**DS 501: STATISTICAL AND MATHEMATICAL METHODS FOR DATA SCIENCE**

Fall 2019

Assignment 1

Due: Tuesday 2nd September, 2019.

**Objective:** Get familiar with R Language, random numbers and frequency distributions, review logarithms

**NOTE:** Fit your answer in the provided space. DO NOT attach extra sheets.

**TODO: Build your background**

1. Install R and R studio
2. Do simple computations at the command prompt, e.g., 2+10, log(10) etc.
3. Read about runif and rnorm commands in R
4. Try these commands at the command prompt: runif(1), runif(10), rnorm(1), rnorm(10) and note that using the parameter 10 gives us 10 random values which in R is stored as a vector, which you can think of as an array. You can store the values in a variable, e.g., x = runif(10) and then you can see any value at a particular index, e.g., x[10]
5. Read about frequency distribution of a set of numbers.
6. Read about the seed of a random number generator. Now try the the commands in step 4 as:

a. set the seed to 10

b. generate 10 random numbers

c. set the seed again to 10

d. generate 10 random numbers

Notice that setting a seed to the same number generates the same set of numbers and that is why it is called a pseudo number generator.

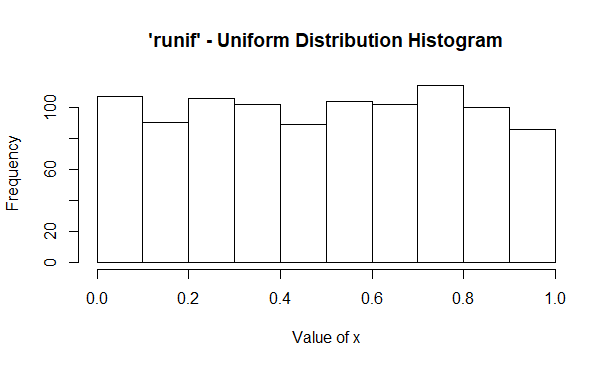
**PROBLEM 1**

1. Set the seed equal to a number which is your roll number. For example 19L1234 would be 191234.
2. Generate 1000 random numbers from the uniform distribution and store in variable x
3. Write the values of the random numbers and make the histogram of these numbers using the command hist(x). Show the histogram you get and do not forget to label each axis:

x[1] = 0.1816757

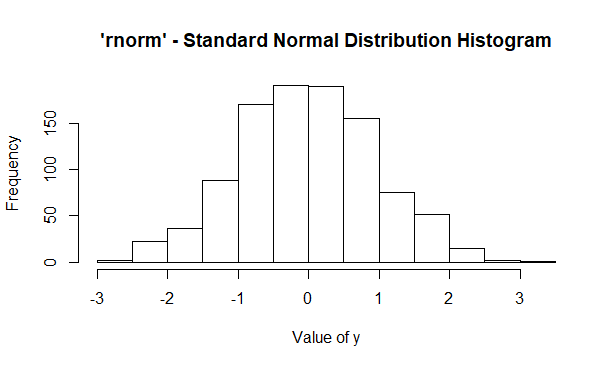
x[1000] = 0.1817416

x[99] = 0.3085808



**PROBLEM 2**

1. Set the seed equal to a number which is your roll number. For example 19L1234 would be 191234.
2. Generate 1000 random numbers from the **standard normal** distribution and store in variable y
3. Write the values of the random numbers and make the histogram of these numbers using the command hist(y). Show the histogram you get and do not forget to label each axis:



y[1] = -0.9089976

y[1000] = 1.395009

y[99] = 0.5670594

**PROBLEM 3**

Write your comments about the difference between the two histograms of problem 1 and problem 2.

Problem 1 – ‘runif’ Uniform Distribution

Problem 2 – ‘rnorm’ Standard Normal Distribution

The height of the bars represent the frequency or the number of occurrences.

Problem 1 being a uniform distribution, shows that it has a finite set of possible values. It has a constant frequency for all the finite values. It has approximately the same number of values in each group or class (represented by a bar).

Problem 2 being a standard normal distribution, shows the distribution is cut in half, with each side being a mirror of the other. It forms a bell-shaped curve to be normal. As we go further from the center, there are fewer and fewer values. The center is in and around the value ‘0’. The tallest bars are around this value.

**PROBLEM 4**: HANDWRITTEN ANSWER ONLY

Prove that log(ab) = log(a)+log(b)