STAT 20 FALL 2018

Introduction to Probability and Statistics

INSTRUCTOR:

Shobhana Murali Stoyanov (shobhana@berkeley.edu)

GSIS:

Amanda Ngo (ngoal@berkeley.edu)
Cara Nolan (cara_nolan@berkeley.edu)
Lei Zhang (lei_zhang@berkeley.edu)
Namita Trikannad (namita_trikannad@berkeley.edu)
Sarang Chaudhary (sarang_chaudhary@berkeley.edu)
Wenlong Mou (wmou@berkeley.edu)
Melody Liu (mengling_liu0223@berkeley.edu)

TIME AND LOCATION:

The lectures will be TuTh from 9:30 AM - 11 AM in 2050 VLSB (Valley Life Sciences Building).

TEXT:

Statistics, 4th edition, by Freedman, Pisani, and Purves.
Supplemental notes will be posted in a couple of weeks, and please make sure to review the lecture slides.

OFFICE HOURS:

Shobhana: Wednesdays from 10 AM - 12 noon, and by appointment, in Evans 333
GSIs: TBA - Note that you can go to *any* GSI's office hours, not just your own.

LAPTOPS & R:

We will be working with the software R to enhance and deepen your comprehension of the concepts that you will be studying, and to provide you with tools that you can use for analyzing data.
You will need to download both R and the environment for R called RStudio. The GSIs will have section during the first Thursday (8/23) to help you with the downloads. Please have both R and RStudio downloaded and ready to go by next week.
Other than when you are explicitly expected to use your laptops for working on problems during class, please do not take out your laptops nor other electronic devices during lecture. You should take notes by hand - research indicates that students who write their notes long-hand tend to have better retention of the material.

CLICKERS:

Please bring your clickers to class, we will have some clicker questions each lecture beginning August 28. Sometimes they will cover concepts studied in previous weeks, and sometimes the current topic. You will be graded on participation, not correctness.

STAT 20 FALL 2018

DISCUSSION FORUM:
We will be using Piazza for discussions. If you have a question (that is not of a personal nature, but about the material) please post it to the class piazza site. The GSIs and I will monitor Piazza, but I encourage you to answer each others' questions. That said, I also want you to think about the problem *before* posting it on Piazza. You don't want to become too reliant on hints.

SECTIONS:
Sections meet on Tuesdays and Thursdays. They are run by the GSIs and attendance is **mandatory**.

HOMEWORK:
There will be little homework assignments assigned after each lecture that will be due by 6am the day of the *next* lecture. These will be just a few multiple choice problems on the most recent lecture material. In addition, you will turn in weekly homework assignments that you will need to upload to Gradescope, which is the website that you will use to submit your homework. The homework will consist on selected problems from the text and some R-programming assignments and will be graded only on completion and **not** on correctness.

QUIZZES AND EXAMS:
There will be 15 minute quizzes during section (almost) **every** Thursday. There will be 11 or 12 quizzes in total, and I will drop the **three** lowest scores while computing your grade. In addition, there will be one **in-class** midterm on **Thursday, October 4**, and a comprehensive final exam on **Tuesday, December 11**, from **3-6 PM** in a location that will be announced later. The quizzes will consist of problems from the text, many of which will have been covered in homework, and also R-related material.

DATA ANALYSIS PROJECT:
You will all turn in a data analysis project at the end of the semester which will use the skills and knowledge you will have developed throughout the semester.

