Name: Yash Daga

Registration Number:20BCE7323

FINAL LAB REPORT

R-Language questions

Name: Yash Daga

Registration Number: 20BCE7323

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

```
> A=matrix(c(1:12),nrow=3,ncol=4,byrow=TRUE)

> A

     [,1] [,2] [,3] [,4]
[1,]    1    2    3    4
[2,]    5    6    7    8
[3,]    9   10   11   12
```

(b)Create same matrix with matrix function.

```
> A=matrix(c(1:12),nrow=3,ncol=4,byrow=FALSE)

> A

     [,1] [,2] [,3] [,4]
[1,]    1    4    7   10
[2,]    2    5    8   11
[3,]    3    6    9   12
```

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

```
> rownames(A) <- c("A", "B", "C")

> A

  [,1] [,2] [,3] [,4]
A    1    2    3    4
B    5    6    7    8
C    9   10   11   12
```

(d)Transpose of the matrix.

```
> t(A)
     [,1] [,2] [,3]
[1,]    1    5    9
[2,]    2    6   10
[3,]    3    7   11
[4,]    4    8   12
```

(e)Use functions cbind and rbind separately to create different matrices.

```
> B=matrix(c(1:6),nrow=3,ncol=2,byrow=TRUE)
> B
     [,1] [,2]
[1,]    1    2
[2,]    3    4
[3,]    5    6
> C=matrix(c(1:3),nrow=3,ncol=1)
> C
     [,1]
[1,]    1
[2,]    2
[3,]    3
> cbind(B,C)
     [,1] [,2] [,3]
[1,]    1    2    1
[2,]    3    4    2
[3,]    5    6    3
> D=matrix(c(6,7),nrow=1,ncol=2)
> D
     [,1] [,2]
[1,]    6    7
```

```
> rbind(B,D)
     [,1] [,2]
[1,]    1    2
[2,]    3    4
[3,]    5    6
[4,]    6    7
```

(f)Use arbitrary numbers to create matrix.

```
E=matrix(sample(1:6,6))
> E
     [,1]
[1,]    2
[2,]    3
[3,]    1
[4,]    4
[5,]    6
[6,]    5
```

(g)Verify matrix multiplication.

```
> 2*A
     [,1] [,2] [,3] [,4]
[1,]    2    4    6    8
[2,]   10   12   14   16
[3,]   18   20   22   24
```

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.

```
> sample(1:40,5)
```

```
[1] 30 18 36 10 25
```

(b)Notice that the default behaviour of sample is sampling without replacement. That is the samples will not contain the same number twice, and obviously cannot be bigger than the length of the vector to be sampled. If you want sampling with replacement, then you need to add the argument replace=T RU E. Sampling with replacement is suitable for modelling coin tosses orthrows of a die. So, for instance, simulate 10 coin tosses.

```
> coin=c("T","H");

> p <- sample(coin,10,replace=T)

> p
```

```
 [1] "H" "T" "H" "T" "T" "H" "T" "H" "H" "H"
```

(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 the prob argument to sample.

```
surgergy =c("Success", "Fail")

> p <- sample(surgergy,10,prob=c(.9, .1),replace=T)

> p
```

```
 [1] "Success" "Success" "Success" "Success" "Success" "Success" "Fail"

 [8] "Success" "Success" "Success"
```

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

```
factorial(40)/factorial(35)*factorial(5)
```

```
[1] 9475315200
```

(e)Find 5!

```
> factorial(5)
```

```
[1] 120
```

Date: 27-02-2021

Lab 1

Q1)

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

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

> x

[1] 2 5 3 7 1 9 6
```

b) Find the number of elements in x.

```
>  length(x)

[1] 7
```

c)Find the last element of x.

```
> x[l]

[1] 6
```

d) Find the minimum and maximum elements of x.

```
> min(x)

[1] 1

> max(x)

[1] 9
```

Q2. Enter the data {1,2,……,19,20} in a variable x.

```
> x=1:20
> x
 [1]  1 2 3 4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
```

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

```
> x[3]
[1] 3
```

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

```
> x[c(3,5)]
[1] 3 5
```

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

```
> x[c (2,5,6,12)]
[1]  2 5 6 12
```

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

```
> rev(x)
 [1] 20 19 18 17 16 15 14 13 12 11  10  9 8  7 6 5  4  3  2  1
```

Name: Yash Daga
Registration Number:
20BCE7323

Date: 27-02-2021

Lab 2

Problem Set 1

Q.1) Few simple statistical measures:

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

```
> x=1:10
```

(b) Find sum of the numbers.

```
> sum(x)
[1] 55
```

(c) Find Mean, median

```
> mean(x)
[1] 5.5
> median(x)
[1] 5.5
```

(d) Find sum of squares of these values.

```
> y
[1]   1   4   9  16  25  36  49  64  81 100
> sum(y)
[1] 385
```

(e) Find the value of $1/n\sum_{i=1}^{n} |x_i - \bar{x}|$, this is known as mean deviation about mean ($MD\bar{x}$).

```
> x=1:10
> z<- abs((x) - mean(x))
> z
```

```
[1] 4.5 3.5 2.5 1.5 0.5 0.5 1.5 2.5 3.5 4.5
> meandev <- (1/length(x))*sum(z)
> meandev
[1] 2.5
```

(f) Check whether $MD_{\bar{x}}$ is less than or equal to standard deviation.

