

BSTA630: Statistical Methods and Data Analysis I

Fall Semester, 2020

Introduction

This course is an introduction to statistical methods and inference for Biostatistics degree candidates. It module covers basic theoretical concepts, estimation, hypothesis testing, one and two-sample tests, analysis of variance (ANOVA), categorical data analysis, and linear regressions. Statistical analyses are implemented primarily using R (<https://www.r-project.org>). A companion SAS (https://www.sas.com/en_us/software/stat.html) codes are available for most lectures.

Note: A working knowledge of calculus and linear algebra and one introductory statistics course are required. Students are expected to have formal training in these areas, and/or to receive permission from the instructor prior to enrolling. In addition, familiarity with concepts in probability is need while these concepts will be only reviewed briefly at the beginning of the course. BSTA 630 is a core course in the Biostatistics graduate program and many students report that it requires a lot of work. Please carefully consider your workload before enrolling in the course.

Description

- Lecture time: 10:30am – 11:50am Tuesdays and Thursdays
- Instructor: Rui Xiao (rxiao@pennmedicine.upenn.edu, 206 BH); Instructor office hours (virtual): by appointment
- TA: Rebecca Deek (Rebecca.Deek@pennmedicine.upenn.edu); TA office hour (virtual): TBD

Textbooks

- Required: Fundamentals of Biostatistics, 8th Edition (Bernard Rosner).

Course Goals

- Become familiar with exploratory data analysis, estimation, hypothesis testing, linear models, and generalized linear models.
- Understand how the development of statistical methodology is motivated by biomedical problems.
- Develop data analytic skills including familiarity with at least one statistical software program.
- Know how to select and evaluate appropriate methods.
- Understand the relationship among methods.
- Be able to interpret statistical results to other professionals.
- Improve and polish writing skills needed to communicate results of data analyses.

Special Instructions

- Statistical methods will be implemented using statistical software R (SAS codes may be available for some lectures).
- Lecture notes, programs and datasets for examples will be available on <https://canvas.upenn.edu>.
- R online resources are available at CRAN: <http://www.r-project.org>.
- SAS online documentation is available at <http://support.sas.com/documentation/>.

Evaluation

Around 8 homework assignments (30%); midterm take-home exam (40%); final data analysis project (30%).

Grading Scale

90– < 92.5 A-	92.5– < 97.5 A	97.5– ≤ 100 A+
80– < 82.5 B-	82.5– < 87.5 B	87.5– < 90 B+
70– < 72.5 C-	72.5– < 77.5 C	77.5– < 80 C+
60– < 70 D	0– – < 60 F	

Homework

- Homework problems are from the following two sources.
 1. Taken directly from the text and meant to reinforce and review concepts presented in class.
 2. Developed by the instructors and may require an extension or application of the methodological concepts developed in class or in the text. They may be based on scenarios encountered by collaborative statisticians and require creative thinking and integration of concepts to solve. At least once during the course we will ask you to construct a simulation to evaluate a statistical approach.
- Submit homework to TA via email, or through Canvas. Homework is due one week after it is assigned or announced by the instructor otherwise. Late homework will NOT be accepted in general.

Data Analysis Project

Each student is required to work on one real data project provided by the instructor. Over the course of the semester the student will determine which techniques introduced in the course are appropriate for describing the data and answering the scientific questions of interest. All students will be required to submit a written report at the end of the semesters.