DSCC / CSC / TCS 462: Computational Introduction to Statistics
Fall 2022
4.0 Credits
TR 4:50 - 6:05 pm
Goergen Hall Room 109

Instructor: Anson Kahng, Ph.D.
Department of Computer Science and Goergen Institute for Data Science
Wegmans 2401
University of Rochester
anson.kahng@rochester.edu
858-353-0550

Office Hours:

Anson Kahng
Tuesdays 9-10 am, Wednesdays 9-10 am, Thursdays 9-10 am
By appointment only
Wegmans 2401 or Zoom (https://rochester.zoom.us/j/93803440042)

Gaurav Chattree
Tuesdays 12-1 pm
Wegmans 1219

Mohamad Ali Kalassina
Wednesdays 2-3 pm
Wegmans 1219

Prerequisites: Discrete math (e.g., MATH 150/150A) and calculus (e.g., MATH 142/161/171)

Course Description
This course will cover foundational concepts in descriptive analysis, probability, and statistical inference. Topics to be covered include data exploration through descriptive statistics (with a heavy emphasis on using R for such analyses), elementary probability, diagnostic testing, combinatorics, random variables, elementary distribution theory, statistical inference, and statistical modeling. The inference portion of the course will focus on building and applying hypothesis tests and confidence intervals for population means, proportions, variances, and correlations. Non-parametric alternatives will also be introduced. The modeling portion of the course will include ANOVA, and simple and multiple regression and their respective computational methods. Students will be introduced to the R statistical computing environment.

This course follows the College credit hour policy for four-credit courses. A four-credit hour course has four hours of instruction (including asynchronous recordings or any other activity in which the instructor is directly engaged in teaching students) and eight hours of out-of- class activity per week, or 12 hours total per week, for the entirety of the semester. Out- of-class activities may include course reading, papers, problem sets, or other associated "homework" for the course.

Course Aims and Objectives
This course aims to provide students with basic statistical competency, focusing on theory, application, and computational methods for analyzing data. By the end of this course, students will be able to:

- Appropriately summarize and display data
- Describe relationships between multiple variables
- Apply probability axioms to random variables
- Determine and apply the appropriate statistical test for a hypothesis

- Build and interpret regression models
- Analyze data using R, both through built-in and custom functions
- Compile results into cohesive reports using RMarkdown.

Course Policies and Expectations
Please keep an open mind and be as flexible with me as possible, and I will be as flexible as I can. While we have become a bit more comfortable with it, we are still in the midst of a pandemic, and there will undoubtedly be ups and downs throughout the semester. I expect that you put in your best effort, stay excited, and communicate with me as much as possible about your situation, your perceptions of the course, and anything else that could help make your experience more worthwhile. I ask that you be as respectful as possible to the learning process.

With any class, there is a risk for academic dishonesty. I am going to do everything possible to prevent academic dishonesty, but I also put my trust in you to do the right thing and follow the rules.

For this course to run well, students need proper technology (discussed below). If there is any issue with getting access, I ask that you get in contact with me as soon as possible. Overall, please stay in contact with me so I know what is going on, and I will keep you updated from my end.

It is expected that you attend lectures in person, attend office hours if you have any questions, and stay up to date with course material. Please be as actively involved in the course as possible. If for any reason you are uncomfortable with attending events in person, please let me know and we can reach a compromise together.

Materials and Access

There is no required text for this course. Course notes and assignments will be placed on Blackboard. Two reference textbooks that are recommended:

1) *Probability and Statistics for Computer Science.* D. Forsyth. ISBN: 9783319644097
2) *An Introduction to R.* William N. Venables, David M. Smith. Available at https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf

R and RStudio will be required for this course. Both are available online for free. In order to download R, visit https://www.r-project.org/ and go to the CRAN link to choose a mirror (any USA mirror will likely do fine) to download the latest stable version. In order to download RStudio, visit https://www.rstudio.com/products/rstudio/download/ to download the free version. We will be using RMarkdown for homework assignments. You also will need to use TeX for some homework assignments; I highly recommend using Overleaf.com, but if you would like your own local version, you can download LaTeX here: https://www.latex-project.org/get/.

A laptop with a webcam is also recommended to watch any lectures that you may miss or to attend virtual office hours.

A calculator is recommended, but not required if you instead use R/RStudio for computations. You may use any type of calculator (basic, scientific, or graphing) as long as it has basic functions, including the square root function and the mathematical constant, e.

