SAMBANDH BHUSAN DHAL

UIN 726006080

# ISEN 613 ( ASSIGNEMENT 1)

**Q 1.1 Generate a vector with 25 elements and each element independently follows a normal**

**distribution (with mean =0 and sd=1)**

```
x<-c(rnorm(25, mean = 0, sd = 1))
```

```
Output:
[1]  0.19287275 -1.01499964  0.28723651  1.04068322  0.34792235 -2.12305180
0.96467543 -0.87301824 -0.65500916
[10] -1.09110768  1.88434731  0.11455972 -0.54075801 -0.86593808 -0.70534078
1.59515509 -0.37671263  0.56921881
[19]  0.81733599  0.30365294  0.71636130  0.19381093  0.06738426  0.19505114
1.55870778

> mean(x)
[1] 0.2017367

> sd(x)
[1] 0.9125052
```

**Q1.2 Reshape this vector into a 5 by 5 matrix in two ways (arranged by row and column)**

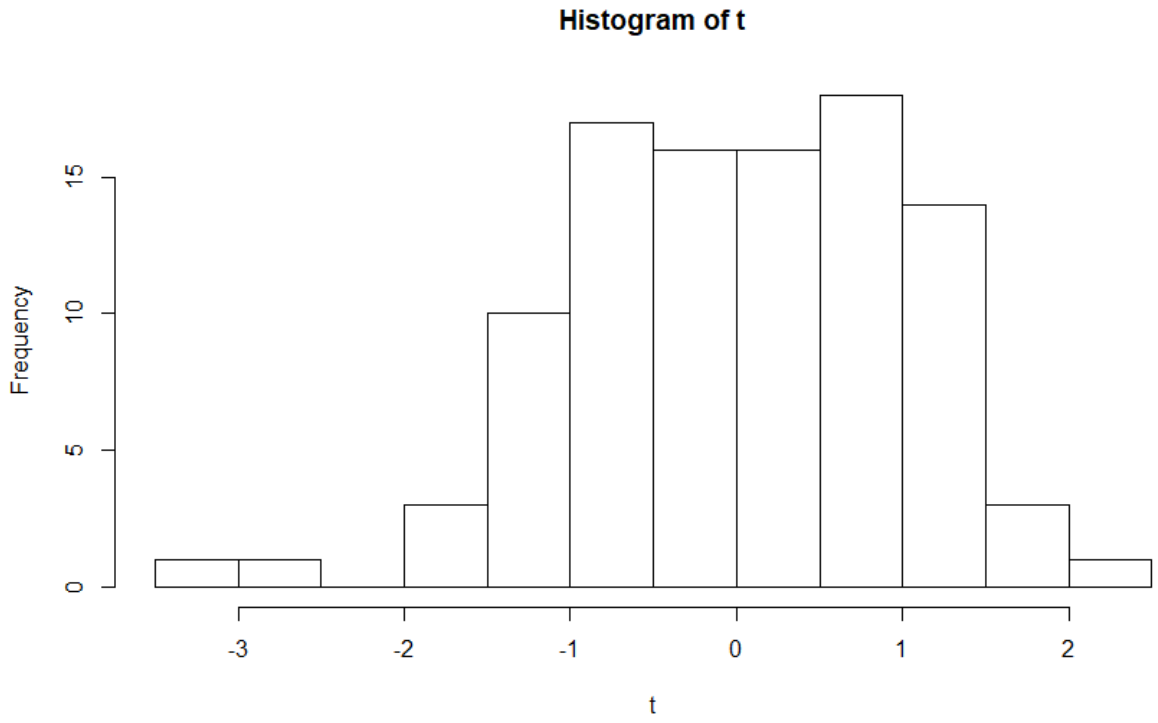
```
>  y=matrix(x,nrow=5,ncol=5,byrow=TRUE)
>  y
           [,1]       [,2]      [,3]        [,4]       [,5]
[1,]  0.09380462  0.1830907 1.3818671  1.21684909 -0.7462179
[2,] -0.33681489 -0.3427827 1.2072827 -1.63290055  1.1145446
[3,] -0.05550044  0.1261510 1.1343358 -0.76292820  2.2549773
[4,]  0.47438153 -0.1787012 0.6366081 -0.03389073 -1.5921784
[5,]  0.09855055  0.4174953 0.1572065  0.82439342 -0.5962048

>  y=matrix(x,nrow=5,ncol=5,byrow=FALSE)
>  y
           [,1]       [,2]        [,3]        [,4]        [,5]
[1,]  0.09380462 -0.3368149 -0.05550044  0.47438153  0.09855055
[2,]  0.18309075 -0.3427827  0.12615097 -0.17870118  0.41749530
[3,]  1.38186714  1.2072827  1.13433580  0.63660812  0.15720648
[4,]  1.21684909 -1.6329005 -0.76292820 -0.03389073  0.82439342
[5,] -0.74621795  1.1145446  2.25497732 -1.59217836 -0.59620483
```

