MET CS555 Foundations of Machine Learning

Instructor Name: Hongsheng Wu
Office Location: TBD

Contact Information: wuh@bu.edu
Office Hours: T 5:00 PM to 5:50 PM

Grader: TBD

Course Dates: T
Course Time/Location:6:00 PM to 8:45 PM
928 Commonwealth Avenue 202
Credits: 4

Course Description

This course provides an overview of the statistical tools most commonly used to process, analyze, and visualize data. Topics include simple linear regression, multiple regression, logistic regression, analysis of variance, and survival analysis. These topics are explored using the statistical package R, with a focus on understanding how to use and interpret output from this software as well as how to visualize results. In each topic area, the methodology, including underlying assumptions and the mechanics of how it all works along with appropriate interpretation of the results, are discussed. Concepts are presented in context of real world examples.

Prerequisite

MET CS 544 or equivalent knowledge, or instructor's consent.

Course Objectives

By completing this course, you will be able to:

- Describe the science of statistics and the scope of its potential applications
- Summarize and present data in meaningful ways
- Select the appropriate statistical analysis depending on the research question at hand
- Form testable hypotheses that can be evaluated using common statistical analyses
- Verify the underlying assumptions of a particular analysis
- Communicate results from analyses performed to others effectively and clearly
- Conduct, present, and interpret common statistical analyses using R

Instructional Format, Course Pedagogy, and Approach to Learning

This course will combine traditional lecturing with hands-on assignments that reinforce the lecture material. In particular, lectures will focus on concepts and ideas, while the assignments will provide substantial experience and skills.

Recommended Books:

Long, J. D. & Teetor, P. (2019). *R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics*, 2nd edition. You can access the free eBook.

Chang, W. (2022) *R Graphics Cookbook,* 2nd edition. You can access the free eBook.

James, G., Witten, D., Hastie, T,. Tibshirani, R. (2021). *An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics)*, 2nd edition. You can access the free eBook.

Courseware

The class has a Blackboard site that contains the syllabus, lectures, assignments, and other course-related materials. You can log in to the Blackboard page at: https://onlinecampus.bu.edu/

Assignments and Grading Criteria

The grade for the course is determined by the following, including both theoretical and algorithmic analysis as well as practical implementation in R:

Graded Items:
- **Assignments**: There are a total of six assignments.
- **Quizzes**: There are six quizzes to evaluate students' understanding of concepts. Students should ensure adequate preparation before the quiz. It requires reviewing the course material in depth and understanding all examples.
- **Final Project**: Students implement the learned knowledge into public datasets
  - 3 students per group
  - Report 5%
  - Interview 5%
- **Final Exam:** There will be one Final Exam in this course.

Overall Grading

Six Assignments	30%
Six Quizzes	30%
Final Project	10%
Final Exam	30%

Letter Grade

100-95.00	A		79.99-77.00	C+
94.99-90.00	A-		76.99-73.00	C
89.99-87.00	B+		72.99-70.00	C-
86.99-83.00	B		69.99-60.00	D
82.99-80.00	B-		below 60.00	F

Class Policies

Assignment Completion & Late Work: We recognize that emergencies occur in professional and personal lives. If one occurs that prevents your completion of homework by a deadline, please share the plan with the instructor. This must be done before the deadline (unless the emergency makes this impossible, of course) and should be accompanied by particulars that back it up. Additional documentation may be requested. **Late submissions without reasons will not accept**. There will be no make-up exam for the final exam. Students who cannot take the final exam on the designated day will receive an incomplete grade. If you have any questions about your grading, you need to contact the grader and cc me **before the next assignment/quiz** (before the final exam for the last assignment/quiz). After that, we will not discuss the grade for that assignment/grade.

Academic Conduct Code: Cheating and plagiarism will not be tolerated in any Metropolitan College course. They will result in no credit for the assignment or examination and may lead to disciplinary actions. See link below
http://www.bu.edu/met/metropolitan_college_people/student/resources/conduct/code.html

Please do not share our class Assignments, Quizzes, and Exams on online websites like Coursehero, Chegg, etc. We are monitoring these sites and sending the providers' takedown requests. Our Class Material has Boston University Copyright.

Disability and Access Services

By university policy, every effort will be made to accommodate students with speech, hearing, vision, or other disabilities. Any student needing accommodation for a documented disability should contact Disability and Access Services at 617-353-3658 or access@bu.edu for review and approval of accommodation requests.

Once students receive their accommodation letter, they must send it to the instructor and/or facilitator each semester. They must also send a copy to the Faculty & Student Support Administrator, who may need to update the course settings to ensure accommodations are in place. Accommodations cannot be implemented if the student does not send their letter.

Tentative Schedule
The following schedule is tentative and subject to change. The online schedule needs to be divided weeks by two

Week	Topics	Lecture	Works due
1	• Fundamental Elements of Statistics • Qualitative and Quantitative Data Summaries	Course_introduction Lecture01	
2	• Normal distribution • Sampling • The Central Limit Theorem	Lecture02	A01 Q01
3	• Statistical Inference • Confidence Intervals • Test of Significance • Stating Hypotheses • Test Statistics and p-Values • Evaluating Hypotheses	Lecture03	
4	• Significance Test "Recipe" • Significance Tests and Confidence Intervals • Inference about a Population Mean • Two-Sample Problems	Lecture04	A02 Q02
5	• Scatterplots • Correlation	Lecture05	
6	• Simple Linear Regression • F-test for Simple Linear Regression • t-test for Simple Linear Regression	Lecture06	A03 Q03
7	• Residual Plots • Outliers and Influence Points • Assumptions of least-square regression	Lecture07	
8	• Equation of multiple linear regression • Interpretation of multiple linear regression • F-test for Multiple Linear Regression • t-tests in Multiple Linear Regression • Cautions about Regression	Lecture08	A04 Q04
9	• One-Way Analysis of Variance	Lecture09	

	• F-test for ANOVA • Evaluating Group Differences • Type I and Type II Errors		
10	• Issues with Multiple Comparisons • Assumptions of Analysis of Variance • Relationship between One-Way Analysis of Variance and Regression • One-Way Analysis of Covariance • Two-Way Analysis of Variance • Two-Way Analysis of Covariance	Lecture10	A05 Q05
11	• One-Sample Tests for Proportions • Significance Tests for a Proportion • Confidence Intervals for a Proportion	Lecture11	
12	• Two-Sample Tests for Proportions • Confidence Intervals for Differences in Proportions • Significance Tests for Differences in Proportions • Effect Measures • Logistic Regression • Multiple Logistic Regression • The area under the ROC Curve	Lecture12	A06 Q06 Final Project
13	Review session	Lecture13	Final Project