**MSDS 6306: Doing Data Science**

# Live session Unit 01 assignment

**Due: Monday 11:59pm Jan 14th .**

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

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

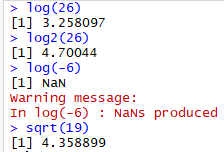
1. (20 points) Basic Math – each question has 5% credits
   1. The log of a positive number.
   2. What is the default base for the log function? Calculate the log of your previous number with a different base.

The default base for the log function is 10.

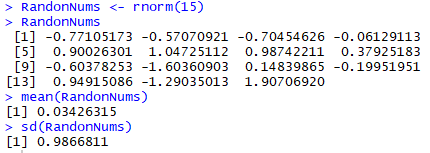
* 1. The log of a negative number. (explain the answer)

The answer was not available because the log of a negative number is undefined.

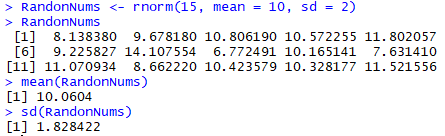
* 1. The square-root of a positive number.



1. (15 points) Random number generation.
   1. Create a vector of 15 standard normal random variables. Calculate its mean and SD (Standard Deviation).



* 1. Change the mean to 10 and the SD to 2 and recalculate the vector of 15 random normal variables. Calculate its mean and SD.



* 1. Why are the means and SD not exactly the same as the means and SDs specified in the function?

The mean and SD are not exactly the same as those specified because rnorm() generates a list of random numbers that should approximate the parameters specified. The larger n gets the closer the mean and SD should be to the specified values.

1. (40 points) Vector Operations
   1. The weights of 6 individuals in kg are 60, 72, 57, 90, 95, 72.

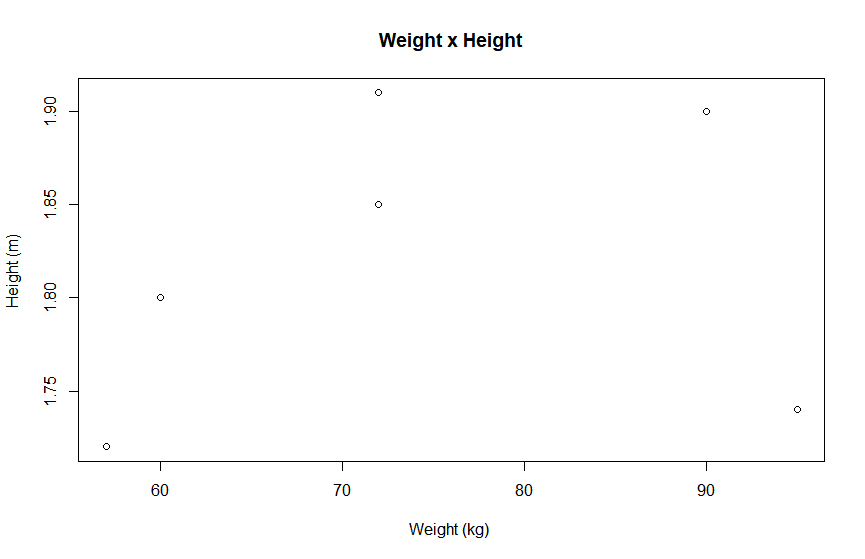
**b.** Their heights (in m) are 1.80, 1.85, 1.72, 1.90, 1.74, 1.91.

1. Enter these vectors into R.



1. Create a scatterplot of weight vs. height. Interpret the scatterplot.





Generally, as weight increases, height increases.

1. Calculate the BMI for each individual. (BMI = weight in kg divided by the square of the height in m)



1. Calculate the mean for weight.



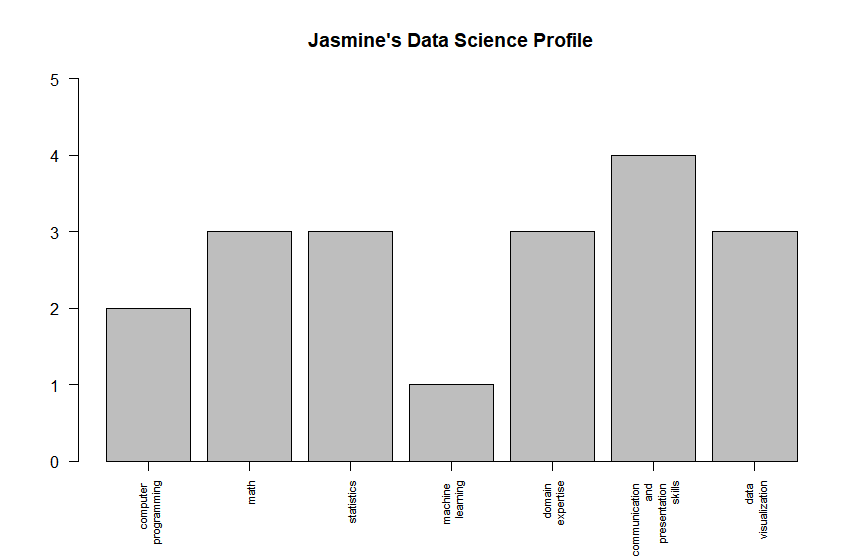
1. Subtract the mean from each value of weight.

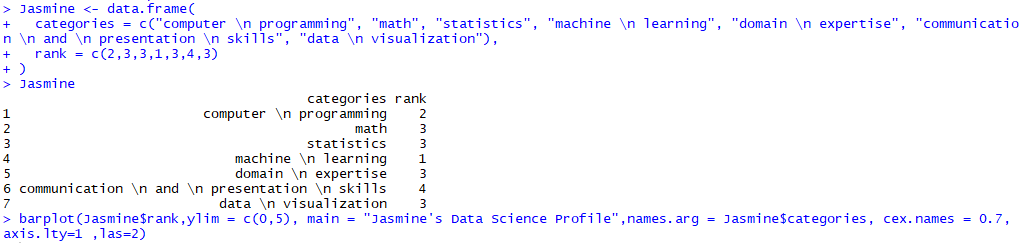


1. Sum the result. Now you know why we square the deviations from the mean to calculate a standard deviation!



1. (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.





Appendix

# Question 1 (20pts)

log(26)

log2(26)

log(-6)

sqrt(19)

# Question 2 (15pts)

RandonNums <- rnorm(15)

RandonNums

mean(RandonNums)

sd(RandonNums)

RandonNums <- rnorm(15, mean = 10, sd = 2)

RandonNums

mean(RandonNums)

sd(RandonNums)

# Question 3 (40pts)

kg <- c(60,72,57,90,95,72)

height <- c(1.80,1.85,1.72,1.90,1.74,1.91)

plot(kg, height, main="Weight x Height",

xlab="Weight (kg) ", ylab="Height (m) ")

BMI <- kg/(height)^2

BMI

kgMean <- mean(kg)

kgMean

kgLessMean <- kg-kgMean

kgLessMean

sum(kgLessMean)

# Question 4 (25pts)

Jasmine <- data.frame(

categories = c("computer \n programming", "math", "statistics", "machine \n learning", "domain \n expertise", "communication \n and \n presentation \n skills", "data \n visualization"),

rank = c(2,3,3,1,3,4,3)

)

Jasmine

barplot(Jasmine$rank,ylim = c(0,5), main = "Jasmine's Data Science Profile",names.arg = Jasmine$categories, cex.names = 0.7, axis.lty=1 ,las=2)