Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

Experiment -1

Title: Introduction: Understanding Data types; importing/exporting data

Procedure :

Installing R on Windows

1. Go to the CRAN website.
2. Click on **"Download R for Windows"**.
3. Click on **"install R for the first time"** link to download the R executable (.exe) file.
4. Run the R executable file to start installation, and allow the app to make changes to your device.
5. Select the installation language.
6. Follow the installation instructions.
7. Click on **"Finish"** to exit the installation setup.

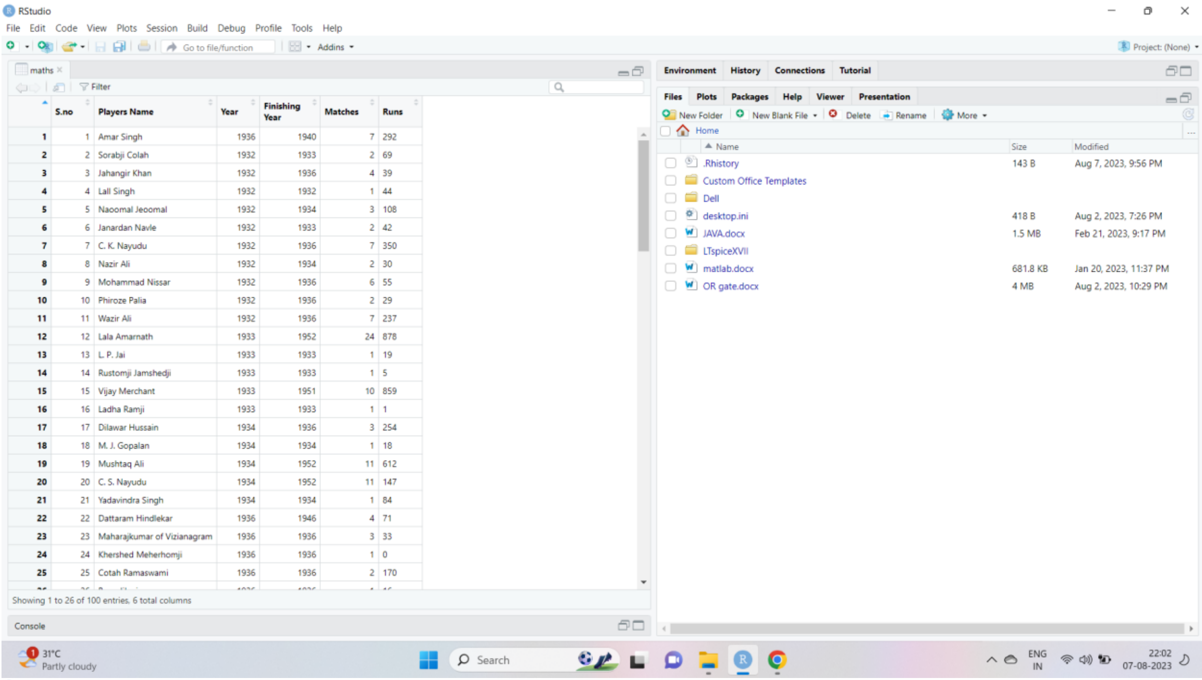
Output:

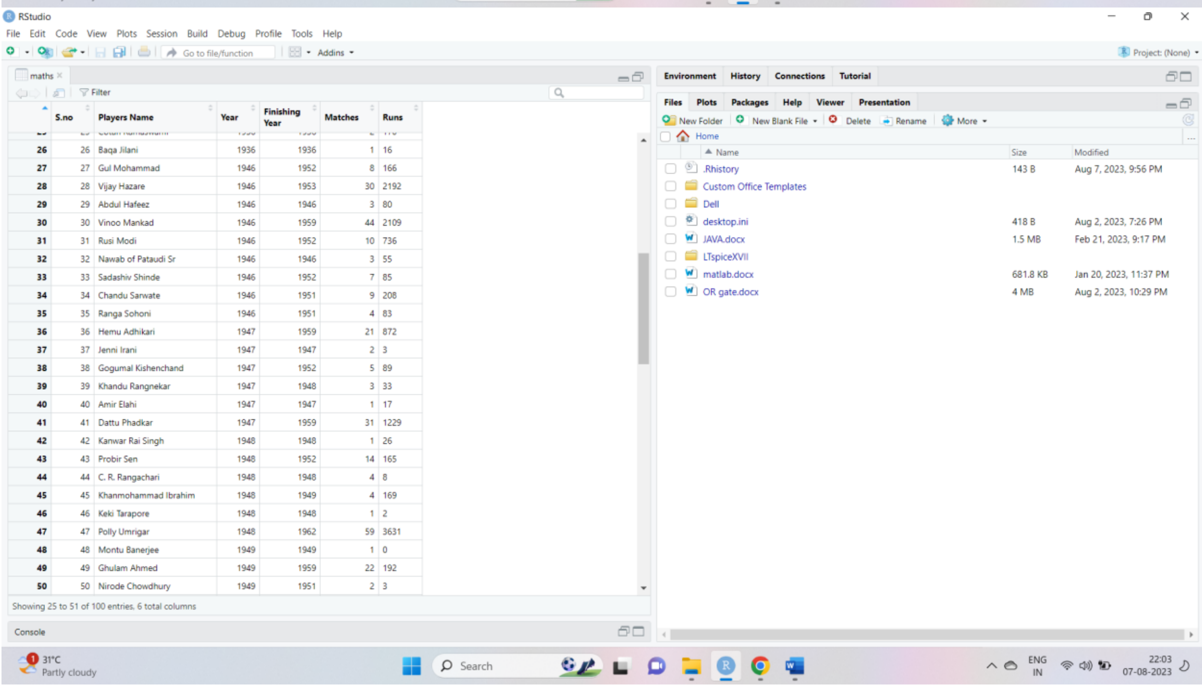
Importing Excelsheet in R.

Procedure:

- Import from the file system or a url
- Change column data types
- Skip columns • Rename the data set
- Select an specific Excel sheet
- Skip the first N rows
- Select NA identifiers

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab





Importing csv file

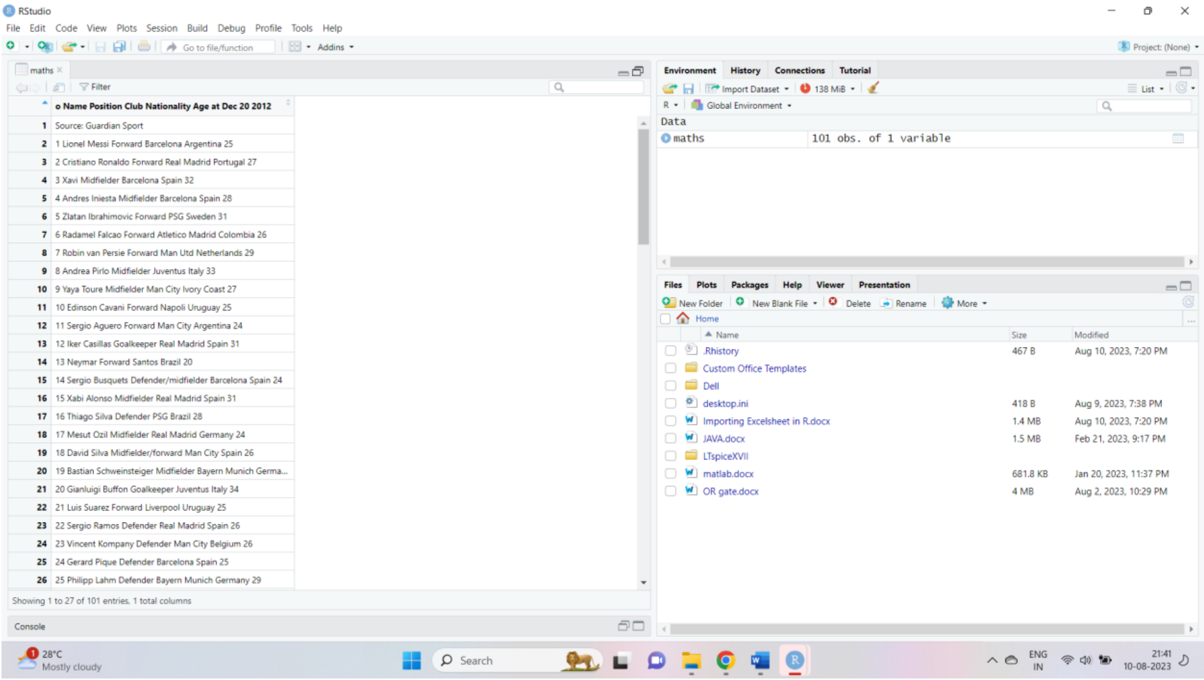
PROCEDURE:

