David Nguyen September 4, 2018

MSDS 6306: Doing Data Science

Live Session Unit 1 Assignment

1.

log(10)

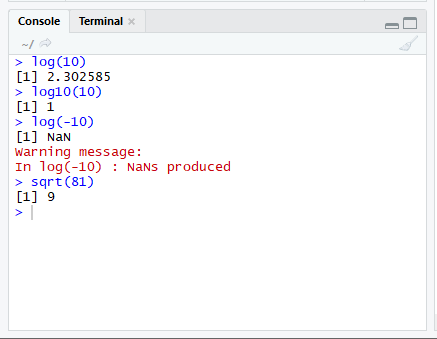
log10(10)

log(-10)

sqrt(81)

The default base for the log function is e.

The log of a negative number is undefined. NaN stands for “Not a Number”.



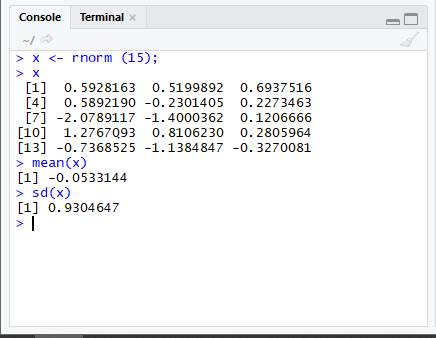
2.

x <- rnorm(15)

x

mean(x)

sd(x)

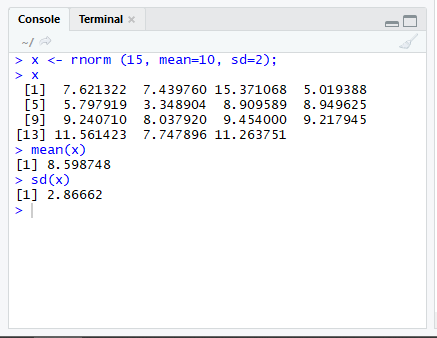


x <- rnorm(15, mean=10, sd=2)

x

mean(x)

sd(x)



The means and the SD are not exactly the same as the means and SDs specified in the function because we are only taking a sample of size 15 and the Central Limit Theorem states that the sample size must be approximately 30 or more to reach the normal distribution.

3.

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

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

weight

plot(weight, height, main= "Weight vs. Height Conpanson", xlab="Weight", ylab= "Height", col="blue")

bmi <- weight/ height\*\*2

bmi

mean(weight)

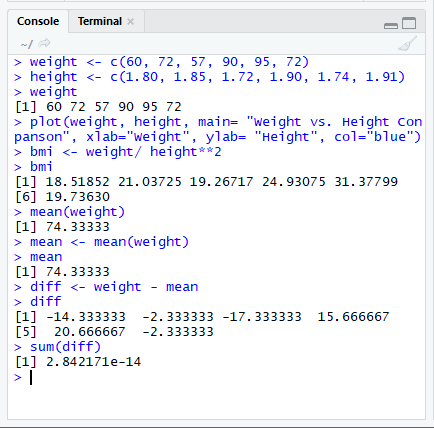
mean <- mean(weight)

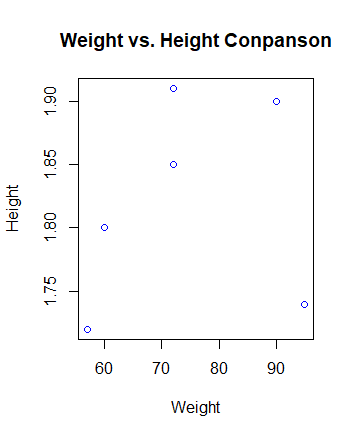
mean

diff <- weight - mean

diff

sum(diff)





As the weight increases, so does the height except for one outlier on the bottom right of the graph.

4.

Category <- c("Computer Programming", "Math", "Statistics", "Machine Learning", "Domain Expertise", "Communication and Presentation Skills", "Data Visualization")

Ranking <- c(1, 5, 4, 1, 2, 5, 4)

David <- data.frame(Category, Ranking)

David

library(ggplot2)

ggplot(data=David, aes(x = Category, y = Ranking))+ geom\_bar(stat='identity')

