BIOS 767: Longitudinal Data Analysis

Spring 2016

Instructor: Bahjat Qaqish

Office: 3105-B McGavran-Greenberg

e-mail: qaqish@bios.unc.edu

office hours: Monday 11-12 and by appointment

Graders:

TBA

e-mail: @unc.edu

Texts:

- Primary Resources
  - Fitzmaurice, Laird, and Ware, Applied Longitudinal Analysis, 2nd edition
  - Relevant papers as assigned
- Supplemental Resources
  - Diggle, Heagerty, Liang, and Zeger, Analysis of Longitudinal Data
  - Gelman and Hill, Data Analysis Using Regression and Multilevel/Hierarchical Models

Software:

- SAS and R (other software allowed but not supported): Sample code will be given. Students are expected to read the software documentation to figure out how to use the various functions and options available.

Location and Time:

Class: 9:05-10:50 Monday and Wednesday, January 11 to April 27, 1301 McGavran-Greenberg Hall

BIOS767 presents modern approaches to the analysis of longitudinal data. Topics include: linear mixed effects models, generalized linear models for correlated data (including generalized estimating equations), computational issues and methods for fitting models, missing data, use of simulation, study design.

Prerequisite is BIOS 762. Knowledge of SAS and/or R software is assumed.

Learning Objectives:

- Gain strong understanding of theory (ideas) of longitudinal data analysis
- Hone skills in fitting and interpreting longitudinal data models for addressing scientific questions that arise in public health and medicine
- Gain competency in study design and power analysis of longitudinal data to complete skills needed to write statistical section of a grant proposal to collect and analyze longitudinal data, or to collaborate on papers that involve analysis of longitudinal data

Copies of presentations used in class will be available online on Sakai. These notes cover some material not contained in the texts and do not cover all material in the texts, so that the assigned readings are very important. You are encouraged to print the course notes and bring them with you to class.

Course note credits: I have drawn heavily on the work of professors and colleagues when creating the course notes, and in particular the notes have evolved from the work of Professors Davidian, Fitzmaurice, Laird, Ware, and Williams while at Harvard; and of Edwards, Helms, Muller, Stewart and Herring at UNC.

Assignments:

Course work will involve homework assignments (HW). Assignments will use problems from the textbook or from other sources, as well as problems developed by the instructor.

For homeworks, students are free to discuss general strategies, but the work they turn in is expected to be their own (i.e., direct copying of computer code or wording of interpretations is not allowed).

For all exams, students are not allowed to discuss the questions with anyone other than the instructor.

Grades

Homeworks, about 5-7% each

Midterm Exam: February 22 (tentative), 25%

Final Exam: 8 am, Thursday, May 5, 25%

The Graduate School uses the grades H (clear excellence), P (entirely satisfactory), L (low pass), and F (failure). Graduate students are expected to earn "P" grades, with remarkable performances rewarded with other grades from the scale as appropriate. Class participation may modify the association between the numeric average and assigned letter grade.

Honor Code:

Students in BIOS767 are expected to abide by the UNC Honor Code. All suspected Honor Code violations will be reported to the UNC Dean of Students, who will investigate the case. These investigations typically involve lengthy hearings of the Honor Court, and as outlined in the Instrument of Student Judicial Governance, "The usual sanction for a first academic violation is definite suspension for at least one academic semester and a grade penalty of an 'F' for the course, a portion of the course, or the assignment."

Course Content: Note: This course content is, to some extent, tentative. Topics may not be covered in a linear order as given. Some topics may be dropped and others added depending on time constraints, student interest and other factors (e.g. adverse weather)

1. Research ethics and reproducibility
2. The BIG picture first
3. What is a model?
4. General Linear Model for Longitudinal Data
5. ANOVA - univariate, multivariate
    - ML and REML estimation
    - EM algorithm
    - General linear mixed-effects model (Laird-Ware)
    - Inference for the random effects: BLUPs, Empirical Bayes, Bayes, Shrinkage
    - Model construction
    - Regression diagnostics
    - Hierarchical (random coefficient) models
    - Multivariate longitudinal data

- Relaxing parametric assumptions: generalized additive mixed model
- Design of randomized and observational longitudinal studies

6. Generalized linear models for longitudinal data
   - Marginal models for binary, ordinal, and count data
   - Random effects models for binary, ordinal, and count data
   - Transition models
   - Likelihood-based models for categorical data
   - GEE
   - Models for mixed discrete and continuous responses
7. Dropouts and missing data
   - Classification of missing data mechanisms
   - Intermittent missing values and dropouts
   - Weighted estimating equations
   - Modeling the dropout process (selection and pattern mixture models)
8. Time-dependent covariates
   - Dangers of time-dependent covariates
   - Lagged covariates
   - Causal inference
9. Joint models for data of different types
10. Joint models for means and correlations
11. Special topics