PSYC148a - Applied Statistical Computing Using R

Course Information

Instructor: Xiaodong Liu, Associate Professor of Psychology (xliu0806@brandeis.edu)

Office Hours/Availability – By Arrangement.

We will use the **private forum** in LATTE for any direct communications. Please send me a message through this forum to arrange a time to speak with me directly with questions, to express concerns, or if you need help with the course material or assignments. I will reply within 48 hours.

Online Meeting Room:

If it is necessary to speak or meet directly, we will set up a time for online conference meeting or online office hours, we will meet through zoom at this address (https://brandeis.zoom.us/j/5613917251)

Term: June 1 - August 8, 2021 | Course Week: Monday to Sunday

Overview

This syllabus contains all relevant information about the course: description, grading criteria, the texts and other materials of instruction, and weekly assignments and due dates. Consider this your roadmap for the course. Please read through the syllabus carefully and feel free to share any questions that you may have. Please print a copy of this syllabus for reference.

Description

This course is designed for students who would like to learn to do statistical computing/programming in R and for students with introductory statistics background who would like to further their understanding of inferential statistics and statistical modeling.
R is the platform for this course. It will be introduced in a way that helps make the connection between data preparation and manipulation, data generation/simulation, data presentation (descriptively and graphically), statistical concepts, and statistical modeling. Additionally, through R programming, students will learn to document the analytic process for reproducibility of research work.

Topics of statistical computing include methods of describing data, numerical summary and graphical presentation of data, computing of probability, density, and quantiles based on given distributions (including Gaussian distribution, t-distribution, F-distribution, Chi-square distribution, and binomial distribution), simulation and data generation, and writing functions for custom statistics. While the main theme of this course is on statistical computing with R, we will emphasize the applied aspect through using data from empirical studies.

Learning Goals/Objectives

By the end of this course students are expected to be able to:

- Document the working process (e.g. statistical analyses that have been done) using R
- understand R environment/workspace and data types in R
- Read data in different formats into R and manipulate data as desired
- Export data from R to desired formats
- Write program/scripts to generate numerical summary of data
- Write program/scripts to present data graphically (and professionally)
- Compute probability or quantiles given a distribution
- Generate (random) data based on a given distribution or hypothesized statistical model
- Write functions to do customized analyses or generate non-standard statistics (e.g. a function to compute adjusted-correlation coefficient based upon sample size), and
- Implement general linear model (e.g. regression model and ANOVA model)

Prerequisites

No prior R experience is required. Some introductory statistics experience (e.g., Psyc51a or an equivalent course) will be helpful.

Course Materials

Required Textbooks/Sources

No required textbooks.
All required readings will be given on LATTE (the course website) two weeks before each week

Recommended Resources/books

These recommended books are for extra readings, students should refer to it as you feel necessary. Other recommended resources may be given under each week.

Pace, L. (2012). Beginning R – an introduction to statistical programming. New York: Apress. (online version is available through Brandeis library. You will need to use your Brandeis email to login to read the book online)

Jones, O., Maillardet, R., & Robinson, A. (2014) Introduction to Scientific Programming and Simulation Using R (2nd ed.). Chapman and Hall/CRC. (advanced reading, covering advanced statistical computing materials) (there might be a PDF version online).

Temple, M. (2016). Simulation for Data Science with R. Packt publishing (advanced reading, devoted to simulation in R, online version available through Brandeis library)

Crawley, M. J. (2013). The R Book. Hoboken, N.J.: John Wiley & Sons Inc. (online version is available through Brandeis library)

Braun, J. & Murdoch, D.J. (2007). A first course in statistical programming with R. Cambridge: Cambridge University Press. (reference only)

Nolan, D. & Lang, D.T. (2015). *Case Studies in Data Science with R*, CRC press. (book website) (advanced reading,).

Required Software

We will use R and RStudio for statistical computing. Both have free versions. Details of getting R and RStudio will be given in first week.

Online Course Content

This course will be conducted completely online using Brandeis' LATTE site, available at this LATTE link (http://latte.brandeis.edu). You will need to use your Brandeis email to login first. The site contains the course syllabus, assigned reading materials, links/resources to other course-related materials, discussion forums, practices and assignments, and due dates. Access information will be emailed to enrolled students before the start of the course.

You should have access to the course contents one week before the course starts. If possible, try to go to the course website earlier to get familiar with the layout and course information. To begin participating in the course, review the Welcome Message and the materials found in the Week 1 block.

