginal models* (taught by JH)		
Oct 7	Quasi-Likelihood and Estimating Equations		Midterm
Oct 9	**Fall Break - No Class**		
Oct 14	GEE 1.0		
Oct 16	GEE Extensions and GEE 2.0	HW5	
Oct 21	Guest Lecture: Quasi-Least Squares		
Section 4	*Generalized linear mixed-effects models* (taught by JH)		
Oct 23	Mixed-Effects Models: Conditional and Maximum Likelihood		
Oct 28	Count Outcomes and GLMM	HW6	HW5
Oct 30	Binary Outcomes and GLMM		
Nov 4	Bayesian Implementation	HW7	HW6
Section 5	*Advanced topics* (taught by JH)		
Nov 6	Time-Dependent Exposures		
Nov 11	Missing Data: Semi-Parametric and Likelihood-Based Methods	HW8	HW7
Nov 13	Transition Models		
Nov 18	Guest Lecture: Confounding in Longitudinal and Clustered Data Analysis		HW8
Nov 20	Marginalized Models		
Nov 25	Machine Learning Models (by RF)		
Nov 27	**Thanksgiving - No Class**		
Dec 2	Review and Summary	Final	
Dec 7	Final Examination		Final