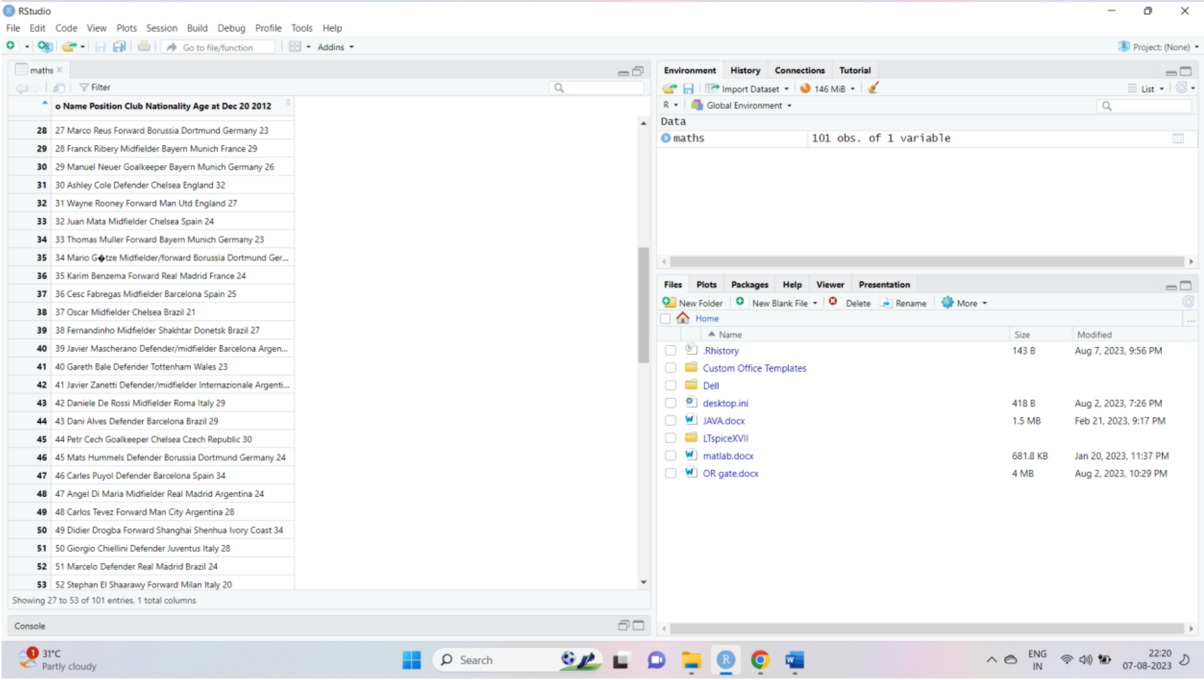
Import from the file system or a url

- Change column data types
- Skip or include-only columns
- Rename the data set
- Skip the first N rows
- Use the header row for column names
- Trim spaces in names

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

- Change the column delimiter
- Encoding selection
- Select quote, escape, comment and NA identifiers

