LAB REPORT

R Programming

Guided by:-

Prof.Ankur singh

presented by:

Valiveti Manikanta Bhuvanesh

19BCD7088

VIT AP

AMARAVATI

Index page

LAB 1	Basics about R
LAB 2	Data Analysis using R
LAB 3	Statistical measure and graph plot for a set of data
LAB 4	Random Sampling and Probability
LAB 5	Binomial Distribution
LAB 6	Poisson Distribution
LAB 7	Normal Distribution
LAB 8	Test of Hypothesis-z test
LAB 9	Test of Hypothesis- t test
LAB 10	Correlation Coefficent
LAB 11	Regression
LAB 12	Real life application

19BCD7088

Lab1

1. Simple Operations

a) Enter the data {2,5,3,7,1,9,6 Enter the data {2,5,3,7,1,9,6} directly and store it } directly and store it in a variable x. in a variable x.

b) Find the number of elements in x, i.e. in the data list.

c) Find the last element of x.

d) Find the minimum element of x.

e) Find the maximum element of x.

```
x <- c(2,5,3,7,1,9,6)

length(x)

max(x)

min(x)
```

2. Enter the data {1, 2, …. ,19,20} in a variable

a) Find the 3rd element in the data list.

b) Find 3rd to 5th element in the data list.

c) Find 2nd, 5th, 6th, and 12th element in the list.

d) Print the data as {20, 19, …, 2, 1} without again entering the data

```
l <- c(1:20)

l

l[3]

l[3:5]

p<-c(l[2],l[5],l[6],l[12])

p

rev(l)
```

Lab 2

Q.1) Few simple statistical measures:

(a) Enter data as 1, 2, 3 . . . , 10.

(b) Find sum of the numbers.

(c) Find Mean, median.

(d) Find sum of squares of these values.

(e) Find the value of 1 n Pn i=1 | xi − x¯ |, this is known as mean deviation about mean (MDx¯).

(f) Check whether MDx¯ is less than or equal to standard deviation.

```
x<-c(1:10)

x

sum(x)

mean(x)

median(x)

sum(x^2)

y<- x-mean(x)

y

MD<-(1/10)*sum(abs(y))

MD

sd<-sd(x)

sd

MD<=sd
```

Q.2) Create a file as follows and store as a:-

	price	FloorArea	Rooms	Age	CentralHeating
1	52.00	1225	3	6.2	YES
2	54.75	1230	3	7.5	NO
3	57.50	1200	3	4.2	NO
4	57.50	1000	2	4.8	NO
5	59.75	1420	4	1.9	YES
6	62.50	1450	3	5.2	YES
7	64.75	1380	4	6.5	NO
8	67.25	1510	4	9.2	NO
9	67.50	1400	5	0.0	NO
10	69.75	1550	6	5.7	NO
11	70.00	1720	6	7.3	YES
12	75.50	1700	5	4.5	NO
13	77.50	1660	6	6.8	YES
14	78.00	1800	7	0.7	YES
15	81.25	1830	6	5.6	YES
16	82.50	1790	6	2.3	NO
17	86.25	2010	6	6.7	YES
18	87.50	2000	6	3.4	NO
19	88.00	2100	8	5.6	YES
20	92.00	2240	7	3.4	YES

(a) How many rows are there in this table? How many columns are there?

(b) How to find the number of rows and number of columns by a single command?

(c) What are the variables in the data file?

(d) If the file is very large, naturally we can not simply type 'a', because it will cover the entire screen and we won't be able to understand anything. So how to see the top or bottom few lines in this file?
(e) If the number of columns is too large, again we may face the same problem. So how to see the first 5 rows and first three columns?

(f) How to get 1st, 3rd, 6th, and 10th row and 2nd, 4th, and 5th columns? (g) How to get values in a specific row or a column?

Price<-c(52.0,54.75,57.50,57.50,59.75,62.50,64.75,67.25,67.50,69.75,70.00,75.50,77.50,78.00,81.25,82.50,86.25,87.50,88.00,92.00)

FloorArea<-c(1225,1230,1200,1000,1420,1450,1380,1510,1400,1550,1720,1700,1660,1800,1830,1790,2010,2000,2100,2240)

Rooms<-c(3,3,3,2,4,3,4,4,5,6,6,5,6,7,6,6,6,6,8,7)

Age<-c(6.2,7.5,4.2,4.8,1.9,5.2,6.5,9.2,0.0,5.7,7.3,4.5,6.8,0.7,5.6,2.3,6.7,3.4,5.6,3.4)

CentralHeating<-c("YES","NO","NO","NO","YES","YES","NO","NO","NO","NO","YES","NO","YES","YES","YES","NO","YES","NO","YES","YES")

TABLE<-data.frame(Price,FloorArea,Rooms,Age,CentralHeating)

TABLE

```
> TABLE
   Price FloorArea Rooms Age CentralHeating
1  52.00      1225     3 6.2            YES
2  54.75      1230     3 7.5             NO
3  57.50      1200     3 4.2             NO
4  57.50      1000     2 4.8             NO
5  59.75      1420     4 1.9            YES
6  62.50      1450     3 5.2            YES
7  64.75      1380     4 6.5             NO
8  67.25      1510     4 9.2             NO
9  67.50      1400     5 0.0             NO
10 69.75      1550     6 5.7             NO
11 70.00      1720     6 7.3            YES
12 75.50      1700     5 4.5             NO
13 77.50      1660     6 6.8            YES
14 78.00      1800     7 0.7            YES
15 81.25      1830     6 5.6            YES
16 82.50      1790     6 2.3             NO
17 86.25      2010     6 6.7            YES
18 87.50      2000     6 3.4             NO
19 88.00      2100     8 5.6            YES
20 92.00      2240     7 3.4            YES
```

nrow(TABLE)

ncol(TABLE)

c(nrow(TABLE),ncol(TABLE))

names(TABLE)

head(TABLE)

tail(TABLE)

TABLE[1:5,1:3]

TABLE[c(1,3,6,10),c(2,4,5)]

TABLE[2]

TABLE[,3]

```
> head(TABLE)
  Price FloorArea Rooms Age CentralHeating
1 52.00      1225     3 6.2            YES
2 54.75      1230     3 7.5             NO
3 57.50      1200     3 4.2             NO
4 57.50      1000     2 4.8             NO
5 59.75      1420     4 1.9            YES
6 62.50      1450     3 5.2            YES
> tail(TABLE)
   Price FloorArea Rooms Age CentralHeating
15 81.25      1830     6 5.6            YES
```

```
R Console

16 82.50      1790     6 2.3             NO
17 86.25      2010     6 6.7            YES
18 87.50      2000     6 3.4             NO
19 88.00      2100     8 5.6            YES
20 92.00      2240     7 3.4            YES
> TABLE[1:5,1:3]
  Price FloorArea Rooms
1 52.00      1225     3
2 54.75      1230     3
3 57.50      1200     3
4 57.50      1000     2
5 59.75      1420     4
> TABLE[c(1,3,6,10),c(2,4,5)]
   FloorArea Age CentralHeating
1       1225 6.2            YES
3       1200 4.2             NO
6       1450 5.2            YES
10      1550 5.7             NO
> TABLE[2]
   FloorArea
1       1225
2       1230
3       1200
4       1000
5       1420
6       1450
7       1380
8       1510
9       1400
10      1550
11      1720
12      1700
13      1660
14      1800
15      1830
16      1790
17      2010
18      2000
19      2100
20      2240
> TABLE[,3]
 [1] 3 3 3 2 4 3 4 4 5 6 6 5 6 7 6 6 6 6 8 7
```

Q.3) Calculate simple statistical measures using the values in the data file.

(a) Find means, medians, standard deviations of Price, Floor Area, Rooms, and Age.

(b) How many houses have central heating and how many don't have?

(c) Plot Price vs. Floor, Price vs. Age, and Price vs. Rooms, in separate graphs.

