**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.

Frequency distribution is

1. Read about the seed of a random number generator. Now try the commands in step 4 as:

Seed is actually the starting point during random number generation. It is usually used for debugging since if use seed to initialize the random number generation it will generate the same sequence of random numbers

a. set the seed to 10

b. generate 10 random numbers

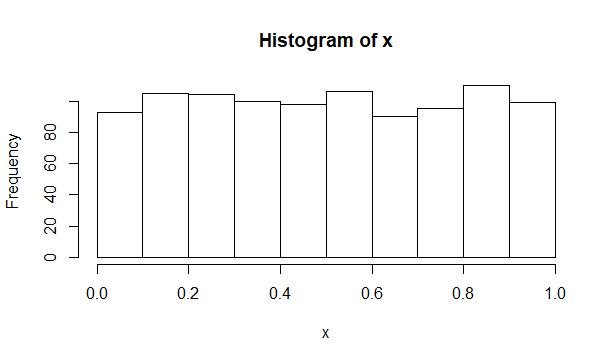
c. set the seed again to 10

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. Note that once you set
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:



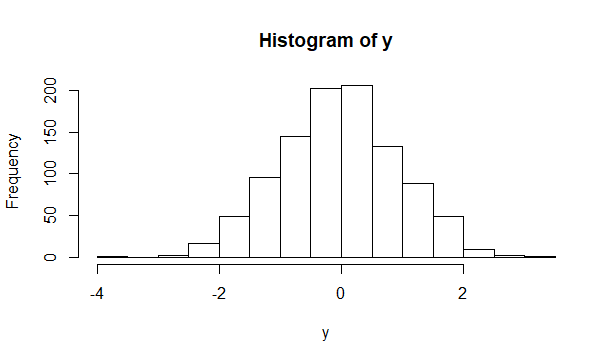
x[1] = 0.3510284

x[1000] = 0.0923837

x[99] = 0.181967

**PROBLEM 2**

1. Set the seed equal to a number which is your roll number. For example 19L1234 would be 191234. Note that once you set
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] = -0.3825454

y[1000] = -0.9342284

y[99] = 0.1275395

**PROBLEM 3**

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

1. Uniform distribution generated the positive values between 0 and 1 while Normal Distribution has both negative and positive values with no range.
2. Uniform distribution has uniform frequencies and all are higher than 80. On the other hand, frequencies are higher in the middle and lower when you move towards the lower or higher end in normal distribution.
3. Normal distribution assumes a “bell shaped” frequency. Uniform distribution a flat shaped picture.

**PROBLEM 4**: HANDWRITTEN ANSWER ONLY

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

**Here's the proof:**

**Let log(ab) = p, log(a) = q, and log(b) = r.**

**Then what we want to prove is that p = q+r**

**Those three logarithm equations by definition of logarithms are equivalent**

**respectively to these three exponential equations:**

**10p = ab, 10q = a, 10r = b**

**Multiply the last two equation (multiply equals by equals):**

**10q10r = ab**

**Multiply on the left by adding exponents of 10:**

**10q+r = ab**

**But also 10p = ab, so**

**10p = 10q+r**

**The bases are the same positive number other than 1, 10.**

**Therefore the exponents must be equal:**

**p = q+r**

**log(ab)= log(a) + log(b)**

**Edwin**