MET CS544 - Foundations of Analytics

Instructor

Suresh Kalathur, Ph.D.
Assistant Professor, Computer Science Dept.
Boston Univeristy Metropolitan College
808 Commonwealth Ave, Room 250
Boston, MA 02215

E-mail: 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. 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. Data populations using discrete, continuous, and multivariate distributions are explored. Errors during measurements and computations are analyzed in the course. Confidence intervals and hypothesis testing topics are also examined. 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 (10%), quizzes (10%), assignments (30%), final project (20%), and a final exam (30%). Assignments are expected to be submitted by their respective due dates. Late submissions carry a penalty.

Course Web Site

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

References

Reference Books

- "*Using R for Introductory Statistics, 2nd edition*", by John Verzani, CRC Press, 2014. ISBN13: 978-1466590731. **(Reference book)**
- "*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

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
    - Big Data Analytics -- Apache Spark
- Module 2 -- Probability
    - Probability
    - Conditional Probability
    - Random Variables
    - Basic Concepts of R -- Programming Constructs
    - Big Data Analytics -- Apache Spark
- Module 3 -- Data Description
    - Univariate Data
    - Bivariate Data
    - Multivariate Data
- Module 4 -- Distributions
    - Discrete Distributions
    - Continuous Distributions
- Module 5 -- Central Limit Theorem, Sampling and Errors
    - Central Limit Theorem
    - Sampling Methods
    - Errors, Measurement of Errors
    - Biases, Noise, Data dredging
- Module 6 -- Estimation of Performance
    - Confidence Intervals for Population Mean
    - Hypothesis Tests, Type-I and Type-II errors
    - Resampling Methods
- **Final Project Presentations**
- **Final Exam**