BIOS 6643
Longitudinal Data Analysis
Fall Semester 2014

Lecture info	M, W	2:30-3:45 PM	ED 2 South, room 2304 (On 8-25 and 9-3, in room 2306)

Instructor info

Instructor	Matt Strand
Office	NJH: BB building, room 222a, Colfax and Colorado, (303) 398-1862 AMC: Building 406, room 101, (303) 724-6861
e-mail	strandm@njhealth.org Note that I usually respond to e-mailed questions within a day.
Office hours	I will be on campus M and W. We can arrange hours to meet in the morning at my office in Building 406, or before or after class at the classroom (or in adjacent office). I can also arrange appointments at my office at NJH on other days.

Approach to teaching the course

You will be provided the set of lecture notes for the course. The textbooks mentioned below may be useful, but are considered optional and not required. You can decide individually whether you think these will be of help to you or not. In class, we will use sets of slides that roughly follow the lecture notes. I will post these to the course web site in advance of lectures. Please plan to review these slides before the lecture, and come prepared with questions. I would like to make the lecture time as interactive as possible.

A brief outline of the main topics for the course (see the calendar for more details)

(1) Introduction to longitudinal and clustered data
(2) Review of general linear models
(3) Linear mixed models
(4) Models for non-continuous outcomes
(5) Nonparametric (and semi-parametric) regression models for longitudinal data.
(6) Additional topics, as time allows.

Texts

BIOS 6643 course notes, Strand, 2013; will be sent to students by e-mail.

Some other references (asterisks on key texts):
**Longitudinal Data Analysis*, Hedeker and Gibbons, Wiley, 2006.
Hedeker's course web site: http://www.uic.edu/classes/bstt/bstt513.
Hedeker's program and data set web site: http://www.uic.edu/~hedeker/long.html.
**Linear Mixed Models for Longitudinal Data*; Verbeke and Molenberghs; Springer; 2000.
Analysis of Longitudinal Data; Diggle, Liang and Zeger; Oxford; 1996.
Statistical Methods for the Analysis of Repeated Measurements; Davis; Springer; 2002.
Modeling Longitudinal Data; Weiss; Springer; 2005.
Applied Longitudinal Analysis, 2nd ed.; Fitzmaurice, Laird and Ware; Wiley; 2011.
Longitudinal data analysis, Fitzmaurice, Davidian, Verbeke and Molenberghs, editors; 2009.
Linear / Matrix algebra: see GLM notes for references.

Web site for the course: http://dccweb.njhealth.org/vc/biom/bios6643/MainPage.cfm

Prerequisites: BIOS 6611, 6612 or consent of instructors. (BIOS6631, 6632 recommended.)

Computer programming

The course will mainly use SAS software. We will also use R on occasion or at least see how R differs from SAS for fitting longitudinal (or clustered) data models.

Course objectives

The main objective of the course is to learn about and be able to use longitudinal or clustered data models in appropriate situations. Primary objectives related to this main objective are listed below.

- To understand the theoretical foundations of general linear models and how they apply to models for longitudinal or clustered data.
- To have a basic understanding of classical methods of longitudinal (or clustered) data analysis.
- To understand and be able to use linear mixed models for modeling of longitudinal or clustered data.
- To become familiar with models for correlated non-normal outcomes.
- To recognize the differences between parametric and nonparametric models for longitudinal or clustered data.

Course competencies

- To understand basic matrix theory for general linear models (GLMs) and linear mixed models (LMMs).
- To know how to estimate effects and conduct hypothesis tests for functions of parameters in longitudinal data models.
- To understand the mechanics of how LMMs are fit.
- To be able to adequately specify a LMM for given longitudinal or clustered data and fit it using SAS or R software.
- To know general modeling approaches for different types of outcomes in longitudinal or clustered data.
- To comprehend the impact that experimental designs (longitudinal versus factorial) can have on estimates for fixed effects (i.e., beta parameters).
- To understand different approaches for modeling correlated binary or count outcomes, including generalized linear models (GLzM) with generalized estimating equations (GEE), and generalized linear mixed models (GLMM).
- To be able to read and comprehend recently published literature for longitudinal (or clustered) data methods, as well as write reports that involve analysis of longitudinal or clustered data.
- To have basic understanding of the following: nonparametric or semi-parametric models for longitudinal or clustered data; measurement error models and regression calibration.
- To be able to quantify and interpret effects that are unique to longitudinal models (e.g., between versus within-subject effects, cumulative versus multiple-lag effects).

