Dear Students,
I am looking forward to introducing you to the wonderful world of learning from data during this summer session. Public health is in the public awareness in a way that is unprecedented in our lifetimes. I consider it a privilege to teach at Berkeley where there is such a diverse cohort of students and where the school has a commitment to social justice and is making concrete steps towards incorporating anti-racist pedagogy across all courses. My goal for the course is to introduce you to concepts in probability and biostatistics, but more importantly to give you the skills to interpret information presented in academic and popular communications, to use data to answer your own questions and to communicate those clearly to others.

We continue to teach and learn under extremely challenging situations. The pandemic, economic uncertainty, unemployment and shelter-in-place as well as political conflict and an ongoing reckoning with racism and disparity add many challenges and stressors to daily life and learning. This larger context can also create dangerous situations for individuals and households. These challenges are not spread evenly across our students. Some of you know each challenge more personally than others and time and again there have been disproportionate effects on Black, Indigenous, and other communities of color and students who are essential workers themselves or live with them.
We recognize that the current context can be especially detrimental to mental health in a time when many of us are cut off from our usual sources of social support. We as a teaching team will do our best to keep these challenges in mind and to adapt as our collective situation evolves. Please also take advantage of the supports the university is working to make available: Berkeley Health Services SPHUG

Changes to the course in light of the pandemic and virtual learning:
Barring emergent internet issues or power shut-offs, the teaching team will offer multiple sessions each week synchronously, with the option to watch the recorded asynchronous version of each. Please note that because of processing, captioning and uploading time, synchronously recorded materials take time to post. We will do our best to upload these materials as soon as we receive them, but there is often a delay of several days before we receive the processed videos for posting.
In light of the challenges of virtual learning we have also made major changes to the grading scheme: lab sessions will be graded on completion only and none of the problem sets will be graded. Solutions will be available to you to check your work. The problem sets will give you the opportunity to practice applying the concepts from lecture and programming skills from labs. We have also added short timed quizzes that will be available for at least 24 hours. The goal of the quizzes is to give you a low stakes way to check your understanding and give you a chance to practice answering questions that are similar in format to those you will see on the timed portions of the exams – many of the quizzes will be also be for completion only. We

have also incorporated participation credits – some of which will be used to provide feedback to the teaching team and test exam systems prior to distribution in order to check for technological or formatting issues. While the entire course can be completed asynchronously, we are requesting that all students meet with their advising GSI at least 3 times during the course to discuss the data project. We will also have an active discussion board on piazza and numerous office hours so that you have opportunities to get live support for the course.

There will be an initial needs assessment as one of our first participation activates to help us prepare and to hopefully better meet student needs. If you are in need of support around technology, the university has also created a program to lend devices, peripherals and hotspots.
Please review the attached syllabus that covers our course goals, structure, policies, resources and sources of support available to you, the timing of any submitted material, and the course outline. I look forward to the next few weeks and hope we will have a good experience learning together!
See you soon,

Prof. Mi-Suk Kang Dufour, PhD, MPH

PH142: Introduction to Probability and Statistics in Biology and Public Health

Course Syllabus (Summer 2021)

Table of Contents

Course Information

Course Meeting Dates/Times: Note that you should sign up for a DIS and LAB session with matching numbers, i.e. if you are in DIS 102, please sign up for LAB 102A	Lecture MTWThF 9:30-11:00am Labs: **DIS 101** MTuW 11:00A-11:59A **DIS 102** MTuW 11:00A-11:59A **DIS 103** MTuW 11:00A-11:59A **DIS 104** MTuW 11:00A-11:59A **DIS 105** MTuW 5:00P-5:59P **DIS 106** MTuW 11:00A-11:59A **LAB 101A** ThF 11:00A-11:59A **LAB 102A** ThF 11:00A-11:59A **LAB 103A** ThF 11:00A-11:59A **LAB 104A** ThF 11:00A-11:59A **LAB 105A** ThF 5:00P-5:59P **LAB 106A** ThF 11:00A-11:59A
Course Location:	Offered remotely on zoom
Instructor:	Mi-Suk Kang Dufour (she/her/hers)
Email:	mi-suk@berkeley.edu
Instructor Availability:	https://mi-suk.youcanbook.me/
GSIs:	Chandler Beon (Technical GSI), (he/him/his) Mitchell Leus (he/him/his) Antonia Gibbs (she/her/hers) Alma Juarez (she/her/hers) Jessica Thaj Yunzhe Zhou
Course Readers:	Shazia Nooruddin Yunwen (Wendy) Ji, (she/her/hers)
Course Email:	142gsi@berkeley.edu

