COURSE NAME AND NUMBER: IS381 Statistics and Probability with R
SEMSTER:
CREDITS: 3
PREREQUISITE(S): IS 210, IS 211, IS 361 AND IS 362

INSTRUCTOR:
EMAIL:
GITHUB:
OFFICE HOURS:

COURSE DESCRIPTION:

This course covers basic techniques in probability and statistics applied using the R statistical programming language. The course starts with introducing students to R for data import, manipulation, and visualization. Statistical topics include descriptive statistics, sampling techniques, discrete probability models, sampling, statistical distributions, correlation, and null hypothesis testing.

PROGRAM LEARNING OUTCOMES ADDRESSED BY THIS COURSE:

1. Describe how information is collected, stored, managed, classified, retrieved, and disseminated
2. Analyze data to solve problems in practical scenarios
3. Apply skills used to program applications, manage systems, and protect data in complex/heterogeneous computing environments
4. Apply analytical and statistical methods to retrieve, manipulate, analyze, and visualize data for decision-making

COURSE LEARNING OUTCOMES:

1. Effectively use R for conducting analysis, creating reports, and presenting results
2. Estimate predictive models using both parametric and non-parametric models.
3. Communicate the accuracy of predictive models using a variety of fit statistics
4. Have strategies for handling missing data in the predictive modeling pipeline.
5. Effectively communicate the results of a predictive models

STUDENTS WILL BE ABLE TO:

- Effectively use R for conducting analysis, creating reports, and presenting results.
- Understand the foundations of probability theory and perform basic probability calculations.
- Build basic stochastic models for commonly encountered data science.
- Explore and summarize data using descriptive statistics.
- Test hypotheses using classical and modern computational techniques.
- Calculate and define the relationship between multiple variables.

REQUIRED TEXTBOOKS:

Introduction to Modern Statistics by Mine Çetinkaya-Rundel and Johanna Hardin. Available for free at https://openintro-ims.netlify.app

R for Data Science 2nd edition by Hadley Wickham, Mine Çetinkaya-Runde, and Garrett Grolemund. Available for free at. https://r4ds.hadley.nz

ADDITIONAL RESOURCES:

- R Software – Download from https://cran.r-project.org
- RStudio Desktop – Download from https://posit.co/downloads/
- Windows users should also download and install RTools from https://cran.r-project.org/bin/windows/Rtools/
- Mac users should also download and install Xcode and gfortran. Directions are available here: https://mac.r-project.org/tools/

ASSIGNMENTS AND GRADING:

Data Project (35% Total; Proposal 15%, Final Presentation 20%)

The purpose of the data project is for you to conduct a reproducible analysis with a data set of your choosing. There are two components to the project, the proposal, which will be graded on a pass/fail basis, and the final report. The outline for each of these are provided in the templates. When submitting the assignments, include the R Markdown file (change the name to include your last name, for example LASTNAME-Proposal.Rmd and LASTNAME-Project.Rmd) along with any supplementary files necessary to run the R Markdown file (e.g. data files, screenshots, etc.). Suggestions for possible data sources are included below, however you are free to use data not listed below. The only requirement is that you are allowed to share the data. Projects will be shared with others on this website, so they should be presented in a way that other students can reproduce your analysis.

Homework Problems (20%, 2.5 points each): This assignment aims to provide an opportunity for you to actively engage in the content you are learning in class. Homework problems are associated with each class topic (see Course Outline) and must be completed once a topic has been covered in class. Each homework assignment will include 5-10 questions that are carefully selected from the textbook. The answers to some of these questions can be found in the back of the textbook – these are good "self-check" questions to ensure you are on the right track. Assignments are graded based on completion, accuracy, and thoroughness; that means you must show your work. Doing so will help us understand where potential misunderstandings lie.

Labs (25%, 5 points each): R is the statistical software you will use for this course. The labs aim to provide an opportunity for you to apply your statistical content knowledge in the context of problems to solve in R, thus also providing you the opportunity to practice and become familiar with the R platform and language. The labs will be guided; thus, step-by-step procedures will be laid out for you to follow in order to get the desired outputs. Interpretations of results are just as important as the results themselves, so once you have the results, interpret them in the context of the problems. Labs are graded based on completion, accuracy, and thoroughness of results and interpretations.