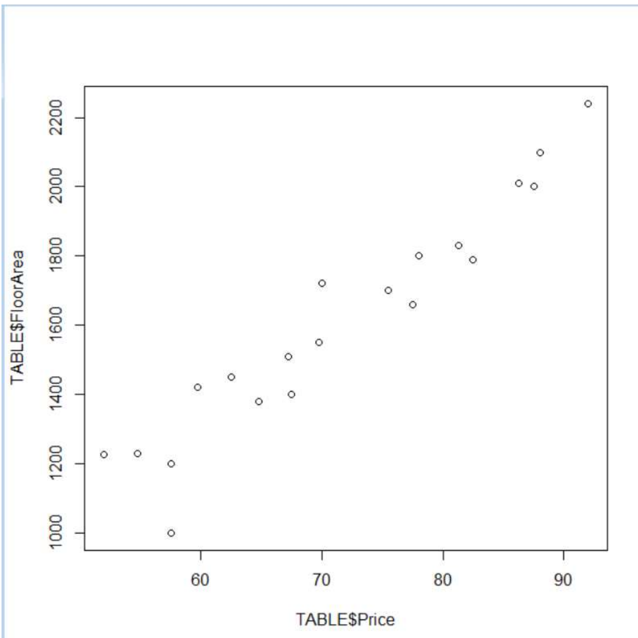
(d) Draw histograms of Prices, Floor Area, and Age.

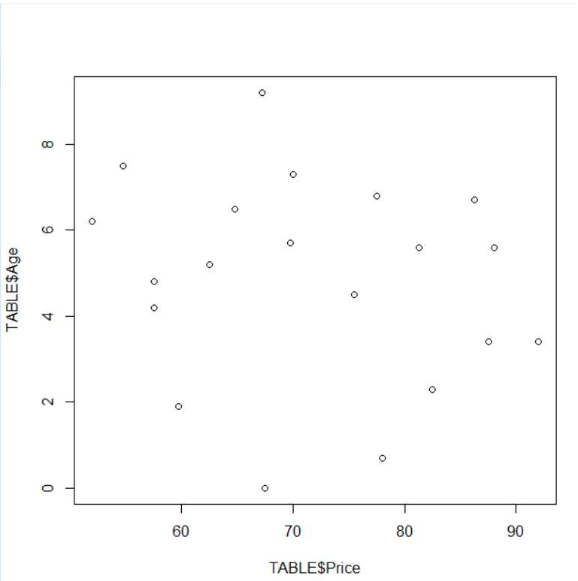
(e) Draw box plots of Price, Floor Area, and Age.

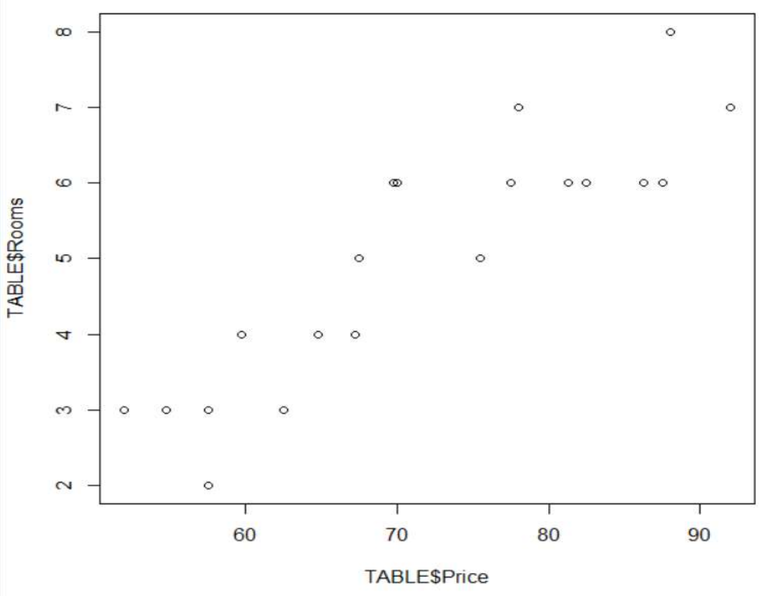
(f.) Draw all the graphs in (c), (d), and (e) in the same graph paper.

```
Price<-
c(52.0,54.75,57.50,57.50,59.75,62.50,64.75,67.25,67.50,69.75,70.00,75.50,77.50,78.00,81.25,82.50,
86.25,87.50,88.00,92.00)

FloorArea<-
c(1225,1230,1200,1000,1420,1450,1380,1510,1400,1550,1720,1700,1660,1800,1830,1790,2010,200
0,2100,2240)

Rooms<-c(3,3,3,2,4,3,4,4,5,6,6,5,6,7,6,6,6,6,8,7)

Age<-c(6.2,7.5,4.2,4.8,1.9,5.2,6.5,9.2,0.0,5.7,7.3,4.5,6.8,0.7,5.6,2.3,6.7,3.4,5.6,3.4)

CentralHeating<-
c("YES","NO","NO","NO","YES","YES","NO","NO","NO","NO","YES","NO","YES","YES","YES","NO","YE
S","NO","YES","YES")

TABLE<-data.frame(Price,FloorArea,Rooms,Age,CentralHeating)

mean(TABLE[,1])

mean(TABLE[,2])

mean(TABLE[,3])

mean(TABLE[,4])

median(TABLE[,1])

median(TABLE[,2])

median(TABLE[,3])

median(TABLE[,4])

sd(TABLE[,1])

sd(TABLE[,2])

sd(TABLE[,3])

sd(TABLE[,4])

sum(TABLE$CentralHeating=="YES")

sum(TABLE$CentralHeating=="NO")
```

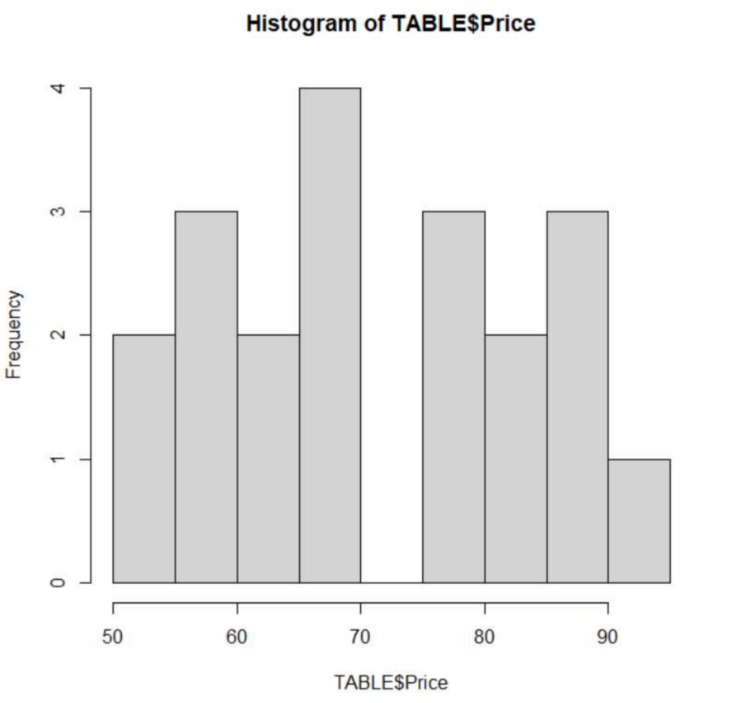
```
plot(TABLE$Price,TABLE$FloorArea)
plot(TABLE$Price,TABLE$Age)
plot(TABLE$Price,TABLE$Rooms)
hist(TABLE$Price)
hist(TABLE$FloorArea)
hist(TABLE$Age)
boxplot(TABLE$Price)
boxplot(TABLE$FloorArea)
boxplot(TABLE$Age)
```