Course bCourses link:	https://bcourses.berkeley.edu/courses/1505299
Course Units:	4
Contributing Instructors:	Corinne Riddell (Fall PH142 instructor) Sophia Fuller and Sarah Johnson (Spring 2021 GSIs – lab videos)

Course Description

This course is an introduction to statistics and data science, primarily for MPH and undergraduate public health majors, and others interested in public health topics. The course can be divided into three parts. In Part I, we will focus on learning to use R to explore and summarize univariate and bivariate distributions. Specifically, we will use the dplyr and ggplot2 packages to manipulate and visualize data sets in R. Part II of the course introduces classical problems in probability and the Normal, binomial, and Poisson distributions. The most important topic we will cover in Part II is the Central Limit Theorem. In Part III, we introduce statistical inference, the process of estimating statistics from samples to make inference about populations. During all parts of the course we will use real and simulated data sets to gain experience conducting biostatistical analyses using R. We will follow the PPDAC model, which stands for "Problem, Plan, Data, Analysis, and Conclusion".

Prerequisites

There are no courses required as pre-requisite to PH142. No experience with computer programming is assumed. The mathematical level required for this class is roughly equivalent to High school algebra.

Course Learning Objectives

After successfully completing Part I of the course, you will be able to:

- Extract relevant statistical information from published articles in the scientific and popular press
- Describe distributions of variables visually and calculate summary statistics for measures of centrality and spread
- Determine the appropriate graphic to plot distributions and provide R code to manipulate and visualize data frames
- Identify basic sampling strategies and study designs used in Public Health
- Describe core concepts of ethics in Public Health
- Perform basic data manipulation in R
- Interpret output from a simple linear regression model

After successfully completing Part II of the course, you will be able to:

- Compute probabilities using the general rules
- Identify and describe binomial and Poisson random variables
- Compute probabilities using basic properties of the Normal distribution
- Express epidemiologic measures as probabilities
- Describe the central limit theorem
- Write R code to compute probabilities for the Normal, binomial, and Poisson distributions

After successfully completing Part III of the course, you will be able to:

- Estimate means, proportions, and differences between means and proportions, compute their confidence intervals and perform statistical tests
- State the assumptions and importance of the assumptions for statistical tests
- Perform a simple chi-squared test
- Perform a matched t-test
- Describe and check the assumptions for simple linear regression. Interpret the confidence interval and statistical test of regression intercept and slope coefficients
- Describe ANOVA, including the null and alternative hypotheses, and interpret output
- Describe when bootstrapping can be used
- Describe a permutation test
- Demonstrate knowledge that has been used throughout the term, in terms of data visualization and data manipulation
- Write R code snippets to perform hypothesis tests and calculate p-values

Methods of Instruction

Lectures on weekdays 9:30 to 11am, they will be a mix of lectures delivered synchronously and recorded. Daily lab/discussion sections, offered synchronously and pre-recorded.

Instructor Information

Dr. Kang Dufour is an epidemiologist and biostatistician with appointments at University of California Berkeley Divisions of Biostatistics and Epidemiology. Her work focuses on implementation research and evaluation of public health programming for infectious diseases including HIV, STDs, and Malaria.

Course Format

Course Schedule – see course website

Course Grading

Grading is based on the following:

- **Participation**. With such a fast-paced course, we want to ensure that you check your understanding frequently as we move through the material. Each lecture topic will have an associated comprehension check – these will graded for completion only. We are also asking that you participate in meetings with your GSI to discuss your data project in the first, third, and fifth weeks of the course (to get feedback prior to each project deadline) these meetings will also count as participation points. Throughout the term you will be also asked to provide feedback, practice with/test course systems and participate in course discussions. Participation assignments will be announced on Piazza and listed on the course website. These will be marked for completion only. You may miss two lecture comprehension activities without penalty.

