

STAT 535: Statistical Computing (Fall 2026)

Course Description:

This course will introduce computing tools needed for statistical analysis including data acquisition from database, data exploration and analysis, numerical analysis and result presentation. Advanced topics include parallel computing, simulation and optimization, and package creation. The class will be taught in a modern statistical computing language. *Prerequisites:* STAT 516/490S/416 and COMPSCI 121/INFO 190S/CICS 110. Prior knowledge of statistical methods and programming experience (STAT 525 or equivalent) is strongly recommended for this course.

Course Objectives:

At the end of the course, you should be able to:

1. Design and implement an end-to-end statistical analysis of a data set,
2. Connect notions of computational complexity to statistical methods for data analysis,
3. Collaborate effectively in a team to develop a solution to real-world problems using statistics,
4. Evaluate and critique the choice data structure for an algorithm, and
5. Communicate statistical analysis results using mathematics, verbal and visual means.

This course aims to not only give you the opportunity to learn fundamental computer science concepts critical for understanding machine learning algorithms, but also brings you into contact with real data and allows you opportunities to make meaningful quantitative contributions to problems in genetics, social science, and other disciplines.

Administrivia:

- Time: TTh 2:30PM-3:45PM
- Location: LGRC A203
- Learning Management System: Canvas
- Instructor: Maryclare Griffin
- Instructor Office: LGRT 1446, Zoom ID 848 427 5519 (as needed)
- Instructor Office Hours: Mondays 3:00-4:00, Wednesdays 11:00-12:00, also by appointment
- Instructor E-mail: mgriffin@math.umass.edu or maryclaregri@umass.edu
- Textbook: Class Lecture Notes.
- Reference Books (Freely Available)
 - The Art of R programming: Tour of Statistical Software Design, 2011, by Matloff. ISBN-13: 978-1593273842. This resource is available online via UMass library's website.
 - R for Data Science: Import, Tidy, Transform, Visualize, and Model Data (2e), 2023, by Wickham, Cetinkaya-Rundel, and Golemund. Freely available at: <https://r4ds.hadley.nz/>
- The computing in this course will be conducted in R (<https://www.r-project.org>) using RStudio (<https://rstudio.com>), both of which are freely available software available for multiple platforms.

Aspirational Schedule and Key Dates:

Week	T	Th	
1	9/8	9/10	Introduction to Coding and Integrated Development Environment (IDE)
2	9/15	9/17	Data Types/Structures, Introduction to Computational Complexity
3	9/22	9/24	Flow Control and Functions
4	9/29	10/1	Exam 1 9/29, Strings and Regular Expressions
5	10/6	10/8	Transformations of Data, SQL
6	10/13	10/15	Visualizing Data
7	10/20	10/22	Exam 2 10/20, Functional Coding
8	10/27	10/29	Optimization, and Model Selection
9	No class!	11/5	Random Number Generation and Sampling
10	11/10	11/12	Monte Carlo Methods
11	11/17	11/19	Exam 3 11/17, Bootstrapping
12	No class!	No class!	
13	12/1	12/3	Binding Lower Level Code, Parallel Computing
14	12/8	12/10	Final Presentations
15	12/15	No class!	Final Presentations

- 12/21: In person final Exam, 3:30pm-5:30pm in LGRC A203.

Grading:

- Problem Sets (Homework) 20%
- In Class Exams 30%
- Project Presentation 10%
- Final Exam 20%
- Participation 20%

Letter grades are typically as follows:

F	D*	D+*	C-*	C	C+	B-	B	B+	A-	A
<55	55+	59+	63+	67+	71+	75+	79+	83+	87+	90+

*Graduate students enrolled in undergraduate courses may receive these grades.

Problem Sets (Homework):

Weekly assignments are due every week before class on Thursdays on Canvas. Your homework must be submitted in with two files: (1) Quarto .qmd file that compiles to .pdf; (2) a .pdf file. Other formats will not be accepted. Your responses must be supported by both textual explanations and the code you generate to produce your result. No late homework will be accepted. Discussion of homework assignments with fellow students is encouraged. Use of LLM's is not prohibited but is strongly discouraged. The code and the final submission must be your own.

In Class Exams:

There will be three in class exams during the semester. They will test concepts from class and homework. They are worth 30% of your grade, which can be thought of as 30 total points. Exams will be worth 10 points each and weighted according to your performance. Let $e^{(1)}$, $e^{(2)}$, and $e^{(3)}$ refer to the highest, second highest, and lowest exam scores, respectively. The total in class exam score will be given by $\frac{1}{2}e^{(1)} + \frac{1}{3}e^{(2)} + \frac{1}{6}e^{(3)}$.

Project Presentation:

For the final project presentation, you will be presenting a topic of interest, related to statistical computing, such as a new computing package, a simulation of an advanced statistical model, or some of your own research results demonstrated using a statistical software or method, according to the provided instructions.

Final Exam:

The final exam will have two equally weighted parts worth a total of 20 points. One part will be like the in

class exams and will test concepts from class and homework. The other part will be an open ended writeup on the material your final project was based on.

Participation:

I will measure participation by taking attendance. Total attendance will be given by $\max\{20, d\}$ where d is the number of distinct class sessions a student attended.

Excused Absences:

Students who are absent due to a university-approved conflict (such as religious observance, athletic event, field trip, performance), health reasons, family illness, or other excusable extenuating circumstances remain responsible for meeting all class requirements and contacting me in a timely fashion about making up missed work. In legitimate and documented extenuating circumstances, please contact me and we will make reasonable arrangements.

Required Statements:

<https://www.umass.edu/senate/book/responsible-employee-required-syllabus-statements>