

MSDS 6306: Introduction to Data Science (401/402)

Live Session Assignment 01

Due: September 6th, 2017 (Wednesday)

Use R to calculate the following. Please submit code and cleaned output on the Learning Platform BEFORE NEXT SESSION. Be sure to COMMENT which questions are which before giving an answer. Make your graders happy people.

1. Basic Math

- 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. Random number generation.

- a. Create a vector of 15 standard normal random variables. Calculate its mean and SD.
- 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. 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. Your data science profile. Enter your data science profile into R as a data frame with two columns. Assign it to 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. Install swirl package and complete modules 1-7. Please upload a R script (or txt file) containing your code from modules 1-7.