HST-190: Introduction to Biostatistics
August 2018

Instructors:

Sebastien Haneuse, Ph.D.
Associate Professor of Biostatistics
Harvard T.H. Chan School of Public Health
Building 2, Room 407
shaneuse@hsph.harvard.edu

Victoria Liublinska, Ph.D.
Sr. Research Analyst
Office of Institutional Research
Harvard University
Victoria_Liublinska@harvard.edu

Teaching Fellow:

Tom Madsen
PhD Candidate
Harvard Biostatistics Department
tmadsen@g.harvard.edu

Summary of Course:

This course will present the fundamentals of biostatistics with the aim of training students to comprehend, critique and communicate findings from the biomedical literature. Students will learn how to assess the importance of chance in the interpretation of experimental data. Major topics covered include probability theory, chi-squared and t-tests, ANOVA, linear and logistic regression, survival analysis, as well as how to perform statistical analysis using MATLAB. These topics will be introduced together with critical readings of studies published in the medical literature.

Meeting Times:

MWF 9:00-12:00, Medical Education Center (Bldg. E), Room 209 (see syllabus for exceptions)

Office Hours:

Sebastien: By appointment
Victoria: Wednesdays after class, noon-1pm, lobby outside of Room 209
Tom: Mondays & Fridays, noon-1pm, lobby outside of Room 209

MATLAB:

Assignments and lectures throughout the course will require the use of (fairly light) statistical programming. The language that will be used for instruction and that will be strongly recommended for homework assignments will be MATLAB. As such, please ensure that you have a valid copy of MATLAB installed on your computer (Student Version is freely available via either Harvard or MIT).

For those with prior programming experience: If you have a strong preference to use another language for your assignments (R, STATA, SAS, Python, etc.), we will not expressly forbid you from doing so. However, do keep in mind that MATLAB is the official language of the course and will be the language used in all demos and instruction.

Required Text Book:

Rosner, *Fundamentals of Biostatistics,* seventh (or higher) edition, Duxbury Press, Boston, 2010.

Additional References:

- Pagano and Gauvreau, *Principles of Biostatistics*, second edition, Duxbury Press, Boston, 2000.
- B. Baldi and D. Moore, *The Practice of Statistics in the Life Sciences*,
- M. Whitlock and D. Schlucter, *Analysis of Biological Data*
- Lock×5., *Statistics, Unlocking the Power of Data*

Course Web Page:

We will be using the Canvas HMS course webpage hosted at https://canvas.hms.harvard.edu/ to host course materials, submit homework assignments, and send out announcements.

Course Components:

- Lectures & Reading: MWF 9am-12pm. Each lecture has a reading associated with it, which is due **before** the lecture.
- Group work: MWF around 10am-11am. During a 30-50 min. break between lectures students will be given assignments to complete in small groups.
- Problem Sets: There will be 7-8 homework assignments to complete individually.
- Final project: The final project will be completed in small groups (2-3 students). You will apply the methods covered in this class to a data set of your choice. This is your opportunity to pursue a direction of academic or personal interest and receive feedback from the teaching staff. Project requirements include creating short presentation and presenting it during the penultimate meeting.
- Final Exam: a 3-hour in-class exam will take place during the last meeting.

Note that lecture attendance is mandatory. A final score (P/D/F) will consist of the following components with the corresponding weights: homework-30%, project-20%, exam-30%, class participation (discussion during lectures, group exercises, Piazza)-20%.

Communication:

We will be using Piazza to answer questions and generate online class discussions. The platform is designed to help you get assistance quickly and efficiently from the teaching staff as well as your classmates. Rather than emailing questions, we encourage you to post your questions on Piazza. Follow instructions on the Canvas course page on how to register and login.

Brief Piazza etiquette:

1. Do not post answers to homework problems, or give strong hints to solutions.
2. It helps to ask questions! It may save hours of confusion or fruitless work.
3. We encourage you to answer questions on Piazza (but remember #1).

Collaboration Policy and Academic Integrity:

You are encouraged to (orally) discuss problem sets with your classmates, but each student must write up solutions separately. Be sure that you have worked through each problem yourself and that the answers you submit are the results of your own efforts. You also may not submit output from another student's computer session. If you ever have a question about whether your level of collaboration is appropriate, ask the instructor! You are required to explicitly acknowledge collaborators by writing their names at the top of your homework assignments.

	Lecture Date	Lecture Topic	Group activity
1	Mon, July 30	Random variables, probability, Bayes theorem, sensitivity, specificity, ROC. Common probability distributions (Bernoulli, Binomial, Normal), definition of expected value and variance.	Exercise solving
2	Wed, Aug 1	Population vs. sample. Descriptive statistics, Estimation, Confidence intervals for population mean. Hypothesis testing. One-sample inference: t-test.	Exercise solving
3	Thur, Aug 2 **1pm-4pm Room: TMEC 227**	Likelihood and MLE, Power of a test and sample size. Inference for two groups: two-sample t-test. Special types of t-tests.	Exercise solving
4	Mon, Aug 13	Nonparametric tests. Data types, study design principles and terminology. Clinical trials: structure, endpoints, (all methods will be illustrated using clinical trials publications); instruction on critical and close reading of statistical aspects of the medical literature.	Exercise solving
5	Wed, Aug 15 **8am-11am**	Correlation (Pearson and non-parametric), Simple and multiple linear regressions for continuous data.	Exercise solving
6	Fri, Aug 17	R-squared statistic, ANOVA, Mutliple Comparisons, Bonferroni correction, PCA.	Exercise solving
7	Mon, Aug 20	Inference for categorical data: 2x2 tables, RxC tables, odds-ratio, relative risk, Mantel-Haenszel.	Exercise solving and project discussion
8	Wed, Aug 22	Logistic regression for binary and ordinal data; Survival analysis: Kaplan Meier estimation, logrank test, Cox proportional hazards model, more complex survival data (truncation, interval censoring).	Exercise solving (optional) and project discussion
9	Fri, Aug 24	Linear mixed-effects models, model checking, variable selection.	Project discussion
10	Mon, Aug 27	Presentations of group projects.	
11	Wed, Aug 29	In-class exam.	