Introduction to Statistical Computing (Statistics 3701)

Zack W. Almquist
Spring Semester, 2018

Class Schedule

Lecture:	M/W	01:25PM - 2:15PM	Mechanical Engineering 18
Lab	F	12:20PM - 1:10PM	Ford Hall 110
Lab	F	01:25PM - 2:15PM	Bruininks Hall 530A

URL: `https://canvas.umn.edu`
Note: Requires UMN login and registration in class to access.
URL: `http://www.github.com`
Note: Requires github login.

Professor

Name:	Zack W. Almquist
Office:	1168 Social Science Building
Office Hours:	M 11:00AM - 1:00PM
Email:	almquist@umn.edu
Telephone:	612-624-4300 (not recommended)

Teaching Assistant

Name:	Mingxuan Han
Office:	495 Ford Hall
Office Hours:	TU 1:00-3)) PM
Email:	hanxx862@umn.edu

Course Objectives

This course is an introduction to statistical computing and will be taught in R. It is assumed students have basic exposure to computer programming, statistics up through linear models and basic knowledge of calculus (i.e., differential and integral calculus). The main focus of the class will be on statistical programming, graphics, elementary Monte

Carlo, simulation studies, and elementary optimization in R. Students will cover classic methods of the bootstrap through modern Bayesian inference. At the end of this course students should have a broad grasp of R and its capabilities, and a basic understanding of key computational methods in statistics such as optimization, numerical analysis, MCMC techniques, and simulation methods.

Prerequisites

It is assumed that the student has taken Math 1272 or 1372 or 1572 (i.e., differential and integral calculus), CSCI 1113 (introduction C/C++ Programming for Scientists and Engineers) and Stat 3032 (Regression and Correlated Data).

Course Requirements

Computers

It is not required that students bring their computers/laptops to lecture and lab (if one is owned), but it is *highly* recommended since both lecture and lab will make extensive use of the computer software `R`. Computer labs are available on campus, please consult with the TA if you have trouble finding the various locations of campus computer labs.

Readings

Weekly readings assignments can be found on the course syllabus. All readings are assumed to be completed before each lecture/seminar. You are expected to read over the class notes each week and make sure you are familiar with the material as the course progresses. Questions are encouraged.

Homework

Homework assignments will normally be administered on a bi-weekly basis and will be due on every other Wednesday. Homework assignments are meant to achieve three results: (1) provide practice with the statistical concepts discussed in class, (2) provide practice with the computational and statistical programing language `R`, and (3) provide a chance to demonstrate your mastery of material and highlight areas where more work is needed. You may work in a group, but all write-ups must be done independently. All collaborators should be appropriately cited in your write up and any detailed R code should also be provided.

Quizzes

To assess mastery of course material, three short quizzes will be administered over the course of the semester. The quizzes will cover all material presented in lecture, assigned readings and labs. Students are advised to keep up-to-date on reading assignments, and to attend lectures regularly, so as to be prepared for quizzes. Students must be present for the quiz to obtain credit; quizzes are not rescheduled due to travel or other considerations. Quizzes are due on canvas at the end of the class in R markdown format (.Rmd) and must compile – no credit will be given to documents that don't compile. All work must be your own.

Lab

Individuals are expected to attend every course, to have completed every reading, and to participate with questions and discussion on each topic as presented. If you plan on missing any class period you are responsible for all material and for contacting the instructor in a timely manner.

Grading

Labs:	20%
Homework:	40%
Quizzes:	40%

Lectures, readings, labs, and review sessions are provided for each student's benefit. It is the responsibility of the student to take advantage of these opportunities to acquire and demonstrate mastery of course material, so as to achieve his or her desired grade.

Letter grade assignment

A	93%+
A-	90-92.99%
B+	87-89.99%
B	83-86.99%
B-	80-82.99%
C+	77-79.99%
C	73-76.99%
C-	70-72.99%
D	60-69.99%
F	<59.99%

Required Texts

- Wickham, Hadley (2015). *R packages*. O' Reilly press.
  - Available for free at `http://r-pkgs.had.co.nz/`.
- Grolemund, Garrett and Hadley Wickham. (2017). R for Data Science. O'Reilly Press.
  - Available for free at `http://r4ds.had.co.nz/`.
- Geyer, Charles (2017). "Stat 3701 Lecture Notes."
  - Modified by Zack Almquist and made available through canvas.
- Introduction to R by the R Core Team.
  - Available for free at `http://cran.us.r-project.org/manuals.html`.

Readings

Be prepared to discuss all readings assigned at anytime in lecture/lab.

Required Software

We will be using the `R` statistical programming language. `R` can be downloaded at `http://www.r-project.org/`.

RStudio IDE (Integrated Development Environment) is a software application which facilitates interaction with the `R` statistical programming language. It is often preferred to the GUI (Graphic User Interface) made available through CRAN. You can download it at `http://www.rstudio.com/`.

(*Not required but recommended*) Latex is a word processor and a document markup language. It can be downloaded and installed on Windows (`http://miktex.org/`), OSX (`https://tug.org/mactex/`) or Linux (use the package manager of your choice).

(*Not required but recommended*) A github account is recommended. One can register for a github account at `https://github.com/`. You can find information about how github works with Rstudio at `http://z.umn.edu/rstudiogit`, and github maintains a quite good help-system at `https://help.github.com/`.

R Markdown

All assignments, labs and quizzes will be written in R markdown format. A basic homework, quiz and lab template will be released at appropriate intervals.

Course Policies

Missing Class, etc.

It is expected that each member of the class will attend every lecture/discussion. If there is an appropriate reason to miss class it is expected that the individual will email or discuss in person with the instructor at least one week in advance. For any medical issues please see the UMN website for university policies.

Cheating, etc.

