

MET CS 544 Foundations of Analytics with R Spring 2019

Instructor: Dr. Shengzhi Zhang
Email: shengzhi@bu.edu
Office: 1010 Commonwealth Ave., Rm. 302, Boston, MA 02215

Course Description: The goal of this course is to provide students with the mathematical and practical background required in the field of data analytics. Starting with an introduction to probability and statistics, the R tool is introduced for statistical computing and graphics. Different types of data are investigated along with data summarization techniques and plotting methods. Data populations using discrete, continuous, and multivariate distributions are explored. Sampling methods and errors during measurements and computations are analyzed in the course. String manipulations and data wrangling methods are examined in detail. The concepts covered in the course are demonstrated using R. Laboratory Course.

Course Prerequisites: MET CS 546 Quantitative methods for Information Systems or equivalent

Course Grading Policy: The course grade will be based on active class participation (5%), quizzes (10%), assignments (25%), final project (20%), Midterm exam(20%), final exam (20%). Assignments are expected to be submitted by their respective due dates. Late submissions carry a penalty.

Course Website: <https://learn.bu.edu>

References Books:

- "Introduction to Probability and Statistics Using R", by G. Jay Kerns, 2010. ISBN13: 978-0-557-24979-4.(Reference book)
<http://cran.r-project.org/web/packages/IPSUR/vignettes/IPSUR.pdf>
- "Using R for Introductory Statistics, 2nd edition", by John Verzani, CRC Press, 2014. ISBN13: 978-1466590731. (Reference book)
- "R for Everyone: Advanced Analytics and Graphics, 2nd Edition", by Jared P. Lander, Addison-Wesley Professional, 2017. ISBN13: 978-0134546926. (Reference book)

Student Conduct Code:

Please review the academic conduct code

Tentative Course Schedule:

- Module 1 -- Introduction
 - Introduction to Statistics and Probability
 - Basic Concepts of R -- Data Types and Structures
- Module 2 -- Probability
 - Probability

- Conditional Probability
- Random Variables
- Basic Concepts of R -- Programming Constructs

- Module 3 -- Data Description

- Univariate Data
- Bivariate Data
- Multivariate Data

- Midterm Exam

- Module 4 -- Distributions

- Discrete Distributions
- Continuous Distributions

- Module 5 -- Central Limit Theorem, Sampling and Errors

- Central Limit Theorem
- Sampling Methods
- Errors
- Resampling Methods

- Module 6 -- Strings, Data Wrangling

- Strings and Regular Expressions
- Data Wrangling -- dplyr and tidy

- Final Exam