STAT 327-001 Statistical Computing - Winter 2019

Instructor: Keshav Pokhrel, Ph.D.
Office: 2087 CB
Phone: (313) 593-5165
Email: kpokhrel@umich.edu
Class Time: 9:30 AM- 10:45 AM TR
Meeting Location: 2048 CB

URL: `www-personal.umd.umich.edu/~kpokhrel`

Office Hours:

Monday: 10:00 - 11:00AM

Tuesday: 11:00 - 12:00PM

Thursday: 11:00AM - 12:00PM

Required Textbooks:

- Norman Matloff, The Art of R Programming: A Tour of Statistical Software Design
- Phil Spector, Data Manipulation with R
- Paul Teetor, The R Cookbook

Recommended Textbooks:

- John M. Chambers, Software for Data Analysis: Programming with R
- Winston Chang, The R Graphics Cookbook
- Maria L. Rizzo, Statistical Computing with R

Course Description

Importance of computing is becoming increasingly essential for almost every academic discipline. A competent statistician must have strong computing knowledge together with statistics. On the other hand, a competent data scientist must have sound statistics knowledge together with computing skills. This class is introductory statistical programming targeted for statistics majors. The course does not assume any computational background from the students, however it is always helpful to have some background in computing.

Students will learn fundamental ideas of statistical programming- data types, data structures, importing and exporting data; debugging; and logical design- through writing code, visualizing and analyzing the given data. Students are also expected to learn how to design reproducible statistical analysis and write maintainable codes. In addition, they will learn about fundamentals of simulation and filter large data sets and transfer large data into workable format.

The class will be taught in the R(`https://www.r-project.org/`) language.

Course Objectives:

- Understand importing, exporting, cleaning and filtering the data.

- Learn to write function and test code for correctness.
- Employ optimization algorithms.
- Develop reproducible documents.
- Learn how to comment and organize code.
- Understand the writing and use of functions.
- Code and run an Markon Chain Monte Carlo (MCMC) algorithm.
- Learn different data visualization techniques.

Homework: There will be homework almost every week and one in-class lab every week. Unless otherwise mentioned in the calender, homework is due on Tuesday at 11:59PM. You need to submit you r-code and explanation of output together. The recommended way to do this is to use Rmarkdown.

Labs: You will be randomly paired with a colleague for in class lab work. Labs are graded using three categories: five points for making a good-faith effort at every part of the assignment; five point for technically-correct, working solutions to each part; and five points for clean, well-formatted, easily readable code.

Datacamp: You are required to sign in to Data Camp(`https://www.datacamp.com/home`) to practice coding exercises. Data camp is a online learning platform at you own pace. Our course has premium access to their homework system. I strongly encourage you to finish some basic courses in R to prepare yourself for this class.

Mid Term Exams

There will be two mid term exam. There will be a coding component involved in the exams and you will be given one and half weeks to finalize your answer.

Final Project:

You are expected to complete a group project (two or three students in a group). The projects are considered to be a major vehicle to carry and reflect all the concepts you learn in the class. I will post separate project guidelines on canvas. Part of your final project grade depends on peer assessment.

Software:

We will use R computing environment. R is a programming language for statistical computing and visualizing the data. It can be downloaded for free from `http://www.r-project.org`. We will use R Studio for regular classroom activities. R studio (`https://www.rstudio.com/`) is an open source Integrated development Environment(IDE) for R.

Grading Policy:

Your performance is measured by the weighted average of homework, exams, labs, data camp coding assignments, and final projects.

Midterm I	20%
Midterm II	20 %
Homework	15%
Lab	15%
Datacamp	5%
Final Exam	25%

Grades on this course are based on your performance: i.e. you must be able to do it not simply better than the rest. There are no make-up exams and no curving. Your letter grade will be based on the following percentage distribution.

Letter Grade	≥ 97	≥ 93	≥ 90	≥ 87	≥ 83	≥ 80	≥ 77	≥ 73	≥ 70	≥ 67	≥ 63	≥ 60	< 60
Percentage	A+	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	E

NOTE: If you have any grade disputes- you need to discuss them with me within a week of the grade being entered. Grades cannot be changed at the end of the semester.

Disability Statement: The University will make reasonable accomodations for persons with documented disabilities. Student need to register with Disability Resources Services (DSR) every semester thay are enrolled for classes. DRS is located in counseling and Support Services, 2157 UC. To be assured of having services when they are needed, students should register no later than the end of add/ drop deadline of each term. Visit the DSR website at: `http://umdearborn.edu/cs_disability`. If you have a disability that necessitates an accomodation or adjustment to the academic requirements stated in this syllabus, you must register with DRS as directed above and notify me. Upon receipt of your notification, we will make accomodation as directed by DRS.