- **Problem sets/Assignments** will be distributed as R markdown files on datahub. Problem sets will not be submitted for marks and you are encouraged to work on it in groups if that is how you learn best. Completing the problem sets is excellent preparation for the exams. Assignments have some automated checks build in to the markdowns for immediate feedback. Full solutions will be posted on datahub a few days after the assignments are made available.

- **Quizzes** will be available from 9 am on the day they are listed until noon on the following day. Quizzes will be relatively short and are meant to provide a chance to check your understanding and to practice with exam-style questions. Once opened, you will have 1 hour to complete the quiz. Your lowest graded quiz score will be dropped.

- **Lab exercises** are intended to practice concepts from lecture in a practical programming environment. You can complete and submit these during the lab section, or on your own. Students often find it much more helpful to complete this in lab rather than independently, but we understand students learn differently, so feel free to do what works best for you. Lab exercises are graded on correct completion, so you must complete the lab fully, passing all tests, in order to receive credit for the assignment. Since we provide all tests for correctness of your code, grading is all-or-nothing. You may miss one lab without penalty.

- **Midterms**. There are three midterms. Midterms will be offered in two portions, an online timed portion and a "take home" portion. Appropriate accommodations for the midterm will be made for those with disabilities (please refer to the "Disabilities" section, below) Please note that only in extremely rare circumstances such as illness (with a doctor's note) will the midterm be given to individual students after the scheduled examination date. Exams will cover the material presented in lecture, supplemental videos, discussion, and lab sections, including R coding syntax, unless otherwise noted.

- **Exam policies.** The exams will be open book. This means you **can use** electronic or hard copies of notes and the course textbook and additional resource list. You **may not** use the internet to search for the answers or inform your answers. Using the internet is **strictly prohibited** and any evidence of this may result in a 0 on the exam. While you take the exam, you are prohibited from discussing the test with anyone other than the PH142 instructional team. Exams will consist of a timed portion which will be offered on gradescope. Similar to the quizzes the timed portion of the exam will be available for 24 hours, but must be completed within one hour upon being opened. There will also be a take home portion of the exam which will be released as a fillable pdf and will be due after 24 hours. You are strongly encouraged to submit early and email the GSI account if you have unanticipated technical challenges affecting submission. We will strive to return graded examinations within one week of the exam date.

- **Data skills demonstration group project.** The purpose of the group project will be to use public health or biological data that you find or have access to and use it to demonstrate the statistical concepts that you've learned throughout the course. You have three options for the data project; 1) you may choose to do the project alone (this is not recommended unless you have a circumstance such as a large time difference that makes it difficult to coordinate with other students), 2) you may self-select into a group of up to 5 students, 3) we can assign you to a group if that is preferable for you. You will be asked about your preferences at the beginning of the term. Each group/project will be assigned to a supporting GSI. The data project will be completed in three parts, you will be asked to meet with your supporting GSI in the first, third, and fifth weeks of the course to check in about your project.

- **Extra credit assignments.** There are three optional extra credit assignments available to you this summer. Each is worth up to 1 percentage point on your overall grade – you may submit up to 2 extra credit assignments. More detailed instructions for each will be posted on the course website. Briefly the choices are as follows:

- **EC 1:** You will be randomly assigned to a topic and week in the course. For extra credit and to show your creativity, you can make a sample question and solution guide corresponding to topics covered during lecture or lab in your assigned week.

- **EC 2:** You may choose to create a meme or cartoon that demonstrates understanding of one of the course concepts. To receive credit the submission should include an image and an explanation of the concept and how this is reflected in the image.

- **EC3:** You may choose a recent study that you find compelling enough that you would change your behavior. You will be asked to critique the study and explain why the methods and findings are compelling.

Final grades will be assigned according to the following percentages:

Quizzes	15%
Participation	10%
Lab completion	10%
Midterm 1	15%
Midterm 2	15%
Group project (in 3 parts)	20%
Midterm 3	15%

S/U (satisfactory/unsatisfactory) grading is permitted for graduate students in this course. P/NP (pass/no pass) grading is permitted for undergraduate students.