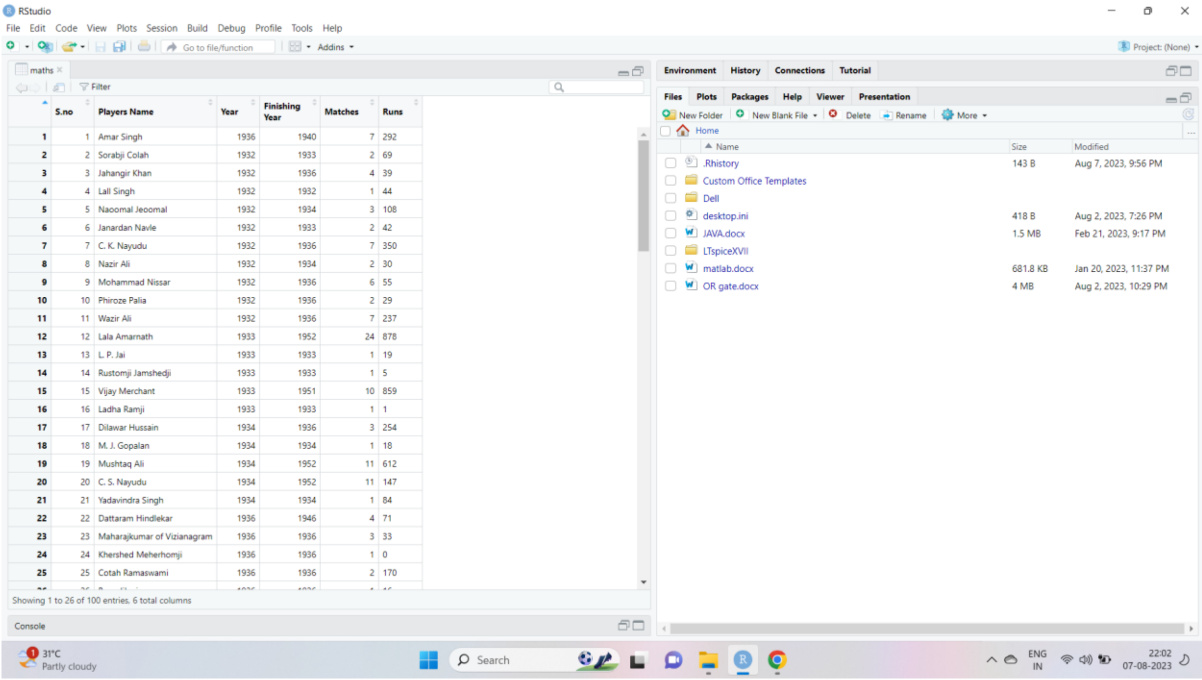




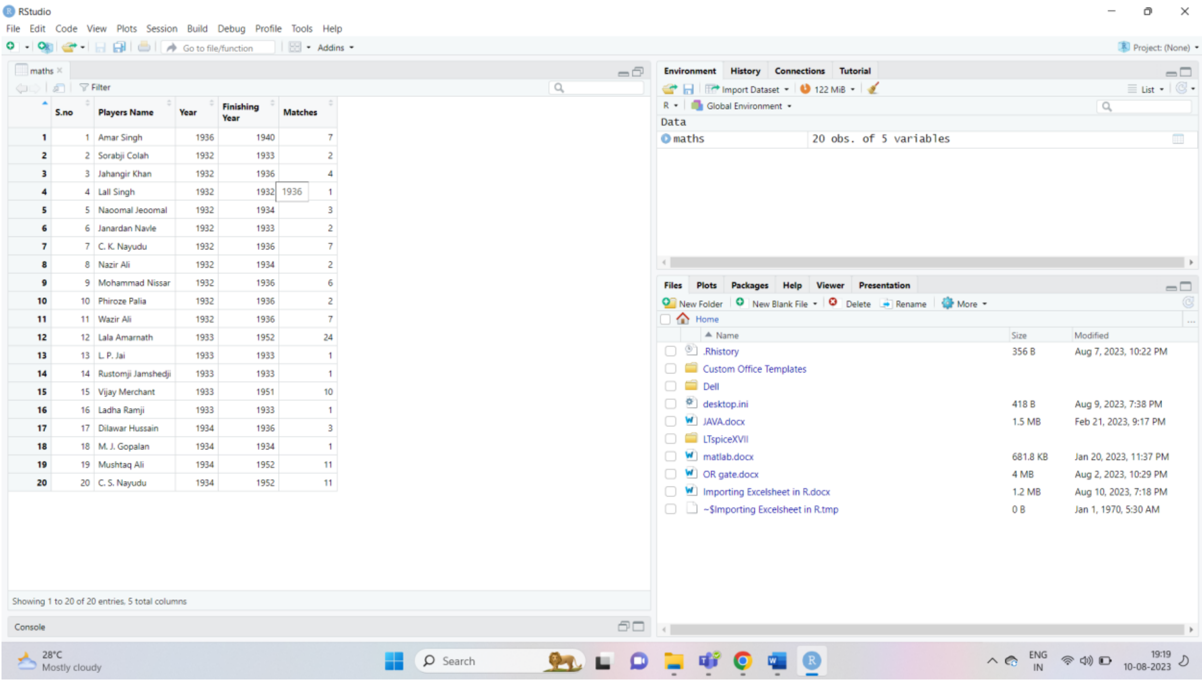
Deleting rows and columns

Before:

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab



AFTER:



Conclusion: Installation, input, output, import and various arithmetic operations have been explored in R

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

Experiment - 2

Title: Computing Summary Statistics /plotting and visualizing data using Tabulation and Graphical Representations

Code:

```
emid=c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15)

age=c(30,37,45,32,50,60,35,32,34,43,32,30,43,50,60)


gender=c(0,1,0,1,1,1,0,0,1,0,0,1,1,0,0)

status=c(1,1,2,2,1,1,1,2,2,1,2,1,2,1,2)

empinfo= data.frame(emid,age,gender,status)

empinfo$gender=factor(empinfo$gender,labels=c("male","female"))

empinfo$status=factor(empinfo$status,labels=c("staff","faculty"))
male=subset(empinfo,empinfo$gender=="male")

female=subset(empinfo,empinfo$gender=="female")
female
summary(empinfo)
summary(male)

summary(female)

summary(age)

table2=table(empinfo$status)

table3=table(empinfo$gender,empinfo$status)

plot(empinfo$age,type="l",main="Age of employees",xlab="emid",ylab="age in
years",col="blue")

pie(table1)
```

```
barplot(table3,beside=T,xlim=c(1,15),ylim=c(0,5),col=c("blue","red"))

boxplot(empinfo$age~empinfo$status,col=c('red','blue'))
```

Output:

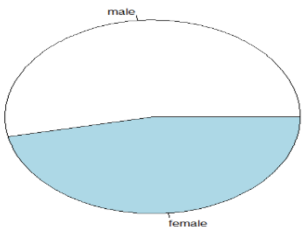
```
> emid
 [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15

> age
 [1] 30 37 45 32 50 60 35 32 34 43 32 30 43 50 60

> gender
 [1] 0 1 0 1 1 1 0 0 1 0 0 1 1 0 0

> status
 [1] 1 1 2 2 1 1 1 2 2 1 2 1 2 1 2

   emid age gender status
1     1  30      0      1
2     2  37      1      1
3     3  45      0      2
4     4  32      1      2
5     5  50      1      1
6     6  60      1      1
7     7  35      0      1
8     8  32      0      2
9     9  34      1      2
10   10  43      0      1
11   11  32      0      2
12   12  30      1      1
13   13  43      1      2
14   14  50      0      1
15   15  60      0      2
>
empinfo$gender
 [1] male   female male   female female female male
 [8] male   female male   male   female female male
[15] male
Levels: male female

empinfo$status
 [1] staff   staff   faculty faculty staff   staff
 [7] staff   faculty faculty staff   faculty staff
[13] faculty staff   faculty
```

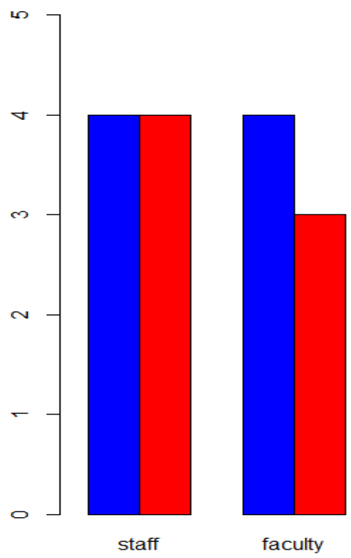
```
Levels: staff faculty

male
   emid age gender  status
1     1  30   male   staff
3     3  45   male faculty
7     7  35   male   staff
8     8  32   male faculty
10   10  43   male   staff
11   11  32   male faculty
14   14  50   male   staff
15   15  60   male faculty

female
   emid age gender  status
2     2  37 female   staff
4     4  32 female faculty
5     5  50 female   staff
6     6  60 female   staff
9     9  34 female faculty
12   12  30 female   staff
13   13  43 female faculty
>
summary(empinfo)
      emid          age          gender      status
 Min.   : 1.0   Min.   :30.00   male  :8   staff  :8
 1st Qu.: 4.5   1st Qu.:32.00   female:7   faculty:7
 Median : 8.0   Median :37.00
 Mean   : 8.0   Mean   :40.87
 3rd Qu.:11.5   3rd Qu.:47.50
 Max.   :15.0   Max.   :60.00
>
> summary(male)
      emid            age          gender
 Min.   : 1.000   Min.   :30.00   male  :8
 1st Qu.: 6.000   1st Qu.:32.00   female:0
 Median : 9.000   Median :39.00
 Mean   : 8.625   Mean   :40.88
 3rd Qu.:11.750   3rd Qu.:46.25
 Max.   :15.000   Max.   :60.00
     status
 staff  :4
 faculty:4

> summary(female)
      emid            age          gender
```

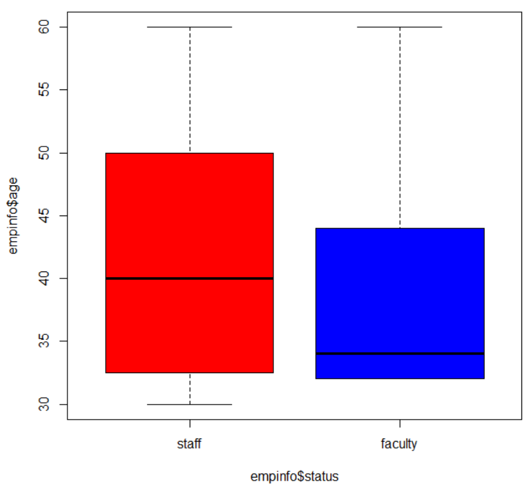
Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

```
 Min.   : 2.000   Min.   :30.00   male  :0
 1st Qu.: 4.500   1st Qu.:33.00   female:7
 Median : 6.000   Median :37.00
 Mean   : 7.286   Mean   :40.86
 3rd Qu.:10.500   3rd Qu.:46.50
 Max.   :13.000   Max.   :60.00
    status
 staff  :4
 faculty:3

summary(age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  30.00   32.00   37.00   40.87   47.50   60.00
>
table 1
  male female
     8      7
>

table2
  staff faculty
      8       7
>

> table3
         staff faculty
  male       4       4
  female     4       3
```

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab







Conclusion: Different alignment of data set and various graphical representations in R have been explored and execute

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

Experiment -3

Title: Applying correlation and simple linear regression model to real data set; computing and interpreting the coefficient of determination

