Course Description

Computation is the beating heart of modern statistics. Large data sets, complex models, and novel data collection methods all necessarily require heavy computation, but computation also allows applications of elegant theories that defy pen and paper solutions. In this course, we will explore the complementary relationship of theory and application provided by modern computational tools and techniques using the R programming language. Theoretical results will be illustrated with a wide variety of applied research.

A detailed [Schedule of Topics (https://umich.instructure.com/courses/621159/pages/schedule-of-topics)](https://umich.instructure.com/courses/621159/pages/schedule-of-topics) gives the weekly material.

Instructors

- Lead Instructor: Dr. Mark Fredrickson
  - Email: mfredric@umich.edu
  - Zoom Office Hours using **[Office Hours Queue](https://officehours.it.umich.edu/queue/132)** ⤷ **(https://officehours.it.umich.edu/queue/132)**
    - Wednesday: 2 - 4pm
    - Thursday: 3 - 5pm
    - Friday: 9 - 11am
  -
  - GSI: Mingyi Tang, mycandy@umich.edu **[(mailto:yangly@umich.edu)](mailto:yangly@umich.edu)**
  - GSI Office Hours: Thursday: 8 - 11am (https://umich.zoom.us/j/6403253329 Passcode: 916138)

Meeting Times

- Tuesday and Thursday, 1:00 to 2:20pm, Angel Hall Auditorium D

Software

We will R for our programming language and recommend RStudio as the environment to interact with R (you can also use just R by itself or Visual Studio, but RStudio provides some useful tools).

We will use OneNote to share course notes.

You can get directions for installation here: **Software Installation (https://umich.instructure.com/courses/621159/pages/software-installation?wrap=1)**

Textbooks

There is no required textbook for the course. The course notes will be made available and are considered our primary reference. There are, however, several recommended books that cover the material in the course. We will most closely follow the first two references below.

- **Statistical Computing with R by Maria L. Rizzo (https://ebookcentral-proquest-com.proxy.lib.umich.edu/lib/umichigan/detail.action?docID=5731927)**
- **R for Data Science by Garrett Grolemund and Hadley Wickham (https://r4ds.had.co.nz/)**
- **Introduction to Data Science by Rafael A. Irizarry (https://rafalab.github.io/dsbook/)**
- **Foundations of Linear and Generalized Linear Models by Alan Agresti (https://search.lib.umich.edu/catalog/record/016787603)**
- **Handbook of Computational Statistics (various authors) (https://search.lib.umich.edu/catalog/record/016787603)**

If you would like additional material on the background statistics required for this course, I recommend **Wasserman's All of Statistics. (https://search.lib.umich.edu/catalog/record/016738825)** Chapters 1, 2, 3, 6, 7, 9, and 10 cover the core ideas we use in this course.

Course Structure and Grading

The course is built around 200 points:

- Eleven homework assignments, 10 points, lowest score dropped.
- Final research and data analysis project with
  - Research Statement/First Draft, 10 points
  - Final Draft, 90 points

Late homework will not be accepted. Please email the professor (**mfredric@umich.edu (mailto:mfredric@umich.edu)**) if you have extenuating circumstances requiring an extension.

Extra Credit Seminars

You may complete up to three extra credit assignments, each worth 2 points (1 percentage point on your final grade), by viewing a live or prerecorded seminar related to computational statistics. To complete the assignment, write a brief paragraph summarizing the seminar and submit one of the **extra credit seminar assignments (https://umich.instructure.com/courses/621159/assignments)** . Be sure to include:

- The speaker's name
- The title of the talk

- The time and date
- Two or three key points, particularly connections to STATS 406 material
- URL if available

For a list of approved seminars, **see the seminars page (https://umich.instructure.com/courses/621159/pages/approved-extra-credit-seminars)** . If you would like to use a different seminar, email the instructor (**mfredric@umich.edu (mailto:mfredric@umich.edu)**).

Final Project

A primary goal of this course is to prepare students to perform their own investigation using computational techniques. Working in groups of three, students will select a research topic, find a relevant data set, and use class techniques to describe, visualize, and analyze the data. Three starter projects will be available that include a published paper and a related data set. In either case, critical to this project is asking and answering a clear question.

You are encouraged to discuss your ideas project ideas with the professor as early as possible. If you are having difficulty thinking of a project, re-analyzing an existing study is often a good place to start.

The final project will be graded in two parts. The first draft should present the research question and data, and also at least sketch out methods that will be used for the analysis. The final draft will complete the analysis and discuss the results in the larger context of the problem.

Students may select their own groups or be randomly assigned into groups. Groups will be finalized no later than Oct 1.

See the full **Final Project (https://umich.instructure.com/courses/621159/pages/final-project)** description for additional details.

Course Policies

Working with other students on homework assignments is encouraged! All students should contribute to all work. Student groups should not simply copy and paste solutions from one student's assignment to another. We expect solutions to be similar for students working together, but students should make sure that they understand and can competently explain any solution if asked.

You can find additional course policies here: **Course Policies (https://umich.instructure.com/courses/621159/pages/course-policies?wrap=1)**