There are no differences in the course requirements or the grading for students who choose an S/U or P/NP option. "S" will appear on transcripts for grades of "B-" or above. "P" will appear on the transcripts for grades of "C-" or above.

A grade estimation tool has been provided on the course website. This tool allows you to generate a rough estimate of the letter grade you can expect to receive based on the received or projected grades you input.

Course Materials

Courses website

To access the course website, go to https://ph142-ucb.github.io/su21/.

Here you will find links to required and optional readings, the syllabus, assignment

descriptions and additional course resources. The course website will contain the most up to date schedule and assignment information.

Required Materials

We will be using **R**, a statistical programming language, and **RStudio**, an integrated development environment on **datahub**, a cloud computing environment created at Berkeley. Use of R, RStudio, and datahub is required for homework assignments and lab exercises and requires an internet connection and web browser. You will learn how to use R, RStudio, and datahub during the first week of classes. You can access datahub from the links on the course website.

Optional Materials

The course textbook is "The practice of statistics in the life sciences" by Brigitte Baldi and David S. Moore. The 4th edition is the latest one, but previous editions are fine. The textbook has been digitized and can be checked out from the UC Berkeley library. You can also purchase or rent the book here. We rely on it more during Part II and III of the course than we do in Part I. It is possible to complete the course using course materials and we note that the instructors differ in some places from the opinions presented in the text. Where the textbook and course materials differ the course materials will take precedence.

Other resources

In addition, here are some free online resources available as supplementary material. We link to these specific resources in the lecture slides when applicable:

- Learning statistics with R: https://learningstatisticswithr.com/lsr-0.6.pdf
- OpenIntro Statistics: https://drive.google.com/file/d/0B-DHaDEbiOGkc1RycUtIcUtIelE/view
- A ModernDive into R and the Tidyverse: https://moderndive.com/9-hypothesis-testing.html#ht-infer
- Statistical Thinking for the 21st Century: https://statsthinking21.github.io/statsthinking21-core-site/ci-effect-size-power.html#statistical-power
- R for Data Science: https://r4ds.had.co.nz/data-visualisation.html

Announcements

Course announcements will be sent out through a once-weekly email blast from Piazza.

Course EMail

We strive to reduce email as much as possible. All questions about course material should be asked on Piazza. This allows us to reduce email and also allows other students

to benefit from the questions and answers. We will not answer any questions about course material via email. Email the GSI account to: make DSP accommodations for tests or homework or request an assignment extension (see "Late Assignments" below). Email the instructor for personal concerns or disruptions that affect your performance in the course or during an emergency that will result or has resulted in a missed test.

Policies

Grace Period

All assignments, unless stated otherwise, are due on the specified day at 10:00pm. Due to the nature of electronic submission, we understand that some students may experience technical difficulties with submission close to the deadline. Therefore, we are offering a grace period of two hours, until 11:59pm, to account for these submission issues. If you are having issues during this time, please email your submission to the GSI email at (142gsi@berkeley.edu) before 10:30pm. We will not be accepting requests regarding submission errors after 11:59 on the due date. The grace period applies by default, you do not need to notify us to use it.

Grading and requests for re-grades

Grading for this course is done through gradescope to allow blinded grading of questions and to provide consistent rubric information for GSIs and students. Gradescope also allows us to use some AI features to speed up the grading turn-around time. We will do our best to return graded submissions within a week of the due date. While we will do our best to ensure timely and consistent grading, we know that there are possibilities for both human and machine-reading errors. Regrades will be allowed on quizzes and midterm exams, requests for regrades will be open for three school days after the grades are released using Gradescope. Note that if you request reconsideration of a graded question, instructors may reconsider grades on the entire assignment. Due to the short turn around time for final grade submissions we generally cannot accommodate re-grade requests for the final exam.

Late Submissions

We will allow 50% credit for assignments submitted within 24 hours after the due dates. Extensions can be made for DSP students but should be requested ideally before the due date by emailing the GSI email account. Anyone else requesting an exemption for late submission should email the GSI account explaining their situation. If an emergency event prevents submitting an assignment by the deadline, please contact the GSI email account as soon as reasonably possible, including documentation with your request for extension.