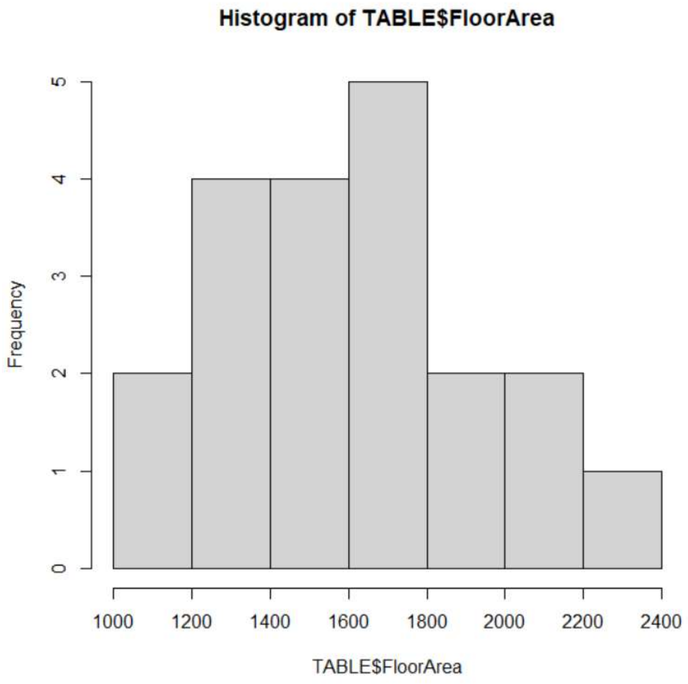


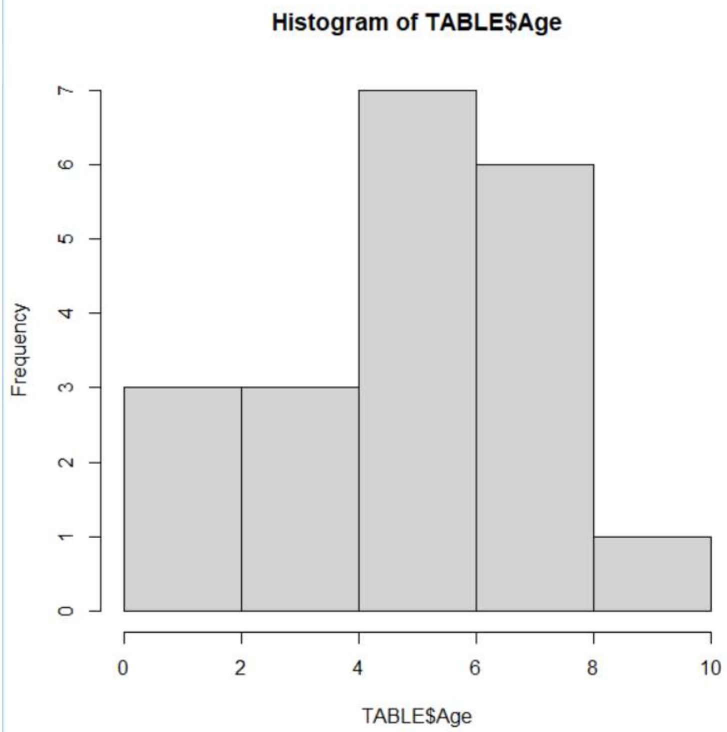


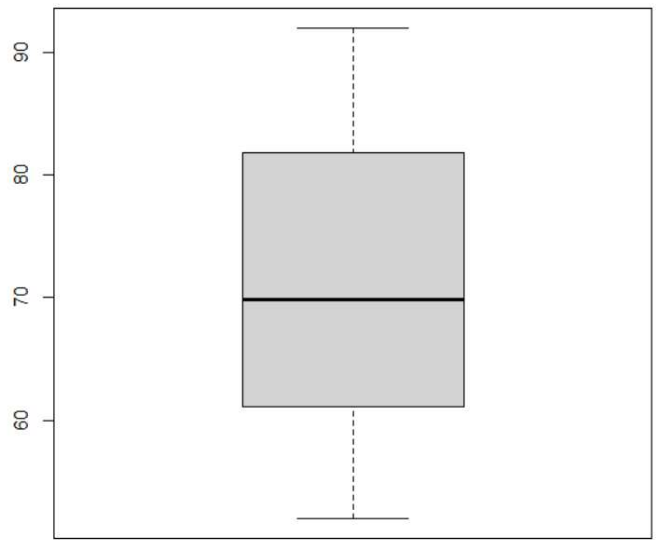
TABLE$Rooms
TABLE$Price



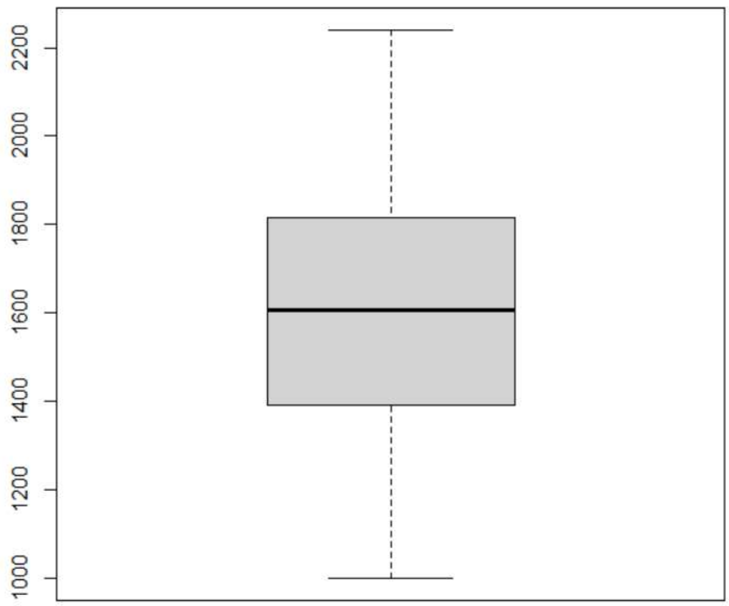
Histogram of TABLE$Price
Frequency
TABLE$Price


Histogram of TABLE$FloorArea
Frequency
0
1
2
3
4
5
1000
1200
1400
1600
1800
2000
2200
2400
TABLE$FloorArea

