STAT 6520: Applied Statistical Analysis with Missing Data

Autumn 2017 (Updated 8/23/17 and 8/28/17)

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

Office Hours: Monday 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 missing data in practice.

Course Objectives:
After the completion of this course, the students are expected to

1. Understand the missing data mechanism and underlying assumptions and identify different patterns of missing data
2. Understand the difference in statistical analysis between the missing data problem and complete data problem (including weighted methods)
3. Be able to perform simple missing data analysis; comprehend its weakness
4. Be able to implement likelihood-based analysis with ignorable missing response; implement the 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 Text:
Statistical Analysis with Missing Data, Little and Rubin, 2nd edition, Wiley. ISBN: 978-0-471-18386-0. Limited copies available online via and one copy will be on reserve at the Ohio State library.

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

Applied Missing Data Analysis, Enders, C.K. The Guilford Press.

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, and standard output from R may appear on exams.

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, Baker Systems 184

Evaluation:

Homework	140 points
Midterm Exam 1	200 points
Midterm Exam 2	300 points
Final Project	360 points
Total	1000 points

These points will be translated into the standard Ohio State letter grading scheme via percentages. The instructor reserves the right to adjust letter grades.

Homework:
There will be 7 graded homework assignments. No late assignments will be accepted. If you find yourself in extraordinary circumstances where you cannot submit your homework on time, please contact the instructor as soon as possible, and before the due date whenever feasible.

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***.

You may work together in solving homework problems, but your submitted work must be your own. You may not copy any of your submission from another student, the text, or any other source without proper attribution (citation). Remember that your grade will be based on your work that you submit. Feel free to ask the instructor for help after your first attempt at any homework problem.

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 11 and 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 is provided in the Project Guidelines.

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.

Your written assignments should be your own original work. In formal assignments, you should formally cite the ideas and words of your research sources. You are encouraged to ask a trusted person to proofread your assignments before you turn them in, but no one else should revise or rewrite your work.

In general, you are prohibited in university courses from turning in work from a past class to your current class, even if you modify it. If you want to build on past research or revisit a topic you've explored in previous courses, please discuss the situation with the instructor.

All research you will conduct in this course is intended to be a learning experience; you should never feel tempted to make your results or your research look more successful than it was. The course evaluation emphasizes the methodological choices over the actual results.

The course includes many opportunities for formal collaboration with your classmates. Study groups and peer-review are encouraged, but you should only turn in work that is yours. If you're unsure about a particular situation, please feel free just to ask ahead of time.

This course includes group projects, which can be stressful for students when it comes to dividing work, taking credit, and receiving grades and feedback. I have attempted to

make the guidelines for group work as clear as possible for each activity and assignment, but please let me know if your group is struggling.

Each student is responsible for reading and understanding the contents of the Code of Student Conduct (http://studentlife.osu.edu/csc/). This document describes activities, including copying colleagues' work and other forms of plagiarism, and violating exam rules, that are unacceptable behavior at the University. The University Rules (Faculty Rule 3335-5-487) mandate that I report any suspected academic misconduct to the Committee on Academic Misconduct (COAM), who will adjudicate each case. Any sanctions will be imposed by COAM, and can range from no action to expulsion from the University. I intend to obey this University Rule for this course. If you have any questions about what constitutes academic misconduct in this course, please contact the instructor.

Added 8/28/17: Official ASC statement on academic misconduct (https://asccas.osu.edu/curriculum/asc-syllabus-elements):
"It is the responsibility of the Committee on Academic Misconduct to investigate or establish procedures for the investigation of all reported cases of student academic misconduct. The term "academic misconduct" includes all forms of student academic misconduct wherever committed; illustrated by, but not limited to, cases of plagiarism and dishonest practices in connection with examinations. Instructors shall report all instances of alleged academic misconduct to the committee (Faculty Rule 3335-5-487). For additional information, see the Code of Student Conduct http://studentlife.osu.edu/csc/."

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 098 Baker Hall to coordinate reasonable accommodations for students with documented disabilities.

Added 8/28/17: Official ASC statement on disability services (https://asccas.osu.edu/curriculum/asc-syllabus-elements):
"Students with disabilities (including mental health, chronic or temporary medical conditions) that have been certified by the Office of Student Life Disability Services will be appropriately accommodated and should inform the instructor as soon as possible of their needs. The Office of Student Life Disability

Services is located in 098 Baker Hall, 113 W. 12th Avenue; telephone 614- 292-3307, slds@osu.edu; slds.osu.edu."

Tentative Schedule:

Days	Content	Reading
8/23/17-8/25/17	Welcome; Review of Maximum Likelihood and Bayesian Inference; Overview of Typical Bayesian Computation	Text Sections 6.1.1, 6.1.2, 6.1.4, 6.1.5; Casella and George, 1992 (doi:10.2307/2685208)
8/28/17-9/1/17	Introduction to Missing Data; Missing Data Patterns; Missing Data Mechanisms; Visualization	Text Sections 1.1-1.3; Templ, Alfons and Filzmoser, 2012 (doi:10.1007/s11634-011-0102-y)
9/6/17	Bayesian Approach	Text Section 10.1
9/8/17-9/13/17	Multiple Imputation, Bayesian Foundation (univariate and multivariate)	Text Sections 10.2, 6.1.3, 5.4
9/15/17	Full Conditional MI	TBA
9/18/17-9/25/17	MI shortcuts: Hot Deck, Improper, Monotone, Subset, Resampling	TBA
9/27/17-9/29/17	MI special topics: congeniality and sensitivity analyses	TBA
10/2/17	Single Imputation	Text Chapter 4
10/4/17-10/6/17	Variance Estimation for Single Imputation	Text Chapter 5
10/9/17	Special Topic: Longitudinal Data	TBA
10/11/17	Midterm Exam	NA
10/16/17-10/20/17	Likelihood-based approach	Text Sections 6.2-6.3
10/23/17-10/27/17	Likelihood-based approach	Text Sections 7.1-7.4, 9.1
10/30/17-11/3/17	EM algorithm and large sample inference	Text Sections 8.1-8.4 (skim 9.2)
11/6/17	Contingency Tables	Text Chapter 13
11/8/17-11/17/17	Complete Data Analysis; Available Case Analysis; Weighting Adjustments	Text Chapter 3
11/20/17	Midterm Exam	NA
11/27/17	Nonignorable Missingness Intro and	Text Sections 15.1, 15.2, 15.4
11/29/17-12/1/17	Poster Presentations	NA
12/4/17-12/6/17	Nonignorable Missingness Methods and Sensitivity Analyses	Text Sections 15.4. 15.5