Syllabus - Computational Statistics

Course Information

Course Title: MATH 4750/MSSC 5750: Computational Statistics
Meeting Time: TuTh 2:00pm - 3:15pm
Location: David Straz Nursing 402 (Microsoft Teams)

Instructor Details

- **Name:** Mehdi Maadooliat, Ph.D.
- **Office:** CU 351
- **Office Hours:** TuTh 12:25 - 1:55pm in CU 351 or by e-mail

Course Description

Computational statistics and statistical computing are two areas that employ computational, graphical, and numerical approaches to solve statistical problems. Competent statisticians must not just run existing programs but understand the principles behind them. This class introduces statistically-oriented programming targeted at statistics majors, with no extensive programming background assumed.

This course covers traditional core material of computational statistics with an emphasis on using R through an example-based approach.

Learning Outcomes

By the end of the course, students will: - Understand core programming concepts, including data structures, functions, iteration, debugging, logical design, and abstraction. - Write maintainable code, debug, and test code for correctness. - Set up and run stochastic simulations in parallel. - Fit basic statistical models and assess the results. - Work with and filter large datasets. - Comment and organize code effectively.

Prerequisites

- COSC 1010
- MATH 3100
- MATH 4700 or MATH 4780

Textbooks

- **Main:** *Statistical Computing with R*, 2nd edition by Maria L. Rizzo. Chapman and Hall/CRC, 2019. ISBN: 1466553324.
- **Supplementary:** *Computational Statistics*, 2nd Edition by Geof H. Givens and Jennifer A. Hoeting. Wiley, 2012. ISBN: 9780470533314.

Grading Breakdown

- **Class Participation:** 10%
- **Homework:** 30%
- **Team Project:** 30%
- **Final Exam or HW/Project (Instructor will decide):** 30%

Grading Scale

Grade	Range
A	95 - 100%
A-	89 - 94.99%
B+	84 - 88.99%
B	78 - 83.99%
B-	73 - 77.99%
C+	67 - 72.99%
C	62 - 66.99%
C-	56 - 61.99%
D+	51 - 55.99%
D	45 - 50.99%
F	$< 44.99\%$

Notes:

- For students in MSSC 5750, there will be extra questions in Homework.

- No late homework will be accepted, and no make-up work will be allowed unless in the case of an emergency. Submit incomplete work if necessary, but ensure scanned PDFs are legible before submission.

Make-up Policy

No make-up exams or homework unless there is an emergency.

Attendance

Attendance is required and subject to the College of Arts and Sciences policy.

Academic Honesty

Students are expected to follow the University's policy on academic honesty as outlined in the Bulletin.

TL;DR: Don't cheat!

Please abide by the following as you work on assignments in this course:

- **Collaboration:** Only work that is clearly assigned as team work should be completed collaboratively.

  - The homework assignments must also be completed individually and you are welcomed to discuss the assignment with classmates at a high level (e.g., discuss what's the best way for approaching a problem, what functions are useful for accomplishing a particular task, etc.). However you may not directly share answers to lab questions (including any code) with anyone other than myself and the teaching assistants.

  - The reading quizzes must be completed individually with absolutely no communication with classmates.

  - For the projects, collaboration within teams is not only allowed, but expected. Communication between teams at a high level is also allowed however you may not share code or components of the project across teams.

  - On individual assignments you may not directly share code with another student in this class, and on team assignments you may not directly share code with another team in this class.

- **Online resources:** I am well aware that a huge volume of code is available on the web to solve any number of problems. Unless I explicitly tell you not to use something, the course's policy is that you may make use of any online resources (e.g., StackOverflow) but you must explicitly cite where you obtained any code you directly use (or use as inspiration). Any recycled code that is discovered and is not explicitly cited will be treated as plagiarism.

- **Use of generative artificial intelligence (AI)**: You should treat generative AI, such as ChatGPT, the same as other online resources. There are two guiding principles that govern how you can use AI in this course:[1] (1) *Cognitive dimension:* Working with AI should not reduce your ability to think clearly. We will practice using AI to facilitate—rather than hinder—learning. (2) *Ethical dimension***:** Students using AI should be transparent about their use and make sure it aligns with academic integrity.

  - **AI tools for code:** You may make use of the technology for coding examples on assignments; if you do so, you must explicitly cite where you obtained the code. Any recycled code that is discovered and is not explicitly cited will be treated as plagiarism. You may use these guidelines for citing AI-generated content.

  - **AI tools for narrative:** Unless instructed otherwise, you may not use generative AI to write narrative on assignments. In general, you may use generative AI as a resource as you complete assignments but not to answer the exercises for you.

  You are ultimately responsible for the work you turn in; it should reflect *your* understanding of the course content.

If you are unsure if the use of a particular resource complies with the academic honesty policy, please ask the instructor.

Course Topics

- Managing Input/Output in R
- Data structures in R
- Data transformations: strings, factors, dates
- Statistical summaries and basic tests in R
- Graphics in R
- Monte Carlo Integration & Variance Reduction
- Bootstrap and Jackknife
- Density Estimation
- Regularization:

  - Ridge and Lasso
  - Smoothing

- Dimension Reduction and Clustering

- K-Means and Hierarchical Clustering
- Principal Component Analysis (PCA)

- Numerical Methods in R

  - Optimization
  - Expectation-Maximization (EM) Algorithm

- Advanced R Programming

  - Introduction to Parallel Computing
  - Object Oriented Programming in R
  - Developing an R Package
  - Developing an Interactive Shiny App
  - Functions and loops in C and using C in R

- Deep Learning in R

  - Basics of Neural Network
    * Implementation using TensorFlow and Keras
  - Fully Connected Network (FCN)
  - Convolutional Neural Networks (CNN)

Important dates

- **Monday, January 13:** Classes begin
- **Tuesday, January 21:** Drop/add ends
- **Saturday - Sunday, March 8 - 16**: Spring Break
- **Friday, April 11:** Last day to withdraw with W
- **Saturday, May 3**: Classes end
- **Monday, May 5, 10:30 am - 12:30 pm**: Presentations

For more important dates, see the full MU Academic Calendar.

Important Note

The syllabus may be modified throughout the course. Any substantial modifications will result in a reissued syllabus.