STAT 20 FALL 2020

Introduction to Probability and Statistics

INSTRUCTOR:
Shobhana Murali Stoyanov (shobhana@berkeley.edu)

GSIS:
TBA on Bcourses

LECTURES:
The synchronous lectures will be held on MWF from 12-1 using Zoom, and posted on bcourses for students who are not able to attend. Details on bcourses. I encourage you to attend the live lectures if you are able to. There are attendance points for lecture, which you can get by attending the lectures OR by watching them later and filling out a google form for attendance, which will be available on bcourses.

THE 15-MINUTE RULE FOR ZOOM CRASHES, POWER OUTAGES ETC:
In case there is a problem with the lecture (I can't connect, or I get kicked off the lecture, or we have a power outage etc), we will use the **15 minute rule**. Don't leave the Zoom lecture immediately since I will be trying to get back in. If I can't get back within 15 minutes, you may assume that the live lecture is canceled for the day due to a failure of technology. I will post a video of the lecture as soon as I am able to do so.

SECTIONS:
Sections meet on Tuesdays and Thursdays. They will be run (remotely) by the GSIs. In order to get your attendance points from section, you need to synchronously attend your section OR complete the worksheet that the GSIs will cover that day, scan and email it to **your** GSI **within 24 hours**. We ***may*** change some section times if there is interest in having different times (especially for students in very different time zones).

TEXT:
I am currently writing up notes for the course which will be in the form of an online textbook, to be made available during the semester. A recommended (but not required) text is *Statistics*, 4th edition, by Freedman, Pisani, and Purves.

OFFICE HOURS:
The GSIs and I will offer virtual office hours via Zoom. These office hours allow for synchronous interaction with the instructor and GSIs and are a good opportunity to discuss your questions relevant to the course. The times are listed below, and the Zoom links will be provided on bcourses. We will also have tutors and their times will be announced later.
Prof. Stoyanov: Mondays and Thursdays 3 - 5 PM.
GSI office hours: TBA - Note that you can go to *any* GSI's office hours, not just your own.

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. Please have both R and RStudio downloaded and ready to go by next week. The GSIs can help you on Thursday (8/27/2020) during the first section.

When I use R in lecture, you may want to follow along in RStudio and run the posted code during lecture. Alternatively, you might want to try it later, and then you can tweak the code and see what happens.

DISCUSSION FORUM (PIAZZA):

We will be using Piazza, a web-based discussion board, for discussions (I will add you to it). Please use Piazza to post questions (not of a personal nature) that are relevant to the entire class. This can include questions about the course materials and topics or mechanics around assignments. The instructor and GSIs will monitor this discussion, but you should also feel free to answer questions posted by other students. This helps to create a general FAQ so that all students in the course may benefit from the exchange. 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. The GSIs and I will usually respond within 24 hours. This means that if you ask a question at 5 pm about homework that is due that night, you may not get a response in time. If your question is private, it is best to use email or a private message on Piazza.

COURSE MATERIALS AND TECHNICAL REQUIREMENTS:

Each week you will find the assigned material posted on bcourses. Please make sure to check in on Monday morning to see what is coming up that week. This course is built on a Learning Management System (LMS) called Canvas and UC Berkeley's version is called bCourses. You will need to meet these computer specifications to participate within this online platform.

LEARNING ACTIVITIES:

You are expected to fully participate in **all the course activities** described here.

- Complete the weekly reading/listening/watching assignments.
- Attend the synchronous Zoom lectures (or watch them later and fill out the attendance form) AND watch the pre-recorded instructor lectures, if any.
- Take part in the kahoots in lecture on Fridays.
- Participate in online discussions on Piazza.
- Watch and listen to any additional media provided for the week.
- Complete the lecture quizzes, turn in the homework and the writing assignments, and any other assignments.
- Complete the weekly quizzes, midterm, and final exams.

STAT 20 FALL 2020

HOMEWORK:
You will turn in weekly homework assignments, due (most) Fridays by 11 pm. You will need to upload to Gradescope, which is the external website that you will use to submit your homework, and where we will post your weekly quizzes and exams. The homework will consist of both conceptual and computational problems as well as R-programming assignments. Homework will be graded on completion and **not** on correctness. The lowest **two** homework grades will be dropped. Late homework (beyond 1 hour) will not be accepted.

POST-LECTURE QUIZZES:
There will be quizzes on **bcourses** that will be assigned after each lecture. They will be due by **11 am** the day of the *next* lecture. These will be just a few multiple choice problems based on the most recent **lecture**.

