MET CS544 A4 (Fall 2022) - Foundations of Analytics and Data Visualization (Wednesdays, 6 PM)

Instructor

Hong Pan, Ph.D.
Instructor, Computer Science Dept.
Boston University Metropolitan College
Bld SOC Rm B57

Email: hongpan@bu.edu
Phone: 917-439-2996

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 assignments (20%), in-class quizzes (30%), mid-term exam (20%), and final exam (30%).

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 (9/7, 9/14, 9/21)**

  Introduction to Statistics
  Basic Concepts of R -- Data Types and Structures
  Assignment1 Due:9/28, In Class Quiz1 Due:9/21

- **Module 2 – Probability (9/21, 9/28, 10/5)**

  Probability
  Conditional Probability
  Basic Concepts of R -- Programming Constructs
  Assignment2 Due, In Class Quiz2 Due: ...

- **Module 3 -- Data Description & Visualization (10/12, 10/19)**

  Univariate Data, Bivariate Data, Multivariate Data
  Visualization using Base R
  Using Plotly & ggplot2 for Visualization
  Assignment3 Due, In Class Quiz3 Due: ...

- **Module 4 – Distributions (11/2, 11/9)**

  Discrete Distributions
  Continuous Distributions
  Assignment4 Due, In Class Quiz4 Due: ...

- **Module 5 -- Central Limit Theorem, Sampling, Dashboards (11/16, 11/30)**

  - Central Limit Theorem
  - Sampling & Resampling
    Methods Errors
  - RMarkdown and Dashboards
  - *Assignment5 Due, In Class Quiz5 Due: ...*

- **Module 6 -- Data Wrangling, Strings (12/7, 12/14)**

  - Data Wrangling -- dplyr and tidyr
  - Summarizing, Grouping, and Joining
  - Data Strings and Regular Expressions
  - *Assignment6 Due, In Class Quiz6 Due: ...*

- **Mid-term Exam (10/26)**
- **Final Exam (12/21)**