ST 732, SPRING 2007
APPLIED LONGITUDINAL DATA ANALYSIS
COURSE INFORMATION

GOAL OF COURSE

To introduce students to statistical models and methods for the analysis of longitudinal data, i.e. data collected repeatedly on experimental units over time (or other conditions).

INSTRUCTOR

Marie Davidian, 220-F Patterson Hall, 515-1940, davidian@stat.ncsu.edu
Office Hours T 10:00-11:00 am, or by appointment

MEETING TIMES

TTh 8:30 am – 9:45 pm, 1202 Burlington

REQUIRED TEXT

Lecture notes prepared by the instructor (see below). These may be purchased at the NC State Bookstores.

TEACHING ASSISTANT

TBA
Office Hours TBA in 9 Patterson Hall

CLASS WEB PAGE

http://www.stat.ncsu.edu/people/davidian/st732/

PREREQUISITE

ST 512, Experimental Statistics for Biological Sciences II, or equivalent. Thus, students should be familiar with basic notions of probability, random variables, and statistical inference, analysis of variance, and (multiple) linear regression. Familiarity with matrix algebra is also useful. We will review matrix algebra at the beginning of the course and make considerable use of matrix notation and operations throughout. ST 512 involves the use of the SAS (Statistical Analysis System) software package; thus, students are expected to have had some exposure to the use of SAS. The course is meant to be accessible both to non-majors and majors. The underlying mathematical theory will not be stressed, and the main focus will be on concepts and applications. Please see the instructor if you have questions about the suitability of your background.

COURSE CONDUCT

Lectures: We will follow mimeographed lecture notes prepared by the instructor. In-class lectures will follow the notes; the notes will be projected and discussed, so you will want to bring your copy so that you may follow along.

Homework: Homework will be assigned approximately once every two weeks; see the class web page for tentative due dates. As the focus of the class is on the practical application of methods for longitudinal data analysis, many of the problems will involve using statistical software to carry out analyses on real data sets. To implement the analyses, we will use SAS; examples of the use of this software are included in the lecture notes and will be discussed in class

(see below). The principles underlying the methods we will learn involve much philosophy and subjectivity; thus, each homework will also include "essay"-type questions to help you to clarify your thinking on these matters.

Homework will be collected **at the beginning of class** on the **date it is due**. It should be **neat**, **all work should be shown**, and **no late homework accepted unless prearranged with the instructor**. **There will be no exceptions to this policy**.

For problems where SAS programs are assigned, **both the program and its output** should be included. Students are encouraged to consult one another on homework problems, but everyone should turn in their own homework, and no "blind copying" permitted. All students should do their own programming!

Computing: Students may use SAS on the campus network. Alternatively, students with access to SAS somewhere else (PCs in your department or at home, at work, etc) may use whatever system is most convenient. Please see the instructor if you have any questions. All data sets to be analyzed as part of homework assignments will be available on the class web page.

Tests: We will have 2 tests during the semester. These will be **in-class**. The tests are tentatively scheduled as follows:

1. Test 1, Tuesday, March 13
2. Test 2, Thursday, May 3 (the final exam day).

Although it will take place on the scheduled final exam day, Test 2 will **not** be a cumulative final exam; rather, it will involve material covered since Test 1.

Data Analysis Project: Students will carry out an analysis of data collected in a study that will be described in detail in the assignment using methods covered in the class (which methods are relevant is to be determined by the student). Students will need to formalize the scientific questions posed by the investigator, carry out the appropriate analyses, interpret the results, and write a comprehensive report for the investigator reporting on all of these activities. The assignment will be handed out on March 27 and will be due on the last day of class, Thursday, April 26.

Final Exam Date: Thursday, May 3, 2007, 8:00 – 11:00 am. As discussed above, there will be no in-class final exam; Test 2 will be held at this time.

Grading: All homeworks and tests will be graded on a scale of 0 to 100. A numerical final score on this scale will be determined according to the following breakdown:

Homework	15%
Test 1	30%
Test 2	30%
Project	25%

Conversion of these scores into letter grades will be made according to the following (the upper score in each range belongs to the next highest grade): A, 92–100; A−, 90–92; B+, 86–90; B, 79–86; B−, 75–79. C, 70–75. Scores below 70 will be handled on a case-by-case basis. The grade of A+ will be given at the discretion of the instructor for truly stellar individual

performance. Depending on overall class performance, these ranges may be adjusted (but only downward – criteria will only become easier, not harder).

Academic Integrity: The instructor expects that students will abide by the University policy on academic integrity found in the Code of Student Conduct Policy (POL11.35.1), available at `http://www.ncsu.edu/policies/student_services/student_discipline/POL11.35.1.php`. Students are expected to uphold the standards of honesty set forth in this Code in completion of all class assignments.

As noted above, students may consult with one another on homework for points of clarification and for discussion, similar to how real data analysts 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 Data Analysis Project. This *includes* having another person review your data analysis report for English usage and grammar. Students are expected to complete all aspects of this project completely independently and to sign the Honor Pledge to this effect. Similarly, students are expected to work on tests completely independently and to sign the Honor Pledge to this effect. Signing the Honor Pledge indicates that you affirm that you have neither given nor received aid unauthorized under the above policies in completing the project or test. Academic integrity violations will be referred to the proper authorities, and no credit will be given for the course.

OUTLINE OF COURSE TOPICS

1. Preliminaries:
   Introduction
   Review of matrix algebra
   Random vectors, multivariate normal distribution, review of linear regression
   Introduction to modeling longitudinal data, exploring covariance structure

2. Classical methods for normally distributed, balanced repeated measurements:
   Univariate repeated measures analysis of variance
   Multivariate repeated measures analysis of variance
   Drawbacks and limitations of classical methods

3. Methods for normally distributed, unbalanced repeated measurements:
   General linear models and models for correlation
   Random coefficient models
   Linear mixed effects models

4. Methods for non-normally distributed, unbalanced data:
   Probability models for discrete and continuous nonnormal data and generalized linear models
   Generalized estimating equations for population-averaged models

5. Advanced topics (brief overview):
   Generalized linear mixed effects models
   Nonlinear mixed effects models
   Missing data mechanisms

REFERENCE TEXTS

Students may find the following texts useful for more in-depth mathematical treatment of course topics or alternative perspectives, although none of these is required:

Crowder, M.J. and Hand, D.J. (1990) *Analysis of Repeated Measures.* London: Chapman and Hall/CRC Press.

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. (2004) *Applied Longitudinal Analysis.* New York: Wiley.

Hand, D. and Crowder, M. (1996) *Practical Longitudinal Data Analysis.* New York: Chapman and Hall/CRC Press.

Lindsey, J.K. (1993) *Models for Repeated Measurements.* New York: Oxford University Press.

Littell, R.C., Milliken, G.A., Stroup, W.A., and Wolfinger, R.D. (1996) SAS System for Mixed Models, SAS Institute, Cary NC.

Longford, N.T. (1993) *Random Coefficient Models.* New York: Oxford University Press.

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

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

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.