STAT 20 FALL 2018

Introduction to Probability and Statistics

INSTRUCTOR:
Hank Ibser (hankibser@berkeley.edu)

GSIS:
Victoria Le <vle@berkeley.edu>,
Eileen Johnson <eileenjohnson@berkeley.edu>,
Yuchen Zhang <yuchen_zhang@berkeley.edu>,
Jason Wu <wjasonu@berkeley.edu>,
Kevin Kiane Jacques Benac <benac@berkeley.edu>,
Rui Chen <rc687@berkeley.edu>,
George Shan <gshan@berkeley.edu>,
Beth Jump <elizabethjump@berkeley.edu>

TIME AND LOCATION:
The lectures will be MWF 1-2pm in 155 Dwinelle.

TEXT, RESOURCES:
Statistics, 4th edition, by Freedman, Pisani, and Purves.
Supplemental notes will be posted periodically, as well as lecture slides for some lectures.
Lectures are screen cast and can be accessed through Course Capture on bcourses.

OFFICE HOURS:
Hank: Mondays 2-3:30, Tuesdays 10-11:30 in 349 Evans.
GSIs: TBA on bcourses - 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. Instructions can be found in the announcement "Next week in Stat 20 - R!" on bcourses. If you want to follow along in RStudio during lecture, you are welcome to do so. Often it is nice to change the code I'm doing slightly to see what happens. However, please restrict yourself to course related activities on your computer during lecture, it can be distracting to those around you. If you absolutely can't stop yourself, please sit in the back of the class.

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. You may use the Reef app instead of getting a clicker.

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. Please don't post answers on piazza. We do our best to respond within 24 hours, but if you post in the evening for a HW due at 11pm, you should not expect to get a response.

SECTIONS:
Sections meet on Mondays and Wednesdays. They are listed as lab sections but I will usually refer to them as discussion or just sections. A brief quiz will be given at the beginning of section on Wednesday, please be on time! (see below)

HOMEWORK:
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 of 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 20 minute quizzes during section **every other** Wednesday (weeks 3,5,9,11,13,15). There I will drop the single lowest score while computing your grade. In addition, there will be one **in-class** midterm on **Wednesday, October 3**, and a comprehensive final exam on **Wednesday, December 12**, from **7-10 PM** in a location that will be announced later. The quizzes will consist of problems like those from the text 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. We will discuss specifics later in the semester.

GRADING:
- Clicker grade (the lowest **four** will be dropped): **3%**
- Weekly homework sets: 9**%** (the lowest **three** will be dropped)
- Biweekly 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
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	Intro to R, Randomization, probability	13
3	More Probability (Mon is holiday)	14-15, notes
4	Probability mass functions, Box Models, Subsetting, Location and Spread	16.2-4, 4
5	Histograms, Random variables, EV, SE	3, 18.1-2,16.1
6	Normal Curve, Normal Approximation, EV and SE for Sample % and Average	17,18.3-6, 5
7	Review, and MIDTERM on Wednesday, Friday lecture on Sampling	19 (Friday)
8	Sampling	19, 23.1, 20
9	Estimation, Confidence intervals, Correlation	21,23.2-4,8
10	More Correlation, Visualization (more ggplot), Regression	9-10
11	Accuracy of Regression, Regression Line	11-12
12	Hypothesis testing	26,29.1
13	HT, Two samples (Mon is a holiday)	27,29.2
14	Chi-squared goodness of fit test (Wed and Fri are holidays)	28.1-2
15	Chi-squared independence, Inference for Regression Line	28.4, notes