ST 790, Longitudinal Data Analysis
Spring 2018

Course:

Lectures:	TTh, 3:00 pm - 4:00 pm, SAS 5270
Website:	`http://www.stat.ncsu.edu/people/davidian/courses/st790/`
Prerequisites:	ST 702, ST 705

Instructor:

Marie Davidian

Office:	5124 SAS Hall
Phone:	919-515-1940
Email:	davidian@ncsu.edu
Website:	`www4.stat.ncsu.edu/~davidian`
Office Hours:	Th, 1:00 pm - 2:00 pm, and by appointment

Teaching Assistant:

Seonghyun Jeong

Office:	1101 SAS Hall (Statistics Tutorial Room)
Email:	sjeong4@ncsu.edu
Office Hours:	M, 2:30 - 3:30 pm

Goal: To provide an overview of fundamental statistical models and methods for the analysis of longitudinal data, including key theoretical results, presented at a heuristic level, and implementation in R and SAS. Upon completion of the course, students will have a comprehensive understanding of the properties and use of modern methods for longitudinal data analysis and will be able to pose scientific questions within the context of appropriate statistical models and carry out and interpret analyses. A complete overview of the course is given at the URL above.

Text: Lecture notes prepared by the instructor. These will be available on the course website.

Grading: Final grade will be determined by the Final Score = $0.35\times$H+ $0.30\times$M+ $0.25\times$F + $0.10\times$D, where H is the homework average, M is the midterm score, F is the score on the final project, and D is instructor discretion score, where each is scored out of 100. The instructor's discretion portion will be based on attendance, participation in class, and instructor's assessment of mastery of the material. There will no final exam. Conversion of this score into a letter grade will be made according to the following tentative grading scale (the upper score in each range except A+ belongs to the next highest grade): A+, 100; A, 92-99; A-, 90-92; B+, 88-90; B 82-88; B-, 78-82. Scores below 70 will be handled on a case-by-case basis.

Chronic absenteeism will result in at least a 5 point reduction in the Final Score, as determined by the instructor. If you you must miss a lecture due to illness, job interview, etc, the instructor would appreciate being informed in advance of the lecture.

Auditors must attend class and attempt and turn in analytical problems on the homeworks. Auditors are not required to complete data analysis problems, the midterm, or the final project.

Homework: There will be five (5) homework assignments. Homeworks will involve a combination of analytical problems, data analyses, and simulation studies, where the latter two will involve programming and getting up to speed on SAS and R software on the part of the student. For problems

involving programming, both the program and its output should be turned in, along with interpretation of the results as dictated by the problem. Unexcused late homework will be discounted by 50%. Only selected problems on each homework will be graded; of course, these will not be announced in advance.

Students are permitted and even encouraged to work together on homework; however, each student must turn in his/her own solutions. Blind copying of the work of other students demonstrates that the student doing the copying is not serious about developing the independence required for a PhD and has obvious disadvantages for the midterm and final project, not to mention mastery of the material.

Tentative assignments/due dates are as follows; definitive information will be posted on the course website. Homework will be collected *at the beginning of class* on the due date.

Homework 1	Chapters 1-3, due Thursday, January 25
Homework 2	Chapters 4-5, due Tuesday, February 13
Homework 3	Chapter 6, due Thursday, February 22
Homework 4	Chapters 6-8, due Thursday, March 29
Homework 5	Chapters 8-9, due Tuesday, April 24

Computing/Software: We will demonstrate implementation of the methods discussed in the notes in both SAS and R. Thus, familiarity with SAS and R and more generally with computing environments and programming is desirable. The majority of the time in the lectures will be focused on the methodological developments and arguments and interpretation of analyses, with relatively less time devoted to the details of implementation. Accordingly, students are expected to develop implementation skills independently through homework assignments and the project.

Final Project: The final project will involve a substantial modeling and analysis task using methods covered in the course (as well as any other models and you methods you see fit to use). The nature and format of the project will be determined later in the semester.

Tentative Schedule: Spring Break is 03/05 - 03/09. There will be no class on Tuesday, 03/27 (ENAR Spring Meeting).

