Stats 102A Course Syllabus

Introduction to Computational Statistics with R

Miles Chen, PhD

2024 Summer Quarter - Session A

Contents

Course Information

Summary

This class is an introduction to the subject of computational statistics and data science. The course is thematically split into two parts. The first part will focus on learning the tools and the necessary skills for computational statistics. Students will learn intermediate R programming and usage of some of its functions and packages. The second part of the class will focus on some foundational methods in computational statistics. This includes numerical methods such as root finding, and mathematical optimization. It will continue on to cover random variables, simulation, and bootstrap methods to answer statistical questions.

Learning Outcomes:

- Students will demonstrate mastery of R programming fundamentals including the writing of functions, conditional statements, and loops.
- Students will identify and utilize appropriate packages in R to create code to prepare data for analysis.
- Students will demonstrate proficiency at "scraping" data from the internet and preparing the data for analysis.
- Students will create advanced data visualizations to address statistical questions that arise from a data set.
- Students will write code to carry out randomization and permutation tests, bootstrapping, and kernel estimation.

Prerequisites:

Stats 10 (or equivalent), Stats 20, Math 33A

Lecture Times

- Lectures will be provided in a series of asynchronous recorded videos.
- Each week, several videos will be posted. The total time of the video content will be approximately 220 minutes each week.

Office Hours

Questions and issues will be resolved the fastest in office hours.

- Mondays 4 - 4:50PM
- Tuesdays 4 - 4:50PM
- Wednesdays 4 - 4:50PM
- Thursdays 4 - 4:50PM
- Zoom link:
  - https://ucla.zoom.us/j/98887074323?pwd=bjFqd2ZCNnRqQmMvU3dnYkVad0RqQT09
  - Meeting ID: 988 8707 4323
  - Passcode: 916280
  - Please **turn on your cameras**

Also general office hours (Open to students in any class and to former students):

- Saturdays 10AM-11AM.
  - https://ucla.zoom.us/j/96806536111
  - Meeting ID: 968 0653 6111
  - Passcode: 023194

No office hours on holidays.

Exam Dates:

- Midterm Exam:
  - Take Home. Released Week 3 Wednesday July 3. Due Week 3 Saturday July 6 11:59PM.
- Final Exam:
  - Week 6 Wednesday, July 31 .
  - Exam will be proctored via Zoom.
  - Please make sure you are available to take the exam at the designated time.

Content

1. Week 1: Data structures, subsetting, flow control, writing functions, scope, environments
2. Week 2: readr, web scraping, tidyr
3. Week 3: dplyr, Regular expressions, Object oriented programming, ggplot
4. Week 4: Numerical methods: floating point arithmetic, Root finding methods
5. Week 5: Numerical optimization
6. Week 6: Randomization tests, permutation tests, bootstrapping

Homework Assignments and Tentative Due Dates

1. HW1: Introductory concepts. Due 6PM Mon Jul 1, 2024
2. HW2: Chutes and Ladders. Due 6PM Mon Jul 8, 2024
3. HW3: Webscraping and Regular Expressions. Due 6PM Mon Jul 15, 2024
4. HW4: Object-oriented programming. Due 6PM Mon Jul 22, 2024
5. HW5: Numeric methods. Due 6PM Mon Jul 29, 2024
6. HW6: Randomization / Bootstrap hypothesis testing. Due 6PM Fri Aug 2, 2024
   - Note that the final homework assignment is due *after* the final exam

Course website

All course materials, announcements, and assignments will be posted via Canvas / Bruin Learn http://bruinlearn.ucla.edu

Course discussion forum on Campuswire. You should have received an email with an invitation to join.

- Campuswire forum: https://campuswire.com/c/G7BD19A50/feed

The professor will participate on Campuswire and will occasionally provide answers to questions and will occasionally endorse good answers by students. I do not actively monitor every single question and student response on Campuswire. It is possible that an incorrect answer gets upvoted by other students as the 'best' answer despite being factually incorrect. I will only stand by the content I write myself and the answers that I endorse.

Office Hours:

Office hours are my preferred method of contact. Questions and issues will be resolved much more quickly via office hours.