Aim: To understand the simple correlation and linear regression with computation and interpretation

Code:

```
data=cars
cars
summary(data)
v1=var(data$speed)
v1
v2=var(data$dist)
v2
covariance=cov(data$speed,data$dist)
covariance
corr=covariance/(sd(data$speed)*sd(data$dist))
corr
cor.test(data$speed,data$dist)
cor.test(data$speed,data$dist,method="spearman")
plot(data$speed,data$dist)

regression1=lm(data$speed~data$dist)
regression1
abline(regression1)

summary(regression1)
regression2=lm(data$dist~data$speed)
regression2
abline(regression2)
summary(regression2)
```

OUTPUT:

```
cars
  speed dist
1     4    2
2     4   10
3     7    4
4     7   22
5     8   16
6     9   10
7    10   18
8    10   26
```

```
9     10   34
10    11   17
11    11   28
12    12   14
13    12   20
14    12   24
15    12   28
16    13   26
17    13   34
18    13   34
19    13   46
20    14   26
21    14   36
22    14   60
23    14   80
24    15   20
25    15   26
26    15   54
27    16   32
28    16   40
29    17   32
30    17   40
31    17   50
32    18   42
33    18   56
34    18   76
35    18   84
36    19   36
37    19   46
38    19   68
39    20   32
40    20   48
41    20   52
42    20   56
43    20   64
44    22   66
45    23   54
46    24   70
47    24   92
48    24   93
49    24  120
50    25   85

    speed           dist
 Min.   : 4.0   Min.   :  2.00
 1st Qu.:12.0   1st Qu.: 26.00
```

```
 Median :15.0  Median : 36.00
 Mean   :15.4   Mean   : 42.98
 3rd Qu.:19.0   3rd Qu.: 56.00
 Max.   :25.0     Max.   :120.00
```

v1
[1] 27.95918

v2
[1] 664.0608

covariance
[1] 109.9469

corr
[1] 0.8068949

cor.test(data$speed,data$dist)
OUTPUT:

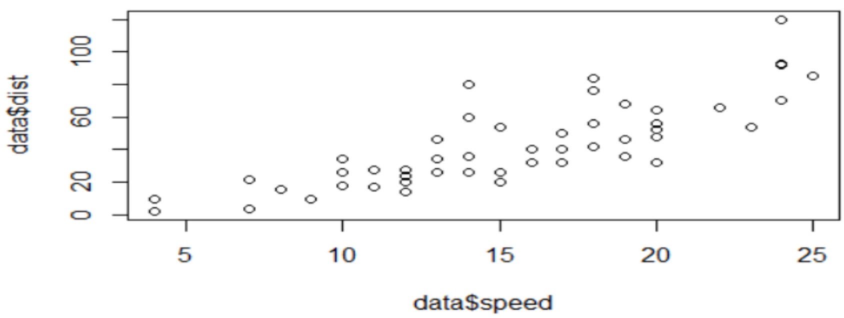
```
	Pearson's product-moment correlation

data:  data$speed and data$dist
t = 9.464, df = 48, p-value = 1.49e-12
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.6816422 0.8862036
sample estimates:
      cor
0.8068949
```

cor.test(data$speed,data$dist,method="spearman")
OUTPUT:

```
	Spearman's rank correlation rho

data:  data$speed and data$dist
S = 3532.8, p-value = 8.825e-14
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.8303568
```

plot(data$speed,data$dist)

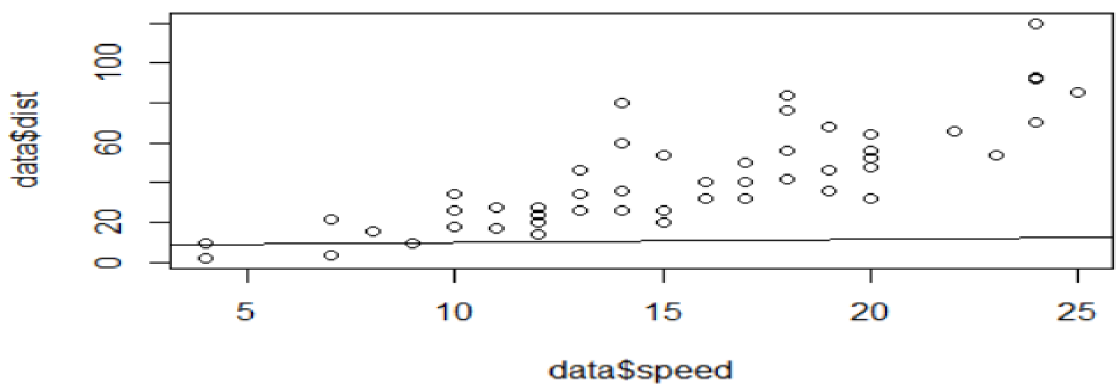
Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab



regression1
Call:
lm(formula = data$speed ~ data$dist)

Coefficients:
(Intercept)   data$dist
     8.2839   0.1656

abline(regression1)



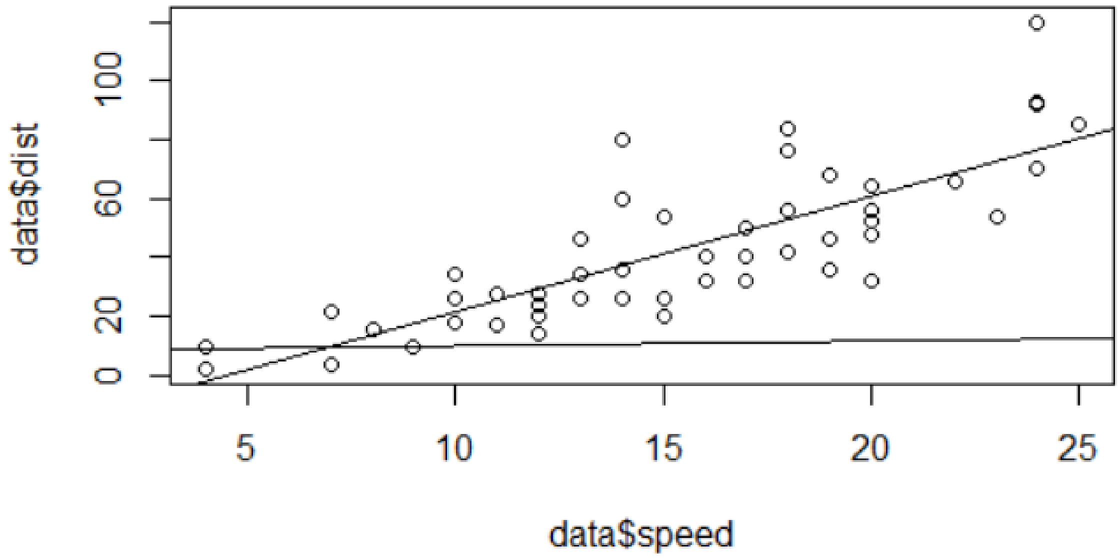
summary(regression1)
Call:
lm(formula = data$speed ~ data$dist)

Residuals:
    Min   1Q Median   3Q   Max
-7.5293 -2.1550 0.3615 2.4377 6.4179

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 8.28391   0.87438   9.474 1.44e-12 ***
data$dist   0.16557   0.01749   9.464 1.49e-12 ***

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.156 on 48 degrees of freedom
Multiple R-squared:  0.6511,  Adjusted R-squared:  0.6438
F-statistic: 89.57 on 1 and 48 DF,  p-value: 1.49e-12

regression2
Call:
lm(formula = data$dist ~ data$speed)

Coefficients:
(Intercept)   data$speed
    -17.579        3.932

abline(regression2)
```




```
summary(regression2)
Call:
lm(formula = data$dist ~ data$speed)

Residuals:
    Min      1Q  Median      3Q     Max
-29.069  -9.525  -2.272   9.215  43.201

