COURSE NUMBER: APSTA-GE.2352

Course Title: Practicum in Statistical Computing

LOCATION: Lecture: GCASL[1], Room 361 (in person)
Lab: Bobst, LL143 (in person)

MEETING TIMES: **Tuesdays, 10:00 AM - 11:40 AM (lecture)**
Tuesdays 12:30PM-1:20PM (lab section 1)
Tuesdays 1:30PM-2:20PM (lab section 2)

Number of Credits: 1 or 3 (details below)

1 credit (version 1: open to some doctoral students): meet first 7 weeks of lecture and lab
1 credit (version 2: advanced placement): meet last 7 weeks of lecture and lab (you join class on week 8; first assignment due that week; week 8-14 lab work is required)
3 credit: meet all 14 weeks, attend lab, do all assignments

Instructor: Marc Scott (lect.) Office hours: TBA; Marc will post a selection of weekly "time slots" on Brightspace.

Course Description:

This course will introduce the student to modern statistical programming and simulation using the language R. The core skills are oriented around first understanding variables, data structures, program flow (e.g., conditional execution, looping) and functional programming, then applying these skills to answer interesting statistical questions involving the comparison of groups, which is core to statistical practice. Most statistical analysis will be motivated via simulations, rather than mathematical theory. The course content (programming and data analysis) requires significant outside reading and programming.

Course Notes:

- **Topics**, listed at the end of this document, are a rough guide. Actual lectures and labs will most closely correspond to the weekly posted .R files and Shiny apps.
- Lab session is in-person and attendance will be taken. It roughly follows this format:
    - **a.** Concept review using code snippets from the book and some supplemental material; you are expected to review the text and concept portion of the Rshiny apps posted on Brightspace **before lab.**
    - **b.** Lab assignment (connected to assigned homework). These build the skills for all homework and projects. Usually, the last part of the shiny app.

- Class sessions (in-person) will consist roughly of four distinct parts:

[1] GCASL is the Global Center for Academic and Spiritual Life, entrance at 238 Thompson Street.

    a. Introduction of a programming concept
    b. Relating that concept to solving a statistical question
    c. Group programming exercise
    d. Discussion / Question and Answer
- We require this course for MS-A3SR students who have not had formal instruction in a computer science course such as "Introduction to Programming in Java or C" or who have no experience with the program language R.
- A natural sequel to this course is APSTA-GE 2017, Educational Data Science Practicum.
- The course provides foundational skills that will be used in nearly all APSTA graduate courses.

Course Co-requisites:

- If the student has little prior experience with statistics, they must take APSTA-GE 2003 concurrently.
- OPTIONAL: R "courses" via Datacamp (you should have received an invitation in late August). These are helpful learning tools for the R language.

Expectations:

- Programming, and particularly debugging, requires substantial **persistence** and **creative** exploration and problem-solving skills. For the student who is new to this type of work, we suggest ***spending some time prior to the first class*** exploring basic programming (any language) with online tutorials such as those developed by the Khan Academy.
- Be open to **making mistakes**! Programming requires many attempts, patience, and perseverance.

Learning Objectives:

By the end of the course, students will be able to:

1. Analyze a statistical question involving the comparison of groups using modern statistical simulation tools.
2. Use modern, structured programming techniques, as well as self-documenting code.
3. Specific projects involving density estimation, smoothing, power analysis, and bootstrap inference.

Course Format: (Lab; Lecture); Fall offering

Course Requirements

There will be 3 short projects, all involving writing R code. Students are encouraged to work together to learn concepts. Assignments are handed in separately and the submitted work must be completed individually.

Evaluation for this course will be weighted depending on credits:

1-credit, **first 7 weeks**:

- Project (due week 8) 40%

- Homework 50%
- Participation (lab) 10%

3-credit, **all 14 weeks**:

- 4 Projects 65%
- Homework 25%
- Participation (lab) 10%

1-credit, **last 7 weeks**:

- 4 Projects (including Tidyverse) 70%
- Homework 20%
- Participation 10%

.

ASSIGNMENT AND GRADING DETAILS

Naming convention and homework notes – **important**:

1. **NAME**: Do not use hw1.R to name homework1. Use LAST_NAME_FIRST_INITIAL_HW1.R, such as Scott_M_HW1.R. This way your homework will not be confused with someone else's. **YOU WILL GAIN PARTICIPATION POINTS IF YOU NAME YOUR FILES WELL!**
2. Put your name in the body of the file (either in comments #LAST_NAME or as part of a title, if a Markdown file)
3. Your homework should run without modification. If you load a file, try to use a reference like "~/filename.filetype" which refers to the 'home' directory, rather than, "C:\\My Documents\\StatComp\\Assignments\\filename.filetype."
4. Do not use "install.packages" in your code. You only need to install.packages once. Instead, use require or library to 'load' a package and we can install them locally if for some reason we don't have them. Use packages referenced in the class and limit use of any other packages.

The grading system is as follows:

A: 94.5% and above	A-: 89.5-94.49%	
B+: 86.5-89.49%	B: 82.5-86.49%	B-: 79.5-82.49%
C+: 76.5-79.49%	C: 72.5-76.49%	C-: 69.5-72.49%
D+: 67-69.99%	D: 64-66.99%	D-: 60-63.99%
F: less than 60%		

Projects:

The projects will be assessed for excellence in: quality of the code (well-commented; functional); organization of the code/writing; and reproducibility/flexibility/extendibility of the code (how modular is the design? Could the structure be reused for a slightly different problem?). To receive maximum credit for each project, satisfaction of all three requirements is required.

