MET CS544 -
Foundations of Analytics and Data Visualization

Instructor

Vijai Diwania, MS
Adjunct Professor, Computer Science Dept.
Boston University Metropolitan College
1010 Commonwealth Ave
Boston, MA 02215

Email: vijaid@bu.edu
Phone: 978-609-6336

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. It is a laboratory course.

Course Prerequisites

MET CS 546 - Introduction to Probability and Statistics, or equivalent

Course Grading Policy

The course grade will be based on class participation (10%), in-class labs (20%), in-class quizzes (20%), mid-term exam (20%), and final exam (30%).

Course Web Site

- **https://learn.bu.edu** or **https://onlinecampus.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://ipsur.r-forge.r-project.org/book/download/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 (1/21, 1/28)**
  - Introduction to Statistics
  - Basic Concepts of R -- Data Types and Structures
  - *In class labs, and quiz1*
- **Module 2 -- Probability (2/4, 2/11)**
  - Probability
  - Conditional Probability
  - Basic Concepts of R -- Programming Constructs
  - *In class labs & quiz2*
- **Module 3 -- Data Description & Visualization (2/25, 3/4)**
  - Univariate Data, Bivariate Data, Multivariate Data
  - Visualization using Base R
  - Using Plotly & ggplot2 for Visualization
  - *In class labs & quiz3*
- **Module 4 -- Distributions (3/25, 4/1)**
  - Discrete Distributions
  - Continuous Distributions
  - *In class labs & quiz4*
- **Module 5 -- Central Limit Theorem, Sampling, Dashboards (4/8, 4/15)**
  - Central Limit Theorem
  - Sampling & Resampling Methods
  - Errors
  - RMarkdown and Dashboards
  - *In class labs*
- **Module 6 -- Data Wrangling, Strings (4/22, 4/29)**
  - Data Wrangling -- dplyr and tidyr
  - Summarizing, Grouping, and Joining Data
  - Strings and Regular Expressions
  - *In class labs & quiz5*
- **Mid Term Exam (3/18)**
- **Final Exam (5/6)**