STAT 20 SPRING 2020

Introduction to Probability and Statistics

INSTRUCTOR:
Hank Ibser (hankibser@berkeley.edu)

GSIS:
Will be announced in bcourses.

TIME AND LOCATION:
The lectures will be TTh 6:30-8pm in 150 Wheeler.

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: TBA, first week of classes Wed and Thurs 10-12 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 will be posted 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.

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.

STAT 20 SPRING 2019

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. Generally HW will be due Fridays 11pm. The first week will have a questionnaire.

QUIZZES AND EXAMS:

There will be six 20 minute quizzes during section **every other** Wednesday (Feb 5, 19, Mar 18, Apr 1, 15, 29). I will drop the single lowest score while computing your grade. In addition, there will be one **in-class** midterm on **Thursday, March 5**, and a cumulative final exam on **Friday, May 15**, from **11:30-2:30 PM** in a location that will be announced later. If you have a conflict you should consider taking the other lecture of stat 20 or the final exam for the other lecture, the material and exams will be very much the same. The quizzes will consist of problems like those from the text and also R-related material. The exam problems will tend to be a little more in depth, especially integrating material from different parts of the course. You'll get old exam problems with solutions as we get closer to the exams.

DATA ANALYSIS PROJECT:

You will do a group (4-6 students) 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, forming groups right after the midterm.

GRADING:

- Weekly homework sets: **10%** (the lowest **two** will be dropped)
- Fortnightly Quizzes: **15%** (the lowest quiz score will be dropped)
- Data analysis project: **10%** (group project due at end of semester, start after midterm)
- Midterm: **20%**
- 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): top 30% gets some kind of an A (roughly a third get A+, A, A-, perhaps a bit more stingy with A+), next 40% some kind of B, next 20% some kind of C, and lowest 10% D/F. Note that especially with the lowest 10% these are guidelines, not certain. Slightly more or fewer students will get any particular grade.

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.

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. You will not be required to use calculus in this course. 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 will do some 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. 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, you will gain a better understanding of the world around us and be able to perform some sophisticated data analysis.

Students at UC Berkeley are often trained (and screened through the admissions process) to be excellent at memorizing formulas and plugging numbers into them. This course is focused on going deeper, and your study habits will benefit from some tweaks. Rather than doing lots and lots of problems, it is better to spend your time doing the problems with some careful thinking. Even after you get the answer to a problems, spend some time thinking about questions like: "Why is that the right answer?" "Under what circumstances can this method be used, and when is it not appropriate?" "In what ways is this problem similar and different from other problems I've done?" "How do I recognize that this is the right method for this problem?" "If I change the setup of the problem a little, how does that change the answer, and is the method still valid?" Questions like this will help you to understand the material more deeply and excel on quizzes and exams.

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 communicate with others during the quizzes and exams. You are welcome to discuss the homework problems, both from the text and coding problems, with other students, but 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 with the DSP proctoring services.