Note: There could be some other competencies added to the list above. Hopefully this is a good start.

Integration of this Course with other Biostatistics courses:

This course was previously taught in a sequence of courses (BIOS7711, which lead into BIOS7712 and 13). Although it has been taken out of that sequence, longitudinal and clustered data methods may be examined in BIOS7712, 7713 or other higher level courses. In this course, GLM theory will first be reviewed, building on what students learned in BIOS6611 and 12.

Course Policies:

Academic Conduct Policy

All students are expected to abide the honor codes of the Colorado School of Public Health. Unless otherwise instructed, all of your work in this course should represent completely independent work. Students are expected to familiarize themselves with the Student Honor Code that can be found at (URL goes here). Any student found to have committed acts of misconduct (including, but not limited to cheating, plagiarism, misconduct of research, and breach of confidentiality) will be subject to the procedures outlined in the Honor Code.

Honor Code

All aspects of this course are conducted under the honor system. All students should have developed the qualities of honesty and integrity, and each student should apply these principles to his or her academic and subsequent professional career. All students are expected to have achieved a level of maturity, which is reflected in appropriate conduct at all times. Related to academic honesty, all work done on exams or other assignments is to be done independently, unless specific instruction to the contrary is provided by the course instructor.

Disability Policy

For students requesting accommodations, contact the Office of Disability Resources and Services. Their staff will assist in determining reasonable accommodations as well as coordinating the approved accommodations. Phone number: (303) 724-5640. Location: Building 500, Room W1103. The address is 13001 E. 17th Place.

Student evaluation

Grading

- Homework – 30%
- Quizzes – 2 at 15% each (75 minutes, in class)
- Data analysis project – 30%
- In class participation – 10% (to be discussed in class)

Homework

- There will be about 7 homeworks assigned.
- Detailed work is expected for full credit.
- I encourage discussion of homework exercises with classmates if it helps, but do your own work.
- Please turn in work that is concise but complete. Put abbreviated SAS code and output at the end of the homework unless otherwise mentioned. (Abbreviated means trying to keep it down to a few pages rather than turning in 20 to 30 pages of output.)
- All homework will be posted on the course web site.

Quizzes

- These will either be closed book or with 1 sheet of notes allowed.
- The quizzes will focus on checking understanding of methods and concepts discussed in class and practiced with homework.

Data analysis project

- You will first find a real data set (e.g., from your work or hobby), analyze it using methods learned in this course, prepare and turn in a short report, and finally, give a brief presentation of your project.
- You will need to find your data by about the 2nd or 3rd week of class that involves clustered data and preferably a continuous outcome. There will be questions on roughly 3 to 4 homework assignments to guide you with the analysis.
- If you are having trouble finding a data set, see me.
- The presentation will be about 5 to 10 minutes per person. It might be helpful to prepare about 5 PPT slides for your talk.

BIOS 6643 Longitudinal Data Analysis Fall Semester 2014

Calendar with tentative outline of topics

August 25, 27: *Longitudinal introduction*
Longitudinal vs. time series data; types of longitudinal data; longitudinal designs; formats for longitudinal data; case studies; longitudinal analyses with 2 time points; simple experiment with factorial treatment structure involving time ('GLM data'); graphs for longitudinal data; variance at different time points versus variance of difference between time points

September 1: *Labor Day*

September 3, 8, 10: *Linear models (LM, or GLM for General Linear Model).*
GLM data and analysis; basic review of the GLM – model and assumptions; least squares – calculus, geometry and linear algebra; computing estimates in a GLM; one-way and two-way effects models; time as a class versus continuous variable; distribution theory; estimability; full-rank versus less-than-full-rank models; some properties of estimators; inference in the GLM (confidence intervals and hypothesis tests)

