PSYC148A Summer 2016

Applied Statistical Computing using R

Department of Psychology
Brandeis University

Dr. Xiaodong Liu

Instructor: Xiaodong Liu, xliu0806@brandeis.edu, Brown 106, 781-736-3244 (O).
Class time: 8:30 – 10:50 M. Tu. Th. (6/6-7/6)

Course Description

This course has two main goals: 1) it is designed for undergraduate and graduate students with background in descriptive and inferential statistics who would like to further their understanding of statistical modeling and its application in applied and academic research; 2) it is designed for students who would like to learn to do statistical programming in R. R will be introduced in a way that helps make the connection between statistical concepts, modeling, and their implementation. In addition, through R programming, students will learn to document their analysis process for reproducibility of research work. Topics of statistical computing include methods of describing data, numerical summary and graphical presentation of data and statistics, computing of probability or quantiles based on given distributions (including Gaussian distribution, t-distribution, F-distribution, Chi-square distribution, binomial distribution, and Poisson 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.

Course Credits:

This is a four-credit course. Success in this four-credit hour course is based on the expectation that students will spend a minimum of 9 hours of study time per week in preparation for class (readings, papers, discussion sections, preparation for exams, etc.).
Learning Objectives and Expected Skill Development

Students who successfully complete this course will be able to set up and use R for a variety of tasks. Specifically, students will 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
- 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)
- implement a general linear model (including ANOVA model) in R

Prerequisites

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

Class Format

Class meetings will consist of lectures and demonstration, students’ presentation on project-based learning modules, and in-class programming practices.

Texts (recommended):


Course requirements and assessment:

Students enrolled in this course will be expected to:

- contribute to the course through class participation (attending classes, asking/answering questions, offering comments) (10%)
- actively participate in project-based learning modules (15%)
- complete all assignments by the designated deadline (50%) (the assignment is mainly R programming and related short memo of interpreting the statistical output)
- complete a final take-home project (25%)
- All the assignments have to be submitted by the due date, late submission will not be accepted (unless with documented emergency).

Guideline for letter grade:

<table>
<thead>
<tr>
<th>Grade</th>
<th>95 &amp; above</th>
<th>90 - &lt; 95</th>
<th>85 - &lt; 90</th>
<th>80 - &lt; 85</th>
<th>75 - &lt; 80</th>
<th>70 - &lt; 75</th>
<th>65 - &lt; 70</th>
<th>60 - &lt; 65</th>
<th>Below 60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Letter</td>
<td>A</td>
<td>A-</td>
<td>B+</td>
<td>B</td>
<td>B-</td>
<td>C+</td>
<td>C</td>
<td>C-</td>
<td>Fail</td>
</tr>
</tbody>
</table>

Academic integrity

Academic integrity is central to the mission of educational excellence at Brandeis University. Each student is expected to turn in work completed independently, except when assignments specifically authorize collaborative effort. **It is not acceptable to use the words or ideas of another person – be it a world-class philosopher, your fellow classmate, or your online helper – without proper acknowledgement of that source.** This means that you must use author citations, endnotes, and, where appropriate, quotation marks to indicate the source of any phrases, sentences, paragraphs, or ideas found in published volumes, on the internet, or created by another student.

Violations of University policies on academic integrity may result in failure in the course or on the assignment, or in suspension or dismissal from the University. If you are in doubt about the instructions for any assignment in this course, it is your responsibility to ask for clarification.

Collaboration

Students are welcome to discuss the course’s materials, practices, and homework assignments with each other (for better understanding and learning). **But the work submitted for grading MUST be your own.** For example, you must write your own code, run your own data analyses, and explain the results in your own words and with your own graphical presentations. You may
not submit the same or similar work to this course that you have submitted or will submit to another course or for another project. You may not provide or make available solutions to practices/homework assignments to other students who are taking this course or may take this course in the future.

Accommodations

Students with a documented disability on record at Brandeis University and wishing to have a reasonable accommodation made for this class should let me know the needs in the first week of the course. Students should then provide me a proposed accommodation with appropriate document, if any. A mutually agreed arrangement will then be documented and be followed for this course.

Rules of using computer/cell phone in class time

The use of cell phone is prohibited during class time. Computer (including laptop, ipad etc.) can only be used for taking class notes. Computer can be used for statistical computing only when I ask you to do so.

I assume that you are here because you want to learn. Using a cell phone or laptop, or computer to talk, text, email, or surf the web on non-course related matters is disrespectful to me and to your fellow students. I will ask you to turn off the computer or leave the classroom if the computer or phone is used for any work unrelated to the class in session (including homework for this course). No exceptions.
Course topics (subject to change)

<table>
<thead>
<tr>
<th>Topic</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic 01:</td>
<td>Introduction to R environment and R programming</td>
</tr>
<tr>
<td></td>
<td>Getting Started Using R</td>
</tr>
<tr>
<td></td>
<td>Using the R Help System</td>
</tr>
<tr>
<td>Topic 02:</td>
<td>Data structure and data type, basic R programming and convention</td>
</tr>
<tr>
<td>Topic 03:</td>
<td>Graphics in R</td>
</tr>
<tr>
<td>Topic 04:</td>
<td>Random Variables and Probability</td>
</tr>
<tr>
<td>Topic 05:</td>
<td>Discrete Distributions and Continuous Distributions</td>
</tr>
<tr>
<td></td>
<td>QQ Plot, PP Plot, and Comparison of Distributions in R</td>
</tr>
<tr>
<td>Topic 06:</td>
<td>Generation of Random Numbers and Patterns</td>
</tr>
<tr>
<td></td>
<td>Simulation in R</td>
</tr>
<tr>
<td>Topic 07:</td>
<td>Dataframe in R: import and export data, data manipulation</td>
</tr>
<tr>
<td>Topic 08:</td>
<td>Basic statistical methods</td>
</tr>
<tr>
<td></td>
<td>Describing data in R</td>
</tr>
<tr>
<td>Topic 09:</td>
<td>Functions and self-defined functions in R</td>
</tr>
<tr>
<td>Topic 10:</td>
<td>General linear modeling and its implementation in R</td>
</tr>
<tr>
<td>Topic 11:</td>
<td>Analysis of Variance in R</td>
</tr>
<tr>
<td>Topic 12:</td>
<td>Bayesian Statistics in R (optional)</td>
</tr>
</tbody>
</table>