COURSE DESCRIPTION

This one-semester course introduces basic applied descriptive and inferential statistics. The first part of the course includes elementary probability theory, an introduction to statistical distributions, principles of estimation and hypothesis testing, methods for comparison of discrete and continuous data including chi-squared test of independence, t-test, analysis of variance (ANOVA), and their non-parametric equivalents. The second part of the course focuses on linear models (regression) theory and their practical implementation. This part also introduces mixed-effects models and concepts of survival analysis.

COURSE LEARNING OBJECTIVES

Students who successfully complete this course will:

- Be able to distinguish among different types of data and correctly apply statistical methods of analysis including summary, descriptive statistics
- Know how to utilize probability distributions and their properties
- Be able to formulate and assess statistical hypotheses
- Have a good understanding of linear regression models, theory and applications, for both fixed and random effects
- Be able to apply various regression techniques to real data projects
- Be familiar with survival analysis concepts and Kaplan-Meier method of estimation
- Be able to use R and/or SAS for data management, analysis and results interpretation

CLASS SESSIONS

Tuesdays and Thursdays, 10:00-11:20am, P&S Building 7th Floor Amphitheatre (AMP 7)
*On Tuesday, Oct 24th 2017, class will be held in P&S Building AMP 1

RECITATION SESSIONS

Mondays and/or Wednesdays, 5:30-6:50pm, Hammer LL203
First day of recitation: Monday, Sept 18th 2017

INSTRUCTOR

Cody Chiuzan, PhD
Department of Biostatistics, 6th floor, room 651
Mailman School of Public Health
Email: cc3780@cumc.columbia.edu
Tel: (212) 305-9107
Office Hours: By appointment (email me)

TEACHING ASSISTANTS

Zilan Chai, email: zc2326@cumc.columbia.edu
Rebecca Deek, email: rad2186@cumc.columbia.edu
Yutao Liu, email: yl3050@cumc.columbia.edu
Shuang Wu, email: sw3140@cumc.columbia.edu
Office Hours: Mondays and Wednesdays, 6:50-7:30pm, Hammer LL203

PREREQUISITES

Students are required to have working knowledge of calculus and linear algebra. Some programming experience with R and/or SAS is recommended, but not required.

TEXTBOOKS

Required:

- Rosner, B. *Fundamentals of Biostatistics*, 8th edition, Books/Cole, Cengage Learning, 2011.
- Kutner M.H., Nachtsheim C.J., Neter J., Li W., *Applied Linear Statistical Models*, 5th edition, McGraw-Hill International, 2013.

Recommended:

- Searle S.R., *Linear Models*, Wiley, 1997.
- Faraway J., *Linear Models with R*, CRC Press, 2005.
- Peng R., *R Programming for Data Science*, 2016.
- Der G. and Everitt B.S., *Applied Medical Statistics Using SAS*, 2nd edition, Taylor & Francis, 2012.
- Hosmer D.W., Lemeshow S., May S., *Applied Survival Analysis: Regression Modeling of Time-to-Event Data,* Wiley & Sons, Inc., 2008.

ASSESSMENT AND GRADING POLICY

Student grades will be based on:

Homework 20%
Exam 1 ..... 25%
Exam 2 ..... 25%
Final Project 25%
Classroom Participation……..... 5%

There will be **5-7 homework assignments** during the semester. Assignments may include derivation of theoretical properties, data analyses and programming. All assignments must be submitted electronically. Theoretical derivations can be handwritten and scanned, but only clear and legible handwriting will be graded. Other work should be typeset, e.g. MS Word, LaTeX, R Markdown, your choice.

The closed-book exams will be held in-class on**:**
Exam 1: Oct 12th, 2017 @ 10:00am and Exam 2: Nov 21st, 2017 @ 10:00am.
If needed, a formula sheet will be provided to you on the day of the exam. Also, bring a calculator, as you will not be allowed to use your cell phones for any reason.