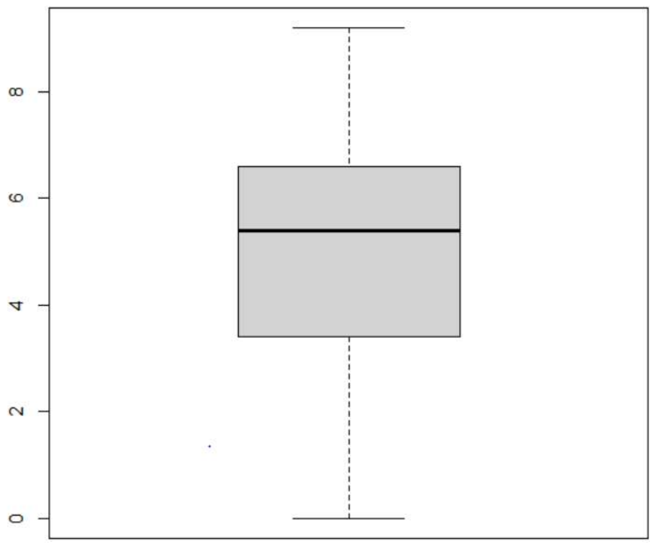

Histogram of TABLE$Age
Frequency
0
1
2
3
4
5
6
7
0
2
4
6
8
10
TABLE$Age



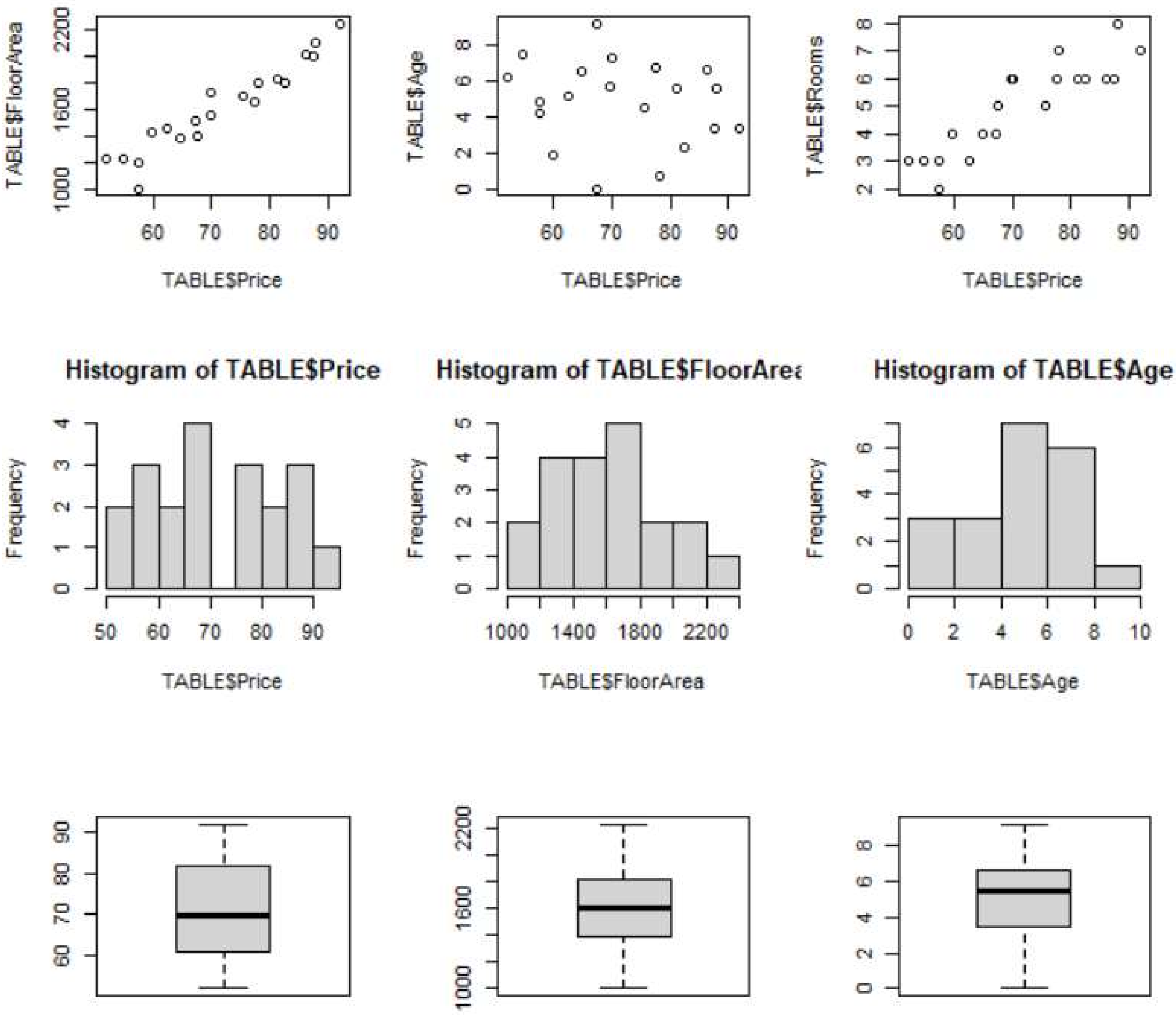
90
80
70
60



2200
2000
1800
1600
1400
1200
1000




```
par(mfrow=c(3,3))
plot(TABLE$Price,TABLE$FloorArea)
plot(TABLE$Price,TABLE$Age)
plot(TABLE$Price,TABLE$Rooms)
hist(TABLE$Price)
hist(TABLE$FloorArea)
hist(TABLE$Age)
boxplot(TABLE$Price)
boxplot(TABLE$FloorArea)
boxplot(TABLE$Age)
```


TABLE$FloorArea
TABLE$Price
TABLE$Age
TABLE$Price
TABLE$Rooms
TABLE$Price
Histogram of TABLE$Price
Frequency
TABLE$Price
Histogram of TABLE$FloorAre
Frequency
TABLE$FloorArea
Histogram of TABLE$Age
Frequency
TABLE$Age

Lab3

Q.1)

(a) Create a data list (4, 4, 4, 4, 3, 3, 3, 5, 5, 5) using ’rep’ function.

(b) Create a list (4, 6, 3, 4, 6, 3, . . . , 4, 6, 3) where there 10 occurrences of 4,6, and 3 in the given order.

(c) Create a list (3, 1, 5, 3, 2, 3, 4, 5, 7, 7, 7, 7, 7, 7, 6, 5, 4, 3, 2, 1, 34, 21, 54) using one line command.
(d) First create a list (2, 1, 3, 4). Then append this list at the end with another list (5, 7, 12, 6, −8). Check whether the number of elements in the augmented list is 11.

```
x<-list(rep(4,4),rep(3,3),rep(5,3))

x

x<-list(rep(c(4,6,3),10))

x

x<-list(3,1,5,3,c(2:5),rep(7,6),c(6:1),34,21,54)

x

x<-list(2,1,3,4)

y<-list(5,7,12,6,-8)

o=append(x,y)

length(o)==11
```

Q.2)

(a) Print all numbers starting with 3 and ending with 7 with an increment of 0 : 0.5. Store these numbers in x.

(b) Print all even numbers between 2 and 14 (both inclusive).

(c) Type 2 ∗ x and see what you get. Each element of x is multiplied by 2.

```
x<-seq(3,7,0.5)

x

seq(2,14,2)

2*x
```

Q.3) Collect atleast 75 students and analyse the data by using descriptive statistics and interpret the results.

```
setwd("C:/Users/User/Documents/R/19bcd7088")

a=read.csv("VMB.csv")
```

```
> a
      id       name gender marks
1  72522     Dewain      M    98
2  20112     Karena      F    69
3  97472 Westbrooke      M    83
4  51525  Charmaine      F    70
5  94042     Lovell      M    82
6  45129       Pete      M    86
7  77777     Justus      M    79
8  56234     Stinky      M    59
9  84631      Benny      F    50
10 59531 Holly-anne      F    68
11 97396      Denis      M    64
12 35698     Pepito      M   100
13 24385        Tan      M    99
14 38691      Benni      F    97
15 79352      Peder      M    51
16 89753      Dante      M    81
17 24551   Stefania      F    57
18 14890     Franky      M    96
19 12103       Yale      M    51
20 25278 Gilbertina      F    77
21 78952    Valaria      F    82
22 93355      Hobie      M    65
23 69532       Stan      M    76
24 72326      Adlai      M    96
25 90043       Iain      M    72
26 61239      Alyse      F    55
27 29246   Dulciana      F    58
28 75573    Carlyle      M    73
29 80011     Deanna      F    51
30 49623      Alena      F    89
31 57036    Jerrold      M    66
```

```
32 78082      Nina    F   61
33 69354     Carey     M   85
34 45891     Neron     M   75
35 79798  Llywellyn     M   94
36 24296       Dag     M   74
37 21653      Pail     M   97
38 35641   Ginnifer     F   97
39 17892    Calypso     F   60
40 51589     Andras     M   81
41 21356     Vonny     F   84
42 99236    Stanton     M   62
43 23646     Carly     M   89
44 52154     Roland     M   69
45 94562      Etan     M   59
46 64235     Tadeo     M   70
47 24521     Lannie     M   50
48 55895     Aldwin     M   59
49 79278     Shelli     F   66
50 66953      Inna     F   50
51 13758    Nicolea     F   73
52 10235      Dunc     M   69
53 79531    Eugenio     M   54

> nrow(a)
[1] 53
> ncol(a)
[1] 4
>
> c(nrow(a),ncol(a))
[1] 53  4
> names(a)
[1] "id"     "name"   "gender" "marks"
```

```
> head(a)
     id       name gender marks
1 72522     Dewain      M    98
2 20112     Karena      F    69
3 97472 Westbrooke      M    83
4 51525  Charmaine      F    70
5 94042     Lovell      M    82
6 45129       Pete      M    86
> tail(a)
      id    name gender marks
48 55895  Aldwin      M    59
49 79278  Shelli      F    66
50 66953    Inna      F    50
51 13758 Nicolea      F    73
52 10235    Dunc      M    69
53 79531 Eugenio      M    54
> a[1:5,1:3]
     id       name gender
1 72522     Dewain      M
2 20112     Karena      F
3 97472 Westbrooke      M
4 51525  Charmaine      F
5 94042     Lovell      M
> a[c(1,3,6,10),c(2,3,4)]
         name gender marks
1      Dewain      M    98
3  Westbrooke      M    83
6        Pete      M    86
10 Holly-anne      F    68
> mean(a[,4])
[1] 73.16981
```

```
> median(a[,4])
[1] 72
> sd(a[,4])
[1] 15.32938
> sum(a$gender=="M")
[1] 34
> sum(a$gender=="F")
[1] 19
```

LAB4

Q.1) Matrices and arrays:

(a)Matrices and arrays are represented as vectors with dimensions:Create one matrixxwith 1 to 12 numbers with 3×4 order.