Final Exam (10%): Exams will consist of conceptual, computational, and application questions, an include a combination of multiple choice, short response questions, as well as a data analysis task. The exams will focus on the material covered during the course of the semester. More detail will be provided about the material assessed by each exam closer in time to the actual exams. It should be noted that most of the statistical skills acquired during this class are constantly building upon earlier learning. This means that even though each exam will focus on the preceding section of the course, students might need to recall skills learned in earlier sections.

Participation (10%): While attendance at synchronous meetups is not required, it is highly encouraged that you do attend: this is where you can ask questions, participate in-situ, and engage with your professor and peers. In addition, announcements and updates relating to coursework will be reviewed during these meetups. With that said, we understand that extenuating circumstances might not allow some of you to attend. Thus, we have built-in diagnostic and weekly formative assessment assignments that will give us an understanding of your current level of knowledge and lingering gaps in knowledge to be completed after attending or watching the recording of every meetup:

1. DAACS SRL (https://cuny.daacs.net) and Google Form (only once, at the beginning of the semester)
2. Weekly One-Minute Papers

You will receive full points upon completion of each of these assignments.

Course Assignments	Points or Percentage of Final Grade
Participation/ Weekly Formative Assessments	10%
Project Proposal	15%
Final Project Presentation	20%
Homework	20%
Labs	25%
Final Exam	10%

CUNY SPS UNDERGRAD GRADING SCALE

Letter Grade	Ranges %	GPA
A	93 - 100	4.0
A-	90 - < 92	3.7
B+	87 - < 90	3.3
B	83 - < 87	3.0
B-	80 - < 83	2.7
C+	77 - < 80	2.3
C	73 - < 77	2.0
C-	70 - < 73	1.7
D	60 - < 70	1.0
F	< 60	0.0

COURSE OUTLINE AND SCHEDULE

Subject to change

Week	Start	End	Topics	Materials	Assignments Due
1			Introduction to R and RStudio		Formative assessment (Google Form link)
2			R coding basics	R4DS Chapter 2	Lab #1: Introduction to R and RStudio
3			Data (importing and structure)	R4DS Chapter 7	HW #1
4			Reshaping data	R4DS Chapter 3 and 5	HW #2
5			Data visualization with ggplot2	R4DS Chapter 9	HW #3

Week	Start	End	Topics	Materials	Assignments Due
6			Exploring categorical data	IMS Chapter 4	Lab #2: Visualization through ggplot2
7			Exploring numerical data	IMS Chapter 5	HW #4
8			Foundation for inference	IMS Chapter 11	Lab #3: Summary statistics
9			Central limit theorem	IMS Chapter 13	HW #5
10			Inference for proportions	IMS Chapter 16 and 17	HW #6
11			Inference for two-way tables	IMS Chapter 18	HW #7
12					Project Proposal DUE
13			Inference for numerical data	IMS Chapter 19 and 20	Lab #4: Categorical Inference
14			Analysis of variance	IMS Chapter 22	Lab #5: Numerical Inference
15			Correlation	IMS Chapter 7.1	HW #8
			Wrap up / Final Presentations		Final data project

ACCESSIBILITY AND ACCOMMODATIONS

The CUNY School of Professional Studies is committed to making higher education accessible to students with disabilities by removing architectural barriers and providing programs and support services necessary for them to benefit from the instruction and resources of the University. Early planning is essential for many of the resources and accommodations provided. Please see: Disability Services on the CUNY SPS Website.

ONLINE ETIQUETTE AND ANTI-HARASSMENT POLICY

The University strictly prohibits the use of university online resources or facilities, including Blackboard, for the purpose of harassment of any individual or for the posting of any material that is scandalous, libelous, offensive or otherwise against the University's policies. Please see: "Netiquette in an Online Academic Setting: A Guide for CUNY School of Professional Studies Students."

ACADEMIC INTEGRITY

Academic dishonesty is unacceptable and will not be tolerated. Cheating, forgery, plagiarism and collusion in dishonest acts undermine the educational mission of the City University of New York and the students' personal and intellectual growth. Please see: Academic Integrity on the CUNY SPS Website.

STUDENT SUPPORT SERVICES

If you need any additional help, please visit Student Support Services: Student Support Services.