Attendance

We encourage attendance at all synchronous sessions as there are opportunities to ask questions to the instructor during lectures and to the GSIs during lab. We do not require attendance. If you do not attend, it is your responsibility to watch the recordings to stay on top of course material. Note that due to the processing and captioning time required most recordings of synchronous sessions will not post until a few days after the session.

Technology

Zoom will be used to conduct lectures and labs. Zoom links will be shared on the calendar embedded on the course website. **You *must* use a Zoom account affiliated with your Berkeley email to access the invite!** Students will be muted and their videos turned off by default since there are so many of you! I encourage you to ask questions, you can do so by unmuting or by using the chat. If you are more comfortable doing so, you may also send a private chat message. This will be visible to the lecturer but not the class. Feel free to turn on your camera when you ask a question orally. Please note that questions asked are part of the lecture/lab recording.

Communication and community

UC Berkeley School of Public Health has a commitment to cultivating a safe, respectful and inclusive community. You can read more about the this in the principles of community statement. Part of fostering this type of community is cultivating respectful communication. We as a teaching team will do our best to communicate in a respectful, compassionate, and professional manner. We ask that you as students do your best to hold these values in your communications with each other and with us.

Questions during lecture and lab are strongly encouraged. If something is unclear to you, it is probably unclear to many others in the room. There may be times, however, when the instructor or the GSI decides that a particular question or discussion is not helpful to the entire class or will take too long to address satisfactorily. In these cases, we may defer the question to be answered after class, on Piazza or during office hours.

We will use **Piazza** for class discussion. The system is highly catered to getting you help fast and efficiently from classmates, the GSIs, and the instructor. In general, you can expect that the GSIs will respond to posted questions within 24 workday hours.

Please do not email course content questions directly to the instructor or GSIs. The instructor and GSIs will not respond to questions about course content by email.

In the interest of supporting work-life balance, the teaching team will not be expected to respond to Piazza questions during holidays and breaks. GSIs will respond to Piazza questions up until 24 hours before exams. However, students may continue to post and answer each other's questions during breaks and in the last 24 hours before an exam. Piazza will be inactivated the day of tests.

For questions and concerns that are not related to course content, please email 142gsi@berkeley.edu GSIs will do their best to respond to the course email account within 1 business day. Email responses may be slower over weekends and breaks.

Anti-racist and inclusive learning environment

As we, at Berkeley Public Health, strive to create an anti-racist learning environment, we commit to teaching this course, to the best of our ability, with an antiracist, racial justice, and equity-minded lens. We invite you to take this journey with us by being fully present. We are interested in your perspectives and in the value and knowledge you bring to help make this an enriching course environment.

We view this syllabus as a dynamic document oriented toward learning and not just coverage of material; thus, We may add or modify topics covered, assignments, and resources (e.g., required readings/videos) slightly based on the needs and interests of students in the course. We welcome feedback and input at any time and invite careful reflection of any modifications that may help improve the course in the future.

As your professor/instructor team, we agree that:

- We will do our best to include course content that include examples relevant to BIPOC communities (e.g., readings; examples; data, etc.)
- Students are the experts of their own experiences. Your world lens is welcomed; and as students, you are invited to lift up information and/or data that is relevant to the course material. Everyone is a teacher and everyone is a student.
- We cannot speak on behalf of all groups, or fully understand the issues, concerns and history of all BIPOC. However, I am/we are willing to listen and learn, admit mistakes and engage in ongoing cultural humility practices.
- We welcome feedback and input at any time during the course without fear of reprisal; if a mid-semester evaluation is conducted, there will be specific language about antiracism teaching practices.

SPH Course Policies

Descriptions of and relevant campus links to SPH school wide course policies on Disability Support Services, Accommodation of Religions Creed, Course Evaluations, Academic Integrity can be found at:
https://berkeley.box.com/s/knh3rbk9ikgvmca4ymy93msgj9bkebq5

Disabilities: The Disabled Students Program (DSP)

The mission of the Disabled Students' Program (DSP) is to ensure that all students with disabilities have equal access to educational opportunities at UC Berkeley. The DSP offers a wide range of services, accommodations, and auxiliary services for students with disabilities. These services are individually designed and based on the specific needs of each student as identified by DSP's Specialists.