Class Participation:

This course is highly interactive, both in terms of working and learning in teams and as a classroom. It is often based on HW or readings as well, so being prepared is important. However, interaction takes a variety of forms, ranging from one-on-one discussions to group presentations, so that different skills are emphasized at different times. **Attendance in your assigned lab section is a key form of participation that will be noted.**

Required Readings and/or Text

Zieffler, Harring, Long (2011). Comparing Groups: Randomization and Bootstrap Methods Using R. Wiley. It should be available through NYU libraries. Look here first (from an NYU computer or VPN):

https://ebookcentral-proquest-com.proxy.library.nyu.edu/lib/nyulibrary-ebooks/detail.action?docID=708887

This external link may work when you are on an NYU local network:

http://onlinelibrary.wiley.com/book/10.1002/9781118063682

There will be a number of readings – particularly user manuals and tutorials available from the web.

Academic Integrity and Group Work Policies:

All students are responsible for understanding and complying with the NYU Steinhardt Statement on Academic Integrity. A copy is available at: http://steinhardt.nyu.edu/policies/academic_integrity.

Unless otherwise indicated, you should work on the assignments by yourself. You may discuss problems with your classmates, the lab assistants, and the professor, but your submitted work must be your own (please see the Academic Integrity statement here). In particular, you may not copy code from your classmates; this will be deemed plagiarism and sanctioned according to NYU guidelines.

Students with Disabilities:

Students with physical or learning disabilities are required to register with the Moses Center for Student Accessibility, 726 Broadway, 2nd Floor, (212-998-4980 and online at http://www.nyu.edu/csd) and are required to present a letter from the Center to the instructor at the start of the semester in order to be considered for appropriate accommodation.

Inclusion:

NYU values an inclusive and equitable environment for all our students. I hope to foster a sense of community in this class and consider it a place where individuals of all backgrounds, beliefs, ethnicities, national origins, gender identities, sexual orientations, religious and political affiliations, and abilities will be treated with respect. It is my intent that all students' learning needs be addressed both in and out of class, and that the diversity that students bring to this class be viewed as a resource and strength. If this standard is not being upheld, please feel free to speak with me.

Course Outline (list of lectures/topics each session)

Week	**Topics**	**Lab Activity**	**Preparation**	**Assignment**
1 Sept 6th, 2021	R as a Mathematical Scratchpad: Scalars, Vectors, Matrices	Topic: Quick arithmetic using vectors and matrices (e.g., sweeping)	Read Chapter 1 **and** have a **laptop** in class!	
2 Sept 13th, 2021	Basic Data Structures; Transforming Variables; Missing Data	Topic: Descriptive statistics (means, variances, boxplots, histograms, scatter plots, prop.table).	Read Chapter 2	HW 1 DUE: Problems 1.1 to 1.4 [Introduced and discussed in lab]
3 Sept 20th, 2021	Functional Programming; Loops and Conditional Operations; Simulation	Unpacking the density function Topic: Extending the plot function based on the objects returned in the density function.	Read Chapter 3	HW 2 DUE Prob 2.2 (abcde) + Additional questions (Shiny App Lab)
4 Sept 27th, 2021	Density Estimation (statistical concepts)	Topic: Write your own "rough" density estimation routine using boxcar weights		**Lab Report 1: Central Limit Theorem (on Brightspace)**
5 Oct 4th, 2021	Median Smoothing	Topic: write your own *median smoother; generalize*		HW 3 DUE Prob 3.2 (abc); 3.3 (abcd)
Oct 11th, 2021	NO CLASS – Legislative day	NO CLASS – Legislative day		NO CLASS – Legislative day
6 Oct 18th, 2021	Tidyverse I: tibbles, pipes, summaries; comparison to	Topic: simple apply functions; summaries;		HW 4 DUE: Prob 3.1 (you are to "play" with various choices for the "bw" parameter to the density function and then write a few

	Base R			sentences about whether you discover a second mode with varying bandwidths (degree of smoothness), NOT A JOURNAL ARTICLE WRITEUP)
7 Oct 25th, 2021	Tidyverse II: Joins; some comparison to Base R	Topic: Tidyverse		
8 Nov 1st, 2021	Simple linear regression	Topic: Simple linear regression (program via optimization & minimizing SSR rather than calculus based); functions calling functions.	Read Chapter 4; **Work through Week8C-snippets shiny app**	**Tidyverse project (all students)**
9 Nov 8th, 2021	Use of … in functions (esp. wrappers); Newton's Method	Topic: More OLS	**Read Newton's Method Lecture notes**	PROJ 1 DUE: Prob 4.1
10 Nov 15th, 2021	alternatives to OLS; ANOVA; sampling variability;	Topic: Comparing Groups – chapter 5 snippets	Read Chapter 5	HW 5 DUE: 5.2
11 Nov 22nd, 2021	Intro to power analysis (t-tests); Randomization Tests (intro & improving Chapter 6 code)	Power analysis via simulation. Introduction to Project 2 Randomization Tests Improving Chapter 6 code	Read Chapter 6	HW 6 DUE: 6.1; 6.2
12 Nov 29th , 2021	Permutations; Power (builds off prior lab)	Lab: more on Power Analysis		
13 Dec 6th , 2021	Bootstrap (Parametric & Nonparametric)	Topic: Bootstrap	Read Chapter 7	PROJ 2 DUE: power analysis (sim & AUC)
14 Dec 13th 2021	Bootstrap (Parametric & Nonparametric)	Topic: Bootstrap	Read Chapter 8	HW 7 DUE: Prob 7.1 (modified) – compare MEANS using parametric t-test & 7.2 (modified) COMPARE VARIANCES using LeveneTest [in library car] NEITHER PROBLEM USES BOOTSTRAP.

15 **FRIDAY** Dec 16th, 2021	*READING WEEK*	**Note this is a Friday**		PROJ 3 DUE: problems 7.1,7.2,7.3,7.4 (USE BOOTSTRAP AS INDICATED)