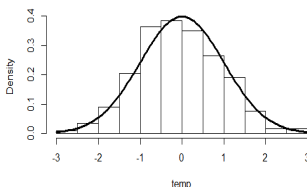




Statistics 128

Statistical Computing

M 6:00pm-8:50pm in Downtown 110



Instructor: Dr. Michelle Norris

e-mail: norris@csus.edu

Access course materials on Canvas at csus.instructure.com

Office: Shasta 258

Office Phone (916) 278-4300

Office Hours: Tu 12-1pm (Shasta 258), Wed 1:30-2:30pm (Brighton 118),
Wed 4:30-5:30 (Shasta 258) and by appointment.

This syllabus is tentative and subject to change; any necessary changes will be announced in class.

Course Description: Computer methods for accessing, transforming, summarizing, graphing, and making statistical inferences from data. Focus is on open-source, command-line software, but menu-driven statistical software may be introduced. Students will learn to apply computer methods to solve problems selected from the areas of modeling, simulation, inference and statistical learning. The intent of this course is to provide students with the software skills needed for statistical work in industry or academia. 3 units.

Prerequisites: Stat 1 or Stat 50, and Math 30 or Math 26A or consent of the instructor.

Optional Recommended Reference Texts: All are available online for free.

The Art of R Programming: A Tour of Statistical Software Design by Norman Matloff (available as an e-book from the campus library website)

R for Data Science: Import, Tidy, Transform, Visualize, and Model Data by Hadley Wickham and Garrett Grolemund (online version available free here: <http://r4ds.had.co.nz/>)

An Introduction to Statistical Learning with Applications in R by James, Witten, Hastie and Tibshirani (downloadable for free at www-bcf.usc.edu/~gareth/ISL/)

Supplies: For the downtown center location, it is **required** that you bring a laptop computer on which you can install R and RStudio to every class (if you cannot, I recommend taking the course when it is offered on the main campus in the computer lab). In fact, learning to install your own software is part of what this course is about. R is free and can be downloaded at www.r-project.org/. RStudio (Desktop version) is a nice user interface for writing, running and debugging R code and can be downloaded here www.rstudio.com/products/rstudio/download/. Be sure to download the versions of R and RStudio that work with your operating system (Windows 8, RedHat 7+, Mac, etc.). R and RStudio can run on Apple, Linux or PC platforms. The internet,

your classmates and your instructor are resources you should use to help you learn to install software.

Course Objectives:

At the completion of this course, students will be able to:

1. Understand the power and importance of computing in modern statistical practice.
2. Input raw data and import data in various common formats such as .csv, .xls, and .txt into a statistical software package.
3. Perform mathematical functions and statistical analyses in a statistical software package using built-in functions.
4. Export analysis output and graphics in an appropriate format from a statistical software package into a document.
5. Identify when built-in functions are not appropriate to the task at hand and write a custom function or program to perform the desired task.
6. Use conditionals, loops, vectorized operations and/or other programming structures appropriate to the statistical software package being used to write custom functions or programs.

Grade Distribution:

Midterm 1	15%
Midterm 2	15%
In-class quizzes and group work	15%
Homework (8-9 assignments)	10%
Project	25%
Final Exam (comprehensive)	20%

In-class quizzes (announced and unannounced) and group exercises will be given frequently (probably about 5-10 throughout the semester). To succeed in this class, you must attend class regularly and keep up with the material. **No makeups on quizzes/group exercises.** Your lowest quiz/group exercise will be dropped at the end of the semester.

The Final Exam will occur **Mon., May 13 from 5:15pm - 7:15pm.**

Letter Grade Distribution:

90 - 100	some form of A
80 - 89	some form of B
70 - 79	some form of C
60 - 69	some form of D
59 or less	F

Course Policies:

- Homework submission: All homework must be submitted to me electronically via Canvas. Hard copy is not required unless specifically asked for in the homework instructions. Code should be commented to clearly indicate its purpose and which homework problem or portion of the analysis it corresponds to. **You should fully understand any code you have written and be prepared to explain it to me if I should ask.**

