MET CS544 A1 (Spring 2022) - Foundations of Analytics with R (Mondays, 6 PM)

Instructor

Suresh Kalathur, Ph.D.
Assistant Professor, Computer Science Dept.
Boston Univeristy Metropolitan College
1010 Commonwealth Ave, Room 304
Boston, MA 02215

Email: kalathur@bu.edu
URL: http://kalathur.com/bu
Phone: 617-358-0006
Fax: 617-353-2367

Course Description

The goal of this course is to provide students with the mathematical and practical background required in the field of data analytics. Probability and statistics concepts will be reviewed as well as the R tool 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 - Introduction to Probability and Statistics, or equivalent

Course Grading Policy

The course grade will be based on quizzes (30%), assignments (20%), group final project (20%), and final exam (30%). Assignments are expected to be submitted by their respective due dates. Late submissions carry a penalty.

Course Web Site

- **https://learn.bu.edu**

References

Reference Books

- "*Introduction to Probability and Statistics Using R*", by G. Jay Kerns, 2010. ISBN13: 978-0-557-24979-4. **(Reference book)**
  https://github.com/gjkerns/IPSUR/blob/master/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
    - Basic Concepts of R -- Data Types and Structures
    - *Assignment1 Due, Quiz1 Due: ...*
- **Module 2 -- Probability**
    - Probability
    - Conditional Probability
    - Basic Concepts of R -- Programming Constructs
    - *Assignment2 Due, Quiz2 Due: ...*
- **Module 3 -- Data Description & Visualization**
    - Univariate Data, Bivariate Data, Multivariate Data
    - Visualization using Base R
    - Using Plotly & ggplot2 for Visualization
    - *Assignment3 Due, Quiz3 Due: ...*
- **Module 4 -- Distributions**
    - Discrete Distributions
    - Continuous Distributions
    - Random Variables using R
    - *Assignment4 Due, Quiz4 Due: ...*
- **Module 5 -- Central Limit Theorem, Sampling, Dashboards**
    - Central Limit Theorem
    - Sampling & Resampling Methods
    - Errors
    - RMarkdown and Dashboards
    - *Assignment5 Due: ...*
- **Module 6 -- Strings, Data Wrangling**
    - Strings and Regular Expressions
    - Data Wrangling -- dplyr and tidyr
    - Summarizing, Grouping, and Joining Data
    - *Assignment6 Due: ...*
- **Group Project Presentations**
- **Final Exam**