BIOST/EPI 531: Statistical Methods for Analysis with Missing Data

MW 10:30-11:50 (SOCC-308)

Gary Chan, PhD,
Associate Professor,
Department of Biostatistics
kcgchan@u.washington.edu

Office hour: Mon 9:00am-10:00am or by appointment (F-653)

Course Web Site
https://canvas.uw.edu/courses/1097539

Course Description

This course introduces statistical methods for missing data analysis, focuses particularly on likelihood and Bayesian methodologies such as EM algorithm and multiple imputation. Inverse probability weighting will also be introduced. Different types of missing mechanisms will be explored. This course is intended for biostatistics PhD students or epidemiology PhD students who are interested in methodological research. Adequate programming experience in R or other statistical software is assumed and extensive simulations and programming exercise will be conducted to study the various statistical methodologies. We will also draw links to important epidemiology literature on missing data analysis.

Course learning objectives

After completing the course you should be able to:

1. Understand different assumptions on missing data mechanism.
2. Understand the difference between inverse probability weighting, maximum likelihood and multiple imputation.
3. Apply suitable analysis based on different model assumptions.
4. Write small programs to perform basic missing data analysis.
5. Critically evaluate the epidemiology literature on missing data.

Optional Textbooks

Molenberghs and Kenward (2007). Missing Data in Clinical Studies. Wiley.

Carpenter and Kenward (2013) Multiple imputation and its application. Wiley.

Little and Rubin (2002). Statistical analysis with missing data. Wiley.

Schafer (1997). Analysis of incomplete multivariate data. Chapman.

Grading

Three assignments (60%), due Wednesdays in class (or through email before class). Final project (40%). Two possible options for a final project. (1) If you have a data set in hand with missing data, you can write an analysis proposal describing the data and how you could analyze the data. Explain the pros and cons of your proposal. (2) You can focus on a published paper with missing data analysis, describe what the authors have done and the underlying assumptions for the analysis. Critique the method.

Tentative Course Schedule

Week	Dates		
1	1/4	Introduction and missing data mechanisms	
2	1/9, 1/11		
3	1/18	Propensity score and weighting	
4	1/23, 1/25		HW 1
5	1/30, 2/1	Maximum likelihood and EM algorithm	
6	2/6, 2/8		HW 2
7	2/13, 2/15	Bayesian inference and Multiple imputation	
8	2/22		
9	2/27, 3/1		HW 3
10	3/6, 3/8	Review	