Major Assessments

Grading Percentages	
25%	Discussion forums: Actively participate in online discussion forum and activities (through posting your thoughts on and solutions to discussion questions, offering comments on the posts from other students, and asking or answering questions from other students). Details of grading criteria for discussion forum will be given below.
65%	Assignments: Complete at least eight assignments by the designated deadline. Each assignment will cover the learning objectives specified in each week and earlier. The points of each assignment may be different depending upon the problems asked, but each will be converted into a 100-scale when computing the final course grade. Each assignment will take the form of a single R Markdown file and a related HTML file (more details in course). Details of grading criteria for each assignment will be given below.
10%	Final Project: The final project will help assess overall to what extent students achieve the learning goals. Details of grading criteria for the final project will be given in the instructions of the final project.

For Graduate Students

Students who have already enrolled in a graduate program and would like to get graduate credit for this course will need to do extra problems given in some assignments and/or in the final project. For graduate credit, the final course grade has to be at least B-).

Late Policy

All the assignments and posting have to be submitted by the given deadline, **late submission will not be accepted** (unless with documented emergency).

Participation in Discussion Forum

The discussion forum in LATTE takes the place of discussions that would occur in a traditional classroom. Keep in mind that postings to the forums will be as rich as you make them. They are required so that you can reflect on the course materials and share your knowledge and ideas while gaining from the insights of your peers as well. The instructor will not lead these discussions but may facilitate through probing questions or comments to which you will be expected to respond.
The discussions are an extension of your academic work, so they will require **a formal style of writing**. They are not the equivalent to discussion postings that you may find in online media sites or in social media (See detailed grading criteria below).

Grading Criteria for Discussions

For each week, you should post to the discussion forum on **at least 3 different days** of the course week.

Initial Post (by Thursday)	10 points of weekly discussion grade	Max. Points
Thoughtful post with supporting evidence (if relevant)	Your post should reflect your effort and careful thoughts, including references or citations to weekly readings or other external references	5
Detailed response completeness	Answers all the question(s) completely	3
Mechanics	Complete sentence, little to no spelling or grammatical errors	2
Responses (by Sunday)	**6 points of weekly discussion grade (3 points each response to a different student's initial post)**	**Max. Points**
Substantive	Follow-on points from your personal insights and/or from the readings, meet minimum requirement specified in each discussion	4
Mechanics	Complete sentence, little to no spelling or grammatical errors	2
Deductions	All posts on only 2 days: -1 point All posts on only 1 day: -2 points	Initial post **1 day late**: -4 points, Initial post **2 days or more late**: -8 points Replies **1 or more days late**: -6 points

Grading Criteria for Each Assignment

Completeness and Correctness 80 %	Deductions will be made based upon points assigned to each problem and whether the scripts and answers are correct.
Knitting of the Markdown file 10%	The Rmd file you submit knits correctly (i.e., if there are no errors and the HTML file is produced when the grader attempts to knit your Rmd file). If your Rmd file fails to knit, it will be returned to you and will be given 24 hours to re-submit your homework. Re-submission of Rmd file will get 50% deduction. If Re-submission of Rmd file still does not work, you will not get this credit and you should submit your R script file. The assignment will be graded based upon the R script file you submitted.
R script styles 10%	Good script styles: Easy to read and follow up; Appropriate and detailed comments; Appropriate indentation; Appropriate use of functions; Reasonable object names; Appropriate use of inline code chunks; No extra or unnecessary scripts Script style is very important. With the exception of first two assignments, 15% deduction if failure on at least 2 criteria; 50% deduction for failure to adhere to 3 or more of the criteria.
Late Deductions	Each assignment is due by 23:50, Sunday of that week (Eastern DST/Boston local time). Your assignment will be automatically time-stamped when you submit your work through LATTE. 20% deduction per 8 hours late (each 8-hour late is from 1 minute passed the due time until 7 hours 59 minutes). 100% deduction with more than 24 hours late.

Details of Weekly Expectations/Schedule

The weekly schedule here is subject to change/adjust by the instructor.

Week 1 - June 1	Introduction
Topics	● Course expectations ● R: download, installation, and interfaces: R Console, R script/document ● RStudio: installation and interfaces ● LATTE: get familiar with components in LATTE
Learning Resources	Reading materials from instructor posted on LATTE
Activities	Introduce Yourself Discussion due Thursday, two responses to other students' discussion due Sunday: Questions/prompts in LATTE Assignment: due Sunday
Week 2 - June 7	R Basics, R Markdown
Topics	● Why using R? ● working directory ● Object ● R basic data type/class ● R Markdown
Learning Resources	Reading materials from instructor posted on LATTE
Activities	Discussion due Thursday, two responses to other students' discussion due Sunday: Questions/prompts in LATTE Assignment: due Sunday
Week 3 - June 14	R Basics: data structures (1)
Topics	● R data structures: vector, matrix ● R built-in function ● functions related to vector, matrix, and list ● Review: descriptive statistics (mean, sd, quantiles)
Learning Resources	Reading materials from instructor posted on LATTE

