

STAT 6520: Applied Statistical Analysis with Missing Data

Autumn 2015

Instructor: Elly Kaizar
221 Cockins Hall
kaizar.1@osu.edu (preferred)
(614) 247-2585

Office Hours: Wednesday 4:00-5:00pm, or by appointment

Course website: <http://carmen.osu.edu>

Course description:

This class will review statistical analysis for complete data and provide an introduction to the models and methods for the dataset with missing values. The course has a significant component of statistical computations dealing with missing data. It is intended for those who already have some experience with standard statistical methods for complete data and want to extend them to handle the missing data in practice.

Course Objectives:

After the completion of this course, the students are expected to

1. Understand the missing data mechanism, the underlying assumptions and identify different patterns of missing data
2. Understand the difference in statistical analysis between missing data problem and complete data problem (including weighted methods)
3. Be able to perform simple missing data analysis with single imputation; comprehend its weakness
4. Be able to implement likelihood-based analysis with ignorable missing response; implement EM algorithm with some statistical package
5. Understand the principle of Bayesian analysis with missing data; implement multiple imputation with some statistical package
6. Understand missing data models in contingency tables
7. Have a basic understanding of the recent development of statistical methods to deal with non-ignorable missing data
8. Be able to implement and interpret statistical methods for missing data in a practical scenario.

Required Texts:

Statistical Analysis with Missing Data, Little and Rubin, 2nd edition, Wiley.

Reference:

Analysis of incomplete multivariate data, Schafer, J.L. Chapman & Hall, London.

Software:

The use of a computer for data analysis is essential for this applied course. The program R will be used for examples, homework and exams. R is a freely available download. Because the implementation of missing data methods is far from standardized across packages, **you will need to use R for this course!** You may use other programs as you prefer, but it is your responsibility to be sure that they are accomplishing the assigned methods.

If you are not already familiar with R programming, there are a plethora of online tutorials available. I

recommend the Swirl tutorial package. Instructions for using Swirl are available here: <http://swirlstats.com/students.html>. I will be providing lots of sample R code, so even beginning R users should be able to complete assignments.

Prerequisites:

STAT 6201, 6302 (623), or 6802 (622), and STAT 6450 (645), 6950, PubHBio 6203, or 703; or permission of instructor. Students *must* be familiar with maximum likelihood estimation and regression analysis.

In-person meetings:

M/W/F 3:00 - 3:55PM, Caldwell Lab 135

Tentative Schedule:

| Week | Content | Readings |
|-------------|--|-----------------------------------|
| 0 | Review of ML and Bayes Estimation and Introduction to Missing Data | Chap 6.1 Chap 1.1-1.2 |
| 1 | Missing data mechanism | Chap 1.3-1.4 |
| 2-3 | Complete-data and Available-case analysis | Chap 3 |
| 4-5 | Likelihood-based approach | Chap 6.2-6.3, 7.1-7.4 |
| 6 | Bayesian approach | Chap 10.1 |
| 7-9 | Imputation introduction Multiple imputation | Chap 4, 5 Chap 10.2 |
| 10 | EM algorithm and large sample inference for ignorable missing data Longitudinal data analysis | Chap 8.1-8.4 (skim 8.5) TBA |
| 11 | Longitudinal data analysis, cont. Large sample ML inference Methods for Contingency tables | Chap 9 Chap 13.1-13.3 |
| 12-13 | Mixed variables Non-ignorable missing data | Chap 14.1, 14.2, 14.4 Chap 15 |
| 13 | Student Presentations | |
| 14 | Recent developments | TBA |

Evaluation:

| | |
|----------------|-----|
| Homework | 15% |
| Midterm Exam 1 | 20% |
| Midterm Exam 2 | 30% |
| Final Project | 35% |

Homework:

There will be 7 graded homework assignments. The lowest grade among these will be dropped from the final grade calculation. The remaining 6 grades will be weighted equally. No late assignments will be accepted. The dropped homework grade is meant to give you flexibility when you have an emergency. If you find yourself in extraordinary circumstances where this allowance does not seem appropriate, please see me asap.

Homework assignments should be electronically to the Carmen Dropbox. Unless otherwise indicated in the assignment, electronic submissions must be a single file in .pdf format. Your name should appear

both *in the file name and at the top of the first page.*

Exams:

There will be 2 in-class exams. The second exam is cumulative. You may bring a calculator (not a cell phone or other communication device) to both. You may bring a single 8.5x11 sheet of paper with your own notes on both sides to the first exam, and two such sheets of paper to the second exam. These sheets may be prepared any way you like, and will not be collected.

Tentative exam dates:

October 9

November 20

Final Project:

The culmination of this course will be a final project that will be completed in groups. Your group must produce:

1. Written proposal (1 page maximum) that describes your idea and data source
2. Final written report first draft
3. Poster presentation
4. Written critique of other groups' project reports and presentations
5. Final written report (5 page maximum, excluding figures and appendices)

More information about the final project will be provided in the second week of the course.

Academic integrity:

Cheating, plagiarism and other forms of academic dishonesty will not be tolerated. Any violation will be prosecuted to the fullest extent as set out in University Rule 3335-31-02.

Accommodation for special needs

Any student who feels they may need an accommodation based on the impact of a disability should contact the instructor privately to discuss your specific needs. You should also contact the Office of Disability Services at 292-3307 or in 150 Pomerene Hall to coordinate reasonable accommodations for students with documented disabilities.