

STAT:7400 Spring 2024

Computer Intensive Statistics

Lectures MWF 11:30 AM –12:20 PM 3 SH

Instructor Boxiang Wang, 261 SH, boxiang-wang@uiowa.edu
Teaching Assistant Zhenhan Fang, zhenhan-fang@uiowa.edu
Department Statistics & Actuarial Science, 241 SH.
<http://www.stat.uiowa.edu>
DEO Professor Kung-Sik Chan, DEO, 319-335-0712,
kung-sik-chan@uiowa.edu

Course Website.

Announcements, homework problems, lecture notes, etc: <http://icon.uiowa.edu>.

Course Schedule.

Important dates and deadlines are posted on the following Google Sheet:

<https://docs.google.com/spreadsheets/d/1L58SNkqMmdzIeb90Fqh6CGVP1URpxDQ45GnBzsSjX9s/edit?usp=sharing>

Course Description and Objectives.

This course is designed for graduate students in statistics and data science. The goal is to develop advanced skills, knowledge, and tools useful in applying modern computationally intensive statistical methods to research in any field. Topics will be selected from design and analysis of simulation experiments, Markov chain Monte Carlo, optimization, machine learning, and deep learning. Other advanced topics may be covered if time permits.

This course is primarily in R and Python, along with substantial exposure in C/C++. Some data science techniques like version control, graphical methods, and parallel computing will be covered. Extensive experience about writing R packages is expected.

A substantial portion of materials in this course are adapted from Professor Tierney's previous teaching.

https://homepage.stat.uiowa.edu/~luke/classes/STAT7400-2023/_book/

Prerequisites.

STAT:5200 (or BIOS:5710) and STAT:3101 (or STAT:3120 or STAT:4101 or STAT:5101) and STAT:5400.

Textbook.

No required textbook.

Optional Reference:

- Maria L. Rizzo (2008). *Statistical Computing with R*. Chapman & Hall/CRC.
- Trevor Hastie, Robert Tibshirani, and Jerome H. Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*. The 2nd Edition, New York: Springer.
https://hastie.su.domains/ElemStatLearn/printings/ESLII_print12.pdf
- Norman Matloff (2011). *The Art of R Programming: A Tour of Statistical Software Design*. No Starch Press.
- Hadley Wickham and Golemund Garrett (2016). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O’Reilly Media, Inc.
<https://r4ds.had.co.nz/>
- Rafael A. Irizarry (2019). *Introduction to Data Science*. Chapman & Hall/CRC.
<https://rafalab.github.io/dsbook/>

Grading.

Your semester grade will consist of the following components:

Homework	50%
Computing Project	40%
Competing Project	10%
<hr/>	
Total	100%

As a *rough* guide, **A,A-** = 90% – 100%, **B+,B,B-** = 80% – 90%, **C+,C,C-** = 70% – 80%, **D+,D,D-** = 60% – 70%. A grade of **A+** represents very exceptional work and is given only in extraordinary situations.

There is no exam in this course. The homework and the computing projects are individual works. You may discuss general issues and approaches with your fellow students, but your work must be your own. See details below. Throughout this course, if you use any references, including solutions to similar problems prepared by other students, you **must** cite and credit your sources.

Homework.

Homework assignments consisting of a mix of computational and theoretical problems will be posted on ICON roughly on every other Friday. Exceptions to this schedule will be announced in class.

Some problems will cover material not addressed in class and may require additional reading. Suggested reading will also be posted on the class web site when appropriate, but you should also seek out and explore relevant references on your own. Homework should be submitted electronically through the ICON submission tools (<https://icon.uiowa.edu>) or GitHub.

Show your work when solving written homework problems. Complete code and output must be submitted for computer problems. Always document your answer. Many students find that these assignments take a long time to complete, so plan your time accordingly.

You are strongly encouraged to study with others as teamwork is an important component of a statistician/data scientist. However, if you work with others on homework, please: a) write up your own assignment and make sure you completely understand all solutions that you submit, and b) write the names of the others in your study group on your assignment.

Note your write-up must be your own. Do not use solutions from previous years.

Late homework is accepted only as required by the university policy, i.e. due to “illness, mandatory religious obligations, or other unavoidable circumstances or University activities.” This is a strict rule; however, you are allowed to drop the lowest homework grade.

Office Hours.

12:20 PM – 3:20 PM Wednesday

Ask questions if you are confused or think a point needs more discussion. Questions can lead to interesting discussions.

Computing Project.

Students registered for this class are required to complete a computing project. You should work on this project on your own. Your project should represent about **40 hours** of work with high quality on a topic of your choice.

You should start to think about the topic as soon as possible. Your project must contain some new ideas with creativity. Below are some ideas:

- **Methodological:** Improve an existing method and support your conclusion with numerical evidence. It is totally acceptable that your proposal fails to outperform the benchmark baselines.
- **Computing:** Produce a new R package. Try to argue why your package brings something new. Ideally you can publish your package on The Comprehensive R Archive Network (CRAN). <https://cran.r-project.org/web/packages/>.
- **Applied:** You may analyze an interesting data set using a relatively new method related to ideas introduced in the class.