	Additional readings: an introduction to R (https://cran.r-project.org/doc/manuals/R-intro.pdf): (Ch. 2: pp.7-12 for vector, Ch. 5: pp. 18-22 for matrix)
Activities	Class Discussion due Thursday, two responses due Sunday: Questions/prompts in LATTE Assignment: due Sunday
Week 4 - June 21	R Basics: data structures (2), getting to know the data
Topics	● R data structures: list and data frame ● functions related to list and data frame ● functions and statistics helping get to know the data
Learning Resources	Reading materials from instructor posted on LATTE Additional readings: an introduction to R (Ch. 6 for list and data frame)
Activities	Class Discussion due Thursday, two responses due Sunday: Questions/prompts in LATTE Assignment: due Sunday
Week 5 - June 28	Factor, Conditional processing, loop, user-defined function
Topics	● Factor ● Conditional processing: if...else…, elseif... ● for loop ● functions: sample( ), rep( ) ● user-defined functions
Learning Resources	Reading materials from instructor posted on LATTE Additional readings: an introduction to R (Ch. 4 for factor)
Activities	Class Discussion due Thursday, two responses due Sunday: Questions/prompts in LATTE Assignment: due Sunday
Week 6 - July 6	Discrete random variable and continuous random variable
Topics	● Random variables: discrete vs. continuous ● Distribution of random variable: Binomial distribution, Normal distribution ● Probability mass function, probability density function

	● R functions on distributions ● Simulation (of data with given distributional property)
Learning Resources	Reading materials from instructor posted on LATTE
Activities	Class Discussion due Thursday, two responses due Sunday: Questions/prompts in LATTE Assignment: due Sunday
Week 7- July 12	Data transformation, Sampling distribution
Topics	● Linear data transformation: deviation, standardization ● Non-linear data transformation: normalization ● Sample statistics as random variables ● Sampling distribution and standard error ● the Law of large numbers, Central limit theorem
Learning Resources	Reading materials from instructor posted on LATTE
Activities	Class Discussion due Thursday, two responses due Sunday: Questions/prompts in LATTE Assignment: due Sunday Instructions for final project posted
Week 8 - July 19	Hypothesis testing, t-test (t-distribution) and F-test (F-distribution) in R
Topics	● Review concepts on hypothesis testing ● functions related to analyzing data with continuous outcome/dependent variable and categorical independent variable (factor) in R ● t-test and F-test
Learning Resources	Reading materials from instructor posted on LATTE
Activities	Class Discussion due Thursday, two responses on final project proposals due Sunday: Questions/prompts in LATTE Assignment: due Sunday Instructions for the format of final project posted
Week 9- July 26	Data input/output in R, Data manipulation (data merging), package ggplot2
Topics	● input and output methods in R ● data merging and concatenation

	● ggplot2 details
Learning Resources	Reading materials from instructor posted on LATTE
Activities	Class Discussion due Thursday, two responses due Sunday: Questions/prompts in LATTE Assignment: due Sunday
Week 10 - August 2	Exploring relation among continuous variable
Topics	● Graphical presentation of relations between continuous variables: scatterplot ● functions related to correlation and regression analyses ● interaction and presentation
Learning Resources	Reading materials from instructor posted on LATTE
Activities	Class Reflection: Questions/prompts in LATTE Final Project due Sunday

Course Policies and Procedures[1]

Asynchronous Work

The weekly readings, discussion, and assignments will be done asynchronously; i.e., students can login to the course, read/download materials, post to the forums, and submit assignments throughout the course week. Please carefully follow the syllabus and the weekly materials on LATTE to help manage your time throughout the course week; once we enter week 2 or 3, students typically become much more comfortable with the pace and flow of the course.

Synchronous Work

Depending upon the needs of each of you, we could schedule to meet online a few times/hours as a class using the Zoom video conferencing link (which will be provided then). These sessions, if any, will be recorded and posted on LATTE so that they can be reviewed at a later date. For group projects, if any, students may use the asynchronous group forums available and, if they choose, they may use synchronous group meeting tools, such as Google Hangout, which is available to all Brandeis students through their Brandeis account. Alternatively, students who live locally may choose to meet in person (but that is your choice).

Work Expectations

Students are responsible to explore each week's materials and submit required work by their due dates. On average, a student can expect to spend approximately 3-4 hours per week reading and practice using the reading materials, and approximately 9-12 hours per week completing assignments and posting to discussions. All assignments are due by the close of each week, Sunday midnight, except specified otherwise for a particular homework.

Late Policies

Late discussion posts are strongly discouraged as the success of class discussion is dependent upon the active engagement of all participants in the course. Late policies related to discussion posts can be found on page 4 (grading criteria on discussion).

Points will be deducted for late assignments (see late policy on page 5: grading criteria on assignments).