QUIZZES AND EXAMS:
There will be **30 minute quizzes** (almost) every **Thursday**, based on the previous week's material, with the first quiz being on **September 3**. There will be **ten** quizzes in total, and I will drop the **two lowest** scores while computing your grade. In addition, there will be one **90 minute** midterm on **Friday, October 16**, and a **three hour** comprehensive final exam on **Thursday, December 18,** from **11:30 AM - 2:30 PM. All the weekly quizzes, the midterm, and the final exams will be posted on Gradescope.** They will be timed, but you will have a window during which they will be available. For the quizzes, midterm and final you will have an 18 hour window in which to submit, and your clock will begin when you first open the quiz/exam on Gradescope. For example, for the quizzes, you can take the quiz whenever you want within the window, but you must submit the quiz within half an hour after starting it. A three hour cumulative final exam will be given in a 18 hour window that will include the scheduled final exam time. I'll announce the exact timing of all of these on bcourses when we get closer. The quizzes will consist of problems like those from the HW, section problems 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 closer to the exams.

DATA ANALYSIS PROJECTS:
You will do **three** group (4-6 students) data analysis projects throughout the semester.The first two (due **9/22** and **10/27**) will be smaller in scale, and the final project (**due 12/8**) will be longer, and will use the skills and knowledge you will have developed throughout the semester. I have created the groups on bcourses and you can sign up for a group. If at the end of the second week, you have not signed up, I will assign you to a group. All the members of a group should be from the same section.

WRITING ASSIGNMENTS:
About every other week, I will assign "readings" that will be related to either class material, or general statistical literacy. These might be articles or podcasts, or videos. You will turn in short write ups (about

500 words) on bcourses about these, perhaps answering some questions. There will be six assignments, and the first one will be due on **September 1**.

GRADING:

- In-lecture google forms + kahoots (the lowest **three** will be dropped): **4%**
- Post-lecture quizzes: **5%** (the lowest **three** will be dropped)
- Writing assignments: **4%** (the lowest **one** will be dropped)
- Weekly homework sets: **7%** (the lowest **two** will be dropped)
- Weekly Quizzes: **15%** (the lowest **two** will be dropped)
- Data analysis mini-projects: 5% each for a total of **10%**
- Final data analysis project: **15%**
- Midterm: **12%**
- Final: **28%** (if you cannot take the final, please **do not** take the class - you will get a failing grade)

INCOMPLETE COURSE GRADE:

Students who have substantially completed the course but for serious extenuating circumstances, are unable to complete the final exam, may request an **Incomplete** grade. This request must be submitted in writing to the instructor. You must provide verifiable documentation for the seriousness of the extenuating circumstances. According to the policy of the university, an Incomplete grade from the Fall semester must be made up by the first day of the *following* Fall semester.

STUDENTS WITH DISABILITIES:

If you require course accommodations due to a physical, emotional, or learning disability, contact UC Berkeley's Disabled Students' Program (DSP). Notify the instructor and GSI through course email of the accommodations you would like to use. You must have a Letter of Accommodation on file with UC Berkeley to have accommodations made in the course.

UC Berkeley is committed to providing robust educational experiences for all learners. With this goal in mind, we have activated the ALLY tool for this course. You will now be able to download content in a format that best fits your learning preference. PDF, HTML, EPUB, and MP3 are now available for most content items. For more information visit the alternative formats link or watch the video entitled, "Ally First Steps Guide."

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 will be

doing 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. 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.

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 may 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 problem, 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 zero on that assessment and will also be reported to the University Office of Student Conduct. In order to guarantee that you are not suspected of cheating, do not communicate with others during the quizzes and exams and do not seek answers online. You are welcome to discuss the homework problems, both from the text and coding problems, with your colleagues, but write them up on your own, so that you learn the material.

DROP DEADLINE:

Please note that this semester, the undergraduate add/drop deadline is on September 16.

TENTATIVE WEEK-BY-WEEK SCHEDULE:

Here is the schedule for the course. Links to readings and assignments will be posted on bcourses.

Week	Date of Monday	Topics	Quiz/Exam
1	8/24	Experiments, Obs. Studies, types of variables	
2	8/31	location & spread, intro to R, subsetting (dplyr)	WQ1 Thursday
3	9/7	histograms, box plots, visualization using ggplot2	WQ2 Thursday
4	9/14	chance processes, box models, probability	WQ3Thursday
5	9/21	counting, random variables, distributions of special random vars	WQ4 Thursday
6	9/28	law of averages, probability histograms, normal approx, CLT	WQ5 Thursday
7	10/5	EV, SE of sample % and average, sampling	WQ6 Thursday
8	10/12	overflow, review for MT	MT Friday
9	10/19	parameter estimation, confidence intervals	No WQ
10	10/26	hypothesis tests	WQ7 Thursday
11	11/2	correlation, ggplot, regression	WQ8 Thursday
12	11/9	regression line, regression effect	WQ9 Thursday
13	11/16	more about regression,	WQ10 Thursday
14	11/23	Thanksgiving	
15	11/30	Chi-square tests	
16	12/7	RRR week	
17	12/14	Final exam on 12/18, 11:30-2:30	Final