We will accommodate disabled students' needs according to DSP documentation; please notify the DSP if you require such accommodation (DSP will then contact the instructor). **Note that this may take several weeks, so please initiate this process ASAP so that any accommodations can be implemented in time for the first midterm exam**. Steps to the application process: https://dsp.berkeley.edu/students/new-students.

If you require DSP accommodations for a test, please email the GSI email account at 142gsi@berkeley.edu with your request and write "DSP accommodation" in the subject heading as soon as you know accommodations are required. If your accommodation allows for extension on take-home assignments, we ask that you discuss your request no later than 24 hours after the assignment is posted.

Mental Health

If you are experiencing stress, anxiety, or other forms of distress during the semester, we hope to be a resource for you–**please don't hesitate to reach out to a GSI or the Professor for support.** You are not alone.

There are also many resources available to you. All registered Berkeley students are eligible to use Counseling Psychological Services. **You do not have to purchase the Student Health Insurance Plan to use these services.** The first five counseling sessions are free for registered Berkeley students. Counselors can provide support in academic success, life management, career and life planning, and personal growth and development.

UC Berkeley, Counseling and Psychological Services

- Please call (510) 642-9494 or stop by the office on the 3rd floor of the Tang Center to make an appointment with a counselor.
- **Drop-in counseling for emergencies:** Monday - Friday, 10:00AM-5:00PM
- **After hours counseling:** In the case of emergencies at night or on weekends, call (855) 817-5667 for free assistance and referrals. Request to speak with a counselor.
- **For emergency support:** Call UCPD 911 or (510) 642-3333

24 Hour Crisis Hotlines

- **Alameda County Crisis Line:** Call 1-800-309-2131 *(offers confidentiality, TDD services for deaf and hearing impaired callers and translation in 140 languages)*
- **National Crisis Help Line:** Call 1-800-273-TALK
- **Crisis Text Line:** Text HOME to 741741
- **National HopeLine Network:** Call 1-800-SUICIDE

We also ask that you look out for your fellow peers. If you see any of the signs below that may indicate your classmate may need assistance, please use the resources above or reach out to any of the GSIs or Professors.

- Withdrawing from other people
- Changes in weight or eating patterns
- Changes in sleeping patterns
- Fatigue or lack of energy
- Increased anxiety or irritability
- Feeling worthless or hopeless

Other Campus Resources:

- **Let's Talk:** *Informal Drop-In Counseling*
- **Self-Help Resources**
- **Be Well at Cal**

Academic Honesty

Learning is hard work–we encourage everyone to work together and support one another. However, while group work is encouraged, with the exception of the group project, **students must submit their own code and answers** for grading. Students can not work together on the quizzes, midterm, or final examinations. **Tests that show evidence of academic misconduct will be immediately flagged and reported to the Center for Student Conduct for review.** This can result in a grade of 0 on an assignment or a harder penalty depending on the degree of the offence. Each term, a few students in this class are reviewed by the Center for Student Conduct as we take cheating very seriously.

Berkeley's code of conduct is **here**. See Section V and Appendix II for information about how UC Berkeley defines academic misconduct. In particular, the sections on cheating and plagiarism are most relevant for this class.

If you are not clear about the expectations for writing a test or examination, be sure to seek clarification from the instructors or your GSI beforehand.

Harassment policy

We are all responsible for creating an environment that is welcoming, civil, safe, and tolerant. UC Berkeley does not tolerate harassment of PH142 students, GSIs, or instructors.

- Instructors and GSIs will act to stop acts of harassment in the classroom.
- Students experiencing harassment can contact the office for the prevention of harassment and discrimination. To file a report, you can email ask_ophd@berkeley.edu or call them at (510) 643-7984. For more information, see: https://ophd.berkeley.edu/.
- Please note that Instructors and GSIs are Responsible Employees and must report incidents of sexual violence and harassment to the Office for Prevention of Harassment and Discrimination. Please see this website for confidential reporting resources: http://survivorsupport.berkeley.edu/Confidential-Resources-Anonymous-Reporting-and-Privacy