

STT 461: Computations in Probability and Statistics

Spring 2018

MWF, 11:30 pm- 12:20 pm,
314 Ernst Bessey Hall

COURSE FACULTY : Hyokyoung G. Hong, PhD
E-mail: hhong@stt.msu.edu

OFFICE HOURS: Mon and Wed: 2:00 pm- 3:00 pm Wells Hall C435

TA OFFICE HOURS: Jialin Qu (qujialin@stt.msu.edu) Fri: 9:00 am-11:00 am Wells Hall C522

COURSE DESCRIPTION: Computer algorithms for evaluation, simulation and visualization. Sampling and prescribed distributions. Robustness and error analysis of procedures used by statistical packages. Graphics for data display, computation of probabilities and percentiles.

LEARNING OBJECTIVES: Computational statistics is rapidly penetrating all areas of statistical applications and is intimately associated with the development of most advances in statistics today. As you read through the topics below you can begin to see the outlines of an emerging approach to statistics that substitutes general purpose computer intensive methods (Monte Carlo) for many of the specialized formulas and tables of other approaches.

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

1. Use the statistical software package R to perform routine statistical analysis.
2. Use R to implement newly developed statistical methods.
3. Use simulation studies to evaluate statistical methods.

COURSE PREREQUISITE: CSE 131 or CSE 230; MTH 314, STT 441.

COURSE WEBSITE: A webpage will be created for this class using www.github.com/youngghhk/STT461. This course website will contain the syllabus, class notes, sample R programs and outputs, datasets (used in lecture and for the homework assignment), and reading materials. Class announcements will also be posted on this site, so it is a good idea to check the site regularly to stay current.

COURSE TEXTBOOK: There is no required textbook. The following are very good textbooks that will guide you learning about statistical analyses in R.

- Introductory Statistics with R (Statistics and Computing) 2nd Edition by Peter Daalgaard.
- Modern Applied Statistics with S. by W.N. Venables and B.D. Ripley.
- An Introduction to Statistical Learning with Applications in R, by J. Gareth, D. Witten, T. Hastie and R. Tibshirani.
- Statistical computing with R. Maria L. Rizzo, Chapman & Hall/CRC, 2007.

HOMEWORK Students are required to submit their homework as both PDF file and hardcopy by the due date of the assignment. Please turn in code separately and electronically. All electronic submissions should be made via Desire 2 Learn (D2L) dropbox and should follow the following naming convention: last name, first name, assignment number, proper extension. So, for example, if John Smith is turning in Homework 1, he would name the file Smith_John1.pdf. The associated code would be Smith_John1.Rmd. If you wish to break up your code into separate files, you may submit them as Smith_John1a.Rmd, Smith_John1b.Rmd, and so on. There will be a 20% penalty per day that your homework is late. Homework in the wrong format will not be given credit.

GRADING Final grades will be based on homeworks (30%) and two exams (30%, 40%).

Any R code must be fully annotated and shown to work to earn points. Tentatively: Exam 1 Part1: W 2-21-18 (closed book), Exam 1 Part 2: F 2-23-18 (open book) Exam 2 Part 1 W 4-25-18 (closed book), Exam 2 Part 2 F 4-27-18 (open book)

The grades will be based on a straight scale roughly according to the following scheme

90-100 %	85- 89.9 %	78- 84.9 %	70- 77.9 %	65- 69.9 %	55- 64.9 %	50- 54.9 %	0 – 49.9 %
4.0	3.5	3.0	2.5	2.0	1.5	1.0	0

R:

The class will be taught in the R language.

R is a free, open-source programming language for statistical computing. Almost all of our work in this class will be done using R. You will need regular, reliable access to a computer running an up-to-date version of R. If this is a problem, let the professors know right away.

RStudio is a free, open-source R programming environment. It contains a built-in code editor, many features to make working with R easier, and works the same way across different operating systems. Use of RStudio is *required* for the labs, and strongly recommended in general.

R Markdown is a variant of Markdown that has embedded R code chunks, to be used with knitr to make it easy to create reproducible web-based reports. The Markdown syntax has some enhancements (see the R Markdown page); for example, you can include LaTeX equations (see Equations in R Markdown).

Approach: Although the focus of the course is on computational methods, for each topic we will first describe the problem from a statistical perspective. If they exist, exact analytical solutions will be discussed and implemented. Otherwise numerical methods will be presented. Derivations will be presented in class and students are expected to take their own notes. Scripts for computations will be developed in class and a summary will be posted in this repository.

Communications: You are responsible for reading emails sent to your MSU account from your professor and the announcements that are placed on the course web site. Information about readings, news events, your grades, assignments and other course related topics will be communicated to you with these electronic methods.

Academic Honesty: The Department of Statistics and 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 web site. It is students' responsibility to keep up with any changed policies and assignments.