

MET CS544 A3 (Fall 2020) - Foundations of Analytics with R (Tuesdays, 6 PM)

Instructor

Suresh Kalathur, Ph.D.
Assistant Professor, Computer Science Dept.
Boston University 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. 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 (10%), quizzes (20%), assignments (20%), 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 (9/8/2020, 9/15/2020)
 - Introduction to Statistics and Probability
 - Basic Concepts of R -- Data Types and Structures
 - **Assignment1 Due: 9/21/2020, Quiz1 Due: 9/21/2020**
- Module 2 -- Probability (9/22/2020, 9/29/2020)
 - Probability
 - Conditional Probability
 - Random Variables
 - Basic Concepts of R -- Programming Constructs
 - **Assignment2 Due: 10/5/2020, Quiz2 Due: 10/5/2020**
- Module 3 -- Data Description (10/6/2020, 10/20/2020)
(No class on 10/13/2020 - Monday Schedule in Effect)
 - Univariate Data
 - Bivariate Data
 - Multivariate Data
 - **Assignment3 Due: 10/26/2020, Quiz3 Due: 10/26/2020**
- Module 4 -- Distributions (10/27/2020, 11/3/2020)
 - Discrete Distributions
 - Continuous Distributions
 - **Assignment4 Due: 11/9/2020, Quiz4 Due: 11/9/2020**
- Module 5 -- Central Limit Theorem, Sampling Techniques (11/10/2020, 11/17/2020)
 - Central Limit Theorem
 - Sampling Methods
 - Errors
 - Resampling Methods
 - **Assignment5 Due: 11/23/2020**
- Module 6 -- Strings, Data Wrangling (11/24/2020, 12/1/2020)
 - Strings and Regular Expressions
 - Data Wrangling -- dplyr and tidyr
 - **Assignment6 Due: 12/7/2020**
- **Final Project Presentations (12/8/2020)**
- **Final Exam (12/15/2020)**