STAT-705-001: Statistical Computing with R

Quarter 1, Fall 2021

(Last Updated: 8/30/2021)

People

Instructor: Yichen Wang   wangyc@wharton.upenn.edu

Teaching assistants:

      Durga Srivatsan   dsri@seas.upenn.edu

      Trevor Mele   meletre@wharton.upenn.edu

Time and Location

Lectures:

    TR 10:15am - 11:35am, 8/31/2021 to 10/7/2021   JMHH F45

Office hours:

Yichen Wang   T 5pm-6pm, 8/31/2021 to 10/19/2021   JMHH F36

Durga Srivatsan   To be announced.

Trevor Mele   To be announced.

Deadlines: the following dates are applicable to MBA students. Students in other programs should confirm with their academic advisor.

    Add & Drop - 9/7, Grade Type Change - 9/16, Withdrawal - 9/20.

Course Description

"The goal of this course is to introduce students to the R programming language and related ecosystem. This course will provide a skill set that is in demand in both the research and business environments. In addition, R is a platform that is used and required in other advanced classes taught at Wharton, so that this class will prepare students for these higher-level classes and electives."

(https://statistics.wharton.upenn.edu/programs/mba/course-descriptions/)

Prerequisite

Any of Stat 102/112/430/431/613/621, the MBA statistics waiver, or equivalent. Students unsure of their prerequisite coursework should consult the instructor as soon as possible.

Software and Books

The following **software** are required and must be installed by the start of Class 2 on 9/2.

- The R software:

  download at https://www.r-project.org/.

- RStudio, an integrated development environment (IDE) for R:

  download the "Desktop" version at https://www.rstudio.com/products/rstudio/.

No **book** is required, but the following books can be useful as technical references.

- De Vries, Andrie, and Joris Meys. *R for Dummies*. John Wiley & Sons, 2015.
- Grolemund, Garrett. *Hands-on Programming with R: Write Your Own Functions and Simulations*. O'Reilly Media, Inc., 2014.

Grade Components

Homework (60%): homework will be assigned weekly during weeks 1-6. Homework 0 is ungraded. Each of Homework 1-5 accounts for 12% of the final grade. There is no dropping of the lowest score for homework assignments.

Students may discuss homework assignments with each other, but each student must write their own submission and attribute all external sources they have consulted. Copying without attribution is cheating and will be subject to disciplinary actions. (See also "Academic Integrity" on page 5.) Late homework submissions are penalized by 25 out of 100 points up to one day late, and 50 points up to two days late. Submissions more than two days late receive a grade of 0.

Quizzes (20%): one 15-minute, closed-book, open-calculator quiz will be given approximately weekly. There will be 4 quizzes in total, with each quiz score accounting for 5% of the final grade. There will be no make-up quizzes.

Final Project (20%): a substantial take-home project is assigned at the end of course and due on 10/26. More details will be given in Class 12, on 10/7.

Course Schedule

The topics are tentative. Barring extraordinary circumstances, the dates of homework dues and quizzes will not change.

	Date	Topics	Assignments
		Unit 1: The Data (3 classes)	
1	8/31	Introduction to the course. R and Rstudio. Preview of main topics. R vectors. R data types.	Homework 0 Posted.
2	9/2	R data structures: matrices, arrays, lists. Indexing, selecting, and modifying data. `apply` functions.	
3	9/7	R dataframes. Loading data files into R. Merging and aggregating dataframes. Case study: R dataframes.	Homework 0 Due at 11:59pm. Homework 1 Posted.
		Unit 2: The Analysis (7 classes)	
4	9/9	Statistical analysis in R: `lm` and `glm`. Intro to user-defined functions in R: `formals`, `body`, and `environment`.	
5	9/14	User-defined functions continued. Quiz 1.	Quiz 1: up to Class 3. Homework 1 Due at 11:59pm. Homework 2 Posted:
6	9/16	Conditional statements: `if`, `else`. Iterations: `for`, `while`. Vectorized operations: revisiting `apply`.	
7	9/21	Case study: writing your own functions. Quiz 2.	Quiz 2: up to Class 6. Homework 2 Due at 11:59pm. Homework 3 Posted.
8	9/23	Intro to stochastic simulation in R. Sampling from probability distributions. Intro to Monte Carlo methods.	
9	9/28	Intro to R packages and ecosystem. Quiz 3.	Quiz 3: up to Class 8. Homework 3 Due at 11:59pm. Homework 4 Posted.
10	9/30	R packages and ecosystem continued. R base graphics.	

		Unit 3: The Presentation (2 classes)	
11	10/5	R graphics continued. The `ggplot2` package for graphics. Quiz 4.	Quiz 4: up to Class 10. Homework 4 Due at 11:59pm. Homework 5 Posted.
12	10/7	Intro to R Markdown and `knitr`. Discussion of the final project. Review of the course.	Final Project Posted.
	10/12		Homework 5 Due at 11:59pm.
	10/26		Final Project Due at 11:59pm.

Communication

Canvas: all lecture notes, assignments and announcements will be posted on Canvas. Students are expected to check the course Canvas site every weekday to keep up with updates. The "Discussions" section on Canvas is the preferred platform for asking questions related to course materials, assignments, or any logistical issues.

Email: questions unsuitable for posting on Canvas may be directed to the teaching staff by email; however, questions posted on Canvas will generally be replied to more promptly as they serve a broader audience.

Regrade request: for each homework or quiz, regrade requests should be submitted in writing (hardcopy or email) to the instructor within one week of grade release. The entire assignment in question will be regraded upon receipt of request, meaning that the resultant grade may increase, decrease, or remain unchanged compared to the initial grade.

Classroom Expectations

Attendance is not required, however missing classes will likely affect learning, and missing in-class quizzes will affect the final grade. Students are encouraged to ask questions in class.

Laptops are helpful for following along in-class demonstrations of R code.

Electronic devices should be silenced and are not to be used for activities that distract fellow students or the instructor. No electronic devices, except calculators, are allowed in quizzes.

Academic Integrity

All students are bound by University of Pennsylvania's Code of Academic Integrity. (https://catalog.upenn.edu/pennbook/code-of-academic-integrity/)

Any academic dishonesty will result in failing grade for the course and referral to the Office of Student Conduct. Students in doubt about whether an action constitutes academic dishonesty should discuss with the instructor in advance.

Students with Disabilities

All students are welcome in this course. Students should make requests for any accommodations they need through Student Disability Services (SDS). In SDS's own words:

"University of Pennsylvania provides reasonable accommodations to students with disabilities who have self-identified and been approved by the Office of Student Disabilities Services (SDS). If you have not yet contacted SDS and would like to request accommodations or have questions, you can make an appointment by calling SDS at 215-573-9235. The office is located in the Weingarten Learning Resources Center at Stouffer Commons 3702 Spruce Street, Suite 300. All services are confidential." (https://wlrc.vpul.upenn.edu/sds/)

"Students with disabilities who seek accommodation at Penn are responsible for self-identifying with SDS. Identification may take place upon admission or at any time during the student's course of study. Students who are approved for accommodations must authorize SDS to inform professors about their approved accommodations. They must also make online requests to SDS for individual exam accommodations each semester. Students are encouraged to introduce themselves to professors to initiate a dialogue about their particular needs." (https://wlrc.vpul.upenn.edu/wp-content/uploads/2020/08/Provosts_Memorandum_Pub._7-14-15.pdf)