

STAT 652: Applied Statistical Analysis with Missing Data

Autumn Quarter, 2011

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

Office Hours: Thursday 10:30-11:30am, or by appointment

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 the missing data in practice.

Course Objectives:

After the completion of this course, the students are expected to

1. Understand the missing data mechanism, the underlying assumptions and identify different patterns of missing data
2. Understand the difference in statistical analysis between missing data problem and complete data problem (including weighted methods)
3. Be able to perform simple missing data analysis with single imputation; comprehend its weakness
4. Be able to implement likelihood-based analysis with ignorable missing response; implement 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 Texts:

Statistical Analysis with Missing Data, Little and Rubin, 2nd edition, Wiley.

Reference:

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

Software:

The use of computer in data analysis is essential for this course. There are many different algorithms used in analysis of missing data, and so we will be discussing and using several different analysis packages. However, the software program SAS will be primarily used for examples, homework and exams.

Prerequisites:

PH-BIO 703, STAT 529, or equivalent (or permission of the instructor)
Knowledge of regression and familiarity with a statistical computing package are necessary.

Lectures:

Tu/Th 8:30AM - 10:18AM Journalism Bldg 0371

Tentative Schedule:

Week	Content	Readings
1	Missing data mechanism	Chap 1
2	Complete-data and Available-case analysis; Nonresponse weighting in sample survey	Chap 3
3-4	Single value imputation	Chap 4, 5
5-6	Likelihood based approach and EM algorithm for ignorable missing data	Chap 6, 8
7	Bayesian approach and Multiple Imputation	Chap 10
8	Methods for Contingency tables and Mixed variables	Chap 13, 14
9-10	Recent developments for non-ignorable missing data	Chap 15

Grading:

Homework	30%
Midterm Exam	30%
Final Exam	40%

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 150 Pomerene Hall to coordinate reasonable accommodations for students with documented disabilities.

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.