Assignments and Grading Procedures
Grading and assessment for the course will be broken into the following pieces:

1. Homework Assignments: 35%
   There will be seven homework assignments, each worth 5% of your total grade. All due dates for assignments will be clearly written on each assignment. Assignments will be posted to Blackboard, and students are expected to submit their assignments through Blackboard as well. Generally, homework assignments will involve the use of R and RMarkdown to analyze datasets from a computational perspective.

   Late work will not be accepted unless arrangements are made well in advance or in extenuating circumstances with valid documentation.

   Statistics is a collaborative field, and you are encouraged to discuss with other students, but you are responsible for your own assignment. The work you submit must be your own, and each student must complete and submit their own assignment. If you discuss the assignment with other students, please indicate who on your submitted assignment, and make sure that the completed assignment you turn in is in fact your work and unique to you. Please see the academic honesty section of the syllabus for more information.

2. Participation: 15%
   It is expected that you will attend classes in person. If you cannot attend, please email the professor ahead of time. Please be engaged and responsive (i.e., ask questions, provide feedback) in both class and office hours.

3. Midterm: 30%
   There will be one midterm exam worth 30% of your total grade. It will take place during the normal class time and is noted on the course schedule. More details will be discussed on Blackboard and in class closer to the time of the midterm.

4. Final Project: 20%
   There will be a cumulative final project that will involve an applied data analysis, written as an RMarkdown report. Components of this project include a project proposal, a final video presentation, and a final written report. Exemplary projects will be selected to give a short in-class presentation. Exact details pertaining to the project, datasets, grading, etc. will be posted on Blackboard.

Letter grades will be assigned as follows after rounding your final grade to the nearest integer:
A: 94-100, A-: 90-93, B+: 87-89, B: 83-86, B-: 80-82, C+: 77-79, C: 73-76, C-: 70-72, D: 65-69, E: 0-64

Teaching Assistants
We are lucky to have two great teaching assistants for this course: **Gaurav Chattree** and **Mohamad Ali Kalassina**.

The TAs will hold weekly office hours, grade assignments, and be a resource for you in the course. If you email a TA, you will get a response within 24 hours (keep this in mind if you are waiting until the last minute to start your assignments). In addition to the TAs, you can always feel free to email the professor directly.

Academic Integrity

Academic integrity is a core value of the University of Rochester. Students who violate the University of Rochester University Policy on Academic Honesty are subject to disciplinary penalties, including the possibility of failure in the course and/or dismissal from the University. Since academic dishonesty harms the individual, other students, and the integrity of the University, policies on academic dishonesty are strictly enforced.

Any usage whatsoever of online solution sets or paid online resources (chegg.com or similar) is considered an academic honesty violation and will be reported to the Board on Academic Honesty. In particular, any assignment found to contain content which originated from such sources is subject to a minimum penalty of zero on the assignment and a full letter grade reduction at the end of the semester (e.g., a B would be reduced to a C). This applies even if the unauthorized content was obtained through indirect means (through a friend, for instance) and/or the student is seemingly unaware that the content originated from such sources. If you have any questions about whether resources are acceptable, please check with your instructor.

Proper standards for collaboration in this course are as follows:

- Homework assignments are allowed to be verbally discussed between students. Students are **not** permitted to share code, output, or written summaries with each other. Only verbal discussions of assignments are permitted between students in the course. Students may discuss the assignment with the teaching assistants or professor through any means of communication. My general rule is that the work you submit must be your own, and that if I were to ask you to reproduce such work, you would be able to do. Each student must submit their own assignment that is a reflection of the work they themselves completed.
- Collaboration in small groups is encouraged; use of resources outside of the course is not permitted.
- Guidelines pertaining to collaboration on the final project will be discussed in the final project expectations, to be posted on Blackboard soon.
- Students may generally discuss course content with each other through any means of communication. You are highly encouraged to form study groups and talk through the material together. Teaching each other is often the best way to learn!

If you have any questions, please let me know. In general, all assignments and activities associated with this course must be performed in accordance with the University of Rochester's Academic Honesty Policy, available at http://www.rochester.edu/college/honesty/policy/index.html.

Accommodations for Students with Disabilities
Students needing academic adjustments or accommodations because of a documented disability must contact the Disability Resource Coordinator for the school in which they are enrolled. I am happy to accommodate any and all accommodations, so long as they are documented with the Office of Disability Resources. I am glad to meet to discuss your specific situation or to help ensure you have the support you need. For additional information, please see: https://www.rochester.edu/college/disability/.

Course Schedule
A calendar of course topics, due dates, and other pertinent information is available on Blackboard. The calendar is subject to change, but please always refer to Blackboard for the most updated copy.