September 15, 17, 22: *Introduction to LMMs*
Random intercept models, simple random-effect and mixed-effect models, notation and assumptions for LMMs; repeated measures ANOVA, AIC

Sep 24, 29: *Inference in LMMs*
distributions, general estimation, properties of estimators, inference for fixed effects, impact of modeling correlation on inference for fixed effects, inference for random effects and variance components

October 1, 6: *Modeling covariance structures in LMMs*
Modeling random effects (G matrix), modeling the error covariance structure (R matrix)

October 8, 13 *LMMs: Nesting and crossing*
Basics of nesting and crossing; modeling 3-level data

October 15: *Software and computation notes*

October 20: Quiz I

October 22, 27: *Generalized Linear Models (GzLM)*, Gary
Introduction to GzLMs, overdispersion

November 29, 3, 5: *Models for non-normal correlated outcomes*
Generalized estimating equations (GEE), generalized linear mixed models, mixture distributions, fitting nonlinear functions

November 10, 12: *Nonparametric longitudinal regression*
Spline modeling, kernel regression

November 17: *Interpreting parameters in longitudinal models*

November 19: *Special topics (this lectures might also be used for catch up or review)*
Measurement error models, power and planning, case crossover designs, semi-variograms, multivariate GLM

November 24: *Student consulting sessions (optional)*

December 1, 3: *Student presentations*

December 8 or 10: Quiz II

BIOS 6643 Longitudinal Data Analysis Fall Semester 2014

Topics and associated reading in course notes

(Numbering on left is for slide set; the last topics on the list currently do not have slide sets)

1. Introduction (*1 Introduction* notes)
2. Descriptive (*2 Descriptive* notes)
3. GLM: Introduction, least squares, model forms (*3 GLM*, Sec. 1-3)
4. GLM: Statistical models and estimation (*3 GLM*, Sec. 4)
5. GLM: Tests and estimation: default and custom tests (*3 GLM*, Sec. 5, 'Quiz')
6. LMM: One-sample RM ANOVA, random intercept models (*4 LMM Intro*, Sec. 1, 2.1)
7. LMM: Multi-sample RM ANOVA, random intercept models; simple random effects and mixed effects models; ICC and reliability; (*4 LMM Intro*, Sec. 2.2 through 3)
8. LMM: notation and assumptions for LMMs, AIC (*4 LMM Intro*, Sec. 4-5)
9. LMM: basic theory for mixed models; inference for fixed effects, impact of modeling correlation on inference for fixed effects (*5 LMM: inference*, Ch. 1-3, 6-7)
10. LMM: Inference for random effects, variance components (*5 LMM: inference*, Ch. 4-5)
11. LMM: Modeling random effects (*6 LMM: modeling random effects and error covariance structure*, Ch. 1)
12. LMM: Modeling error covariance structure, (*6 LMM: modeling random effects and error covariance structure*, Ch. 2)
13. LMM: Nesting and crossing (*7 LMM nesting and crossing*)
14. LMM: Software and computational issues (*8 LMM software and computation* notes)
    Non-normal: GzLM (Gary; *9 Non-normal*, Ch. 3)
    Non-normal: GzLM and overdispersion (Gary; *9 Non-normal*, Ch. 3)
15. Non-normal: GEE (*9 Non-normal*, Ch. 4)
16. Non-normal: Generalized linear mixed models (*9 Non-normal*, Ch. 5)
17. Non-normal: Mixture distributions, fitting nonlinear functions (*9 Non-normal*, Ch. 6-7)
18. General: longitudinal nonparametric regression – spline modeling (*11 LMM extended topics*, Ch. 1)
19. General: longitudinal nonparametric regression – kernel regression (*11 LMM extended topics*, Ch. 2)
20. General: Interpreting parameters in longitudinal models (*10 Interpreting parameters in longitudinal models*, Ch. 1-4, 6)

    General: Multivariate GLM, measurement error methods, power and planning for LMMs, semi-variograms for LMMs (*12 Auxiliary material*)