On rare occasions, issues do occur that may warrant an exception to the late policy above. Please notify the instructor at least 24 hours in advance of a due date/time if an issue arises that will make it impossible for you to meet a stated due date. Exceptions, although rare, will be considered on a case-by-case basis. If exceptions are to be made, relevant documents are expected for such rare occasions.

Confidentiality in the "Classroom"

As we proceed throughout our class, I'd like to highlight a point about confidentiality in our online classroom. We can draw on the wealth of examples from our personal experiences in class discussions and in our written work. However, it is imperative that we not share information that

[1] This part is adopted from other online course [reference given by Carol Damm]

is confidential in nature. In addition, we should respect our fellow classmates and work under the assumption that what is discussed here stays within the confines of the classroom.

Finally, for your awareness, members of the University's technical staff have access to all course sites to aid in course setup and technical troubleshooting. Select staff have access to all courses for oversight purposes. Students enrolled in online courses can expect that individuals other than their fellow classmates and the course instructor(s) may visit their course for various purposes. Their intentions are to aid in technical troubleshooting and to ensure that quality course delivery standards are met. Strict confidentiality of student information is maintained.

Grading Standards

Students are graded on demonstration of knowledge, competence, and on effort. Each student is expected to maintain high standards of honesty and ethical behavior. All assignments are meant to represent your own work. I expect students to conduct themselves courteously online. If in my judgment a student's conduct is not courteous, I reserve the right to reduce that student's grade.

Guideline for letter grade:

Course points	Letter grade
94 & above	A
90 - < 94	A-
85 - < 90	B+
80 - < 85	B
75 - < 80	B-[2]
70 - < 75	C+
65 - < 70	C
60 - < 65	C-
Below 60	No credit

Feedback

Feedback will be provided on assignments within 7 days of the due date. If necessary, feedback on discussion will be given by the instructor during the week or no later than 7 days of the close of each week.

[2] B- is the minimum for graduate credits.

University-wide Standards

Accommodations

Brandeis seeks to welcome and include all students. If you are a student who needs accommodations as outlined in an accommodations letter, please talk with me and present your letter of accommodation as soon as you can. I want to support you.

In order to provide test accommodations, I need the letter more than 48 hours in advance. I want to provide your accommodations, but cannot do so retroactively. If you have questions about documenting a disability or requesting accommodations, please contact Student Accessibility Support (SAS) at 781.736.3470 or access@brandeis.edu.

Academic Honesty and Student Integrity

Academic honesty and student integrity are of fundamental importance at Brandeis University and we want students to understand this clearly at the start of the term.

As stated in the Brandeis Rights and Responsibilities handbook, "Every member of the University Community is expected to maintain the highest standards of academic honesty. A student shall not receive credit for work that is not the product of the student's own effort. A student's name on any written exercise constitutes a statement that the work is the result of the student's own thought and study, stated in the students own words, and produced without the assistance of others, except in quotes, footnotes or references with appropriate acknowledgement of the source."

In particular, students must be aware that material (including ideas, phrases, sentences, etc.) taken from the Internet and other sources MUST be appropriately cited if quoted, and footnoted in any written work turned in for this, or any, Brandeis class. Also, students will not be allowed to collaborate on work except by the specific permission of the instructor. Failure to cite resources properly may result in a referral being made to the Office of Student Development and Judicial Education. The outcome of this action may involve academic and disciplinary sanctions, which could include (but are not limited to) such penalties as receiving no credit for the assignment in question, receiving no credit for the related course, or suspension or dismissal from the University.

Students may be required to submit work to TurnItIn.com software to verify originality. TurnItIn is a tool that compares student assignment submissions to internet sources and a comprehensive database of other papers. It creates a report that provides a link to possible matches and a '"similarity score". TurnItIn does not determine whether a paper has been plagiarized; individual faculty will make that judgment. All papers submitted to TurnItIn are kept in a separate reference database of Brandeis work, to be used solely for the purpose of detecting plagiarism in the future. Students retain copyright on their original course work. Allegations of alleged academic dishonesty will be forwarded to the Director of Academic Integrity. Sanctions for academic dishonesty can include failing grades and/or suspension from the university. Citation and research assistance can be found at LTS - Library guides.

Further information regarding academic integrity may be found in the following publications: "In Pursuit of Excellence - A Guide to Academic Integrity for the Brandeis Community", "(Students') Rights and Responsibilities Handbook", AND " Graduate Professional Studies

Student Handbook". A student who is in doubt about standards of academic honesty (regarding plagiarism, multiple submissions of written work, unacknowledged or unauthorized collaborative effort, false citation or false data) should consult either the course instructor or other staff of the Summer School at Brandeis.

University Caveat

The above schedule, content, and procedures in this course are subject to change in the event of extenuating circumstances.