# Assignment 1: Probability and Statistics with R

**Due Date: 28th March,**

**11:55PM**

**General Instructions:**

1) All submissions have to be individual and will be checked through Turnitin for plagiarism.
2) Coding Scheme for this assignment is '2N-•-b'. Honor Code violations are investigated accordingly.
3) Please remember to include your name and ID at the top of the assignment.
4) Describe your solutions in detail (just mentioning the answer is not enough).
5) In your answers, please clearly indicate the question number and sub-•-question number (if any).
6) There are 12 Questions. Attempt all Questions.
7) All the 12 questions will be submitted in one file, with all the codes and outputs clearly mentioned multiple files will not be accepted for evaluation.
8) You would need R markdown to do the assignment and then you would need to convert into HTML/PDF and then submit.

1. If a random variable has the normal distribution with $\mu = 77.0$ and $\sigma = 3.4$, find the probability that random variable is (a) less than 72.6
(b) greater than 88.5
(c) between 81 and84 (d) between 56 and92

2. In a car race, the finishing times are normally distributed with mean 145 minutes and standard deviation of 12 minutes.
(a) Find percentage of car racers whose finish time is between 130 and 160 minutes.
(b) Find percentage of car racers whose finish time is less than 130 minutes.

3. A test-•-taker has recently taken an aptitude test with 15 questions. All of the questions are True-•-False type in nature.
(a) What is the probability that the student got first five questions correct.
(b) What is the probability that the student got five questions correct.

4. 68% of the marks in exam are between 35 and 42. Assuming data is normally distributed, what are the mean and standard deviation?

5. A professor asked students to take two exams. 30% of the class cleared both exams and 55% of the class cleared the first exam. What percentage of class who cleared first exam also cleared the second exam?

6. In India, 82% of all urban homes have a TV. 62% have a TV and DVD player. What is probability that a home has a DVD player given that the home has a TV.

7. You toss a coin three times. Assume that the coin is fair. What is the probability of getting:

    (a) All three heads

    (b) Exactly one head

    (c) Given that you have seen exactly one head, what is the probability of getting at-•-least two heads?

8. A small insurance agency has two sales person who sell policies to retail clients. The amount of insurance claims filed by clients served by first agent in a year(X1) can be approximated using a normal distribution with mean $\mu_1$= INR 1200 and variance = 90000. Similarly, the amount of claims filed by clients served by the second agent(X2) can also be approximated using a normal distribution with mean $\mu_2$= INR 1800 and variance = 160000. What is the probability that the total amount of claims filed by the second agent's clients is lower? Assume that X1 and X2 are independent of each another.
    (a) 1.15%

(b) 88.5%

(c) 11.5%

(d) 98.85% (e) 50%

9. The blue M&M was introduced in 1995. Before then, the color mix in a bag of plain M&Ms was (30% Brown, 20% Yellow, 20% Red, 10% Green, 10% Orange, 10% Tan). Afterward it was (24% Blue, 20% Green, 16% Orange, 14% Yellow, 13% Red, 13% Brown). A friend of mine has two bags of M&Ms, and he tells me that one is from 1994 and one from 1996. He won't tell me which is which, but he gives me one M&M from each bag. One is yellow and one is green. What is the probability that the yellow M&M came from the 1994 bag?

10.    Find the daily stock price of Wal-•- Mart for the last three months. (A good source for the data is http://moneycentral.msn.com or Yahoo Finance or Google Finance (there are many more such sources). You can ask for the three-•-month chart and export the data to a spreadsheet.) (a) Calculate the mean and the standard deviation of the stock prices.
(b) Get the corresponding data for Kmart and calculate the mean and the standard deviation.
(c) The coefficient of variation (CV)is defined as the ratio of the standard deviation over the mean. Calculate the CV of Wal-•-Mart and Kmart stock prices.
(d) If the CV of the daily stock prices is taken as an indicator of risk of the stock, how do Wal-•-Mart and Kmart stocks compare in terms of risk? (There are better measures of risk, but we will use CV in this exercise.)
(e) Get the corresponding data of the Dow Jones Industrial Average (DJIA) and compute its CV. How do Wal-•- Mart and Kmart stocks compare with the DJIA in terms of risk?
(f) Suppose you bought 100 shares of Wal-•-Mart stock three months ago and held it. What are the mean and the standard deviation of the daily market price of your holding for the three months?

*Note: Question referenced from Aczel A., Sounderpandian J., Complete Business Statistics (7ed.)*

11.     For this problem, consider the dataset Prob_Assignment_Dataset.xlsx attached. As the vigilant monitors of the Zappos.Com website, we are obsessed with who is coming to our website and what they do during their visit. This challenge requires you to look at data similar to that which Analytics teams at Zappos would use to assess the overall performance of the business.

The data set is a fictional representation of actual data sets that we work with. Perform unii variate descriptive analysis on the variables, and Submit a short description explaining any insights or trends you discovered. Include graphs, data tables, and other helpful visuals to communicate your discoveries. You do not have to work on all of the variables (select any 3-•- 4 that are of most interest to you).

12.     Consider the same dataset as earlier. Now perform bi-•-variate data analysis as discussed in last class to find out relationships between different variables. Write a short description on what you find in the analyses along with any tables, graphs. You do not have to analyze all variables, any 3-•-4 that interest you most should be the focus here.