The final (group) project will consist of an extended data analysis and a brief, but *well-structured report*. The project will be due on **Dec 15th, 2017 @ 5:00pm.**

Recitation sessions will consist of a combination of practice problems and R/SAS software lab. Even though these sessions are not mandatory, students are strongly encouraged to attend, as the content will complement the class lectures and contain important software procedures.

Classroom participation is important! A continuous dialogue between students and instructor is always beneficial in addressing misunderstandings or raise interesting problems. Sporadically, you will also have class exercises to apply various statistical concepts/methods. Students are expected to solve these exercises and, sometimes, even present their solutions to the entire class.

COURSE REQUIREMENTS

All students are expected to attend class regularly. As a courtesy to both your instructor and your classmates, please DO NOT be late.

Working together on homework is very much encouraged, but all write-ups must be done independently and clearly indicate the submitter's understanding of the material.

The in-class exams should be an individual effort, but the final project will be a group assignment. The instructor will randomly assign groups of 3-4 students and everybody is expected to participate. In order to emphasize the importance of the team-work (essential for a biostatistician), 10% of the project grade will count towards your individual participation to the group project.

All assignments turned in after the due date will not be accepted and 'rewarded' with a zero grade.

COURSE STRUCTURE

The course covers a large amount of material consisting of lecture notes and code examples. This syllabus is designed to give an overview of the course layout and a guide for topics to be covered; **please note that the order and the content of the topics may change** during the semester.

We will use R and/or SAS for data analysis. All course materials (i.e., lecture notes, assignments, data sets) will be posted on **CourseWorks** and on the course website.

MAILMAN SCHOOL POLICIES AND EXPECTATIONS

Students and faculty have a shared commitment to the School's mission, values and oath. mailman.columbia.edu/about/mission-history.

Academic Integrity

Students are required to adhere to the Mailman School Honor Code in the student handbook. mailman-handbook.com/2015/node/165.

Disability Access

In order to receive disability-related academic accommodations, students must first be registered with the Office of Disability Services (ODS). Students who have, or think they may have a disability are invited to contact ODS for a confidential discussion at 212.854.2388 (V) 212.854.2378 (TTY), or by email at disability@columbia.edu. If you have already registered with ODS, please speak to your instructor to ensure that s/he has been notified of your recommended accommodations by Lillian Morales (lm31@columbia.edu), the School's liaison to the Office of Disability Services.

COURSE SCHEDULE

Week	Date	Topic
1	Sep. 5	Introduction to Biostatistics: Types of Data, Study Designs
2	Sep. 7	Descriptive Statistics
3	Sep. 12	Basic Probability Concept and Common Distributions (1)
4	Sep. 14	Basic Probability Concept and Common Distributions (2)
5	Sep. 19	Methods of Inference for One-Sample Mean
6	Sep. 21	Methods of Inference for Two-Sample Means
7	Sep. 26	Methods of Inference for 3+ Sample Means
8	Sep. 28	Methods of Inference for One-Proportion
9	Oct. 3	Methods of Inference for Two-Proportions
10	Oct. 5	Measures of Association for Categorical Data
11	Oct. 10	Review 1
12	Oct. 12	**EXAM 1**
13	Oct. 17	Correlation and Simple Linear Regression (SLR)
14	Oct. 19	Estimation and Inference in SLR
15	Oct. 24	Multiple Linear Regression (MLR)
16	Oct. 26	ANOVA Testing in MLR
17	Oct. 31	Model Diagnostics MLR
18	Nov. 2	Model Selection and Validation in MLR
19	Nov. 7	**NO CLASS**
20	Nov. 9	'Non-Linear' Models: Regression Splines and Polynomials
21	Nov. 14	Full-Rank and Less Than Full-Rank Models Missing Data in Linear Regression Framework
22	Nov. 16	Review 2
23	Nov. 21	**EXAM 2**
24	Nov. 23	**NO CLASS (Thanksgiving)**
25	Nov. 28	Introduction to Mixed Models (1)
26	Nov. 30	Introduction to Mixed Models (2)
27	Dec. 5	Introduction to Survival Analysis (1)
28	Dec. 7	Introduction to Survival Analysis (2) **Last Day of Class**
29	Dec. 15	**Final Project DUE**