BIOST/EPI 531: Statistical Methods for Analysis with Missing Data
Winter Quarter, 2019

Instructor: Mauricio Sadinle
Assistant Professor
Department of Biostatistics, University of Washington, Seattle

Contact Information: E-mail: msadinle@uw.edu
Office: Health Sciences Building F653
Website: http://faculty.washington.edu/msadinle/

Office Hours: Wednesdays 2:30 - 3:30pm and Thursdays 1:00 - 2:00pm, HSB F653.

Class Hours and Location: Mondays and Wednesdays 1:00pm – 2:20pm, South Campus Center (SOCC) 308.

Class Website: Syllabus, slides, assignments, and more will be available at:

https://canvas.uw.edu/courses/1254251

Course Description

This course formally introduces methodologies for handling missing data in statistical analyses. It covers naïve methods, missing-data assumptions, likelihood-based approaches, Bayesian and multiple imputation approaches, inverse-probability weighting, pattern-mixture models, sensitivity analysis and approaches under nonignorable missingness. Computational tools such as the Expectation-Maximization algorithm and the Gibbs' sampler will be introduced. This course is intended for students who are interested in methodological research.

Learning Objectives

Upon successful completion of this course you should be able to:

- Demonstrate the pitfalls of naïve methods for handling missing data, such as mean imputation and complete-case analysis
- Understand different missing-data assumptions and the concept of ignorability
- Implement basic likelihood-based approaches using the EM algorithm and Gibbs samplers
- Implement different versions of multiple imputation, understand their limitations and the requirement of congeniality
- Perform basic statistical analyses using inverse-probability weighting
- Understand the concepts of identifiability, observational equivalence, nonparametric identifiability
- Perform sensitivity analyses
- Critically evaluate the literature on methodologies for handling missing data

Prerequisites

While the UW Course Catalog does not list prerequisites for this course, it will be assumed that you have an intermediate knowledge of statistics. In particular, you should have familiarity with, concurrently be learning, or be willing to quickly catch up on the following topics:

- Basic calculus (e.g., integrals and derivatives)
- Basic probability theory (e.g., density functions, conditional probabilities, conditional independence, expectations, multivariate normal and multinomial distributions)
- Matrix notations and manipulations
- Basic statistics theory (e.g., i.i.d. data, likelihood functions, maximum likelihood estimation)
- Linear regression and logistic regression
- R programming for iterative algorithms

Class Schedule

This is the current schedule, after modifications due to snow days.

	Date	Topics	Homework	Reading
1.	Jan 7	Syllabus, motivating examples	HW0 posted	Chap 1
2.	Jan 9	General setup, missing-data mechanisms		Chap 1
3.	Jan 14	Naïve methods: imputation and complete cases		Chap 2
4.	Jan 16	R session 1	HW0 due, HW1 posted	
	Jan 21	No class (university holiday: M. L. King Day)		
5.	Jan 23	Likelihood-based methods		Chap 3
6.	Jan 28	The Expectation-Maximization algorithm		Chap 3
7.	Jan 30	R session 2	HW1 due, HW2 posted	
	Feb 4	Snow day		
8.	Feb 6	Introduction to Bayesian inference		Chap 4
	Feb 11	Snow day		
9.	Feb 13	Gibbs sampling, data augmentation		Chap 4
	Feb 18	No class (university holiday: Presidents Day)		
10.	Feb 20	Multiple imputation	HW2 due	Chap 4
11.	Feb 25	R session 3	HW3 posted	
12.	Feb 27	Inverse-probability weighting		Chap 5
13.	Mar 4	Doubly robust estimation		Chap 5
14.	Mar 6	R session 4	HW3 due, HW4 posted	
15.	Mar 11	Identifiability, nonignorable missing data		Chap 6
16.	Mar 13	Pattern-mixture models and sensitivity analyses		Chap 6, 7
	Mar 20		HW4 due	

Readings are from Davidian and Tsiatis' lecture notes.

