ST 762, FALL 2009
Nonlinear Models for Univariate and Multivariate Response

Instructor: Marie Davidian, Professor, 5124 SAS Hall, 515-1940, davidian@stat.ncsu.edu

Class meeting: TTh 1:30–2:45 pm, SAS Hall 2229

Instructor's office hours: T 12:00–1:00 pm, in SAS Hall 5214 or Cox 308

Course Description: This course will provide a detailed treatment of regression models and associated inferential methods both for univariate and multivariate (e.g. repeated measures) response. The techniques to be discussed are now an essential part of the modern statistician's toolkit and are widely used in numerous application areas.

The first 1/2 to 2/3 of the course will focus on nonlinear regression models for univariate response, including models for nonconstant response variance. The remainder of the course will be devoted to introduction to extension of the univariate model to two popular types of nonlinear regression models for multivariate response: (i) "population-averaged" models and models for covariance structure will discussed; methods for fitting these models are popularly known in the literature as "generalized estimating equations," and (ii) "subject-specific" models, e.g., generalized linear and nonlinear mixed effects models.

Properties of competing inferential techniques and the effects of model misspecification will be studied via theoretical arguments carried out at a nonrigorous, heuristic level and via simulation exercises on the part of students. Although we will go through theoretical arguments in class in some detail, and students will be expected to understand and be able to carry out similar arguments at the same level, our main objective will be to appreciate the implications of the results for practice rather than the technical details. Implementation of the methods and application to data will be emphasized in the homework assignments.

Prerequisites: ST 512R, ST 552, familiarity with SAS or R/Splus and a scientific computing language (e.g. MATLAB, FORTRAN, C++, SAS IML, etc). Students should have a strong background in probability and inference at the level of ST 521–522 (the prerequisites for ST 552).

Class Web Page: Homework assignments, solutions, and other course material will be available at http://www.stat.ncsu.edu/people/davidian/courses/st762/. Important announcements made in class will also be posted here.

Grading: Grades will be based on completion of the following:

- *Homework:* There will be six homework assignments at roughly two-to-three week intervals. Homeworks will contain both analytical problems and data analysis problems and will be handed out in two parts: Problems to turn in (generally 1 or 2) and extra problems not to be turned in. Students may work with one another, but each student must independently write the programs for the data analysis problems and the final solutions to analytic problems to be turned in. *Although only 1 or 2 problems are to be turned in and graded for each assignment, you are responsible for the material covered in all problems (graded and extra).*

  Tentative homework due dates: 9/8, 9/22, 10/13, 11/3, 11/17, 12/1

  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 programming is required, both the program and its output should be turned in.

- *Take-home data analysis project:* Midway through the course, you will be provided with a description of a challenge facing an investigator and the scientific questions s/he wishes to address. You will carry out a complete analysis and write a formal report (typed) for the investigator. This will be a "closed" assignment; that is, you must work independently, not consulting with one another or, for that matter, with *anyone* (even your mother). Students may of course use the class notes and ask questions of the instructor.

  Tentative dates: Handed out 10/13, due 10/20

- *In-class, closed book test:* A two-hour evening period will be scheduled during which you will complete some analytical and short-answer exercises. Anything covered in the course up to the point of the test will be fair game, and ungraded homework problems from homeworks assigned so far could form the basis for midterm problems.

  Tentative date: Evening of 11/5

- *Final Project:* In early November, students will be placed in groups, and each group will be assigned at random to read a recent paper or related set of papers from the literature involving extensions of material covered in class. There will be two parts to this assignment:

  1. *Brief summary paper:* Each student will prepare *independently* of his/her groupmates a *brief* (no more than 5 double-spaced pages) written summary of the paper(s). The summary should provide, *in the student's own words*, a high-level (no technical details) description of what the paper(s) is(are) about and why the work is important, and a brief discussion of how the work is related to/extends topics covered in class. This paper should be something a person with advanced training in statistics and the material in this course but not necessarily familiar with the topic could go to to get a general idea of what the paper(s) is(are) about. The student's paper should use the notation developed in class.

     Due date: This will be due on the day of the presentations (see next item)

  2. *Group presentation:* Each group will give to the rest of the class a joint oral presentation summarizing the papers. The size of the groups will be determined by the size of the class.

     Due date: The presentations will take place during the scheduled final exam period. We will try to choose an earlier time if possible.

  More details on the project will be provided later in the course.

The course grade will be determined according to the following breakdown: homework, 10%; data analysis project, 30%; in-class test, 35%; summary paper, 10%; oral presentation, 10%; and instructor's discretion, 5%. The instructor's discretion portion will be based on attendance, participation in class, and instructor's assessment of mastery of the material.

Conversion of these scores into letter grades will be made according to the following scheme (except for 100, the upper score in each range belongs to the next highest grade): A, 92–100; A−, 90–92; B+, 88–90; B, 82–88; B−, 78–82. C, 70–78. 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).

Text: We will follow lecture notes by the instructor. **Course notes will be available for purchase at Sir Speedy at the beginning of the semester.** Texts that you may wish to consult for further reading are:

Bates, D.M. and Watts, D.G. (1988). *Nonlinear Regression Analysis and Its Applications*. Wiley.

Carroll, R.J. and Ruppert, D. (1988). *Transformation and Weighting in Regression*. /Chapman and Hall/CRC Press.

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

Diggle, P.J., Heagerty, P., Liang, K.-Y., and Zeger, S.L. (2002). *Analysis of Longitudinal Data, Second Edition*. Oxford University Press.

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

Gallant, A.R. (1987). *Nonlinear Statistical Models*. Wiley.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models, Second Edition*. Chapman and Hall/CRC Press.

Seber, G.A.F and Wild, C.J. (1989). *Nonlinear Regression*. Wiley.

None of the these texts is required.

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 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 data analysis project. This *includes* having another person review your data analysis report for English usage and grammar. This same standard also applies to the summary paper written as part of the final project. 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 test 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 projects or test. Academic integrity violations will be referred to the proper authorities, and no credit will be given for the course.

Course Topics: The following is a list of the topics discussed in the lecture notes. Time will not permit coverage of all topics. Topics for this semester will be drawn from this list.

1. Introduction and motivation

I. Models for univariate response

2. Introduction to nonlinear models
3. Implementation of generalized least squares (GLS), iteratively reweighted least squares
4. Generalized (non)linear models, quasilikelihood
5. Normal theory maximum likelihood (ML)
6. Unknown parameters in the variance function
7. Detecting and modeling nonconstant variance
8. Large sample theory – a casual approach
9. The "folklore" theorem and "optimality" of GLS
10. Linear vs. quadratic estimating equations for the regression parameter
11. Effect of estimating weights in GLS
12. Estimation of unknown parameters in variance function models

II. Models for multivariate response

13. Modeling multivariate response – sources of correlation and "subject-specific" vs. "population-averaged" approaches
14. Generalized estimating equation methods for marginal (population-averaged) models
15. Nonlinear and generalized linear mixed effects (subject-specific) models – approximate and "exact" methods

Course Evaluations: Online class evaluations will be available for students to complete during the last two weeks of class. Students will receive an email message directing them to a website where they can login using their Unity IDs and complete evaluations. All evaluations are confidential; instructors will never know how any one student responded to any question, and students will never know the ratings for any particular instructors.