(b)Create same matrix with matrix function.

(c)Give name of rows of this matrix with A,B,C.

(d)Transpose of the matrix.

(e)Use functionscbindandrbindseparately to create different matrices.

(f)Use arbitrary numbers to create matrix.(g)Verify matrix multiplication.

```
m<-c(1:12)

dim(m)<-c(3,4)

m

m<-matrix(c(1:12),nrow=3,ncol=4)

m

m<-matrix(c(1:12),nrow=3,ncol=4,dimnames=list(c("A","B","C")))

m

t(m)

rbind(1:4,5:8,9:12)

cbind(c(1,5,9),c(2,6,10),c(3,7,11),c(4,8,12))

w<-t(m)

w

m%*%w
```

Q.2) Random sampling

(a)In R you can simulate these situations with the sample function. Pick five numbers at randomfrom the set 1 : 40.

(b)Notice that the default behaviour of sample is sampling without replacement. That is thesamples will not contain the same number twice, and obviously can not be bigger than the lengthof the vector to be sampled. If you want sampling with replacement, then you need to add theargumentreplace=T RU E.

Sampling with replacement is suitable for modelling coin tosses orthrows of a die. So, for instance, simulate 10 coin tosses.

(c)In fair coin-tossing, the probability of heads should equal the probability of tails, but the ideaof a random event is not restricted to symmetric cases. It could be equally well applied to othercases, such as the successful outcome of a surgical procedure. Hopefully there would be a betterthan 50% chance of this. Simulate data with non equal probabilities for the outcomes (say, a 90%chance of success) by using theprobargument to sample.

(d)The choose function can be used to calculate the following expression. (405) =40!/5!35!.

(e)Find 5!

```
x<-1:40

x

sample(x)

sample(x,5)

sample(c("H","T"),10,replace=T)

sample(c("win","lose"),10,replace=T,prob=c(0.9,0.1))

choose(40,5)

factorial(5)
```

LAB 5

Q.1) Five terminals on an on-line computer system are attached to a communication line to the central computer system. The probability that any terminal is ready to transmit is 0.95. Let X denote the number of ready terminals.

(a) Find the probability of getting exactly 3 ready terminals.

dbinom(3,5,0.95)

[1] 0.02143438

(b) Find all the probabilities.

dbinom(c(0:5),5,0.95)

Q.2) It is known that 20% of integrated circuit chips on a production line are defective. To maintain and monitor the quality of the chips, a sample of twenty chips is selected at regular intervals for inspection. Let X denote the number of defectives found in the sample. Find the probability of different number of defective found in the sample?

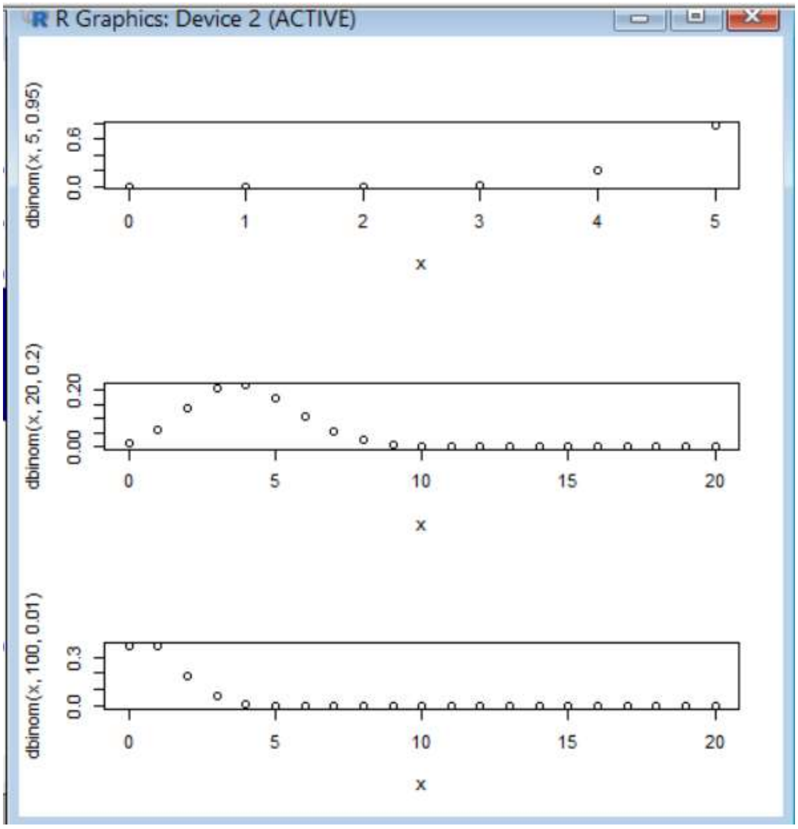
dbinom(c(0:5),20,0.2)

Q.3) It is known that 1% of bits transmitted through a digital transmission are received in error. One hundred bits are transmitted each day. Find the probability of different number of bits found in error each day.?

dbinom(c(0:5),100,0.01)

Q.4) Plot all of the above problems in a single window for random variable and respective Probability distribution.

```
par(mfrow = c(3,1))

x<-0:5

plot(x,dbinom(x,5,0.95))

x<-0:20

plot(x,dbinom(x,20,0.2))

plot(x,dbinom(x,100,0.01))
```



Q.5) For Q.No. 1 Find P(X = 3) and P(X > 3).

For Q. No. 2 Find P(X = 4) and P(X > 4).

Find all the cumulative probabilities and round to 4 decimal places.

```
pbinom(3,5,0.95)

1-pbinom(3,5,0.95)

pbinom(4,20,0.2)

1-pbinom(4,20,0.2)

x<-0:5

p<-pbinom(x,5,0.95)

p

round(p,4)
```

Q.6) The probability that a patient recover from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that

(a) at least 10 survive,

(b) from 3 to 8 survive,

(c) exactly 5 survive?

```
x<-0:20

p<-pbinom(x,20,0.2)

p

round(p,4)

sum(dbinom(c(10:15),15,0.4))

sum(dbinom(c(3:8),15,0.4))

dbinom(3,15,0.4)
```

Lab 6

Q.1) During a laboratory experiment, the average number of radioactive particles passing througha counter in 1 millisecond is 4. What is the probability that 6 particles enter the counter in a givenmillisecond?

dpois(6,4)

Q.2) In a certain industrial facility, accidents occur infrequently. It is known that the probability

of an accident on any given day is 0.005 and accidents are independent of each other.

(a) What is the probability that in any given period of 400 days there will be an accident on oneday?

dbinom(1,400,0.005)

(b) What is the probability that there are at most three days with an accident?

x<-(0:3)

sum(dbinom(x,400,0.005))

Q.3) In a manufacturing process where glass products are made, defects or bubbles occur, occasionally rendering the piece undesirable for marketing. It is known that, on average, 1 in every 1000 of these items produced has one or more bubbles. What is the probability that a random sample of 8000 will yield fewer than 7 items possessing bubbles?

x<-c(0:6)

sum(dbinom(x,8000,0.001))

Q.4) On average, 3 traffic accidents per month occur at a certain intersection. What is the

probability that in any given month at this intersection

(a) exactly 5 accidents will occur?

dpois(5,3)

(b) fewer than 3 accidents will occur?

sum(dpois(0:2,3))

(c) at least 2 accidents will occur?

1-sum(dpois(0:1,3))

Q.5) The potential buyer of a particular engine requires (among other things) that the engine start successfully 10 consecutive times. Suppose the probability of a successful start is 0.990. Let us assume that the outcomes of attempted starts are independent.

(a) What is the probability that the engine is accepted after only 10 starts?
dbinom(10,10,0.990)

(b) What is the probability that 12 attempted starts are made during the acceptance **process?**

dbinom(10,12,0.990)

Q.6) The acceptance scheme for purchasing lots containing a large number of batteries is to test no more than 75 randomly selected batteries and to reject a lot if a single battery fails. Suppose the probability of a failure is 0.001.