Course Materials

Each lecture will have a set of slides that will often draw from the lecture notes developed by Marie Davidian and Anastasios Tsiatis at NC State. To the best of the instructor's knowledge, these lecture notes are the most comprehensive resource for learning about inference with missing data, as existing books typically focus on one particular approach. The lecture notes are available at:

`https://www4.stat.ncsu.edu/~davidian/st790/notes.html`

Dr. Davidian has graciously granted the instructor permission to use her course materials.

Additional resources include, but are not restricted to:

- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Hoboken, New Jersey: Wiley.
- Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.
- Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis With Missing Data*. New York: Wiley.
- Tsiatis, A. A. (2006). *Semiparametric Inference and Missing Data*. New York: Springer.
- Daniels, M. J. and Hogan, J. W. (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Boca Raton, Florida: Chapman & Hall.

Grading

The final grade will be based on five (5) homework assignments, HW0 – HW4. HW0 will be a short assessment of the prerequisites required for successful completion of this course, and it will be graded as 1 if most of your solutions are correct and 0 otherwise. HW1 – HW4 will each be worth 20 points. The final grade will be based on a monotonically increasing concave transformation of $G = g_0 \times (\sum_{i=1}^{4} g_j)$ to the [0,4] interval, where g_j is the grade of HWj. The homework assignments will be posted and due via the Canvas website on the dates indicated in the Class Schedule. You are encouraged to discuss homework assignments with your classmates, but solutions are to be your own. Late assignments will be graded over $\max(20 - 5 \times h/24, 0)$ points, where h is the number of late hours.

Course Policies

Attendance and Participation:

Although attendance and participation in class are not required, they are highly encouraged. Keep in mind the following points:

- Our sessions will occasionally have interactive parts, where you will be asked to comment on or solve certain problems. These problems will be part of the homework assignments, so if you come to class, you will have an advantage. There will also be five (5) interactive R sessions, as indicated in the Class Schedule, where we will implement the methodologies covered in the previous sessions. I encourage you to bring your laptops to the R sessions. Homework assignments will also rely on the code and examples covered in the R sessions.
- Attending and being active in class by asking and answering questions will benefit you and your classmates. If you have a question, do not be afraid to ask! Chances are that others are confused on the same point as well.
- If you miss a lecture it is your responsibility to catch up with the contents of the class.

Electronic Discussion Board and Email:

The Canvas web page contains an electronic discussion board. The board will be used for announcements and questions from the students. If a student has a question about the course that might be of interest to other students, s/he should post the question to the electronic discussion board rather than emailing the instructor. If the question is urgent, then the student may email the instructor in addition to posting on the discussion board. The discussion board can be used to discuss any topic related to the course material. Students are encouraged to answer as well as ask questions on the board.

Access and Accommodations:

Your experience in this class is important to me. If you have already established accommodations with Disability Resources for Students (DRS), please communicate your approved accommodations to me at your earliest convenience so we can discuss your needs in this course. If you have not yet established services through DRS, but have a temporary health condition or permanent disability that requires accommodations (conditions include but not limited to; mental health, attention-related, learning, vision, hearing, physical or health impacts), you are welcome to contact DRS at 206-543-8924 or uwdrs@uw.edu or disability.uw.edu. DRS offers resources and coordinates reasonable accommodations for students with disabilities and/or temporary health conditions. Reasonable accommodations are established through an interactive process between you, your instructor(s) and DRS. It is the policy and practice of the University of Washington to create inclusive and accessible learning environments consistent with federal and state law.

Academic Integrity:

Students at the University of Washington (UW) are expected to maintain the highest standards of academic conduct, professional honesty, and personal integrity.
The UW School of Public Health (SPH) is committed to upholding standards of academic integrity consistent with the academic and professional communities of which it is a part. Plagiarism, cheating, and other misconduct are serious violations of the University of Washington Student Conduct Code (WAC 478-120). We expect you to know and follow the university's policies on cheating and plagiarism, and the SPH Academic Integrity Policy. Any suspected cases of academic misconduct will be handled according to University of Washington regulations. For more information, see the University of Washington Community Standards and Student Conduct website.