All work is assumed to be your own and all individuals are expected to follow the university policy on cheating and misconduct. If you have any questions please consult the UMN website for university policies.

Assignment Details

Homework Assignments

Homework will be assigned on a biweekly basis starting on the second Wednesday of the Semester and will be due two weeks later at the end of class. Homework assignments will be graded on a 100 point basis. Each assignment must be turned in on canvas in R Markdown format (.Rmd). No credit will be given for files that do not compile. Your lowest score will be dropped.

Lab Assignments

Every Friday in class you will have a lab. Labs must be completed in class and submitted by the end of class in canvas as R Markdown file (.Rmd). No credit will be given if they do not compile. Labs will be graded based on effort. Your lowest lab score will be dropped.

Quizzes

Every other Wednesday starting on the second week of the semester you will be given a quiz at the beginning of class. Quizzes must be handed in via canvas by the end of class in a R Markdown file (.Rmd). Your file must compile or zero grade will be assigned. You can make up half the points missed by correcting them in your homework. Your lowest grade will be dropped.

Assignment and Reading Dates

Week 1 (01/15-01/17/18): Basics of R

- *Lab (01/19/18) :*
  - Due at the end of class.
- *Readings:*
  - Introduction to R by the R Core Team.
- *Homeworks:*
  - No homework.
- *Quizzes:*
  - No Quiz.

Week 2 (01/22-01/24/18): R Functions and Packages

- *Lab (01/26/18) :*
  - Due at the end of class.
- *Readings:*
  - Introduction to R by the R Core Team.
- *Homeworks:*
  - First homework handed out.
- *Quizzes:*
  - Quiz 1.

Week 3 (01/29-01/31/18): Computer Arithmetic

- *Lab (02/02/18) :*
  - Due at the end of class.
- *Readings:*
  - Geyer Notes: Computer Arithmetic.
- *Homeworks:*
  - First homework due.
- *Quizzes:*
  - No Quiz.

Week 4 (02/05-02/07/18): Data

- *Lab (02/09/18) :*
  - Due at the end of class.
- *Readings:*
  - Geyer Notes: Data.
- *Homeworks:*
  - Second homework handed out.
- *Quizzes:*
  - Quiz 2.

Week 5 (02/12-02/14/18): Data and the Tidyverse

- *Lab (02/16/18) :*
  - Due at the end of class.
- *Readings:*
  - R for Data Science.
- *Homeworks:*
  - Second homework due.
- *Quizzes:*
  - No Quiz.

Week 6 (02/19-02/21/18): R Packages

- *Lab (02/23/18) :*
  - Due at the end of class.
- *Readings:*
  - R packages.
- *Homeworks:*
  - Third homework handed out.
- *Quizzes:*
  - Quiz 3.

Week 7 (02/26-02/28/18): Statistical Models

- *Lab (03/02/18) :*
  - Due at the end of class.
- *Readings:*
  - Geyer Notes: Statistical Models
- *Homeworks:*
  - Third homework due.
- *Quizzes:*
  - No Quiz.

Week 8 (03/05-03/07/18): Optimization and Solving Equations (Numerical Analysis)

- *Lab (03/09/18) :*
  - Due at the end of class.
- *Readings:*
  - Geyer Notes: Optimization and Solving Equations
- *Homeworks:*
  - Fourth homework handed out.
- *Quizzes:*
  - Quiz 4.

[*Spring Break*]

Week 9 (03/12-03/14/18): Spring Break

- NO CLASS

Week 10 (03/19-03/21/18): Statistical Models, Part II

- *Lab (03/23/18) :*
  - Due at the end of class.
- *Readings:*
  - Geyer Notes: Statistical Models, Part II
- *Homeworks:*
  - Fourth homework due.
- *Quizzes:*
  - No Quiz.

Week 11 (03/26-03/28/18): Statistical Models, Part II

- *Lab (03/30/18) :*
  - Due at the end of class.
- *Readings:*
  - Geyer Notes: Statistical Models, Part II
- *Homeworks:*
  - Fifth homework handed out.
- *Quizzes:*
  - Quiz 5.

Week 12 (04/02-04/04/18): Simulation

- *Lab (04/06/18) :*
  - Due at the end of class.
- *Readings:*
  - Geyer Notes: Simulation
- *Homeworks:*
  - Fifth homework due.
- *Quizzes:*
  - No Quiz.

Week 13 (04/09-04/11/18): Bootstrap

- *Lab (04/13/18) :*
  - Due at the end of class.
- *Readings:*
  - Geyer Notes: Bootstrap
- *Homeworks:*
  - Sixth homework handed out.
- *Quizzes:*
  - Quiz 6.

Week 14 (04/16-04/18/18): Parallel Computation

- *Lab (04/20/18) :*
  - Due at the end of class.
- *Readings:*
  - Geyer Notes: Parallel Computation
- *Homeworks:*
  - Sixth homework due.
- *Quizzes:*
  - No Quiz.

Week 15 (04/23-04/25/18): Bayesian Inference via Markov Chain Monte Carlo (MCMC)

- *Lab (04/27/18) :*
  - Due at the end of class.
- *Readings:*
  - Geyer Notes: Bayesian Inference via Markov Chain Monte Carlo (MCMC)
- *Homeworks:*
  - Seventh homework handed out.
- *Quizzes:*
  - Quiz 7.

Week 16 (04/30-05/02/18): Bayesian Inference via Markov Chain Monte Carlo (MCMC)

- *Lab (05/04/18) :*
  - Due at the end of class.
- *Readings:*
  - Geyer Notes: Bayesian Inference via Markov Chain Monte Carlo (MCMC)
- *Homeworks:*
  - Seventh homework due.
- *Quizzes:*
  - No Quiz.

Week 17 (05/07/18-05/09/18): Finals

- NO FINAL