When you come to office hours, please **introduce yourself**. Say "Hi Miles, I'm Joe Bruin." Do this **every** time you visit me until I start calling you by your name.

I like when students come to office hours with questions about material. I love to explain things and to help students understand.

I am happy to make accommodations for students who face difficult circumstances and may need extensions for assignment deadlines. Please do not hesitate to visit office hours. I also like when students come to office hours to tell me more about themselves and to seek counsel about classes to take or next steps.

I am happy to correct grading mistakes. I do not want to get in arguments with students over points. I do not like arguing whether a particular mistake should be a 5 or 10 point deduction.

Software

In this course, we will use RStudio, an IDE for the software package R. Students are expected to have taken Stats 20 and have a working knowledge of using R prior to enrolling in the course.

Books

We will use excerpts from a handful of different resources: a few chapters from one book, a few from another, and so on. The books I've selected are available for free to UCLA students via our digital library subscriptions. There is no need to purchase the textbooks. I have not requested that any of the books be stocked in the bookstore.

If students wish to have a physical copy of the textbook, they may purchase one at any book retailer of their choice.

[Wickham] *Advanced R* by Hadley Wickham, 2nd Edition

An online version of the textbook can be read for free: https://adv-r.hadley.nz/

[Maillardet et al.] *Introduction to Scientific Programming and Simulation Using R* by Robert Maillardet, Owen Jones, and Andrew Robinson

PDF of the textbook is free for UCLA students connected via the VPN:

https://www.taylorfrancis.com/books/introduction-scientific-programming-simulation-using-owen-jones-robert-maillardet-andrew-robinson/10.1201/9781420068740

[Chang] *R Graphics Cookbook* by Winston Chang

Free for UCLA students at:

https://www.oreilly.com/library/view/r-graphics-cookbook/9781491978597/?ar

Choose "Sign in with your institution" and select "University of California, Los Angeles"

[Zieffler et al.] *Comparing Groups: Randomization and Bootstrap Methods Using R* by Andrew S. Zieffler, Jeffrey R. Harring, and Jeffrey D. Long

PDF of the textbook is free for UCLA students connected via the VPN: http://onlinelibrary.wiley.com/book/10.1002/9781118063682

Connecting to the UCLA VPN:

To access some of the textbooks, you must connect via the UCLA VPN https://www.it.ucla.edu/it-support-center/services/virtual-private-network-vpn-clients

Grading

Grade Breakdown

- 15% Lecture Viewing Quizzes
- 36% Homework (Up to 6 assignments, none are dropped)
- 4% Campuswire Participation
- 20% Midterm Exam
- 25% Final Exam

Letter Grade cutoffs

Letter grades are assigned on a straight scale as follows:

- 59.9 and below: F
- 60.0 - 62.9: D-, 63.0 - 66.9: D, 67.0 - 69.9: D+
- 70.0 - 72.9: C-, 73.0 - 76.9: C, 77.0 - 79.9: C+
- 80.0 - 82.9: B-, 83.0 - 86.9: B, 87.0 - 89.9: B+
- 90.0 - 92.9: A-, 93.0 and up A, top 5% of students: A+

I do not curve grades. The top 5% of students will get an A+. An A+ is the only grade assignment that is assigned based on performance relative to peers.

If you are a majoring in Statistics&DS or Data Theory or pursuing a minor it is important that you take the class for letter grade. Taking the grade as pass/no-pass will not fulfill the requirements for completing the major or minor.

Lecture Viewing Quizzes

Lecture videos will be accompanied with a short quiz. Please be sure to respond to each quiz.

It is considered academic dishonesty to share the answers to the viewing quiz or to gain the answers to the viewing quiz by a method that does not involve watching the lecture.

Homework

There will be up to 6 homework assignments. Each homework assignment is worth the same amount. Each homework assignment will be scored out of a total of 100 points. None of the homework scores will be dropped. It is your responsibility to verify that your homework assignment successfully uploaded by the deadline.

