Course:
Lectures: TTh, 3:00 pm - 4:15 pm, SAS 1108
Website: http://www4.stat.ncsu.edu/~davidian/st790/
Prerequisites: ST 522, ST 552

Instructor:
Marie Davidian
Office: 5124 SAS Hall
Phone: 919-515-1940
Email: davidian@ncsu.edu
Website: www4.stat.ncsu.edu/~davidian
Office Hours: Th, 11:00 am - 1:00 pm, and by appointment

Teaching Assistant:
Suchit Mehrotra
Office: 1101 SAS Hall (Statistics Tutorial Room)
Email: smehrot@ncsu.edu
Office Hour: T, 9:00 - 10:00 am

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

Grading: Final grade will be determined by the Final Score = 0.65 × H + 0.25 × F + 0.10 × D, where H is the homework average, 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 midterm or 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 must miss a lecture due to illness, job interview, etc, the instructor would appreciate being informed in advance of the lecture.

Homework: There will be four (4) homework assignments. Assignments and due dates will be posted on the course website. Tentative due dates are 01/31, 02/21, 03/23, and 04/18. Homeworks will involve a combination of analytical problems, data analyses, and simulation studies, where the latter two will involve programming 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%. Homework will be collected at the beginning of class on the due date.

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 final grade and mastery of the material.
**Final Project:** There will be a final project, the nature of which will be determined by the size of the class. Possible formats for the project include (i) students read a paper from the recent literature (assigned at random) and prepare a 15 minute presentation summarizing the main findings using the notation and framework of the course to the extent possible; (ii) the same, but students will write a short (5–7 page) paper; (iii) students will identify a data set in the public domain, carry out an analysis using methods from the course; and prepare a poster or written report.

**Computing/Software:** We will use some SAS procedures and R packages to demonstrate implementation of certain methods; thus, some familiarity with SAS and R is desirable. There is not an abundance of generic software to carry out missing data analyses; thus, it is often the case that a data analyst must do some specialized programming to implement specific missing data analyses. Accordingly, you will be asked to do some programming, which can be in your favorite programming language (R, SAS, MATLAB, C, C++, Python, etc, even Fortran).

**Tentative Schedule of Course Topics:** We may deviate from this schedule to include one or two guest lectures; this will be determined as the course progresses. In addition to during Spring Break (03/06-10); there will be no class on Tuesday, 03/14 (ENAR Spring Meeting).

01/10 - 01/17 – Introduction and Motivation
- Challenges posed by missing data
- Statistical framework
- Missing data mechanisms
- Review of estimating equations

01/19 - 01/24 – Naïve Methods
- Complete case and available case methods
- Single imputation methods
- Last Observation Carried Forward (LOCF)

01/26 - 02/14 – Likelihood-based Methods Under Missing At Random (MAR)
- Review of maximum likelihood inference for full data
- Factorization of the density of \((R, Z)\)
- Observed data likelihood and ignorability
- Expectation-Maximization (EM) algorithm
- Missing information principle
- Bayesian inference

02/16 - 03/02 – Multiple Imputation Methods Under Missing At Random (MAR)
- Fundamentals of multiple imputation
- Proper versus improper imputation
- Rubin’s variance formula
- Asymptotic results
- Imputation from a multivariate normal distribution
- Multivariate Imputation by Chained Equations (MICE)
03/16 - 04/04 – Inverse Probability Weighted Methods Under Missing At Random (MAR)
- Illustrative examples
- Weighted generalized estimating equations for longitudinal data with dropout
- Inverse weighting at the occasion level
- Inverse weighting at the individual level
- Doubly robust estimation

04/06 - 04/11 - Pattern Mixture Models
- Introduction and rationale
- Modeling strategies

04/13 - 04/20 – Sensitivity Analysis to Deviations from Missing At Random (MAR)
- Challenges under Missing Not At Random (MNAR)
- Estimation of a single mean
- Longitudinal data with dropout

04/25 – Additional Topics
04/27 – Last day of class, Wrap-up

Class Evaluations: Online class evaluations will be available for students to complete from 8:00 am on April 17, 2017 through 8:00 am on May 1, 2017. 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 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/.
Further Resources: There is no textbook for this course; as noted above, we will use lecture notes prepared by the instructor and Dr. Tsiatis, 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 missing data methods, the following books are good resources:


