UMBC

AN HONORS UNIVERSITY IN MARYLAND

# Department of Computer Science and Electrical Engineering

## CMSC 462: Introduction to Data Science

**Spring 2023**

**Study guide – Test 1 (Thursday March 2)**

The Test-1 will consist of mostly short answer type questions. **Closed book, close note. Some problems will require analysis of given R output. Please read the relevant sections from the book.**

1. **Introduction to Data Science**
   - **a.** Data Science – definition.
   - **b.** Big data and four Vs
   - **c.** Data, Information and Knowledge

2. **Introduction to Statistics**
   - **a.** Introduction to statistics and different types of data and how to measure them.
   - **b.** Distribution-
     i. Binomial ii. Poisson iii. Normal
     Given data you should be able to describe how to fit any distribution. You need not memorize the formulas; however, given the formula you should be able to fill in the numbers.
   - **c.** Sampling – what it is has to do with sample size? How to calculate Margin of Error?

3. **Introduction to R:**
   - **a.** R Data structure in detail
   - **b.** DataFrame manipulation

4. **Prediction**
   - **a. Linear regression.**
     - **i.** A complete understanding of different terms in the output
       1. $R^2$, RSE, Coefficient estimates, t statistics, p-values and their meaning
       2. Using the output from a linear regression to make predictions
       3. Predicting individual values or mean response with confidence intervals
   - **b. Multiple regression**
     - i. A complete understanding of different terms in the output
       1. $R^2$, RSE, Coefficient estimates, F statistics, p-values and their meaning
       2. How to interpret the effect of different predictor variables based on their coefficient estimate?

5. **Classification –**
   - i. Naïve Bayes Algorithm