All homework assignments will be posted as an R Markdown file. Students will download the R Markdown file, modify it with their own answers, and submit the rendered PDF file and the Rmd file. The rendered PDF file is submitted to Gradescope. The modified Rmd file is submitted to Bruin Learn.

Deadlines and late penalties are based on the submission time of the PDF file to Gradescope. Penalties will not be applied to Rmd files that are submitted late to Canvas, so long as the output of the rmd file matches the submitted PDF.

Homework assignments must not refer to any resources on the local machine or to files that are not publicly available online. Students should not make manual edits to a data file on their local machines.

Homework submission requirements

The graders have to go through many students' submissions. The grader will not spend more than 5-10 minutes on each submission. The grader does not have the luxury of trying to run code or debug code, or even see what individual lines of code can do.

The grader will look for the printed results, so it is **very important** that students properly print the requested results in their documents.

It is your responsibility to read through the rendered output and to make sure that every problem and task has the required output before submitting your assignment.

Example of bad submission: Problem 1: Use a loop to produce a vector of the first 10 numbers in the Fibonacci sequence

```
x <- c(1, 1)
for(i in 3:10) {
  x <- c(x, x[i - 1] + x[i - 2])
}
```

We have no idea what was produced. You will receive 0 points for this submission.

Example of good submission: Problem 1: Use a loop to produce a vector of the first 10 numbers in the Fibonacci sequence

```
x <- c(1, 1)
for(i in 3:10) {
  x <- c(x, x[i - 1] + x[i - 2])
}
print(x)
  [1]  1  1  2  3  5  8 13 21 34 55
```

This is the exact same submission, but adds the line `print(x)`. The difference, while small, makes a huge difference.

It is your responsibility to read through the rendered output and to make sure that every problem and task has the required output before submitting your assignment.

Gradescope submission requirements Submissions to Gradescope must be a PDF with letter-sized pages that was knit via LaTeX. Students must not knit their output to a Word document or HTML document and then save the file as a PDF. Improperly formatted submissions will lose points.

Students must properly tag the pages of their homework assignments on Gradescope. If the output for a particular problem spans multiple pages, all relevant pages must be tagged for that portion. Improperly tagged submissions will lose points.

Late Policy for Homework

Homework deadlines are based on local time in Los Angeles, California.

There is a 10 minute grace period for homework submissions. Submissions up to 10 minutes and 59 seconds late will be accepted with no penalty.

Homework assignments submitted 11 minutes late or more will be accepted with penalty. There is a deduction of 5 points for being less than one hour late. An additional 5-point deduction will be taken for each additional hour it is late.

Deductions are suspended between the hours of 12 am and 5:59 pm the following day. The penalties resume at 6 pm.

The maximum late penalty is 60 points.

So, if an assignment is due at 6pm, a homework assignment submitted on CCLE at

- 6:10:59 pm will receive no penalty.
- 6:11:00 pm will receive 5-point penalty.
- 6:59:59 pm will receive 5-point penalty.
- 7:00:00 pm will receive 10-point penalty.
- 7:59:59 pm will receive 10-point penalty.
- 8:00:00 pm will receive 15-point penalty.
- 11:59:59 pm will receive 30-point penalty.
- 12:00:00 am (following day) will receive 30-point penalty.
- 5:59:59 pm (following day) will receive 30-point penalty.
- 6:00:00 pm (following day) will receive 35-point penalty.
- 10:59:59 pm (following day) will receive 55-point penalty.
- 11:00:00 pm (following day) will receive 60-point penalty.
- Anything later will be accepted with a 60-point penalty.

I strongly advise uploading the homework to CCLE well in advance of the deadline in case there are connectivity problems or server issues.

After 7 calendar days, late homework assignments will not be accepted. Exceptions beyond the 7-day submission window will be granted in only extreme circumstances and will only be granted via an office hours meeting with the professor.

Homework extensions

I will grant a 72-hour reprieve from the late penalty for homework submissions if the homework is accompanied with documentation. If you need a homework extension but do not have documentation, you can contact the professor to request an extension. The professor's reply granting approval of an extension can then be used as documentation for an extension.

The documentation must be included as part of the homework submission PDF.

