BST / BIST 245: Analysis of Multivariate and Longitudinal Data

Harvard T.H. Chan School of Public Health
Fall 2024

Instructor:	Tom Chen, PhD
TA:	Lee Ding
Lectures:	TuTh 11:30am - 01:00pm in FXB G11
Lab:	F 10:00 am - 11:00 am in Building 2-428
Office Hours:	Instructor: (Physical) Right after class (Zoom) M 10:00am - 11:00am TA: TBD
E-mail:	Instructor: `tchen@hsph.harvard.edu` TA: `lee_ding@g.harvard.edu`
Links:	Canvas, Piazza, Zoom, Google Forms

Course description:

BST 245 is a high-level graduate course in the analysis of dependent multivariate and longitudinal data. In particular, the course will extend linear and generalized linear models to the analysis of dependent data. Emphasis will be placed on parametric and semi-parametric regression-based estimation methods. At the conclusion of the course, students should: (i) understand the theoretical framework for dependent data analysis and the statistical properties of analysis methods; (ii) be able to apply appropriate techniques to summarize and generate inference from dependent data, as well as verify the validity of the approach and correctly interpret the results; and, (iii) be familiar with current research topics in dependent data methodology and applications. Within this framework, specific topics that will be covered include:

- Linear mixed models
- General linear models
- Generalized estimating equations
- Generalized linear mixed models
- Special topics
  - Missing data
  - Survival analysis
  - Time-varying covariates

Prerequisites:

This course is designed for doctoral students in biostatistics. Prerequisites include successful completion of BST 231 (Inference I) and BST 232 (Methods). Students are expected to have a strong understanding of the theory of linear models, generalized linear models, and quasi-likelihood, as well as proficiency in multivariable calculus and linear algebra. If you are not a doctoral student in the Department of Biostatistics, you must obtain permission to enroll in the course.

Evaluation:

Evaluation and the course grade will be based on four components:

Homework	40%
Exam	30%
Final project	30%
Extra credit (see below)	?% - ??%

Homework will be assigned (and due) approximately every 2.5 weeks. All homeworks should be submitted electronically by 11:59pm on the due date. All students are given a freebie of 3 days for late homeworks, no questions asked, allocated however you want (e.g. 1 day for Homework 1, 2 days for Homework 3). You need not (and should not) contact the TA nor instructor for additional time unless you have used up all the freebie days and have an emergency to justify additional days. Any remaining freebie days at the end of the semester is an additional 0.5% extra credit on your total grade (for at most 1.5%). Barring aside emergency circumstances, no late homeworks are accepted.

The exam will be a week-long take home exam which should be worked on individually; exam will also include extra credit problems. Freebie days cannot be used here. The final project will consist of a 4-5 week project in which students will pick a topic in multivariate/longitudinal data that is not covered in class, write up a report ($\sim$ 10 pages) and deliver a 20-minute presentation.

A list of potential topics will be provided, although students will have the option to work on a topic that is not on the list if they get prior approval.

Notes and textbooks:

Electronic copies of course handouts, slides, notes, homework assignments, and datasets will be made available on the course website. While there are no required textbooks, the most closely aligned supplementary text for this course is:

- Diggle P, Heagerty P, Liang KY, Zeger S. *Analysis of Longitudinal Data*. 2nd Edition. Oxford University Press, 2002.

Additional texts that cover topics relevant to this course include:

- Finch WH, Bolin JE, Kelley K. *Multilevel Modeling Using R*. 3rd Edition. Chapman

and Hall/CRC, 2024.

- Muller KE, Stewart PW. *Linear Model Theory: Univariate, Multivariate, and Mixed Models*. John Wiley & Sons, 2006.
- Searle SR, Casella G, McCulloch CE. *Variance Components*. Wiley Series in Probability and Statistics, 1992.

Important dates:

- Nov 26, 28: No classes (Day before Thanksgiving break and Thanksgiving)
- Dec 19 (Thursday): Last day of class