(a) What is the probability that a lot is accepted?

1-sum(dbinom(1:75,75,0.001))

(b) What is the probability that a lot is rejected on the 20th test?

a<-0.001*((1-0.001)^19)

a

(c) What is the probability that it is rejected in 10 or fewer trials?

x<-1:10

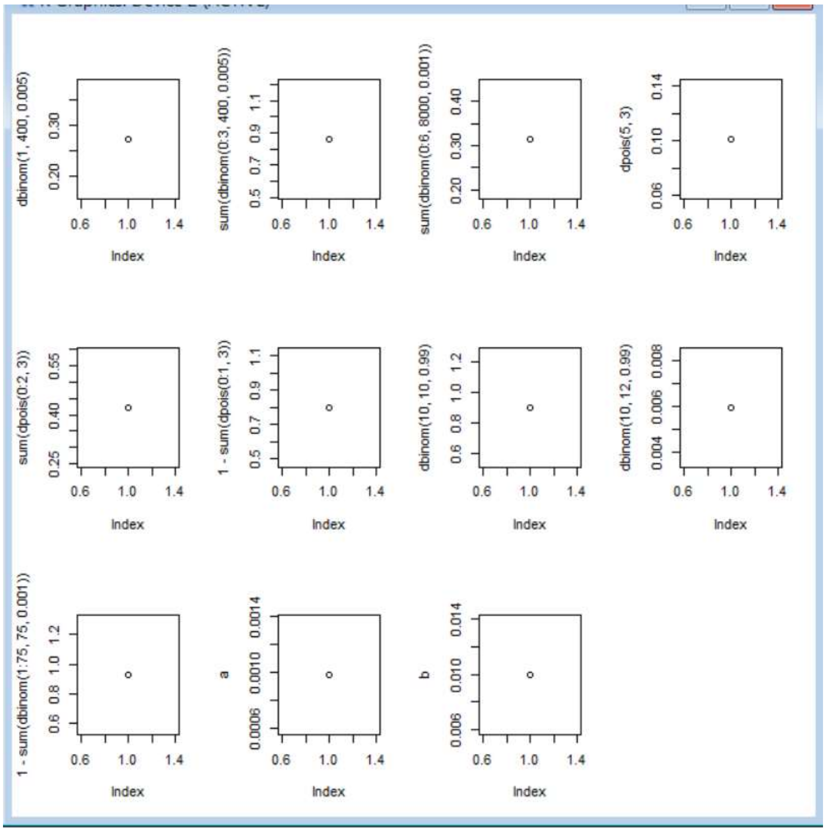
b<-sum(0.001*((1-0.001)^(x-1)))

b

Q.7) Plot the graph for Q. No. 2, 4, 5 and 6 for Random Variable against Probability Distribution function.

```
par(mfrow=c(3,4))

plot(dbinom(1,400,0.005))

plot(sum(dbinom(0:3,400,0.005)))

plot(sum(dbinom(0:6,8000,0.001)))

plot(dpois(5,3))

plot(sum(dpois(0:2,3)))

plot(1-sum(dpois(0:1,3)))
```

```
plot(dbinom(10,10,0.990))
plot(dbinom(10,12,0.990))
plot(1-sum(dbinom(1:75,75,0.001)))
plot(a)
plot(b)
```



Lab 7

Q.1) IQ is a normal distribution of mean of 100 and standard deviation of 15

(a) What percentage of people have an IQ < 125?

(b) What percentage of people have an IQ > 110?

(c) What percentage of people have 110 < IQ < 125?

(d) Find 25% for standard normal distribution.

(e) Find 25% normal distribution with mean and standard deviation 2& 3.

(f) What IQ separates the lower 25% from the others?

(g) What IQ separates the top 25% from the others?

(h) Find 25 percentile for mean 100 and SD 15.

```
pnorm(125,100,15,lower.tail=T)

pnorm(110,100,15,lower.tail=F)

pnorm(125,100,15,lower.tail=T)-pnorm(110,100,15,lower.tail=F)

qnorm(0.25)

qnorm(0.25,2,3)

qnorm(0.25,100,15,T)

qnorm(0.25,100,15,F)

qnorm(0.25,100,15)
```

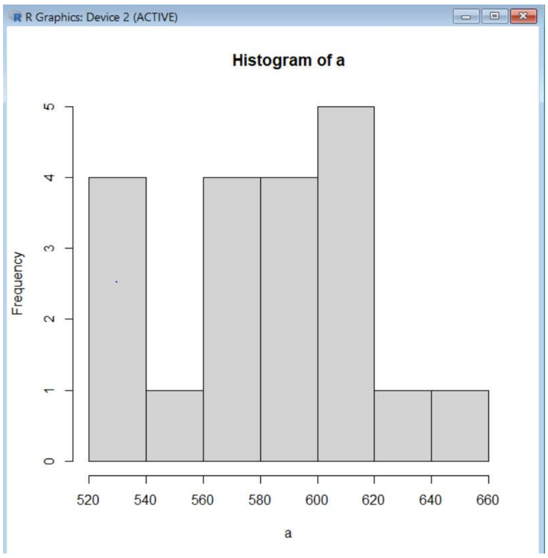
Q.2) Generate the 20 random number for a normal distribution with mean 572 and SD is 51.

Calculate mean and SD of data set.

```
a<-rnorm(n=20,mean=572, sd=51)

a

mean(a)

sd(a)
```

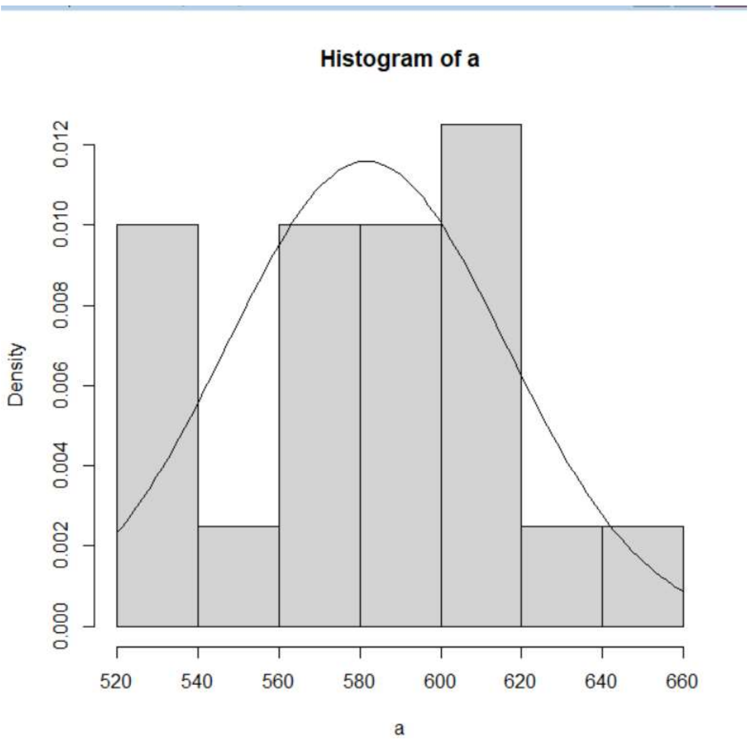
Q.3) Make appropriate histogram of data in above question and visually assume if normal density curve & histogram density estimates are similar.

hist(a)



hist(a,freq=F)

curve(dnorm(x,mean(a),sd(a)),add=T)



Lab 8

Q.1) Test the hypothesis that the mean systolic blood pressure in a certain population equals

140 mmHg. The standard deviation has a known value of 20 and a data set of 55 patients is available. 120,115,94,118,111,102,102,131,105,107, 115,139,115,113,114,105,115,134

,109,109,93,118,109,106,125,150,142,119,127,141,149,144,142,149,161,143,140,148,149,141,146,159,152,135,134,161,130,125, 141,148,153,145,13 7,147,175.

```
x<-
c(120,115,94,118,111,102,131,105,107,115,139,115,113,114,105,115,134,109,109,93,118,109,106,
125,150,142,119,141,149,144,142,149,161,143,140,148,149,141,146,159,152,135,134,161,130,125,
141,148,153,145,137,147,175)

x

n<-length(x)

z<-sqrt(n)*((mean(x)-140)/20)

z

p<-2*(pnorm(z))

p

if(p<0.02){

print("rejected");

}else{

print("accepted");

}
```

