**For this assignment you have to submit the R script files.**

**Q1.** Write a function, called myFunction(), which takes a single argument n, an implements the following:

1. Simulate n independent numbers, denoted by x = x1, x2, x3,.. xn  from the N(0,1) distribution.
2. Calculate the mean, xmean.
3. If xmean >= 0, then simulate n independent numbers denoted by y = y1, y2, y3,.. yn from the exponential density with the mean xmean.
4. If xmean < 0, then simulate n independent numbers denoted by z = z1, z2, z3,.. zn from the exponential density with the mean -xmean.
5. Set y = -z
6. Calculate k which is the number of j with  >
7. Return the list of x, y and k with the names xVec, yVec and count respectively.

**Q2.** Write a function colMeans which will accept a data frame as an input and find out the mean of the data of each column of the dataframe. If the column is a non-numeric field, then the function must display an appropriate message. Use this function on the cricData dataframe that we have created in class.

**Q3.** Write a function primefn() which will accept a integer parameter, n and return the first n prime numbers.

**Q4.** Read the contents of the yearly\_sales files into a dataframe.

Find out the number of records.

Extract the third column of the data frame.

Extract the gender column of the data frame

Retrieve the 1st and the 2nd rows of the data frame

Retrieve the first, third and the fourth columns

Retrieve all the records whose gender is female.

Categorize sales\_totals into three groups, small (<100), medium (100 – 500) and big (>500).

Create a new ordered factor with levels, small, medium and big.

Add the ordered factor to the data frame.

Look the summary of the data frame using the summary function.

Find the following:

1. Correlation between sales\_total and num\_of\_orders
2. Covariance between sales\_total and num\_of\_orders
3. Mean of sales\_total
4. Median of sales\_total
5. Range of sales\_total
6. Standard deviation and variance of sales\_total
7. Find out the standard deviation of the first three columns of the data frame using only one function.

**Q5.** Download and install the “vcd” package. Use the Arthritis data set for the following problems

Draw a vertical bar plot and a horizontal bar plot on the number of “None”, “Some” and “Marked”. Give a title and label the axes.

Consider the cross tabulation of treatment type and improvement status.

Draw the graph as a stacked bar plot and a grouped bar plot.

Draw bar plots for sorted mean and median values.

Draw a spinogram.

**Q6.** Create a vector of values – 10, 12, 4, 16, 8. Let us assume that these are the number of gold medals won by the countries US, UK, Australia, China and France in a particular competition.

Draw a simple pie chart, pie chart with percentages, a 3D pie chart based on the above.

Draw a fan plot based on the above data.

**Q7.** Load the mtcars data

Draw a simple histogram.

Draw another histogram with 12 bins of red color.

Draw another histogram with rug plot and frame.

Draw the fourth histogram with normal curve.

Draw a kernel density plot

Draw a box plot, and then draw a parallel box plot to compare these three groups: four, six and eight cylinders.

**Q8.** Refer to the BankSalaries.csv file. Read the data in a dataframe. Write a function which will accept the dataframe and return the following analysis in the format shown below:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Gender** |  |  |
| **JobGrade** | **Female** | **Male** | **Total** |
| **1** | 34.29% | 17.65% | 28.85% |
| **2** | 20.71% | 19.12% | 20.19% |
| **3** | 25.71% | 10.29% | 20.67% |
| **4** | 12.14% | 16.18% | 13.46% |
| **5** | 6.43% | 17.65% | 10.10% |
| **6** | 0.71% | 19.12% | 6.73% |