01/09 - 01/11 – 1. Introduction and Motivation

- Objectives
- Running examples
- Overview of statistical modeling

01/11 - 01/16 – 2. Modeling Longitudinal Data

- Data structure and notation
- Conceptual framework for continuous response
- Population-averaged vs. subject-specific modeling
- Models for correlation structure; exploring correlation structure
- Discrete response

01/18 - 01/23 – 3. Repeated Measures Analysis of Variance

- Univariate repeated measures ANOVA
- Specialized tests

- Multivariate repeated measures ANOVA

01/25 – 4. Modern Methods: Preliminaries

- Drawbacks of classical methods
- Review of large sample inference and estimating equations

01/25 - 02/08 – 5. Population-Averaged Linear Models for Continuous Response

- Model specification
- Maximum likelihood estimation under normality
- Restricted maximum likelihood
- Large sample inference
- Implications of missing data

02/13 - 02/27 – 6. Linear Mixed Effects Models

- Model specification
- Maximum likelihood and restricted maximum likelihood
- Large sample inference and implications of missing data
- Best linear unbiased prediction and empirical Bayes
- Implementation via the EM algorithm
- Testing variance components

03/01 – Midterm Exam (In class)

03/13 - 03/15 – 7. Review of Generalized and Nonlinear Models for Univariate Response

- Nonlinear and Generalized (non)linear models
- Estimating equations and variance function estimation
- Large sample inference

03/15 - 04/03 – 8. Population-Averaged Models and Generalized Estimating Equations

- Model specification
- Linear and quadratic estimating equations for fixed effects
- Quadratic estimating equations for covariance parameters
- Large sample inference
- Modeling issues
- Implications of missing data

04/03 - 04/19 – 9. Generalized Linear and Nonlinear Mixed Effects Models

- Model specification
- Maximum likelihood
- Empirical Bayes estimators for random effects
- Approximate inference based on individual estimates
- Approximate inference based on linearization
- “Exact” inference

- Implications of missing data

04/12 – Final project assigned

04/19 - 04/26 – 10. Advanced Topics (chosen from the following)

- Multi-level models
- Semi-nonparametric models
- Multivariate longitudinal outcomes
- Relaxing assumptions on random effects

Final exam date (Tuesday, May 1) – Final project due (tentative)

Class Evaluations: Online class evaluations will be available for students to complete from 8:00 am on April 16, 2018 through 8:00 am on April 30, 2018. Students will receive an email message directing them to a website where they can login using their Unity ID and complete evaluations. All evaluations are confidential; instructors will not know how any one student responded to any question, and students will not know the ratings for any instructors.

Academic Integrity: The instructor expects that students will abide by the University policy on academic integrity found in the Code of Student Conduct Policy (POL 11.35.1), available at the website `https://policies.ncsu.edu/policy/pol-11-35-01`. It is the understanding and expectation that a student's signature on the midterm and final project means that the student neither gave nor received unauthorized aid. As noted above, students *may* consult with one another on the homework, similar to how real research statisticians might consult with one another. However, students engaging in direct copying of the work or computer programs of fellow students will be considered in violation of the Code of Conduct. Students *may not* consult with one another or with any other person except the instructor in any way in completing the midterm and final project. This includes having another person review your final project report for English usage and grammar.

Students with Disabilities: Reasonable accommodations will be made for students with verifiable disabilities. In order to take advantage of available accommodations, students must register with Disability Services Office (DSO) at 2221 Student Health Services Building, Campus Box 7509, 515-7653; see `http://dso.dasa.ncsu.edu/`. For more information on NC States policy on working with students with disabilities, please see the Academic Accommodations for Students with Disabilities Regulation (REG02.20.01) (`https://policies.ncsu.edu/regulation/reg-02-20-01/`).

Further Resources: There is no textbook for this course; as noted above, we will use lecture notes prepared by the instructor, and no other books are required. The notes cite publications where further information on the specific developments presented can be found. In addition, if you are interested in more general, further reading on longitudinal data methods, the following books are good resources:

Davidian, M. and Giltinan, D.M. (1995). *Nonlinear Models for Repeated Measurement Data*. London: Chapman and Hall/CRC Press.

Diggle, P.J., Heagerty, P., Liang, K.-Y., and Zeger, S.L. (2002). *Analysis of Longitudinal Data*, 2nd edition. New York: Oxford University Press.

Fitzmaurice, G.M., Laird, N.M., and Ware, J.H. (2011). *Applied Longitudinal Analysis*, 2nd edition. New York: Wiley.

Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G. (2009). *Longitudinal Data Analysis*. Boca Raton: Chapman and Hall/CRC Press.

Molenberghs, G. and Kenward, M. G. (2007). *Missing Data in Clinical Studies*. Chichester, UK: Wiley.

Pinheiro, J.C. and Bates, D.M. (2000). *Mixed Effects Models in S and S-PLUS*. New York: Springer.

Verbeke, G. and Molenberghs, T. (1997). *Linear Mixed Models in Practice: A SAS-Oriented Approach* Lecture Notes in Statistics 126. New York: Springer.

Verbeke, G. and Molenberghs, T. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.

Vonesh, E.F. and Chinchilli, V.M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York: Marcel Dekker.

Weiss, R.E. (2005). *Modeling Longitudinal Data*. New York: Springer.