Biostatistics I

Statistics starts with data (Leo Breiman, 2001).

Fall Semester 2019
Laurel Building Room C1 A&B
402 East 67th Street
Monday & Wednesdays, 1:00-3:00 p.m.

Instructors: Karla V. Ballman, PhD [LA-225, kab2053@med.cornell.edu, (646) 962-8023]
Imaani Easthausen, MS [LA-0003D, ime2003@med.cornell.edu, (646) 962-8060]
Kaylee Ho, MS [LA-0007, kah4001@med.cornell.edu, (646) 962-8024]
Hanchao Zhang, MS [hz1641@nyu.edu, (646) 501-3422]

Office Hours: by appointment with Karla (email her)
TBA with Imaani
TBA with Kaylee
TBA with Hanchao

Course Grade

A+	98-100%	4.3
A	93-97%	4.0
A–	0-92%	3.7
B+	88-89%	3.3
B	83-87%	3.0
B–	80-82%	2.7
C+	78-79%	2.3
C	73-77%	2.0
C–	70-72%	1.7
D	60-69%	1.3
F	Below 60%	0.0

I. Objective

An introduction to the fundamentals of biostatistics with primary emphasis on understanding of statistical concepts behind data analytic principles. This course will also teach R, a freely available software, to explore, visualize and perform statistical analysis with data. Topics covered include: exploratory data analysis; basic concepts of statistics; construction of hypothesis tests and confidence intervals; performance of statistical comparisons; simple modeling; and determination of power and sample size.

II. Course aims and outcomes

The aim of this course is to provide students with a set of foundational statistical and computational techniques that will be the building blocks for the other courses in the program.

Learning outcomes

By the end of the course the student will be able to:

- Summarize data with numerical measures and graphical techniques
- Perform basic data manipulations in R
- Identify relationships in data distributions via visual displays
- Calculate and interpret confidence intervals for population means and proportions, both via standard methods and bootstrap
- Interpret and explain a p-value
- Perform comparisons of two populations with standard, parametric tests
- Perform comparisons of two populations with nonparametric methods (e.g. permutation tests)
- Determine sample size and power for comparisons of two populations
- Fit and interpret linear regression models
- Use R to describe, visualize, analyze, and simulate data

III. Course Format

- there will be two 2-hour lectures each week
- there will be weekly homework exercises related to the lecture
- there will be one mid-term exam and one final exam

IV. Approach

This course will make extensive use of computational techniques using the statistical programming language R. Unlike other introductory Biostatistics courses, we will not start from probability theory; instead we will first introduce the statistical concepts, with the aid of simulations and basic computer programming. We will introduce probabilistic distributions and parametric statistical families when needed. Development of R programming skills will be a key component of the course, and students are expected to be advanced R programmers by the end of the course. The classes will be a mix of lecture and activities, with required readings before each class. Students will also be expected to use Rmarkdown to generate deliverables.

V. Course requirements

a) course prerequisites: college-level calculus and a basic knowledge of the R programming language

b) laptop with R and RStudio installed

Your laptop should be brought to each class.

VI. Course materials

a) Required texts

- Baumer, Kaplan, Horton. *Modern Data Science with R.* CRC Press, 2017.
- Wickham, Grolemund. *R for Data Science*. O'Reilly, 2017. Free: http://r4ds.had.co.nz
- van Belle, Fisher, Heagerty, Lumley. *Biostatistics A Methodology for the Health Sciences, Second Edition*. Wiley & Sons, 2004.

b) Lecture notes available on Canvas

c) Further readings, suggested during class and posted on Canvas

VI. Grading components

- Final Exam: 30%
- Mid-term exam: 20%
- Weekly assignments (7 best out of 8): 20%
- Final group project: 25%
- Class participation: 5%

Note that no late assignments will be accepted. The seven best assignments will count towards the final grade. Class participation will be assessed by the number of questions asked and answered in class and on Canvas. Active engagement in the classroom activities will also be a component of overall class participation.

VI. Academic Integrity

Each student in this course is expected to abide by the Cornell University Code of Academic Integrity. Any work submitted by a student or student group in this course for academic credit will be the student's (group's) own work. For this course, collaboration is allowed on the assignments but only between the students in the group. Assignments can be done individually or in groups of two. If done in a group of two, each person should sign their name to the homework and only submit one homework (both group members will get the same grade for the homework).

You are encouraged to study together and to discuss information and concepts covered in lecture with other students. You can give "consulting" help to or receive "consulting" help from other students. However, this permissible cooperation should never involve one student (or group) copying of all or part of work done by another person (or group), in ANY form (e.g. an e-mail, an e-mail attachment file, a storage device, a hard copy, etc).

Should copying occur, all parties involved will automatically get half the total points for the assignment. Penalty for violation of this Code can also be extended to include failure of the course and University disciplinary action.

During examinations, you must do your own work. Talking or discussion is not permitted during the examinations, nor may you compare papers, copy from others, or collaborate in any way. You are also not permitted to communicate electronically or in any other format. **Any collaborative behavior during the examinations will result in failure of the exam for all parties involved and may lead to failure of the course and University disciplinary action.**

VII. Accommodations for students with disabilities

In compliance with the Cornell University policy and equal access laws, I am available to discuss appropriate academic accommodations that may be required for students with disabilities. Requests for academic accommodations are to be made during the first three weeks of the semester, except for unusual circumstances, so arrangements can be made. Students are encouraged to register with Student Disability Services to verify their eligibility for appropriate accommodations

VIII. Inclusivity Statement

We understand that our members represent a rich variety of backgrounds and perspectives. The Biostatistics and Data Science Program is committed to providing an atmosphere for learning that respects diversity. While working together to build this community we ask all members to:

- share their unique experiences, values and beliefs
- be open to the views of others
- honor the uniqueness of their colleagues
- appreciate the opportunity that we have to learn from each other in this community
- value each other's opinions and communicate in a respectful manner
- keep confidential discussions that the community has of a personal (or professional) nature
- use this opportunity together to discuss ways in which we can create an inclusive environment in this course and across the Cornell community

IX. Tentative Course Schedule (*may change to accommodate guest presenters, instructor's schedule, and times)*

Week	Date	Topic
1	**9/9/19**	course overview; exploratory data analysis (EDA)
	9/12/19	EDA continued; basic probability
2	**9/16/19**	binomial distribution; normal distribution
	9/18/19	sampling distribution of a proportion
3	**9/23/19**	estimation of a proportion
	9/25/19	test of a proportion; bootstrapping a proportion
4	**9/30/19**	tests of significance; hypothesis testing
	10/2/19	sampling; resampling; the bootstrap
5	**10/7/19**	sample size and power
	10/9/19	sampling distribution of a mean; central limit theorem; estimation
6	**10/14/19**	test of a mean / sample size and power

Week	Date	Topic
	10/16/19	relationship between two categorical variables; 2×2 tables
7	**10/21/19**	mid-term exam
	10/23/19	comparison of two proportions
8	**10/28/19**	analysis of $k \times p$ tables
	10/30/19	comparison of two means
9	**11/4/19**	analysis of variance
	11/6/19	relationship between two continuous variables
10	**11/10/19**	simple linear regression
	11/13/19	linear regression models
11	**11/18/19**	linear regression diagnostics
	11/20/19	TBD
12	**11/25/19**	association / causation / applications
	11/27/19	association / causation / applications
13	**12/2/19**	final project presentations
	12/4/19	final project presentations
14	**12/09/18**	final exam (Time and Place TBD)