STT802/EPI 853B: Statistical Computing

Instructor: Gustavo de los Campos (gustavoc@msu.edu)

Time/Place: M/W 3:00-4:20PM (Online synchronous). Short videos will be made available after the class. If you are not in the US and cannot attend the class via zoom in a synchronous mode, please contact me to discuss possible accommodations.

Prerequisites: Mathematical statistics (STT 442 or equivalent) and Linear Algebra (MTH 309 or equivalent).

Grading: In-class assignments (25%), HW (25%), Midterm (25%), Final (25%)

Score (5)	<50	50-59	60-69	70-74	75-79	80-84	85-89	$\geq$90
Grade	0	1	1.5	2	2.5	3	3.5	4

Course Description: In this course, we will cover computational methods commonly used in statistics implemented using the R-language. Course materials will be posted in the Github repository:

https://github.com/gdlc/stat_comp

The course will be offered as an online synchronous. Students will receive a zoom link for the class. Each class will roughly consist of three modules: (1) A ~15 discussions of solutions to the previous in-class assignment, (2) Introduction of a topic followed by a description of the current in-class work (~25 min), and (3) Breakout sessions (~40 min) in which students will collaborate in a small group to solve the in-class assignments. In-class assignments will be due in D2L before the beginning of the next class.

Contents: If time permits, we aim to cover the following modules. A more detailed list of topics, with links to available materials can be found at our Github repository.

Module 1: Introduction to R
- Types
- Basic operations
- Reading/writing data
- Arrays
- Loops and conditional statements
- Descriptive statistics and plots
- Reporting using R-Markdown

Module 2: Linear Algebra
-   Matrix operations in R
-   Matrix factorizations (singular value and QR-decomposition)

Module 3: Least-Squares problems
-   Least Squares (using lm and using matrix operations)
-   Splines

Module 3: Maximum Likelihood
-   Overview of ML estimation and inference
-   ML using general-purpose optimization algorithms (application to logistic regression)
-   The EM-Algorithm (application to mixture models)

Module 4: Sampling Random Variables
-   Univariate: the d,p,q,r functions
-   Multivariate Normal Distribution
-   Gibbs Sampler
-   Metropolis and Metropolis-Hastings

Module 5: Power Analysis
-   Using built-in packages
-   Using Monte Carlos simulations

Module 6: Resampling methods
-   Bootstrap (application: SEs for odds ratios in logistic regression)
-   Permutation tests (p-values for odds ratios in logistic regression)
-   Cross-validation

Module 7: Large scale hypothesis testing (controlling Type-I error and FDR)
-   Controlling type-I error rate
-   False Discovery Rate

Module 8: Variable screening and variable selection
-   Independent screening
-   Forward regression
-   Lasso and Elastic Net

Evaluation: The evaluation will be based on in-class assignments, homework, a midterm, and Final exam

Textbook: We do not have a required textbook. The following are very good textbooks that will guide you to learning about statistical analyses in R.

- Advanced R 2nd Edition by Hadley Wickham [website].

- Introduction to Data Science, by Rafael Irizarry [website].

- The elements of Statistical Learning [website], and Computer Age Statistical Inference [pdf] are not R books, but they excellent books that describe the underlying theory of the methods we will implement.

Academic Honesty: The Departments of Epidemiology & Biostatistics and Statistics & Probability adheres to the policies of academic honesty as specified in the General Student Regulations 1.0, Protection of Scholarships and Grades, and in the All-University of Integrity of scholarship and Grades which are included in Spartan Life: Student Handbook and Resource Guide. Students who plagiarize will receive a grade 0.0 on the homework, exam, or quiz.

ADA: To arrange for accommodation a student should contact the Resource Center for People with Disabilities at http://www.rcpd.msu.edu/ or (517)353-9642

Disclaimer: Changes on the syllabus/important dates will be announced in class and on the course website. It is the students' responsibility to keep up with any changed policies and assignments.