
STAT 20 SPRING 2021

Introduction to Probability and Statistics

INSTRUCTOR:

Hank Ibsen (hankibser@berkeley.edu)

GSIS:

Will be announced in bcourses.

TIME AND LOCATION:

The lectures will be recorded live on zoom at the regularly scheduled times (TuTh 2pm-3:30pm PDT). I encourage you to attend live if possible, though you can also view on bcourses. I will try to post them by 5:30pm. If I ever get kicked off zoom during live lecture, I will attempt to reconnect. If I give up I will post an announcement on bcourses so you know I'm not coming back. If I am ever gone for more than 15 minutes you can assume that I'm unable to come back.

TEXT, RESOURCES:

Required text: *Statistics*, 4th edition, by Freedman, Pisani, and Purves. I originally listed this as a recommended text but I think the course will work more smoothly if this is required.

OFFICE HOURS:

Once we finalize, I'll post office hours in an announcement on bcourses. For the first week of classes, I'll have office hours TuTh 3:30-5pm (right after lecture) and WedThuFri 9-11am.

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. When I lecture about R, I recommend that you follow along and run the posted lecture code while you watch lecture. Often it is nice to change the code I'm doing slightly to see what happens.

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 will not meet on Wednesday January 20 but will meet Monday January 25 and after that. We may change some section times if there is interest (especially for students in very different time zones). Section is optional, mostly will be problem solving sessions. Sections will not be recorded.

STAT 20 SPRING 2021

HOMEWORK:

You will turn in 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 on completion and **not** on correctness. Generally HW will be due **Fridays at 11pm** and will be announced a week ahead of time. For some HW assignments I'll also assign a short reading or podcast about current events for you to comment on.

MINI-QUIZZES

There will be a 10 minute mini-quiz after every lecture, due before the start of next lecture. They will be posted on bcourses and you will have a window between about 5:30pm PDT and the start of the next lecture in which to take the quiz. These will be quick conceptual quizzes that you shouldn't have trouble with if you view the lectures. I will not take attendance for lectures but you should definitely watch the lecture if you expect to get full points on these quizzes, sometimes the questions will be just to check to see if you watched and will not have to do with statistical content.

QUIZZES AND EXAMS:

There will be 30 minute quizzes due **almost every Wednesday at 11pm** (not the first week, midterm week, etc). You will take them online and you'll I will drop the two lowest scores when computing your grade. In addition, there will be a 1.5 hour midterm on **Thursday/Friday March 4/5**. For the quizzes, midterm and final you will have a 24 hour window in which to submit, and you must submit within the time limit starting when you first open the quiz/exam on Gradescope. So 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 (I may also give a little extra time to make the submission). A three hour cumulative final exam will be given in a 24 hour window including the scheduled final exam time, **Monday May 10, 12:30-3:30pm**. 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) projects throughout the semester. The first two will be smaller in scale as practice, and the final project will be longer at the end of the semester, using the skills and knowledge you will have developed throughout the semester.

STAT 20 SPRING 2021

GRADING:

- Homework sets: **10%** (the lowest two will be dropped)
- Mini-quizzes: **5%** (the lowest 5 will be dropped)
- Weekly Quizzes: **10%** (the lowest two quiz scores will be dropped)
- Two smaller group projects: 5% each for a total of **10%**.
- Data analysis final project: **15%** (group project due at end of semester, start after midterm)
- Midterm: **15%**
- Final: **35%**

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 may get any particular grade.

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 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.

STAT 20 SPRING 2021

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](#).

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, 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 other students, but write them up on your own so that you learn the material. Last fall I ended up failing 11 students for cheating and I really hate to do that. Please don't cheat.

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.

STAT 20 SPRING 2021

TOPICS & TENTATIVE SCHEDULE:

Week	Date of Monday	Topics	Ch of text
1	1/18	Expts, observational studies/Intro to R, Location and Spread	1,2/4
2	1/25	Subsetting in R (dplyr package)/Histograms	none/3
3	2/1	ggplot2 package in R/Probability	none/13,14
4	2/8	More Probability, Binomial/Box Models	15/16
5	2/15	EV,SE, Random Variables/Probability Histograms	17/18.1-2
6	2/22	Normal Curve and Approx/EV, SE for Averages and Percents	18.1-3/20.1-3,23.1
7	3/1	Review and MIDTERM on Thurs/Friday 3/4-3/5	
8	3/8	Sampling/Correction Factor, Confidence Intervals	19/20.4-5,21
9	3/15	More CIs, Project Discussion/Hypothesis Testing	23.2-4/26.1-5
10	3/22	Spring Break	
11	3/29	T test/Two samples	26,26.1-2
12	4/5	More two samples/Correlation, More ggplot	27.3-4/8-9
13	4/12	Regression/Vertical Strips	10/11
14	4/19	Regression Line and R/Chi-square test	12.1/28.1-3
15	4/26	End Chi-square test, Wrap up HT	28.4/29
16	5/3	RRR week	Final Proj due
17	5/10	Final exam in window including 5/10, 11:30-2:30pm	Final