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -17.5791     6.7584  -2.601   0.0123 *
```

```
data$speed    3.9324     0.4155   9.464 1.49e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.38 on 48 degrees of freedom
Multiple R-squared:  0.6511,  Adjusted R-squared:  0.6438
F-statistic: 89.57 on 1 and 48 DF,  p-value: 1.49e-12
```

Problem: The body weight and the BMI of 12 school going children are given in the following table.Find the simple regression model BMI on weight and examine the results.

Code:

```
weight=c(15,26,27,25,25.5,27,32,18,22,20,26,24)

bmi=c(13.35,16.12,16.74,16.00,13.59,15.73,15.65,13.85,16.07,12.8,13.65,14.42)

cor(weight,bmi)

model<-lm(bmi~weight)

summary.lm(model)

info=data.frame(weight,bmi)

plot(info$weight,info$bmi)

regression1=lm(info$weight~info$bmi)
regression1

abline(regression1)

summary(regression1)

regression2=lm(info$bmi~info$weight)
regression2

abline(regression2)

summary(regression2)
```

Output:

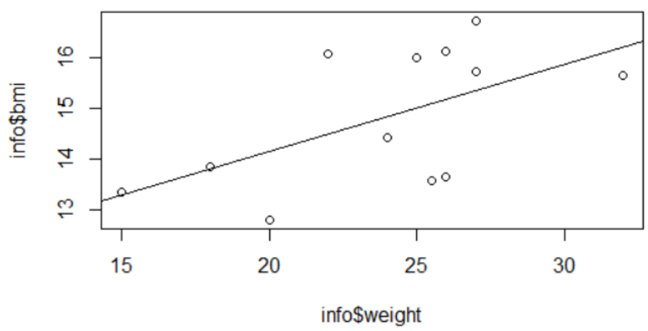
```
[1] 0.5790235
> summary.lm(model)
```

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

```
Call:
lm(formula = bmi ~ weight)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.52988 -0.75527  0.04426  0.95286  1.57397 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 10.73487    1.85405   5.790 0.000175 ***
weight       0.17096    0.07612   2.246 0.048524 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.155 on 10 degrees of freedom
Multiple R-squared:  0.3353,  Adjusted R-squared:  0.2688 
F-statistic: 5.044 on 1 and 10 DF,  p-value: 0.04852

> regression1

Call:
lm(formula = info$weight ~ info$bmi)

Coefficients:
(Intercept)     info$bmi  
     -5.126        1.961  

> summary(regression1)

Call:
lm(formula = info$weight ~ info$bmi)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.0543 -1.9471 -0.2311  1.9525  6.4352 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)  -5.1260    12.9996  -0.394   0.7016  
info$bmi      1.9611     0.8732   2.246   0.0485 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.912 on 10 degrees of freedom
Multiple R-squared:  0.3353,  Adjusted R-squared:  0.2688 
F-statistic: 5.044 on 1 and 10 DF,  p-value: 0.04852
```

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

```
> regression2

Call:
lm(formula = info$bmi ~ info$weight)

Coefficients:
(Intercept)  info$weight
     10.735        0.171


> summary(regression2)

Call:
lm(formula = info$bmi ~ info$weight)

Residuals:
     Min       1Q   Median       3Q      Max
-1.52988 -0.75527  0.04426  0.95286  1.57397

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.73487    1.85405   5.790 0.000175 ***
info$weight  0.17096    0.07612   2.246 0.048524 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.155 on 10 degrees of freedom
Multiple R-squared:  0.3353,  Adjusted R-squared:  0.2688
F-statistic: 5.044 on 1 and 10 DF,  p-value: 0.04852
```



Conclusion: The simple correlation and linear regression equation have been computed and interpreted

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

Experiment - 4

Title: Applying multiple linear regression model to real dataset; computing and interpreting the multiple coefficients of determination

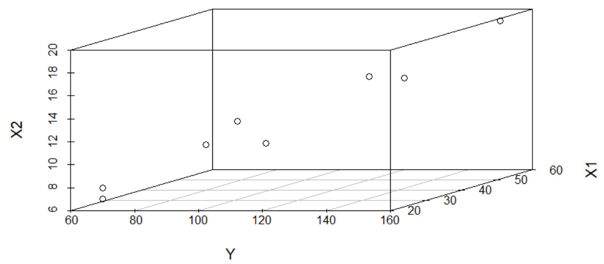
Problems: The sale of a Product in lakhs of rupees (Y) is expected to be influenced by two variables namely the advertising expenditure XI (in 'OOORS) and the number of sales persons(X2) in a region. Sample data on 8 Regions of a state has given the following results

Code:

```
Y=c(110,80,70,120,150,90,70,120)
Y
X1=c(30,40,20,50,60,40,20,60)
X1
X2=c(11,10,7,15,19,12,8,14)
X2
RegModel=lm(Y~X1+X2)
RegModel
summary(RegModel)
library(scatterplot3d)
scatterplot3d(Y,X1,X2)
```

Output:

```
> Y
[1] 110  80  70 120 150  90  70 120
> X1
[1] 30 40 20 50 60 40 20 60
> X2
[1] 11 10  7 15 19 12  8 14
> RegModel

Call:
lm(formula = Y ~ X1 + X2)

Coefficients:
(Intercept)           X1           X2
    16.8314      -0.2442       7.8488


>
```



Part-2

Code:

```
data=mtcars
data
X=mtcars$mpg
X
Y=mtcars$disp
Y
Z=mtcars$hp
Z
RegModel<- lm(Z~X+Y)
RegModel
summary(RegModel)
library(scatterplot3d)
graph=scatterplot3d(X,Y,Z)

graph$plane3d(RegModel)
```

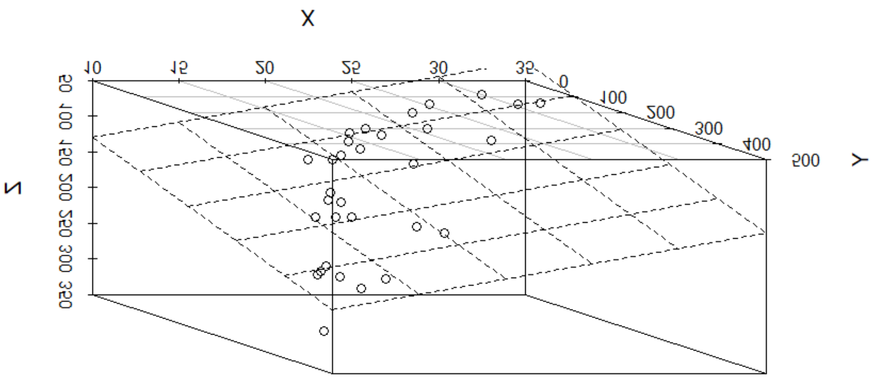
Output:

```
data
                     mpg cyl  disp  hp drat    wt  qsec vs am gear carb
Mazda RX4           21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
Mazda RX4 Wag       21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
Datsun 710          22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
Hornet 4 Drive      21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
Hornet Sportabout   18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
Valiant             18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
Duster 360          14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
Merc 240D           24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
Merc 230            22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
Merc 280            19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
Merc 280C           17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
Merc 450SE          16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
Merc 450SL          17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
Merc 450SLC         15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
Cadillac Fleetwood  10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0    3    4
Chrysler Imperial   14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
Fiat 128            32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
Honda Civic         30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
```

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

```
Toyota Corolla      33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
Toyota Corona       21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
Dodge Challenger    15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
AMC Javelin         15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
Camaro Z28          13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
Pontiac Firebird    19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
Fiat X1-9           27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
Porsche 914-2       26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
Lotus Europa        30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
Ford Pantera L      15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
Ferrari Dino        19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
Maserati Bora       15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
Volvo 142E          21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
> X
 [1] 21.0 21.0 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 17.8 16.4 17.3 15.2 10.4 10.4 14.7 32
[19] 30.4 33.9 21.5 15.5 15.2 13.3 19.2 27.3 26.0 30.4 15.8 19.7 15.0 21.4
> Y
 [1] 160.0 160.0 108.0 258.0 360.0 225.0 360.0 146.7 140.8 167.6 167.6 275.8 275.8 275.8 472
[16] 460.0 440.0  78.7  75.7  71.1 120.1 318.0 304.0 350.0 400.0  79.0 120.3  95.1 351.0 145
[31] 301.0 121.0
> Z
 [1] 110 110  93 110 175 105 245  62  95 123 123 180 180 180 205 215 230  66  52  65  97 150
[23] 150 245 175  66  91 113 264 175 335 109
> RegModel

Call:
lm(formula = Z ~ X + Y)

Coefficients:
(Intercept)            X            Y
   172.2204      -4.2732       0.2614

 > summary(RegModel)

 Call:
 lm(formula = Z ~ X + Y)

 Residuals:
    Min      1Q  Median      3Q     Max
 -48.70  -17.67  -10.16   10.12  148.19

 Coefficients:
             Estimate Std. Error t value Pr(>|t|)
 (Intercept) 172.2204    69.9014   2.464   0.0199 *
 X            -4.2732     2.3027  -1.856   0.0737 .
 Y             0.2614     0.1120   2.335   0.0267 *
 ---
 Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 Residual standard error: 41.01 on 29 degrees of freedom
 Multiple R-squared:  0.6653,    Adjusted R-squared:  0.6423
 F-statistic: 28.83 on 2 and 29 DF,  p-value: 1.279e-07
```

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab



>

Conclusion: Multiple linear regression model has been explored and visualized

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

Experiment - 5

Title: Binomial distribution

Essential R- Lab built-in functions: : dbinom(), sum(), pbinom(), weighted.mean(), plot()

Problems: Four coins are tossed simultaneously. What is the probability of getting

- (i) 2 heads
- (ii) atleast 2 heads
- (iii) atmost 2 heads
- (iv) Expectation of x
- (v) Variance of x
- (vi) Visualize the probability distribution

Code:

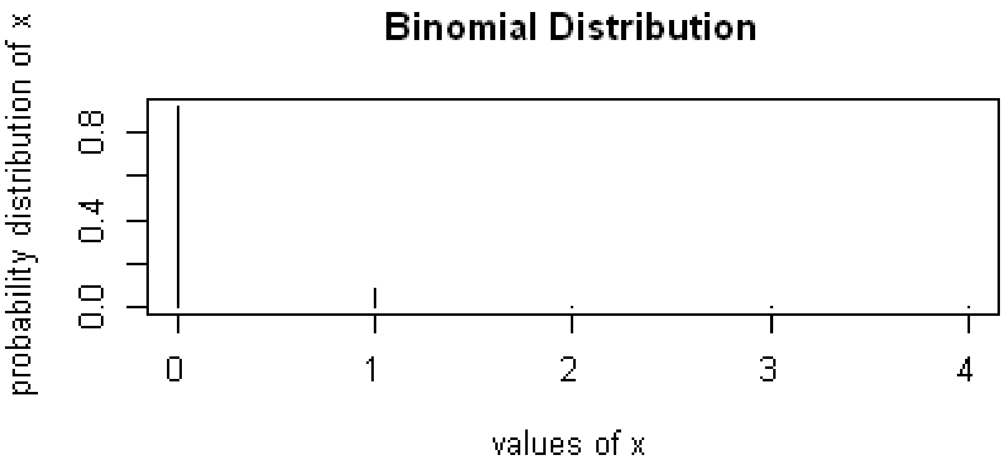
```
n=4
p=0.02
dbinom(2,n,p)
sum(dbinom(2:4,n,p))
1-pbinom(1,n,p)
sum(dbinom(0:2,n,p))
pbinom(2,n,p)
x=0:n
px=dbinom(x,n,p)
Ex=weighted.mean(x,px)
Ex
Varx=weighted.mean(x*x,px)-(weighted.mean(x,px))^2
Varx
plot(x,px,type="h",xlab="values of x",ylab="probability distribution of x",main="Binomial
Distribution")
```

Output:

```
> p=0.02

> dbinom(2,n,p)
[1] 0.00230496

> sum(dbinom(2:4,n,p))
[1] 0.00233648

> 1-pbinom(1,n,p)
[1] 0.00233648
```

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

```
> sum(dbinom(0:2,n,p))
[1] 0.9999685

> pbinom(2,n,p)
[1] 0.9999685


> Varx
[1] 0.0784
```



Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

Experiment - 6

Title: Normal distribution, Poisson distribution

Essential R- Lab built-in functions: sum(), round(), dpois(), plot(), weighted.mean()

Problems: A manufacturer of pins knows that 2% of his products are defective. If he sells pins in boxes of 20 and find the number of boxes containing

(i) at least 2 defective
(ii) exactly 2 defective
(iii) at most 2 defective pins in a consignment of 1000 boxes
(iv) plot the distribution
(v) E(x)
(vi) Variance of X?

Code:

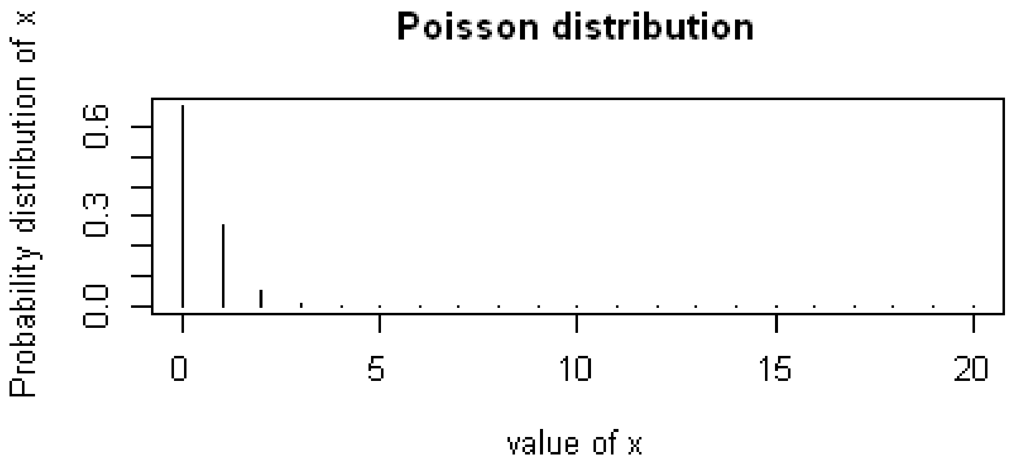
```
m=20
m
ps=0.02
ps
lambda=m*ps
lambda
p1=sum(dpois(2:m,lambda))
p1
round(1000*p1)
p2=dpois(2,lambda)
p2
round(1000*p2)
p3=sum(dpois(0:2,lambda))
p3
round(1000*p3)
x1=0:m
px1=dpois(x1,lambda)
plot(x1,px1,type="h",xlab="value of x",ylab="Probability distribution of x",main="Poisson
distribution")
Ex1=weighted.mean(x1,px1)
Ex1
Varx1=weighted.mean(x1*x1,px1)-(weighted.mean(x1,px1))^2
Varx1
```

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

Output:

```
m
[1] 20

> ps
[1] 0.02

> lambda
[1] 0.4


> p1
[1] 0.06155194

> round(1000*p1)
[1] 62

> p2
[1] 0.0536256

> round(1000*p2)
[1] 54

> p3
[1] 0.9920737

> round(1000*p3)
[1] 992

> Ex1
[1] 0.4

> Varx1
[1] 0.4
```

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab



Part-2
Essential R- Lab built-in functions: dnorm(), pnorm(), seq(), polygon(), data.frame()

Problem : A company finds that the time taken by one of its engineers to complete or repair job has a normal distribution with mean 20 minutes and S.D 5 minutes. State what proportion of jobs take: i. Less than 15 minutes ii. More than 25 minutes iii. Between 15 and 25 minutes iv. Plot the distribution v. Table the distribution

Code:
```
x=seq(0,40)
x
y=dnorm(x,mean=20,sd=5)
y
plot(x,y,type="l")
p1=pnorm(15,mean=20,sd=5)
p1
x2=seq(0,15)
x2
y2=dnorm(x2,mean=20,sd=5)
y2
polygon(c(0,x2,15),c(0,y2,0),col="yellow")
p2=pnorm(40,mean=20,sd=5)-pnorm(25,mean=20,sd=5)
p2
x1=seq(25,40)
x1
y1=dnorm(x1,mean=20,sd=5)
y1
polygon(c(25,x1,40),c(0,y1,0),col="red")
p3=pnorm(25,mean=20,sd=5)-pnorm(15,mean=20,sd=5)
p3
x3=seq(15,25)
x3
```

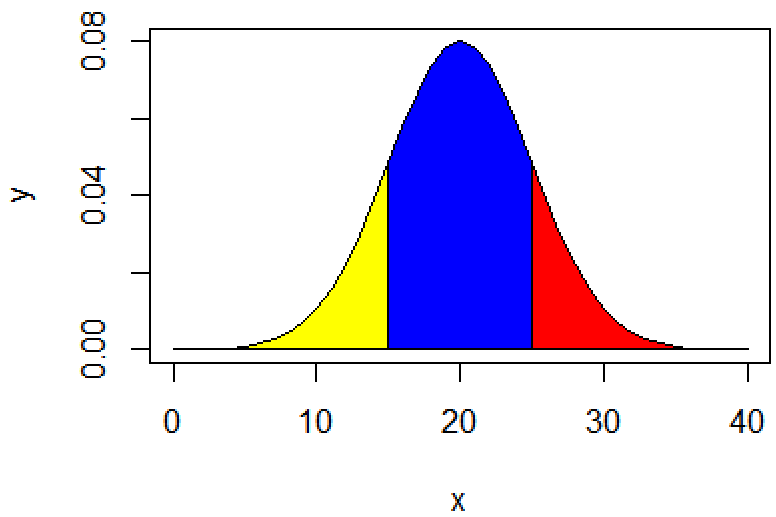
```
y3=dnorm(x3,mean=20,sd=5)
y3
polygon(c(15,x3,25),c(0,y3,0),col="blue")
data.frame(p1,p2,p3)
```

Output:

```
> x=seq(0,40)
>x
 [1]  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
31 32 33
[35] 34 35 36 37 38 39 40
> y=dnorm(x,mean=20,sd=5)
>y
 [1] 2.676605e-05 5.838939e-05 1.223804e-04 2.464438e-04 4.768176e-04 8.863697e-04
1.583090e-03
 [8] 2.716594e-03 4.478906e-03 7.094919e-03 1.079819e-02 1.579003e-02 2.218417e-02
2.994549e-02
[15] 3.883721e-02 4.839414e-02 5.793831e-02 6.664492e-02 7.365403e-02 7.820854e-02
7.978846e-02
[22] 7.820854e-02 7.365403e-02 6.664492e-02 5.793831e-02 4.839414e-02 3.883721e-02
2.994549e-02
[29] 2.218417e-02 1.579003e-02 1.079819e-02 7.094919e-03 4.478906e-03 2.716594e-03
1.583090e-03
[36] 8.863697e-04 4.768176e-04 2.464438e-04 1.223804e-04 5.838939e-05 2.676605e-05
>plot(x,y,type="l")
> p1=pnorm(15,mean=20,sd=5)
> p1
[1] 0.1586553
>x2=seq(0,15)
>x2
 [1]  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15
> y2=dnorm(x2,mean=20,sd=5)
> y2
 [1] 2.676605e-05 5.838939e-05 1.223804e-04 2.464438e-04 4.768176e-04 8.863697e-04
1.583090e-03
 [8] 2.716594e-03 4.478906e-03 7.094919e-03 1.079819e-02 1.579003e-02 2.218417e-02
2.994549e-02
[15] 3.883721e-02 4.839414e-02
>polygon(c(0,x2,15),c(0,y2,0),col="yellow")
> p2=pnorm(40,mean=20,sd=5)-pnorm(25,mean=20,sd=5)
> p2
[1] 0.1586236
>x1=seq(25,40)
>x1
 [1] 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
> y1=dnorm(x1,mean=20,sd=5)
```

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

```
> y1
 [1] 4.839414e-02 3.883721e-02 2.994549e-02 2.218417e-02 1.579003e-02 1.079819e-02
7.094919e-03
 [8] 4.478906e-03 2.716594e-03 1.583090e-03 8.863697e-04 4.768176e-04 2.464438e-04
1.223804e-04
[15] 5.838939e-05 2.676605e-05
>polygon(c(25,x1,40),c(0,y1,0),col="red")
> p3=pnorm(25,mean=20,sd=5)-pnorm(15,mean=20,sd=5)
> p3
[1] 0.6826895
>x3=seq(15,25)
>x3
 [1] 15 16 17 18 19 20 21 22 23 24 25

> y3=dnorm(x3,mean=20,sd=5)
> y3
 [1] 0.04839414 0.05793831 0.06664492 0.07365403 0.07820854 0.07978846 0.07820854
0.07365403 0.06664492
[10] 0.05793831 0.04839414
>polygon(c(15,x3,25),c(0,y3,0),col="blue")
>data.frame(p1,p2,p3)
        p1        p2        p3
1 0.1586553 0.1586236 0.6826895
```



Conclusion: Poisson distribution and Normal distribution have been explored using R

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

Experiment - 7

Title: Testing of Hypothesis - Large Sample mean Test

Essential R- Lab built-in functions: : qnorm(), pnorm(), sqrt(), print()

Problems: Suppose the mean weight of King Penguins found in an Antarctic colony last year was 15.4 kg. In a sample of 35 penguins same time this year in the same colony, the mean penguin weight is 14.6 kg. Assume the population standard deviation is 2.5 kg. At 0 .05 significance level, can we reject the null hypothesis that the mean penguin weight does not differ from last year?

Code:

```
xbar=14.6
xbar
mu0=15.4
mu0
sigma=2.5
sigma
n=35
n
z=(xbar-mu0)/(sigma/sqrt(n))
z
alpha=0.05
alpha
zhalfalpha=qnorm(1-(alpha/2))
zhalfalpha
c(-zhalfalpha,zhalfalpha)
pval=2*pnorm(z)
pval
if(pval>alpha){print("Accept Null Hypothesis")}else{print("Reject Null Hypothesis")}
```

Output:

```
> xbar
[1] 14.6

> mu0=15.4

> mu0
[1] 15.4

> sigma=2.5
```

```
> sigma
[1] 2.5

> n=35

> n
[1] 35

> z=(xbar-mu0)/(sigma/sqrt(n))

> z
[1] -1.893146

> alpha=0.05

> alpha
[1] 0.05

> zhalfalpha=qnorm(1-(alpha/2))

> zhalfalpha
[1] 1.959964

> c(-zhalfalpha,zhalfalpha)
[1] -1.959964  1.959964

> pval=2*pnorm(z)

> pval
[1] 0.05833852

> [1] "Accept Null hypothesis"
```

Title: Testing of Hypothesis - Large Sample proportion Test

Essential R- Lab built-in functions: qnorm(), sqrt(), print()

Problems: The fatality rate of typhoid patients is believed to be 17.26%. In a certain year 640 patients suffering from typhoid were treated in a metropolitan hospital and only 63 patients died. Can you consider the hospital efficient?

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

Code:

```
> n=640

> Sprop=63/n

> Pprop=0.1726

> q=1-Pprop

> z=(Sprop-Pprop)/sqrt(Pprop*q/n)

> E=qnorm(.975)

> c(-E,E)

> Sprop+c(-E,E)*sqrt(Pprop*(1-Pprop)/n)

> if(z>-E && z<E){print("Hospital is not efficient")}else{print("Hospital is efficient@")}
```

Output:

```
> n
[1] 640

> Sprop
[1] 0.0984375

> Pprop
[1] 0.1726

> q
[1] 0.8274

> z
[1] -4.964736

> c(-E,E)
[1] -1.959964  1.959964
```

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

```
> Sprop+c(-E,E)*sqrt(Pprop*(1-Pprop)/n)
[1] 0.06915985 0.12771515

> if(z>-E && z<E){print("Hospital is not efficient")}else{print("Hospital is efficient@")}
[1] "Hospital is efficient@"
```

Conclusion: Testing of hypothesis for large sample tests using R functions has been explored and concluded

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

Experiment - 8

Title: Testing of Hypothesis - Two Sample mean Test

Essential R- Lab built-in functions: qnorm(), sqrt(), print()

Problems: In a random sample of size 500, the mean is found to be 20. In another independent sample of size 400, the mean is 15. Could the samples have been drawn from the same population with S.D 4?

Code:
```
> xbar

> ybar=15

> sigma=4

> n1=500

> n2=400

> z=(xbar-ybar)/(sigma*sqrt((1/n1)+(1/n2)))

> alpha=0.05

> zalpha=qnorm(1-(alpha/2))

> if(z<=zalpha){print("Accept Null Hypothesis")}else{print("Reject Null Hypothesis")}
```

Output:
```
> xbar
[1] 20

> ybar
[1] 15

> sigma
[1] 4

> n1
[1] 500

> n2
[1] 400
```

```
> z
[1] 18.6339

> alpha
[1] 0.05


> zalpha
[1] 1.959964

[1] "Reject Null Hypothesis"
```

Title: Testing of Hypothesis - Two Sample proportion Test

Essential R- Lab built-in functions: qnorm(), sqrt(), print()

Problems: In a large city A, 20% of a random sample of 900 schools boys had a slight physical defect. In another large city B, 18.5% of a random sample of 1600 school boys had the same defect. Is the difference between the proportions significant?

Code: > p1

```
> p2=0.185

> n1=900

> n2=1600

> p=(n1*p1+n2*p2)/(n1+n2)

> q=1-P

> z=(p1-p2)/sqrt(p*q*((1/n1)+(1/n2)))

> alpha=0.05

> zalpha=qnorm(1-(alpha/2))

> if(z<=zalpha){print("Accept Null hypothesis")}else{print("Reject the Null Hypothesis")}
```

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

Output:
```
> p1
[1] 0.2

> p2
[1] 0.185

> n1
[1] 900

> n2
[1] 1600

> p
[1] 0.1904

> q
[1] 0.8096

> z
[1] 0.9169249

> alpha
[1] 0.05

> zalpha
[1] 1.959964

[1] "Accept Null hypothesis
```

Conclusion: Testing of hypothesis for large sample tests using R functions has been explored and concluded

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

Experiment - 9

Title: : APPLYING THE T-TEST FOR INDEPENDENT AND DEPENDENT SAMPLES

Essential R- Lab built-in functions: : t.test(),print()

Problems: Two independent samples of sizes 8 and 7 contained the following values: Sample Is the difference between the sample means significant?
1
2
3

Code:

```
sample1=c(19,17,15,21,16,18,16,14)
sample1
sample2=c(15,14,15,19,15,18,16,20)
sample2
t=t.test(sample1,sample2)
t
cv=t$statistic
cv
tv=qt(0.975,14)
tv
if(cv <= tv){print("Accept Ho")} else{print("Reject Ho")}
```

Output:

```
sample1
[1] 19 17 15 21 16 18 16 14
> sample2
[1] 15 14 15 19 15 18 16 20
> t

        Welch Two Sample t-test

data:  sample1 and sample2
t = 0.44721, df = 13.989, p-value = 0.6616
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.898128  2.898128
sample estimates:
mean of x mean of y
     17.0      16.5

> cv
        t
0.4472136
> tv
[1] 2.144787

> [1] "Accept Ho"
```

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

Part 2- F-Test

Code:

```
sample1=c(9,11,13,11,15,9,12,14)
sample1
sample2=c(10,12,10,14,9,8,10)
sample2
f=var.test(sample1,sample2)
f
cv=f$statistic
cv
tv=qf(0.95,7,7)
tv
if(cv <= tv){print("Accept Ho")} else{print("Reject Ho")}
```

Output:

```
> sample1
[1]  9 11 13 11 15  9 12 14
> sample2
[1] 10 12 10 14  9  8 10
> f

        F test to compare two variances

data:  sample1 and sample2
F = 1.2108, num df = 7, denom df = 6, p-value = 0.8315
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2125976 6.1978188
sample estimates:
ratio of variances
          1.210843

> cv
       F
1.210843
> tv
[1] 3.787044
> source("~/.active-rstudio-document")
[1] "Accept Ho"
```

Conclusion: Student's t-test and F-test have been explored and executed

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

Experiment - 10

Title: Applying Chi-square test for goodness of fit test and Contingency test to real dataset

Essential R- Lab built-in functions: chisq.test(),print()

Code:

```
data<-matrix(c(69,51,81,20,35,44),ncol=2,byrow=T)
data
cv=chisq.test(data)
cv
cv=cv$p.value
cv
if(cv >alpha){print("Attributes are independent")} else{print("Attributes are
not independent")}
```

Output:

```
> data
     [,1] [,2]
[1,]   69   51
[2,]   81   20
[3,]   35   44

>cv
               Pearson's Chi-squared test
data: data  X-squared = 25.629, df = 2, p-value = 2.721e-06

> cv
[1] 2.72114e-06

[1] "Attributes are \nnot independent"
```

Conclusion: Chi-squared test has been explored through R functions

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

Experiment - 11

Title: Performing ANOVA for real dataset for Completely Randomized Design, Randomized Block design, Latin square Design

Essential R- Lab built-in functions: data.frame(),print(),aov()

Code:

```
> x=c(5,6,8,9,7,8,10,11,12,4,7,3,5,4,1)
> y=c('A','A','A','A','A','B','B','B','B','B','C','C','C','C','C')
> male=data.frame(x,y)
> male
> summary(aov(x~y,data=male))
> x=c(51,49,47,49,52,49,49,50,48,52,53,51)
> y=c('A','A','A','A','B','B','B','B','C','C','C','C')
> yield=data.frame(x,y)
> yield
> summary(aov(x~y,data=yield))
```

Output:

```
   x y
1   5 A
2   6 A
3   8 A
4   9 A
5   7 A
6   8 B
7  10 B
8  11 B
9  12 B
10  4 B
11  7 C
12  3 C
13  5 C
14  4 C
15  1 C


> summary(aov(x~y,data=male))
            Df Sum Sq Mean Sq F value Pr(>F)
y            2  63.33   31.67   5.429 0.0209 *
Residuals   12  70.00    5.83
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Regn. No:22BEC1126
Name: M. Hariharan
Slot: Lab 57+58
R- Lab

```
> yield
   x y
1  51 A
2  49 A
3  47 A
4  49 A
5  52 B
6  49 B
7  49 B
8  50 B
9  48 C
10 52 C
11 53 C
12 51 C
> summary(aov(x~y,data=yield))
            Df Sum Sq Mean Sq F value Pr(>F)
y            2      8   4.000   1.286  0.323
Residuals    9     28   3.111
```