GRADING:
- Clicker grade (the lowest **four** will be dropped): **3%**
- Post-lecture multiple choice assignments: **2%** (the lowest **three** will be dropped)
- Weekly homework sets: **7%** (the lowest **three** will be dropped)
- Weekly Quizzes: **15%** (the lowest quiz score will be dropped)
- Data analysis project: **10%**
- Midterm: **18%**
- Final: **45%** (if you can't take the final, please **do not** take the class - you will get a failing grade)

This class is graded on a curve. Your final letter grade is calculated based on your percentile in the class

(more or less) according to the following grading scheme (mandated by the statistics department):
A+ 98-100 percentile / A 90-98 percentile / A- 80-90 percentile (top 20% get some kind of A)
B+ 70-80 percentile / B 60-70 percentile / B- 50-60 percentile (next 30% get some kind of B)
C+ 40-50 percentile / C 20-40 percentile / C- 10-20 percentile (next 40% get some kind of C)
D/F 0-10 percentile
Of course, if students do very well, we will not curve down.

ABOUT THE COURSE & LEARNING GOALS

Stat 20 is an introductory course and does not assume prior knowledge of any probability or statistics. We will discuss examples from various fields, and some mathematical background such as calculus is assumed, mostly to make sure that you have some level of mathematical maturity. It is difficult to succeed in today's world without a solid understanding of basic statistics in the fields of business and economics, or just to be an informed citizen and consumer. This course aims to provide you both with such an understanding *and* with the statistical tools you will need to analyze data. To this end, we are working on redoing this course in order to incorporate programming in R, which is a free software environment for statistical computing and graphics that runs on a wide variety of platforms. We will be using the open-source IDE (integrated development environment) RStudio. Data analysis consists of exploring the data with numerical and graphical summaries, data visualization, and inference - perhaps prediction, or estimation. We hope that by the end of the semester, you will be equipped with the statistical and computational tools you need to draw conclusions about the data you will study. By introducing you to the powerful computational environment R, we hope that you will gain a better understanding of the world around us and be able to perform some sophisticated data analysis.

ACADEMIC INTEGRITY:

Please read the university's statement on academic integrity. You will be held to the UC Berkeley Honor Code.

Cheating: Anyone caught cheating on a quiz or exam will receive a failing grade and will also be reported to the University Office of Student Conduct. In order to guarantee that you are not suspected of cheating, please keep your eyes on your own materials and do not converse with others during the quizzes and exams. You are welcome to discuss the homework problems, both from the text and coding problems, with your colleagues, but do try to write them up on your own, so that you learn the material.

ACCOMMODATIONS FOR STUDENTS WITH DISABILITIES:

Please see me as soon as possible if you need particular accommodations so that we can work out the necessary arrangements for the quizzes and exams. You are responsible for making sure that we know about your accommodations sufficiently in advance to schedule your proctoring with the DSP proctoring services.

SCHEDULING CONFLICTS:

Please notify me in writing by the second week of the term about any known or potential extracurricular conflicts (such as religious observances, graduate or medical school interviews, or sports team activities). I will try my best to help you with making accommodations, but cannot promise them in all cases. In the event there is no mutually workable solution, it may be best to drop the class.

DROP DEADLINE:

Please note that this semester, the undergraduate add/drop deadline will now occur in the fourth week of the semester, on September 12. Note that it is an earlier deadline than previous semesters.

TOPICS & TENTATIVE SCHEDULE:

Week	Topic	FPP chapters
1	Experiments and observational studies	1, 2
2	Randomization, probability	13, 14, 16.1
3	Box models, discrete distributions, probability histograms	15, 16.2-16.4, 18.1
4	Types of variables, measures of location and spread, visualization of data	3, 4
5	Random variables, prob. mass functions, distribution functions, EV, SE, normal, CLT	17, 5, 18.2-18.6
6	Continuous random variables, sampling, distributions of sampling statistics	19, 20, 23.1, 23.2
7	Overflow, review, and MIDTERM on Thursday	
8	Estimation, confidence intervals	21, 23.3, 23.4
9	Surveys and survey package	notes
10	One sample hypothesis tests	26
11	Two sample hypothesis tests, caveats	27, 29
12	Categorical data analysis, chi-square test, contingency tables	28
13	Correlation, association, regression	8, 9, 10, 11
14	Regression line (Thursday is a holiday)	12
15	Residuals, regression errors, inference	notes