**Q 1.3 Similarly, generate another vector with 100 elements and plot its histogram.**

```
> t<-c(rnorm(100,mean=0,sd=1))
> t
  [1] -0.08512105 -1.27910164 -0.69216111 -1.49637624  0.97970061  1.77820037
-0.79685979  0.60505197 -0.03501026
 [10]  0.62918767  0.38044158  0.95765096  0.91799539 -0.12980757  0.41836793
-0.82753721 -1.37435116  1.20477123
```

```
 [19]  1.01086660 -1.21992856  0.28698040 -0.29360693 -0.70866044  1.44220347
-1.18687251  0.81814933 -0.80592018
 [28] -0.78006552 -1.22728735 -2.66099859  1.19497279  1.34966395 -0.50795188
0.04172783 -0.24219434 -0.32913079
 [37]  0.30502130  0.81970188 -1.92985079 -0.03259253  0.97776205  1.05808542
0.17956343  0.76004982 -0.84614757
 [46]  0.91101118 -0.79271607 -0.25829383  0.37723523 -0.62151798  1.39323525
1.30697037  0.73810409 -0.87995810
 [55] -1.27178622 -1.37526009 -0.56574386 -1.21878735 -0.86168213  1.20238148
-1.91298433  0.36682526 -0.09247434
 [64] -0.47915311  1.61683050 -0.04272383  0.38623010  1.99661717  0.56874419
-0.55916447 -0.63576518  0.30460933
 [73]  1.39664236  0.43292899  0.51453268 -0.48958569  1.17250351  0.34097847
0.50641831 -0.09107485  0.97040679
 [82] -0.89286184  0.40228540  1.33453389 -0.02817895  0.12446199 -0.31901952
-3.00148297  0.02794131  0.27565147
 [91]  0.94763174 -0.92511963  0.79208648  1.17349331 -1.34521994 -0.09384127
1.44204939  2.30018540  0.95228366
[100] -1.62251632
```

```
> hist(t)
```




**Q 1.4 Provide screenshots of the R code used for the above questions as well as the plots in the report. Explain the plots in your own words.**

