STAT 535: Statistical Computing (Spring 2025)

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:* Prerequisites: STAT 516 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: MW 2:30PM-3:45PM
- Location: LGRT A201
- Learning Management System: Canvas
- Instructor: Maryclare Griffin
- Instructor Office: LGRT 1446, Zoom ID 848 427 5519 (as needed)
- Instructor Office Hours: Mondays 12-1, Tuesdays 3-4, and 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 Grolemund. 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	M	W	Th	
2	2/3	2/5		Introduction to Coding and Integrated Development Environment (IDE)
3	2/10	2/12		Data Types/Structures, Introduction to Computational Complexity
4	No Class!	2/19	2/20	Flow Control and Functions
5	2/24	2/26		Strings and Regular Expressions
6	3/3	3/5		Transformations of Data, SQL
7	3/10	3/12		Visualizing Data
8	3/24	3/26		Functional Coding
9	3/31	4/2		Optimization, and Model Selection
10	4/7	4/9		Random Number Generation and Sampling
11	4/14	4/16		Monte Carlo Methods
12	No class!	4/23		Bootstrapping
13	4/28	4/30		Binding Lower Level Code, Parallel Computing
14	5/5	5/7		Final Presentations

- 5/16: Final Project Due by 5:00PM ET.

Actual Schedule and Key Dates:

Week	M	W	F	
2	2/3	2/5		Introduction to Coding and Integrated Development Environment (IDE)
3	No Class!	2/12		Data Types/Structures: Vectors
4	No Class!	No Class!		-
5	2/24	2/26		Data Types/Structures: Vectors and Matrices
6	3/3	3/5		Data Types/Structures: Matrices
7	3/10	3/12		Random Number Generation and Sampling
8	3/24	3/26		Functions, Reading in and Manipulating Data
9	3/31	4/2		Databases, SQL
10	4/7	4/9		More SQL, Lists, Strings and Regular Expressions
11	4/14	4/16	4/18	Transformations of Data, Visualization, Optimization
12	No class!	4/23		Monte Carlo Methods
13	4/28	4/30		Binding Lower Level Code, Parallel Computing
14	5/5	5/7		Final Presentations

- 5/16: Final Project Due by 5:00PM ET.

Grading:

- Problem Sets (Homework) 50%
- Project 40%
- Participation 10%

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 Wednesdays 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. However, the code and the final submission must be your own.

Project:
For the final project, 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.

Participation:
Throughout the semester, participation will be assessed in terms of the number of times you ask a question in class. Each unique class session you ask a question in counts for 1 percentage point towards your final grade, with a maximum of 10 percentage points.

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/non-responsible-employee-required-syllabus-statements`