Gradescope accepts only one file per submission for homework assignments. Your homework PDF should include the documentation showing the **date** and reason for your inability to submit the homework on time.

Do not email the professor or your TA with your documentation.

Extensions beyond 72 hours will be granted only via an office hours meeting (Zoom or in-person) with the professor. Requests for extensions must be made prior to the homework deadline. Extension requests made via email will be ignored.

Requests for extensions after the homework deadline will be granted only in extreme situations. The request must be made via office hours. Extension requests made via email will be ignored.

Homework Regrades

Homework assignments are graded by designated graders (graduate students in the Statistics Department). Occasionally a grader may make a mistake in grading your submission.

Both the TA and the professor have the authority to change the grade of a homework assignment.

All requests for regrading a homework assignment must be submitted via Gradescope.

Do not email the TA or the professor for a request to regrade an assignment.

Only submit a regrade request if you are certain that a mistake was made in the grading of your homework.

During a regrade request, the TA or Professor may discover and deduct points for mistakes that were not found during the initial grading. It is possible your regrade request results in a lower grade for the assignment.

Campuswire Grading:

Participation on Campuswire accounts for 4 percentage points of your final grade. I will download Campuswire's reputation report on an arbitrary day during Week 10 and will enter grades as follows:

- Students who remain at noob level (Level 0) will get 0 points.
- Students who achieve starter level (Level 1) will get 1 point.
- Students who achieve starter level (Level 1) AND earn 50 reputation points will get 2 points.
- Students who achieve intermediate level (Level 2 - Red bird) AND 100 reputation points will get 3 points.
- Students who achieve intermediate level (Level 2 - Red bird) AND 150 reputation points will get 4 points (full credit).
- Students who achieve advanced level (Level 3 - Eagle) will get 5 points (full credit plus 1 extra credit point)

Students can check their reputation level in the "Grades" section on Campuswire.

Low quality posts may be removed at the instructor's discretion and will not count towards a student's reputation score.

It is important that when you create your Campuswire account, you enter your first and last name as they appear on Canvas / Bruin Learn.

Make up policy for Exams

If you know in advance that you will miss an exam, contact me and I can schedule you for a make-up exam.

If an emergency arises and you feel you will not be able to perform well on the exam, let me know right away with documentation and I can schedule you for a make-up. Please do not take the exam, perform poorly, and then request a make-up exam. Once you take the exam, I cannot allow you to retake it. If you have an emergency, I will be flexible and will schedule a make-up exam.

If you miss an exam, you will need to provide documentation for your reason for missing the exam.

Make-ups of the exam must be taken within 7 days of the actual exam or a zero will be recorded, unless an exception is made by the professor. (For example, a missed exam originally given on Wednesday must be made up by the following Wednesday.)

Course Policies

Copyright Policy

All course materials posted on the course website or distributed in class (including but not limited to lecture slides, homework assignments, lecture videos, quizzes, exams) are intended for personal use only by students currently enrolled in Stats 102A with Miles Chen. **It is a violation of course policy to post, share, or distribute any course material electronically or physically without permission from the instructor, even after the course is over.**

Academic Integrity

As a student and member of the University community, you are here to get an education and are, therefore, expected to demonstrate integrity in your academic endeavors. All students must uphold University of California Standards of Student Conduct as administered by the Office of the Dean of Students. Students are subject to disciplinary

action for several types of misconduct, including but not limited to: cheating, multiple submissions, plagiarism, prohibited collaboration, facilitating academic dishonesty, or knowingly furnishing false information.

In addition, each student is the sole owner of their own code and work and must NOT do any of the following:

- Submit work that was written by someone other than the student.
- Publish code or solutions online.
- Post the course questions on forums other than the designated course discussion forum (Campuswire). This means students cannot post questions on places like Stack Overflow or other similar places.
- Submit someone else's work or a modification of that work, with or without that person's knowledge.
- Allow someone else to submit their work or a modification of their work.
- Use an AI tool (including but not limited to Chat-GPT, Github Copilot, Google Bard) to write code or responses.
- Contract course work out to others.
- Plan or execute with another student(s) some form of cheating during an exam.
- Make use of unauthorized material during an exam or for preparation of an exam.
- View completed assignments or exams from the current or previous iterations of the course, regardless of their source.