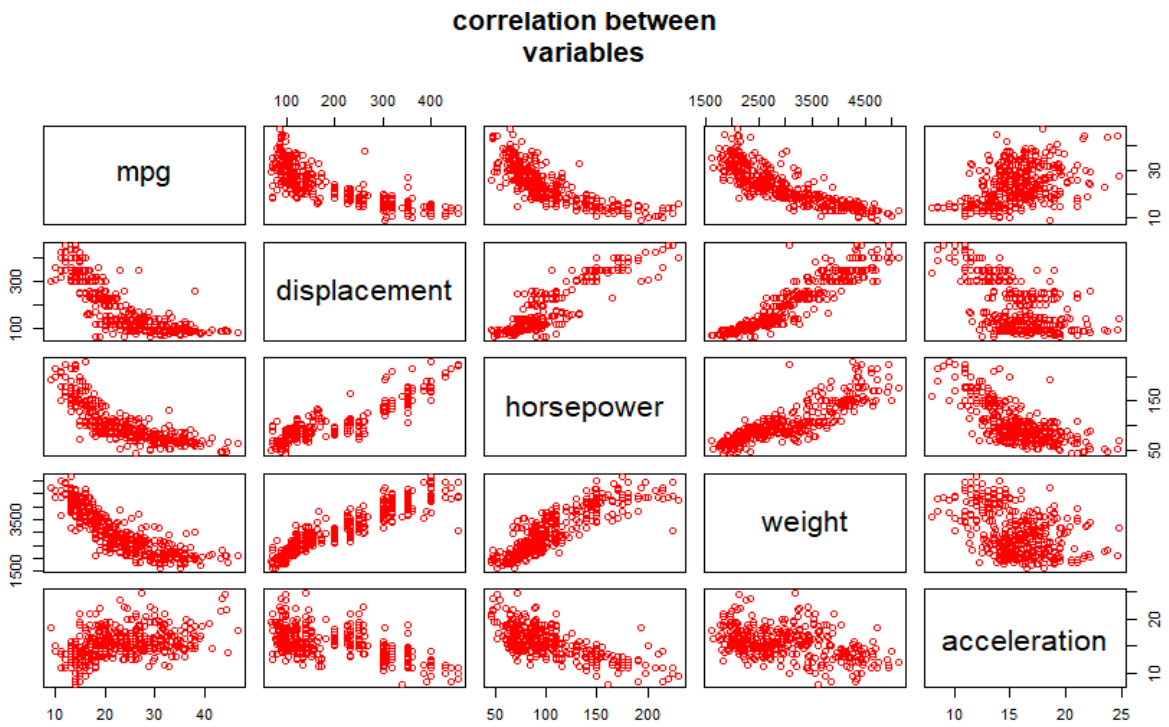
I have already attached the screenshots and the commands used to run each one of them.
Some of the inferences which can be drawn from the histogram is:-
1.The frequency of the numbers lying between 0.5 to 1 is maximum which is more than 15 followed by the number s ranging from -0.5 to -1.
2.There is no number lying between -2.5 to -2 and its frequency is 0.

**Q 2. Upload the Auto data set, which is in the ISLR library. Understand information about this data set by either ways we introduced in class (like "?Auto" and names(Auto))**

```
> library(ISLR)
> fix(Auto)
> dim(Auto)
[1] 392   9

> summary(Auto)
      mpg          cylinders      displacement     horsepower        weight
acceleration        year
 Min.   : 9.00   Min.   :3.000   Min.   : 68.0   Min.   : 46.0   Min.   :1613
Min.   : 8.00   Min.   :70.00
 1st Qu.:17.00   1st Qu.:4.000   1st Qu.:105.0   1st Qu.: 75.0   1st Qu.:2225
1st Qu.:13.78   1st Qu.:73.00
 Median :22.75   Median :4.000   Median :151.0   Median : 93.5   Median :2804
Median :15.50   Median :76.00
 Mean   :23.45   Mean   :5.472   Mean   :194.4   Mean   :104.5   Mean   :2978
Mean   :15.54   Mean   :75.98
 3rd Qu.:29.00   3rd Qu.:8.000   3rd Qu.:275.8   3rd Qu.:126.0   3rd Qu.:3615
3rd Qu.:17.02   3rd Qu.:79.00
 Max.   :46.60   Max.   :8.000   Max.   :455.0   Max.   :230.0   Max.   :5140
Max.   :24.80   Max.   :82.00


     origin                      name
 Min.   :1.000   amc matador       :  5
 1st Qu.:1.000   ford pinto        :  5
 Median :1.000   toyota corolla    :  5
 Mean   :1.577   amc gremlin       :  4
 3rd Qu.:2.000   amc hornet        :  4
 Max.   :3.000   chevrolet chevette:  4
                 (Other)           :365
```



```
> pairs(Auto)
> attach(Auto)
> hist(mpg)
> pairs(~mpg + displacement + horsepower + weight + acceleration, main= "corr
elation between
+ variables", col = "red")
```

