

# MSDS 6306: Introduction to Data Science

## Live session Unit 01 assignment

**Due: 1 hour before the live session for Unit 02**

Calculation Questions (using R) - Use R to calculate the following:

**NOTE:** Submit a word file with the R codes in text format and its screenshot that shows the result.

1. (20 points) Basic Math – each question has 5% credits
  - a. The log of a positive number.
  - b. What is the default base for the log function? Calculate the log of your previous number with a different base.
  - c. The log of a negative number. (explain the answer)
  - d. The square-root of a positive number.
2. (15 points) Random number generation.
  - a. Create a vector of 15 standard normal random variables. Calculate its mean and SD (Standard Deviation).
  - b. Change the mean to 10 and the SD to 2 and recalculate the vector of 15 random normal variables. Calculate its mean and SD.
  - c. Why are the means and SD not exactly the same as the means and SDs specified in the function?
3. (40 points) Vector Operations
  - a. The weights of 6 individuals in kg are 60, 72, 57, 90, 95, 72.
  - b. Their heights (in m) are 1.80, 1.85, 1.72, 1.90, 1.74, 1.91.
  - c. Enter these vectors into R.
  - d. Create a scatterplot of weight vs. height. Interpret the scatterplot.
  - e. Calculate the BMI for each individual. (BMI = weight in kg divided by the square of the height in m)
  - f. Calculate the mean for weight.
  - g. Subtract the mean from each value of weight.

- h.** Sum the result. Now you know why we square the deviations from the mean to calculate a standard deviation!
- 4.** (25 points) Your data science profile. Enter your data science profile into R as a data frame with two columns. Call it by your first name. The categories are computer programming, math, statistics, machine learning, domain expertise, communication and presentation skills, and data visualization. Your ranking for each category 1-5, with 5 as best. Create a bar graph of your data science profile. When you submit your work, please submit your code, including the data entry piece.
- 5.** (50 points) Install the swirl package and complete modules 1-7 (see the document RSwirlUnit01.docx for instructions). Please upload an R script (or txt file) containing your code from modules 1-7.