For more information about academic integrity and student conduct code, please go to http://www.deanofstudents.ucla.edu/Student-Conduct

Gender Discrimination, Sexual Violence, and Harassment

Title IX prohibits gender discrimination, including sexual harassment, domestic and dating violence, sexual assault, and stalking.

http://www.sexualharassment.ucla.edu/

http://www.sexualviolence.ucla.edu/

> The University of California is committed to creating and maintaining a community dedicated to the advancement, application and transmission of knowledge and creative endeavors through academic excellence, where all individuals who participate in University programs and activities can work and learn together in an atmosphere free of harassment, exploitation, or intimidation.

If you, or someone you know, have been a victim of sexual harassment or sexual violence, please seek help.

Any type of harassment of any other student will not be tolerated. Please be aware that your TA and Lecturer are legally compelled to report any information in regards to sexual harassment or sexual violence to the Title IX officer.

Confidential support and advocacy is available for students who have experienced sexual harassment or sexual violence at the CARE Advocacy Office for Sexual and Gender-Based Violence, John Wooden Center West 1st Floor, (310) 206-2465, CAREadvocate@careprogram.ucla.edu. Sexual violence or sexual harassment can also be reported directly to the University's Title IX Coordinator, 2241 Murphy Hall, titleix@conet.ucla.edu, (310) 206-3417.

Reach a confidential UCLA advocate, 24 hours a day, 7 days a week by calling the CAPS 24/7 Crisis Hotline (310) 825-0768

Mental Health and Student Well-Being

> Counseling and Psychological Services (CAPS) is a multidisciplinary student mental health center for the UCLA campus. CAPS supports the academic and student development missions of the University and the Division of Student Affairs. We're here to support your mental health needs as you pursue your academic goals. Our services are designed to foster the development of healthy well-being necessary for success in a complex global environment.

http://www.counseling.ucla.edu

Attending and completing university coursework is difficult and can become burdensome both physically and psychologically. Take care of your mental well-being and seek support from some of the many excellent resources offered at CAPS.

Accommodations for students with disabilities

Students needing academic accommodations based on a disability must contact the Center for Accessible Education (CAE) at (310) 825-1501 or present in person at Murphy Hall A255. As the professionals delegated authority from the campus to determine reasonable disability accommodations, CAE will assess all requested accommodations and communicate appropriately with faculty. In the event that a student has approval for proctoring arrangements during exams, please inform your respective professors and/or Teaching Assistant before date of exams. When possible, students should contact the CAE within the first two weeks of the term as reasonable notice is needed to coordinate accommodations.

For more information visit http://www.cae.ucla.edu

Course Grade Changes

After course grades have been submitted to the registrar, grades are final. Grade changes will only be considered if there has been a clerical or procedural mistake. Students have one quarter to make requests for a grade change. Graded exams and other materials will be kept for one quarter. After one quarter, course grade changes will not be made.

General Syllabus Policy

The instructor reserves the right to make changes to the syllabus. All changes to the syllabus will be announced to the class.

If any policy in the syllabus requires interpretation, the instructor will clarify the policy and the instructor's interpretation of the policy is final.

Additional Information

Undergraduate Mentoring

I am a mentor for undergraduates for the UCLA Department of Statistics. This means that you may visit me during office hours (or by appointment) for an informal meeting where you can ask questions unrelated to course content, such as research opportunities, graduate studies, career paths, or any other topic pertinent to your education. You are welcome to visit and meet even after the course ends.

Requests for Letters of Recommendation

If you think you might request a letter of recommendation in the future, please make an effort to make yourself known to me. I am not able to write strong letters of recommendation for students I do not know well.

I also do not have a good memory, so please visit office hours frequently and repeatedly introduce yourself to me. Please attend class in person and sit near the front of the classroom where I will see and recognize your face. Actively participate in class.

For a full guideline, please see this page on my personal website: https://smileschen.com/lor.html