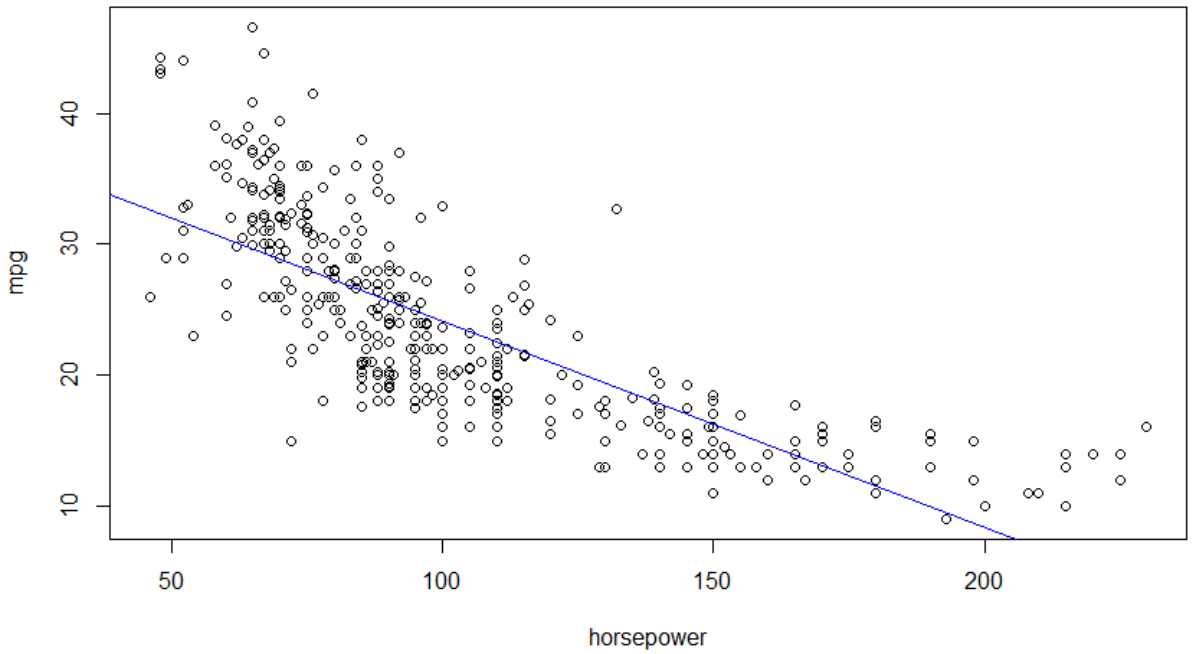



The correlation between the variables are as follows:

1. mpg and displacement have a negative correlation among them. They are inversely proportional to each other.
2. mpg and horsepower have the same negative correlation among them.
3. mpg and weight even share the same negative correlation.
4. mpg and acceleration do not have any correlation whatsoever among them.
5. Displacement, horsepower and weight have a positive correlation among them. When one increases, the other factor increases and vice-versa.
6. Displacement and acceleration however do not have any correlation among them as we can see from the graph.
7. Horsepower and weight have positive correlation among them where as horsepower and acceleration have negative correlation among them.
8. Weight and acceleration however do not have any correlation among them.

**Q 4. Draw a line on the scatterplot of mpg vs. horsepower to represent relationship between the two variables**.

```
> plot(horsepower,mpg)
> abline(lm(mpg~horsepower),col="blue")
```




**Q 5. Is there a better way to represent their relationship rather than the linear model you just drew? (No need to use mathematical formula. Just draw something on the figure)**

**We can use polynomial regression instead of linear regression for this model and I have used paint to represent this.**