Complex Data Analysis

Syllabus
Spring 2018

A. Instructor:

Jaroslaw Harezlak, Associate Professor

Contact information:
Department of Epidemiology and Biostatistics
Indiana University School of Public Health-Bloomington
1025 E 7th St., C107
Bloomington, IN 47405

Email: harezlak@iu.edu

Office hours: By appointment

B. Combined lectures by Zoom (Skype) and in-class

February 19th, 2018 till June 15th, 2018

C. Course Description:

This course covers modern methods for the analysis of repeated measures, correlated outcomes and longitudinal data, including the unbalanced and incomplete data frequently encountered in biomedical research. Topics include an introduction to the analysis of correlated data, repeated measures analysis of variance (ANOVA), random-effects and growth-curve models, generalized linear models for correlated data, including generalized estimating equations (GEE), and generalized linear mixed models (GLMMs).
Class presentations and homework assignments will focus on data analysis in **R** (www.r-project.org using fucntions *lme()* and *gls()* from the *library(nlme).*

D. Prerequisites:

Linear regression course. Some familiarity with R statistical computing environment is assumed. Students who are uncertain about their level of preparation are encouraged to contact the instructor.

E. Required and Recommended Texts:

The course will use the material from following textbooks:

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2011) *Applied Longitudinal Analysis* Wiley & Sons (FLW), 2nd edition.

Laird, N.M. (2004) Analysis of Longitudinal and Cluster-Correlated Data, Institute of Mathematical Statistics

We will supplement the textbook with instructors' own lecture notes.

F. Educational objectives:

At the end of the course students will be able to:

- use modern statistical methods for the analysis of repeated measures, correlated outcomes and longitudinal data
- utilize R software for the analysis of such data

Format: Lectures with periodic labs to demonstrate and explore statistical methods and their implementation in R.

G. Evaluation and Grading:

Students will be evaluated based on their performance on the homework assignments (30%), the mid-term exam (30%), and the final project/presentation (40%).

H. Course Outline

Part I: ANALYSIS OF VARIANCE AND INTRODUCTION TO CORRELATED DATA

1. Lecture: Introduction
   Reading: FLW, Chapter 1

2. Lecture: Introduction to Correlated Data
   Reading: FLW, Chapter 2

Part II: LINEAR MODELS FOR LONGITUDINAL DATA

3. Lecture: Longitudinal Data – Linear Models
   Reading: FLW, Chapter 3

4. Lab: Using functions gls() and lme() in R

5, 6. Lectures: Statistical Basis of Repeated Measures Analysis
   Reading: FLW, Chapter 4

7. Lecture: Parallel Groups Repeated Measures Analysis
   Reading: FLW, Chapter 5

8. Lab: Longitudinal Data Analysis using gls() + lme() in R

9. Lecture: Statistical Modeling of Repeated Measures
   Reading: FLW, Chapter 6

10. Lecture: General Linear Model for Longitudinal Data
    Reading: FLW, Chapter 6

11. Lecture: Mixed Effects Model for Longitudinal Data
    Reading: FLW, Chapter 8

12. Lecture: Prediction of Subject-Specific Regression Coefficients
    Growth Curve Models and Two-Stage Analysis
    Reading: FLW, Chapter 8

13. Lab: Linear Mixed Models

14. Lecture: Selection of Model for Covariance, Empirical Variance Estimation,
    Reading: FLW, Chapter 7
15. Lecture: Drop-outs, missing values and mis-timed values
    Reading: FLW, Chapter 17

16. **Mid-Term Exam**

17. Lecture: Design and sample size estimation
    Reading: FLW, Chapter 20

Part III: GENERALIZED LINEAR MODELS FOR LONGITUDINAL DATA

18. Lecture: Review of Logistic and Poisson Regression

19. Lecture: Introduction to Generalized Linear Models
    Reading: FLW, Chapter 11

20. Lecture: Generalized Linear Models for Longitudinal Data

21-22. Lectures: Marginal Models and Generalized Estimating Equations
    Reading: FLW, Chapters 12-13

23. Lab: GEE using R functions

24. Lecture: Generalized Linear Mixed Models
    Reading: FLW, Chapter 14

25. Lecture: Contrasting Marginal and Mixed Effects Models
    Reading: FLW, Chapter 16

26. Lab: GLMM using R function: gam() from the library(mgcv)

27. Lecture: Multilevel Models
    Reading: FLW, Chapter 22

28. Lab: Multilevel Models

29. Student presentations