MSDS 6306: Doing Data Science

Live session Unit 01 assignment

Max Moro

August 27, 2018

## 1. (20 points) Basic Math – each question has 5% credits

1. The log of a positive number.

log(10)



1. What is the default base for the log function? Calculate the log of your previous number with a different base.

**The log function has a natural base (e)**

log10(10)



**log10 has base 10**

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

log(-10)



**log of a negative number is undefined, so R returns NaN (undefined, or ‘Not a Number’)**

1. The square-root of a positive number.

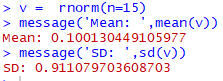
sqrt(1701)



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

v = rnorm(n=15)  
message('Mean: ',mean(v))

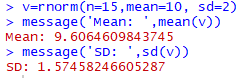
message('SD: ',sd(v))



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.

v=rnorm(n=15,mean=10, sd=2)  
message('Mean: ',mean(v))

message('SD: ',sd(v))



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

**Because the *rnorm* function creates random values from a distribution with the specified SD and Mean. If the number of observations (n) is small, the result will not converge to the specified SD and mean, a larger n will get closer to the specified values.**

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

w = c(60, 72, 57, 90, 95, 72)  
print(w)



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

h=c(1.80, 1.85, 1.72, 1.90, 1.74, 1.91)  
print(h)

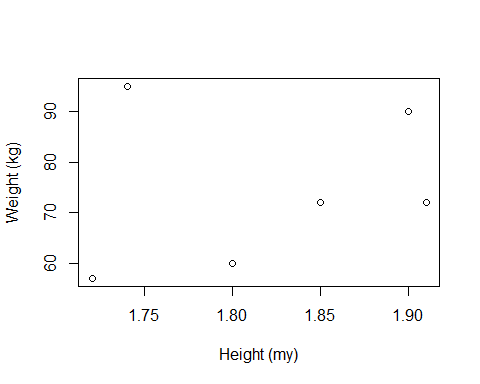


1. Enter these vectors into R.



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

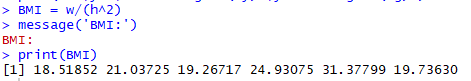
plot(x=h,y=w,xlab='Height (my)',ylab='Weight (kg)')



**From the plot we can see that, generally, as the the Height increases, the Weight increases. We have somes outlier on this correlation.**

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

BMI = w/(h^2)  
message('BMI:')  
print(BMI)



1. Calculate the mean for weight.

message('Mean of weight: ',mean(w))



1. Subtract the mean from each value of weight

weightDiff = w-(mean(w))  
print(weightDiff)



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

sum(weightDiff)

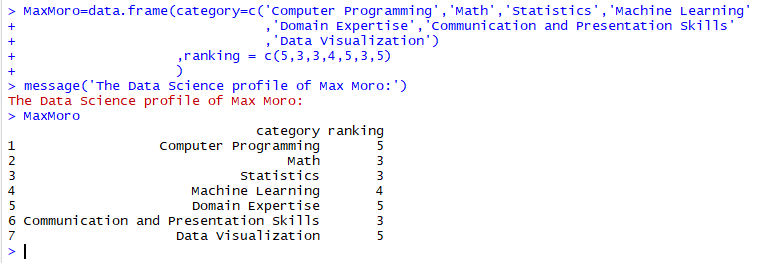


**Yep.. it is Zero, we need to square to calculate the SD**

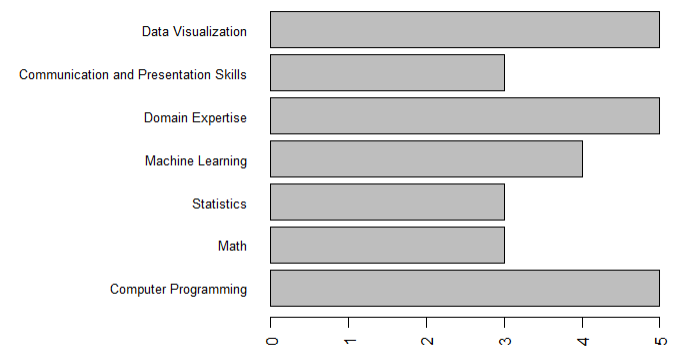
1. (25 points) Your data science profile. Enter your data science profile into R as a data frame with two columns. Call it by your first name. The categories are computer programming, math, statistics, machine learning, domain expertise, communication and presentation skills, and data visualization. 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 submityour code, including the data entry piece.

MaxMoro=data.frame(category=c('Computer Programming','Math','Statistics','Machine Learning','Domain Expertise','Communication and Presentation Skills','Data Visualization')  
 ,ranking = c(5,3,3,4,5,3,5)  
 )  
message('The Data Science profile of Max Moro:')

print(MaxMoro)



par(mar=c(5,15,1,0))  
barplot(height=MaxMoro$ranking,names=MaxMoro$category,horiz=T,las=2,cex.names=.8)



Or vertical