There are many possible choices for the topic of your project. Identifying a suitable topic is an important part of your task. The project should represent new work, **not something you have done for another course or as part of your thesis**. Note that this course brings you a gold opportunity to take risk of attempting some ideas

that may fail to work, so feel free to be creative in this project.

The computing project will be carried out in three stages.

- **Computing Project Proposal.** Due at 11:30 AM on Mar 4. The proposal has a one-page limit (not restrictively enforced). This is a detailed description of what you plan to do, including questions to be addressed, software to be used, and methods to be investigated or applied, etc.
- **Computing Project Midterm.** Slides due at 11:30 AM on Apr 1. We will schedule individual sessions during that week to give your personalized suggestions on your slides and presentation. You will give a 15-minute presentation on Apr 8, Apr 10, or Apr 12, with focus on the background, proposed approach, and preliminary results.
- **Computing Project Final Showcase.** You will give a 5-minute speed presentation with a very brief review on the background and showcase your final results.

Competing Project.

You will work in a group of three or four, depending on the final enrollment of this course, on a real-world data competition. You can choose a competition from Kaggle <https://www.kaggle.com/competitions> or any other similar competition you are interested in.

The Competing Project Plan is due on Mar 18. The plan should simply include your team members and which competition you will attend. All the teams will showcase your results on Apr 26.

Attendance.

Attendance is required. Please email the instructor if you are going to be absent from class. See instruction for Absence from Class due to illness, emergence, or extenuating circumstances in <https://registrar.uiowa.edu/absence-class> .

See COVID-related questions in <https://coronavirus.uiowa.edu/>.

A student who is absent for more than five days may request the UI Service Center, 17 Calvin Hall, registrar@uiowa.edu, to notify me of the reason for the absence.

Students with UI-authorized activities must discuss their absences with the instructor as soon as possible. Religious obligations must be communicated within the first three weeks of classes. See also Exam Policies.

Copyright.

The lectures, slides, homework, quiz and exam questions are for use only by students attending STAT:7400 (0001). It is illegal to share with anyone else the course materials

without instructor's permission. See also Sharing of Class Recordings

Academic Honesty and Misconduct.

All students in CLAS courses are expected to abide by the CLAS Code of Academic Honesty. Undergraduate academic misconduct must be reported by instructors to CLAS according to these procedures. Graduate academic misconduct must be reported to the Graduate College according to Section F of the Graduate College Manual.

Student Complaints.

Students with a complaint about a grade or a related matter should first discuss the situation with the instructor, and finally with the Director or Chair of the school, department, or program offering the course.

Undergraduate students should contact CLAS Undergraduate Programs for support when the matter is not resolved at the previous level. Graduate students should contact the CLAS Associate Dean for Graduate Education and Outreach and Engagement when additional support is needed.

Drop Deadline for this Course

You may drop an individual course before the deadline; after this deadline you will need collegiate approval. You can look up the drop deadline for this course here. When you drop a course, a "W" will appear on your transcript. The mark of "W" is a neutral mark that does not affect your GPA. Directions for adding or dropping a course and other registration changes can be found on the Registrar's website. Undergraduate students can find policies on dropping CLAS courses here. Graduate students should adhere to the academic deadlines and policies set by the Graduate College.

Communication: UI Email.

Students are responsible for all official correspondences sent to their UI email address (uiowa.edu) and must use this address for any communication with instructors or staff in the UI community. For the privacy and the protection of student records, UI faculty and staff can only correspond with UI email addresses.

Mental Health Resources and Student Support.

Students are encouraged to be mindful of their mental health and seek help as a preventive measure or if feeling overwhelmed and/or struggling to meet course expectations. Students are encouraged to talk to their instructor for assistance with specific class-related concerns. For additional support and counseling, students are encouraged to contact University Counseling Service (UCS). Information about UCS, including resources and how to schedule an appointment, can be found at counseling.uiowa.edu. Find out more about UI mental health services at mentalhealth.uiowa.edu.

Student Care and Assistance provides assistance to University of Iowa students who are experiencing a variety of crisis and emergency situations, including but not limited

to medical issues, family emergencies, unexpected challenges, and sourcing basic needs such as food and shelter. More information on the resources related to basic needs can be found at this website. Students are encouraged to contact Student Care & Assistance in the Office of the Dean of Students (Room 135 IMU, dos-assistance@uiowa.edu, or 319-335-1162) for support and assistance with resources.

University Policies On this website, you will find the university course policies and resources for students. You will find information about the following topics:

Basic Needs and Support for Students

Classroom Expectations

Exam Make-up Owing to Absence

Free Speech and Expression

Mental Health

Military Service Obligations

Non-discrimination

Religious Holy Days

Sexual Harassment/Misconduct and Supportive Measures

Sharing of Class Recordings

Accommodations for Students with Disabilities.

The University is committed to providing an educational experience that is accessible to all. If a student has a diagnosed disability or other disabling condition that may impact the student's ability to complete the course requirements as stated in the syllabus, the student may seek accommodations through Student Disability Services (SDS). SDS is responsible for making Letters of Accommodation (LOA) available. The student must provide an LOA to the instructor as early in the semester as possible, but requests not made at least two weeks prior to the scheduled activity for which an accommodation is sought may not be accommodated. The LOA will specify what reasonable course accommodations the student is eligible for and those the instructor should provide. Additional information can be found on the SDS website.