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

Course Syllabus (Summer 2023)

Table of Contents

SYLLABUS SUBJECT TO CHANGE

PH142 Summer 2023 Course Syllabus
Dr. Kang Dufour

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 Online
Instructor:	Mi-Suk Kang Dufour (she/her/hers)
Email:	mi-suk@berkeley.edu
Instructor Office Hours :	https://mi-suk.youcanbook.me/
GSIs:	
Course Readers:	
Course Email:	142gsi@berkeley.edu
Course Canvas/bCourses link:	

Course Unit:	4
Contributing Instructors:	Corinne Riddell (Fall ph142 instructor)
	Sophia Fuller and Sarah Johnson (Spring 2021 GSIs - lab materials)

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

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 asynchronously. Daily lab/discussion sections, offered synchronously and pre-recorded. Daily sections are delivered synchronously. Office hours will also be offered daily at multiple times to accommodate a range of schedules.

SYLLABUS SUBJECT TO CHANGE

PH142 Summer 2023 Course Syllabus
Dr. Kang Dufour

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, CMV, STDs, and Malaria.

Course Format

Course Schedule – see course website for most up to date schedule and resources. Lecture slides will be available prior to lecture and code used in lecture will be made available in rmarkdown format through the datahub. Lectures will be a mix of synchronous zoom sessions and pre-recorded video. Synchronous sessions will also be recorded and made available.

Course Grading

Grading is based on the following:

- **Participation**. Required meetings with your GSI to discuss your data project will 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 the ed discussion board. These will be marked for completion only. You may miss two participation activities without penalty.

- **Homework Assignments** will be distributed as R markdown files on datahub. **Homework 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 homework is excellent preparation for the exams.** All solutions will be posted on datahub a few days after the homework was made available.

- **Quizzes** will be available from 9 am on the day they are listed 5pm two days later. Quizzes will be relatively short and meant to encourage you to keep on top of the lecture content. Once opened, you will have 30 minutes to complete the quiz. **Your lowest 2 graded quiz scores 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 beforehand. Students 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.

- **Exams**. There are three exams. Exams 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. Evidence of cheating may result in a 0 on the test or further disciplinary action. Exams will consist of a timed portion which will be offered on gradescope during the scheduled class/exam time window and a take home portion which will be available for at least 24 hours. **We will not accept late submissions, so 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 assignment.** For those who have missed questions on the first two exams there will be an opportunity to recover points through an extra credit assignment. To recover up to 3 points on the exam the assignment is to create a short explanation of the concept for which you lost points, and to create questions to test the knowledge of the concept including a rubric and solution guide corresponding. Details will be provided

following the first exam. The assignment will be available for one week following the relase of exam grades.

Final grades will be assigned according to the following percentages:

Weekly quizzes	15%
Participation	10%
Lab completion	10%
Midterm 1	15%
Midterm 2	15%
Group project (in 3 parts)	20%
Final exam	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.

Course Materials

Courses website

To access the course website, go to https://ph142-ucb.github.io/su23/.

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 textbook is in the queue for digitization at the university library and will be made available online to you when it's ready.** The 4th edition is the latest one, but previous editions are fine. 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

We try not to flood you with messages. Please set your notifications for discussions according to your preferences. Course announcements will mostly be sent out through a once-weekly announcement posting from Ed.

Course EMail

We strive to reduce email as much as possible. All questions about course material should be asked on the discussion board. 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.

Regrades

Regrades will be allowed on quizzes, data projects parts 1 and 2, and the first two midterm exams, and must be submitted **within 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 or final submission for the data project.

Late Submissions

Assignments submitted 24 hours after the due date will be penalized by 50%. 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 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 recordings may not post until a few days after the synchronous 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!** Lecture will be recorded each week and one lab section will be recorded each week and made available as soon as we receive the processed videos (generally 2-3 days). Students will be muted and their videos turned off by default since there are so many of you! I encourage you to ask questions using the chat or by using the raise your hand

feature. 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.

Correspondence

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 on Piazza.

We will use **Ed** 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.

GSIs will not respond to Ed discussion questions during holidays and breaks. GSIs will respond to Ed 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. Ed will be inactivated the day of tests.

For questions and concerns that are not related to course content, please email ph142@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, I/we commit to teaching this course, to the best of my/our ability, with an antiracist, racial justice, and equity-minded lens. I/we invite you to take this journey with me/us by being fully present. I/we am/are interested in your perspectives and in the value and knowledge you bring to help make this an enriching classroom environment.

I/We view this syllabus as a dynamic document oriented toward learning and not just coverage of material; thus, I/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. I/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, I/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.

- I/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.

- I/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 ph142@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