Stat 610: Statistical Computing

Meeting time: Tuesdays and Thursdays, 11:30-12:45pm — Meeting location: HH 1000
Website: jfukuyama.github.io/teaching/stat610

Instructor: Prof. Julia Fukuyama — jfukuyam at iu dot edu
Office hours: Thursdays 2-4pm

Associate Instructor: Mr. John Koo — johnkoo at iu dot edu
Office hours: Mondays 4-6pm

Course Overview

As a statistician, you will need to manipulate data, optimize, and simulate. You will also need to know enough about how the methods you use work to diagnose problems when they arise and to be able to implement modified versions when the standard implementations don't suit your purposes.

You also need to write accurate, clean, maintainable, demonstrably correct code. To that end, the first half of the class will be devoted to how to program well, with statistical tasks giving us the computational problems.

Once we have the software engineering down, we will move on to the algorithms used in applied statistics. These can be roughly broken up into optimization methods and stochastic simulation methods.

Textbooks

The primary textbook for the first half of the course with be *The Art of R Programming*, by Norman Matloff. *The R Cookbook*, by Paul Teetor, will also be useful. The primary textbook for the second half of the course will be *Numerical Analysis for Statisticians*, by Kenneth Lange. Additional readings will be posted on the course website.

Course format and covid considerations

The course will be available either in person or online. Lectures will be held in HH 1000, which has a capacity of 180 under normal circumstances and which should be large enough to accommodate all students in the course with plenty of room for social distancing. The lectures will all be recorded and posted on Kaltura for students who cannot or wish not to be on campus. The assignments will all be posted online and turned in online, and there will be no required in-person elements to the course.

Class Schedule

Dates and topics subject to change.

Week 1

- Data types and data structures
- Flow control and looping
- Reading: Matloff Chapters 1-6, Chapter 7.1

Week 2

- Text
- Regular expresions
- Reading: Matloff Chapter 11
- Homework 1 out, due September 10

Week 3

- Writing and calling functions
- Code refactoring
- Reading: Matloff Sections 7.6, 7.9, 14.1-14.3
- Homework 2 out, due September 17

Week 4

- Split/apply/combine
- Reading: Matloff Chapter 10, R Cookbook Chapter 6
- No new homework

Week 5

- Shape changing/merging/transformations
- Debugging
- Reading: Matloff Chapter 13
- Homework 3 out, due October 1

Week 6

- Testing/top-down design
- Reading: Matloff Section 7.6
- Homework 4 out, due October 8

Week 7

- Version control/git
- Homework 5 out

Week 8

- Performance enhancement and code profiling
- Fitting and using statistical models
- Reading: Matloff Chapter 14

Week 9

- Gradient methods/Newton's method
- Homework 6 out, due November 5

Week 10

- Convex optimization and the EM algorithm
- Homework 7 out, due November 19

Week 11

- Random number generation
- Monte Carlo integration
- Homework 8 out

Week 12

- ABC

Week 13

- Markov chains/Metropolis Hastings

Week 14

- Machine learning, briefly

Week 15

- Final project presentations/code review

Assessment

Assessment will be based on a combination of homework and a final project. Final grades will be based on:

- 50% homework
- 40% final project
- 10% participation

Participation will have an idiosyncratic meaning this semester, which we will figure out together.

There will be 8 homeworks over the course of the semester, generally graded out of 5 points, with one point for a good-faith effort at all the problems, 5 points for correct answers with clean code, and an intermediate number of points otherwise.

Homeworks will be assigned on Sundays and due the following Tuesday (9 days later). At the time the homework is assigned, we will generally not have covered all the material needed to complete the homework, but we will have covered everything by the Thursday before the due date. The idea is to give you the homework early enough that you can think about it while the material is being covered in lecture. Therefore, it will generally be a good idea to take a look at the homework when it is assigned even if you aren't able to complete all the problems yet.

Course Policies

Late Policy

Each student has three "free" late days to use on assignments. After that, homework is penalized at one point (out of five, remember!) per 24 hours. Special accommodations may be granted if you ask very early or if there are extenuating circumstances.

Academic Integrity

You are expected to abide by the guidelines of the IU Code of Student Rights, Responsibilities, and Conduct (http://studentcode.iu.edu/responsibilities/academic-misconduct.html) regarding cheating and plagiarism. Any ideas or materials taken from another source must be fully acknowledged and cited.

Disability Accommodation

Please contact me if you require assistance or academic accommodations for a disability. You should establish your eligibility for disability support services through the Office of Disability Services for Students in Wells Library W302, 812-855-7578.

Student Rights

Any student who believes another person in a class is threatening the safety of the class by not wearing a mask or observing physical distancing requirements may leave the class without consequence.

Attendance

The student responsibility form requires that you take your temperature every morning and that you refrain from attending class if you have a temperature of 100.4 or other symptoms of illness. In order to ensure that you can do this, attendance will not be a factor in the final grade. Attendance may still be taken to comply with accreditation requirements.

Assigned Seating

In order to ensure we can contact you in the event you are exposed to COVID-19, you must remain in your assigned seat for the entire semester.