Q.2) A coin is tossed 100 times and turns up head 43 times Test the claim that this is a fair coin.Use 5% level of significance to test the claim.

```
p<-0.5

x<-43/100

z<-(x-p)/0.05

p<-2*(pnorm(z))

p

if(p<0.05){

print("rejected");

}else{

print("accepted");

}
```

Q.3) A manufacturer of sports equipment has developed a new synthetic fishing line that the company claims has a mean breaking strength of 8 kilograms with a standard deviation of 0.5 kilogram. Test the hypothesis that μ=8 kilograms against the alternative that μ is not equal to 8 kilograms if a random sample of 50 lines is tested and found to have a mean breaking strength of 7.8 kilograms. Use a 0.01 level of significance.

```
z<-sqrt(50)*((7.8-8)/0.5)
p<-2*(1-pnorm(abs(z)))
p
if(p<0.01){
print("rejected");
}else{
print("accepted");
}
```

Lab 9

Q.1) An outbreak of salmonella-related illness was attributed to ice produced at certain factory. Scientists measured the level of salmonella in 9 randomaly sampled batches ice cream. The levels (in MPN/g) were:

0.593

0.142

0.329

0.691

0.231

0.793

0.519

0.392

0.418

Is there evidence that the mean level of salmonella in ice cream greater than 0.3 MPN/g.

x<-c(0.593,0.142,0.329,0.691,0.231,0.793,0.519,0.392,0.418)

t.test(x,alternative="greater",mu=0.3)

Q.2) Suppose that 10 volunteers have taken an intellegence test; here are the results obtained. The average score of the entire population is 75 in the entire test. Is there any significant difference(with a significance level of 95%) between the sample and population means, assuming that the variance of the population is not known.

Scores: 65, 78, 88, 55, 48, 95, 66, 57, 79, 81.

x<-c(65,78,88,55,48,95,66,57,79,81)

t.test(x,alternative="greater",mu=75)

Q.3) Comparing two independent sample means, taken from two population with unknown variance.

The following data shows the heights of the individuals of two different countries with unknown population variances. Is there any significant difference between the average heights of the two groups.

A: 175, 168, 168, 190, 156, 181, 182, 175, 174, 179

B: 185, 169, 173, 173, 188, 186, 175, 174, 179, 180

p<-c(175,168,168,190,156,181,182,175,174,179)

q<-c(185,169,173,173,188,186,175,174,179,180)

t.test(p,q)

Lab 10

Q.1) It is important that scientific researchers in the area of forest products be able to study correlation among the anatomy and mechanical properties of trees. For the study Quantitative Anatomical Characteristics of Plantation Grown Loblolly Pine (Pinus Taeda L.) and Cottonwood (Populus deltoides Bart. Ex Marsh.) and Their Relationships to Mechanical Properties, conducted by the Department of Forestry and Forest Products at Virginia Tech, 29 loblolly pines were randomly selected for investigation. Table shows the resulting data on the specific gravity in grams/cm3 and the modulus of rupture in kilopascals (kPa). Compute and interpret the sample correlation coefficient.

Specific Gravity, x (g/cm^3)	Modulus of Rupture, y (kPa)	Specific Gravity, x (g/cm^3)	Modulus of Rupture, y (kPa)
0.414	29,186	0.581	85,156
0.383	29,266	0.557	69,571
0.399	26,215	0.550	84,160
0.402	30,162	0.531	73,466
0.442	38,867	0.550	78,610
0.422	37,831	0.556	67,657
0.466	44,576	0.523	74,017
0.500	46,097	0.602	87,291
0.514	59,698	0.569	86,836
0.530	67,705	0.544	82,540
0.569	66,088	0.557	81,699
0.558	78,486	0.530	82,096
0.577	89,869	0.547	75,657
0.572	77,369	0.585	80,490
0.548	67,095		

x<-c(0.414,0.383,0.399,0.402,0.442,0.422,0.466,0.500,0.514,0.530,0.569,0.558,0.577,0.572,0.548,0.581,0.557,0.550,0.531,0.550,0.556,0.523,0.602,0.569,0.544,0.557,0.530,0.547,0.585)

y<-c(29186,29266,26215,30162,38867,37831,44576,46097,59698,67705,66088,78486,89869,77369,67095,85156,69571,84160,73466,78610,67657,74017,87291,86836,82540,81699,82096,75657,80490)

xi<-(x-(mean(x)))

yi<-(y-(mean(y)))

a<-(xi*yi)

a<-sum(a)

a1<-sum(xi^2)

b1<-sum(yi^2)

p<-(a1*b1)

p<-sqrt(p)

```
r<-a/p
r
```

Q.2) Compute and interpret the correlation coefficient for the following grades of 6 students selected at random:

Mathematics grade	70	92	80	74	65	83
English grade	74	84	63	87	78	90

```
x<-c(70,92,80,74,65,83)
y<-c(74,84,63,87,78,90)
xi<-(x-(mean(x)))
yi<-(y-(mean(y)))
a<-(xi*yi)
a<-sum(a)
a1<-sum(xi^2)
b1<-sum(yi^2)
p<-(a1*b1)
p<-sqrt(p)
r<-a/p
r
```

Q.3) Assume that x and y are random variables with a bivariate normal distribution.

Calculate r.

Individual	Arm Strength, x	Dynamic Lift, y
1	17.3	71.7
2	19.3	48.3
3	19.5	88.3
4	19.7	75.0
5	22.9	91.7
6	23.1	100.0
7	26.4	73.3
8	26.8	65.0
9	27.6	75.0
10	28.1	88.3
11	28.2	68.3
12	28.7	96.7
13	29.0	76.7
14	29.6	78.3
15	29.9	60.0
16	29.9	71.7
17	30.3	85.0
18	31.3	85.0
19	36.0	88.3
20	39.5	100.0
21	40.4	100.0
22	44.3	100.0
23	44.6	91.7
24	50.4	100.0
25	55.9	71.7

```
x<-c(17.3,19.3,19.5,19.7,22.9,23.1,26.4,26.8,27.6,28.1,28.2,28.7,29.0,29.6,29.9,29.9,30.3,31.3,36.0,39.5,40.4,44.3,44.6,50.4,55.9)

y<-c(71.7,48.3,88.3,75.0,91.7,100.0,73.3,65.0,75.0,88.3,68.3,96.7,76.7,78.3,60.0,71.7,85.0,85.0,88.3,100.0,100.0,100.0,91.7,100.0,71.7)

xi<-(x-(mean(x)))

yi<-(y-(mean(y)))

a<-(xi*yi)

a<-sum(a)

a1<-sum(xi^2)

b1<-sum(yi^2)

p<-(a1*b1)

p<-sqrt(p)

r<-a/p

r
```

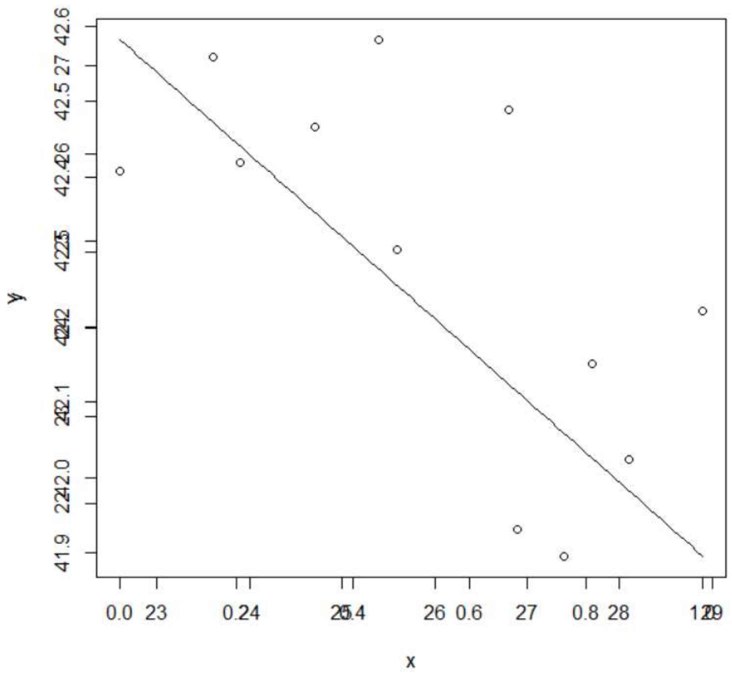
Lab 11

Q.1) In a certain type of metal test specimen, the normal stress on a specimen is known to be functionally related to the shear resistance. The following is a set of coded experimental data on the two variables:

Normal Stress, x	Shear Resistance, y
26.8	26.5
25.4	27.3
28.9	24.2
23.6	27.1
27.7	23.6
23.9	25.9
24.7	26.3
28.1	22.5
26.9	21.7
27.4	21.4
22.6	25.8
25.6	24.9

- Estimate the shear resistance for a normal stress of 24.5.
- Plot the data; does it appear that a simple linear regression will be a suitable model?

```
x<-c(26.8,25.4,28.9,23.6,27.7,23.9,24.7,28.1,26.9,27.4,22.6,25.6)
y<-c(26.5,27.3,24.2,27.1,23.6,25.9,26.3,22.5,21.7,21.4,25.8,24.9)
n<-(length(x))
xi<-sum(x)
yi<-sum(y)
a1<-sum(x^2)
b1<-sum(y^2)
p<-sum(x*y)
b<-((n*p)-(xi*yi))/((n*a1)-(xi*xi))
b
a<-(yi-(b*xi))/n
a
Y=function(X)((b*X)+a)Y
Y(24.5)
plot(Y)
par(new=TRUE)
plot(x,y)
```



Q.2) A study was made by a retail merchant to determine the relation between weekly advertising expenditures and sales.

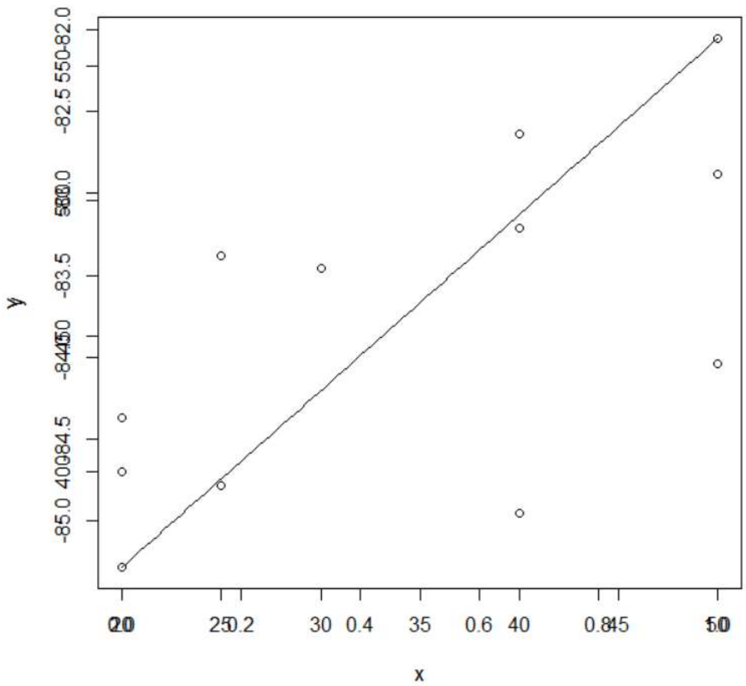
(a) Plot a scatter diagram.

(b) Find the equation of the regression line to predict weekly sales from advertising expenditures.

Advertising Costs ($)	Sales ($)
40	385
20	400
25	395
20	365
30	475
50	440
40	490
20	420
50	560
40	525
25	480
50	510

```
x<-c(40,20,25,20,30,50,40,20,50,40,25,50)

y<-c(385,400,395,365,475,440,490,420,560,525,480,510)

n<-(length(x))

xi<-sum(x)

yi<-sum(y)

a1<-sum(x^2)
```

```
b1<-sum(y^2)

p<-sum(x*y)

b<-((n*p)-(xi*yi))/((n*a1)-(xi*xi))

b

a<-(yi-(b*xi))/n

a

Y=function(X)((b*X)+a)Y

plot(Y)

par(new=TRUE)

plot(x,y)
```



Lab 12

1

On the occasion of diwali a shopkeeper wants to give cupons to the customers for the next purchase.Depends on how much the customer purchased on current day according to the following tabel

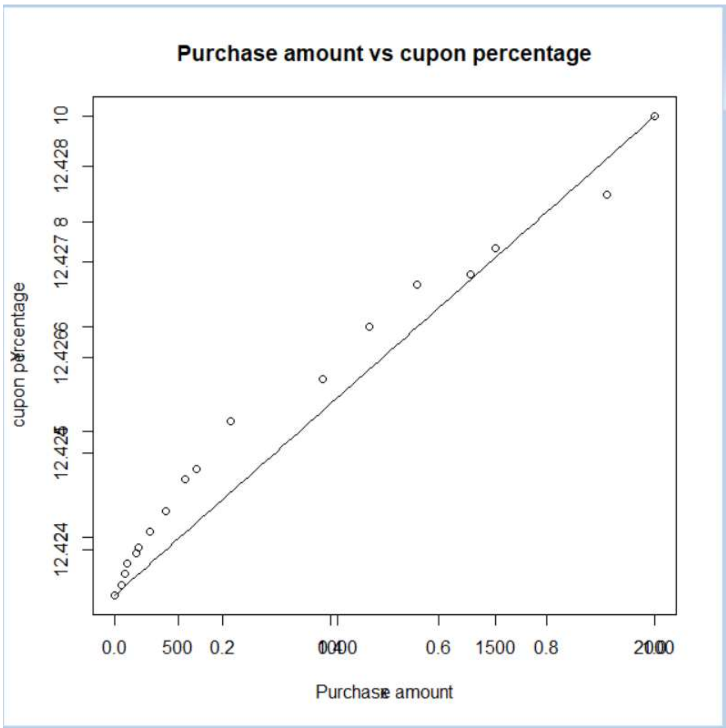
Purchased amount	Cupon percentage
300	0.9
320	1.1
331	1.3
340	1.5
368	1.7
376	1.8
410	2.1
460	2.5
520	3.1
556	3.3
665	4.2
954	5
1100	6
1250	6.8
1420	7
1500	7.5
1850	8.5
2000	10

Study the relationship between the purchased amount and cupon percentage

A) Find the equation of the straight line that fits the data best

B)Plot the above data and line,give approriate labeling to the axes

```
x<-c(300,320,331,340,368,376,410,460,520,556,665,954,1100,1250,1420,1500,1850,2000)
y<-c(0.9,1.1,1.3,1.5,1.7,1.8,2.1,2.5,3.1,3.3,4.2,5,6,6.8,7,7.5,8.5,10)
n<-(length(x))
xi<-sum(x)
yi<-sum(y)
a1<-sum(x^2)
b1<-sum(y^2)
p<-sum(x*y)
b<-((n*p)-(xi*yi))/((n*a1)-(xi*xi))
b
a<-(yi-(b*xi))/n
a
Y=function(X)((b*X)+a)
Y
plot(Y)
par(new=TRUE)
plot(x,y,xlab="Purchase amount",ylab="cupon percentage",main="Purchase amount vs cupon percentage")
```



2

Test the hypothesis that the mean maths marks of the class in a certain population is equals to 73.56.The standard deviation has a known value of 15 and maths marks of 35 students is available

83,75,69,45,48,23,68,99,23,55,68,74,25,67,54,47,33,77,76,91,95,83,50,45,41,59,79,100,63,60,70,63,67,34,42

Can we conclude that all the students crossed mean population? Perform 0.05 level of significance to check.

```
x<-c(83,75,69,45,48,23,68,99,23,55,68,74,25,67,54,47,33,77,76,91,95,83,50,45,41,59,79,100,63,60,70,63,67,34,42)

x

n<-length(x)

z<-sqrt(n)*((mean(x)-73.56)/15)

z

p<-2*(pnorm(z))

p

if(p<0.05){

print("rejected");

}else{

print("accepted");
```