ST 790, Statistical Methods for Analysis With Missing Data
Spring 2015

Course:
Lectures: TTh, 1:30 pm - 2:45 pm, SAS 5270
Website: http://www.stat.ncsu.edu/people/davidian/courses/st790/
Prerequisites: ST 522, ST 552

Instructors:
Butch Tsiatis
Office: 5144 SAS Hall
Phone: 919-515-1928
Email tsiatis@ncsu.edu
Website: www4.stat.ncsu.edu/~tsiatis
Office Hours: T, 12:00 pm - 1:00 pm

Marie Davidian
Office: 5124 SAS Hall
Phone: 919-515-1940
Email: davidian@ncsu.edu
Website: www4.stat.ncsu.edu/~davidian
Office Hours: T, 12:00 pm - 1:00 pm

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

Grading: Final grade will be determined by the Final Score = 0.80×HWKAVG + 0.20×PROJ, where HWKAVG is the homework average and PROJ is the score on the final project, where each is scored out of 100. Unexcused absences will not be tolerated; each unexcused absence after the first will result in a 10 point reduction in the Final Score. If you must miss a lecture due to illness, job interview, etc, you must inform us in advance of the lecture for the absence to be excused. There will no midterms or final exam.

Homework: There will be 4 to 5 homework assignments. Assignments and due dates will be posted on the course website. 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. Unexcused late homework will be discounted by 50%.

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++, etc, even Fortran).
**Policy on Academic Integrity:** The University policy on academic integrity is spelled out in the NCSU Code of Student Conduct. For a more details see the NCSU Office of Student Conduct website http://www.ncsu.edu/student_conduct/. For this course, you may work together on homework assignments. However, copying another student’s work or programs will not be tolerated; you must turn in your own solutions and programs. There will be no collaboration permitted on the final project; work on the project must be entirely your own, with no consultation with any individual other than the instructors for this course.

**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/.

**Tentative Schedule of Course Topics:** In addition to during Spring Break (03/09-13); there will be no class on Tuesday, 03/17 (ENAR Spring Meeting), and Thursday, 04/02 (Easter Break).

01/08 - 01/15 – Introduction and Motivation

- Challenges posed by missing data
- Statistical framework
- Missing data mechanisms
- Review of estimating equations

01/20 - 01/22 – Naïve Methods

- Complete case and available case methods
- Single imputation methods
- Last Observation Carried Forward (LOCF)

01/27 - 02/12 – 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
02/17 - 03/03 – 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
- Multiple Imputation by Chained Equations (MICE)

03/05 - 03/31 – 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

04/07 - 04/09 - Pattern Mixture Models

- Introduction and rationale
- Modeling strategies

04/14 - 04/21 – 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

04/23 – Last day of class, Wrap-up
Further Resources: There is no textbook for this course; as noted above, we will use lecture notes prepared by the instructors, 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:


