BSTA 782
Statistical Methods for Incomplete Data
Spring, 2022
Blockley 235

Instructor:	Qi Long, Blockley 201, (215)573-0659, qlong@upenn.edu
Lectures:	March 14 – April 27, Mon and Weds, 1:45-3:15 pm, Blockley 235
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).
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 methodology of incomplete data, covering missing data patterns, missing data mechanisms (MCAR, MAR, and MNAR), 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.

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: Likelihood-based Methods; EM algorithm.

- Part 5: Inverse Probability Weighting (IPW) and Augmented IPW Methods.

- Part 6: Methods for Handling Missing Not At Random (pattern mixture models; selection models; sensitivity analysis)

Grading Policy:

- Attendance and Participation @ 30%

- Homework @ 40% (2-3 assignments)

- Final Presentation @ 30%

Grades:

- $(85, 100] \approx A$

- $(75, 85] \approx B$

- $(59, 75] \approx C$

- +/- grades will be given accordingly.

Final Presentation: The final presentation will entail a review of 1-2 papers related to analysis of incomplete data and is tentatively scheduled for May 2 at 1-3pm. There will be 15 minutes for each presentation and 2 minutes for Q&A. Adherence to the time limit will be a factor in grading.