- Cell phones: Please refrain from using cell phones in class. I find it distracting to see students using their cell phones while I am teaching.
- Absences: If you are absent, it is your responsibility to find out what you missed (announcements, hw assignments, notes, etc) from one of your classmates. Write the names and contact information of a few classmates below:
- No makeup exams except in cases of documented medical emergency or death in immediate family. You must notify me before the exam (e-mail or phone OK) and provide supporting documentation in person during my office hours. Makeups will be under the conditions set by the instructor and may be an oral or written exam.
- Accommodations: If you have a disability which requires accommodations for this course, please let me know during the first week of class.
- Grades will be maintained on Canvas.
- Drops: The Drop Policy for the Mathematics/Statistics courses is available here: <https://www.csus.edu/math/documents/policies/Drop%20Policy.pdf> You are responsible for reading it. (In a nutshell, it says you can drop during the first 6 weeks of class pretty easily, but, after the 6th week, dropping is only permitted under compelling medical or job-related circumstances and if you are earning a C or better in the course.)

Academic Honesty Policy Summary:

Academic Integrity is of utmost importance to me. Do not cheat. Cheating incidents will be reported to the Dean of Student Affairs and appropriate punitive action will be taken. Punitive action may include: a permanent record of the incident in your academic file, a failing grade in the course, suspension or expulsion from the university. Its better to fail with your integrity intact than to pass without it.

Copying solutions or blocks of code found on the internet or in texts without properly citing the reference constitutes plagiarism. Don't do it. Copying solutions or code from other students is likewise prohibited. If you are unsure whether a resource is permitted, don't hesitate to ask. Some of the assignments will require you to write code which may already exist in a package or command. If the intent of the assignment is for you to write your own code, write your code from scratch and refrain from using other sources. For other assignments, the intent may be for you to use existing functions to analyze data. In this case, you may use internet and text resources to find appropriate commands and their syntax. You should understand the commands you are using. I will work to make the intent of each assignment clear, but ask if you are confused. Using R help files is always allowed as well as user group websites like stackoverflow.com. You should use these resources to help find appropriate commands and their syntax, not to copy verbatim large blocks of code. Again, if in doubt, ask me.

The complete Academic Honesty Policy of Sacramento State University is here: <http://www.csus.edu/umannual/student/STU-0100.htm>.

Online submission of, or placing one's name on an exam, assignment, or any course document is a statement of academic honor that the student has not received or given inappropriate assistance in completing it and that the student has complied with the Academic Honesty Policy in that work.

Tentative Course Outline:

The weekly coverage might change as it depends on the progress of the class.

Week	Content
Week 1	*****No Class due to the Martin Luther King, Jr. Holiday*****
Week 2	<ul style="list-style-type: none">• Installing R and packages, overview of objects in R, Vectors• Suggested Reading: Ch. 1 and 2, Appendix A and B, Matloff
Week 3	<ul style="list-style-type: none">• R: Matrices and lists, user-defined functions• Suggested Reading: Ch. 3-4, Matloff
Week 4	<ul style="list-style-type: none">• R: Dataframes, Factors and Tables• Suggested Reading: Ch. 5-6, Matloff
Week 5	<ul style="list-style-type: none">• R: Control structures – looping, conditionals• Suggested Reading: Ch. 7, Matloff
Week 6	<ul style="list-style-type: none">• EXAM 1• More on recursion and looping, debugging• Extracting data from a website (webscraping)
Week 7	<ul style="list-style-type: none">• R: Built-in Math and Statistical functions• Suggested Reading: Ch. 8, Matloff
Week 8	<ul style="list-style-type: none">• R: Graphics (Base)• Suggested Reading: Ch. 12, Matloff
March 18, 2019	*****SPRING RECESS, NO CLASS*****
Week 9	<ul style="list-style-type: none">• Graphics - ggplot2 package
Week 10	*****No Class due to Cesar Chavez Day Holiday*****
Week 11	<ul style="list-style-type: none">• Applications in R (machine learning - decision trees and cross-validation)
Week 12	<ul style="list-style-type: none">• EXAM 2• More decision trees or clustering
Week 13	<ul style="list-style-type: none">• Applications in R (Linear regression or simulation)
Week 14	<ul style="list-style-type: none">• Topic TBA: Depends on preference of the class. Some options are cluster analysis with mixed data, introduction to databases, interfacing R to SQL databases, using the datatable package for large datasets, more on linear regression, Intro to Bayesian Statistics, Simulating random variables in R, creating apps with the R Shiny package, RSweave, and RMarkdown.
Week 15	<ul style="list-style-type: none">• Topic TBA
Week 16	<ul style="list-style-type: none">• FINAL EXAM