Academic Honesty: The University of Michigan-Dearborn values academic honesty and integrity. Each student has a responsibility to understand, accept, and comply with the University's standards of academic conduct as set forth by the Code of Academic Conduct at: `http://umdearborn.edu/policies_st-rights`, as well as policies established by each college. Cheating , collusion, misconduct, fabrication, and plagiarism are considered serious offenses, and may be monitored using tools including but not limited to TurnItIn. Violations can result in penalties up to and including expulsion from the University. At the instructor,s direction, the penalty may be a grade zero on the assignment up to and including recommending that student be expelled from the University. It is the sole responsibility of the student to understand and follow academic guidelines regarding plagiarism. The University of Michigan-Dearborn has an online academic integrity tutorial that can be accessed at: `http://umdearborn.edu/umemergencyalert`.

Safety: All students are encouraged to program 911 and UM-Dearborn's University Police phone number (313) 593-5333 into personal cell phones. In case of emergency, first dial 911 and then if the situation allows call University Police. The Emergency Alert Notification (EAN) system is the official process for notifying the campus community for emergency events. All students are strongly encouraged to register in the campus EAN, for communications during an emergency. The following link includes information on registering as well as safety and emergency procedures information: If you hear a fire alarm, class will be immediately suspended, and you must evacuate the building by

using the nearest exit. Please proceed outdoors to the assembly area and away from the building. Do not use elevators. It is highly recommended that you do not head to your vehicle or leave campus since it is necessary to account for all persons and to ensure that first responders can access the campus. If the class is notified of a shelter-in-place requirement for a tornado warning or severe weather warning, your instructor will suspend class and shelter the class in the lowest level of this building away from windows and doors. If notified of an active threat (shooter) you will Run (get out), Hide (find a safe place to stay) or Fight (with anything available). Your response will be dictated by the specific circumstances of the encounter.

Important Dates:

Exam 1 ..... Thursday, February 14th
Exam 2 ..... Tuesday, April 2nd
Final Exam ..... Thursday, April 25(8:00 AM-11:00AM)

Tentative Academic Calender:

Week 1, 01/07-01/11: Introduction to Statistical Computing

- First day of class
- Go over syllabus
- R basic and data types
- R markdown

Week 2, 01/14- 01/18: R Basics and Data Types

- Function in R
- Working with Dplyr and tidyR
- Vector , Matrices, Arrays and data frames

Week 3, 1/21-1/25: Cleaning Data

- Cleaning data with dplyr and tidyR
- Working with dplyr and tidyR

Week 4, 01/28-02/01: Flow Control

- Writing and Debugging Functions in R
- Required Reading

  - Matloff Chapter 12-14

- Recommended reading
  - Jones Chapter 7-9
  Jones Chapter 20
  Rizzo Chapter 3

Week 5,02/04- 02/08, Writing Function

- Basic of Logic
- Loops and other control
- Required Reading

  - Matloff Chapter 7

  - Jones Chapter 5

Week 6, 02/11-02/15, Writing Function Contd.. , Exam 1

- Exam 1
- Loops and other control
- Required Reading

  - Matloff Chapter 7

  - Jones Chapter 5

- Recommended reading
  - Jones chapter 7-9
  Teetor chapter 13

Week 7, 02/18-02/22, Fitting and Using Statistical Models

- Random Number Generation
- Parametric Distributions
- Maximum Likelihood Estimates

  Reading for the week: R Cookbook, chapter 11

Week 8, 02/25-03/01, Transformation

- Sorting, Ordering Data Frames
- Merging Data Frames
- Reshaping Dataframes
  -Wide to long or long to wide

Week 9, 3/4-3/8, Spring Recess

Week 10 , 3/11-3/15, Optimization

- Maxima and Minima
- Gradient Decent and Newton's Method
- Scaling and big O notation
- - Curve Fitting by Optimization
  Recommended Reading: 13.1–13.2 in The R Cookbook

Week 11, 3/18-3/22, Databases

- Databases and R syntax.
- Structured query language(SQL).
- R/SQL translation.
- - Accessing Databases through R.
  Recommended Reading: 13.1–13.2 in The R Cookbook

Week 12, 3/25-3/29, Data Bases Contd., Exam II

- Review Databases
- Review Optimization
- - Exam II Recommended Reading: 13.1–13.2 in The R Cookbook

Week 13, 4/1-4/5, Simulation I

- Random variable generation
- Markov Chain Monte Carlo
  Reading: Matloff, chapter 8; R Cookbook

Week 14, 4/8-5/12, Simulation II

- Random variable generation
- Markov Chain Monte Carlo
  Reading: Matloff, chapter 8; R Cookbook

Week15, 4/14-4/19, Graphics

- ggplot2 in R
- Review- Simulation
- Final Project Discussion Reading: Matloff, chapter 8; R Cookbook
- Recommended Reading
  Jones Chapter 20