

**BSTA 782**  
**Statistical Methods for Incomplete Data**  
**Spring, 2020**  
**701 Blockley**

---

- Instructor:** Qi Long, 201 Blockley, (215)573-0659, qlong@upenn.edu
- Lectures:** January 16 – April 28, Tue and Thu, 10:30-11:50 am, 701 Blockley  
No lectures on March 10 and 12 (Spring Break)  
Canceled lectures: February 13; March 3 and 24
- Office Hours:** Qi Long, by appointment
- Textbooks:** Statistical Analysis with Missing Data, 3rd Edition, by Little, R.J.A., and Rubin, D., John Wiley & Sons (2019).  
Suggested: Semiparametric Theory and Missing Data, by Tsiatis, A.A., Springer (2007).
- Prerequisites:** BSTA 621/622, BSTA 632, BSTA 651, or their equivalents; permission of instructor. Knowledge about Bayesian modeling, though not required, can be helpful.
- Course Description:** This course reviews the theory and methodology of incomplete data, covering missing data patterns, missing data mechanisms including MCAR, MAR, and MNAR, potential impacts of missing data on data analysis; imputation methods; likelihood-based methods for handling missing data; computational methods such as the EM algorithm and its extensions; semiparametric methods for missing data such as IPW and AIPW; methods for MNAR and nonignorable missingness including sensitivity analysis. If time permits, it will also cover other advanced topics on handling incomplete data, such as in high-dimensional data or distributed analysis setting.

## Outline of Lectures

- Part 1: Introduction (missing Data Patterns; missing Data mechanisms; overview of missing data methods).
- Part 2: Ad Hoc Methods for Handling Missing Data (complete-case analysis; available-case analysis; LOCF).
- Part 3: Single and Multiple Imputation Methods.
- Part 4: Inverse Probability Weighting (IPW) and Augmented IPW Methods.
- Part 5: Likelihood-based Methods; EM algorithm.
- Part 6: Methods for Handling Missing Not At Random (pattern mixture models; selection models; sensitivity analysis)

## Grading Policy:

- Attendance and Participation @ 20%
- Homework @ 35%
- Final Project Report @ 25%
- Final project Presentation @ 20%

## Grades:

- $(85, 100] \approx A$
- $(75, 85] \approx B$
- $(59, 75] \approx C$
- +/- grades will be given accordingly.

**Final Project:** Options are: A) analysis of a real study that involves incomplete data; B) conduct simulation studies on incomplete data methods (e.g., comparing current imputation methods in simulations); and C) refine existing methods for incomplete data. For all options, students are encouraged to discuss potential topic(s) with the instructor.

Tentative Schedule: Each student should select the topic for the final project, submit and discuss his/her project proposal with the instructor in person by April 7; the proposal needs to be approved by the instructor. Each student will give an oral presentation for the final project on April 28 at 10:30am-12:10pm (Location: 701 Blockley). There will be 18 minutes for each presentation and 2 minutes for Q&A. Grading for the final presentation is based on the 18-min presentation only and adherence to the time limit will be a factor in grading. The written report for the final project is due by noon on April 30.