```
> meandev<=sd(x)
[1] TRUE
```

Name: Yash Daga
Registration Number:
20BCE7323

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

Ans 5 rows and 20 columns

>nrow(roomsinfo)

[1] 20

>ncol(roominfo)

[1] 5

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

Ans > matrix(roomsinfo)

    [,1]

[1,] Numeric,20

[2,] Numeric,20

[3,] Numeric,20

[4,]Numeric,20

[5,]Character,20

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

Ans Variable: price,FloorArea,Rooms,Age,CentralHeating,roomsinfo

> ls(roomsinfo)
[1]”Age”   “CentralHeating” “FloorArea”   “price”

[5]”Rooms”

(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?

Ans roomsinfo [1,]

  price FloorArea Rooms Age Centralheating

1   52   1225   3   6   Yes

roomsinfo [20,]

  price FloorArea Rooms Age Centralheating

20   92 1225   7   3   Yes

(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?

Ans >roomsinfo[1:5,1:3]

	price	floorArea	rooms
1	52.00	1225	3
2	54.00	1230	3
3	57.50	1200	3
4	57.50	1000	2
5	59.75	1420	4

(f)How to get 1st, 3rd, 6th, and 10th row and 2nd, 4th, and 5th columns?

>roomsinfo[c(1,2,6,10),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

(g)How to get values in a specific row or a column?

>roomsinfo [5,]

	Price	FloorArea	rooms	age	CentralHeating
5	59.75	1420	4	1.9	YES

>roomsinfo [,4]

[1] 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

[20] 3.4

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.

Ans

```
>mean(roomsinfo$price)

[1] 71.55

>median(roomsinf$price)

[1] 69.875

>sd(roomsinfo$price)

[1] 12.2664

>mean(roomsinfo$age)

[1] 4.875

>median(roomsinfo$age)

[1] 5.4

>sd(roomsinfo$age)

[1] 2.366182

>mean(roomsinfo$floorArea)

[1] 1610.75

>median(roomsinfo$floorArea)

[1] 1605

>sd(roomsinfo$floorArea)

[1] 331.9649

>mean(roomsinfo$rooms)

[1] 5

>median(roomsinfo$rooms)

[1] 5.5

>sd(roomsinfo$rooms)
```

[1] 1.6543

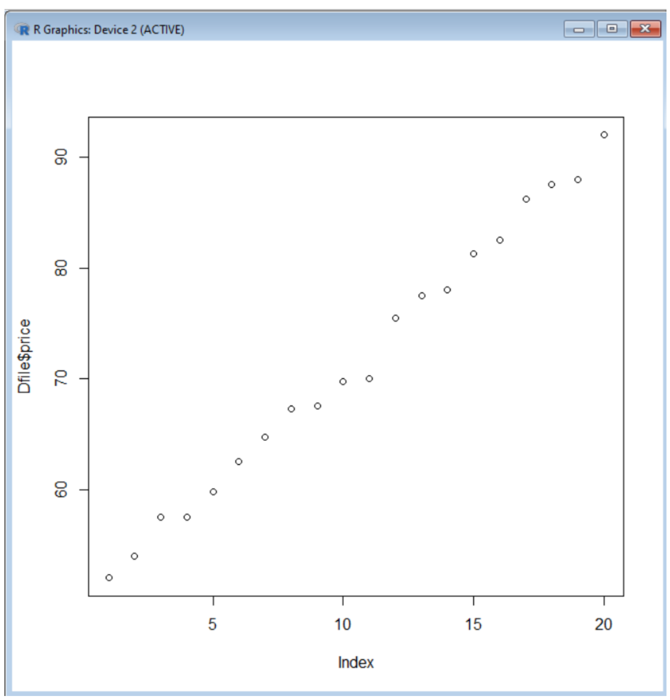
b)How many houses have central heating and how many don't have?

Ans

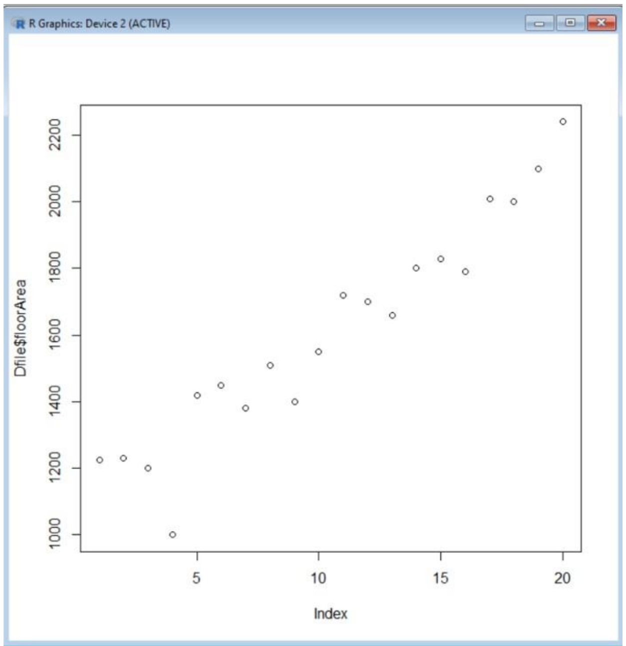
```
>sum(roomsinfo$centralHeating=="YES")
[1] 10
>sum(roomsinfo$centralHeating=="NO")
[1] 10
>
```

c) Plot Price vs. Floor, Price vs. Age, and Price vs. Rooms, in separategraphs.

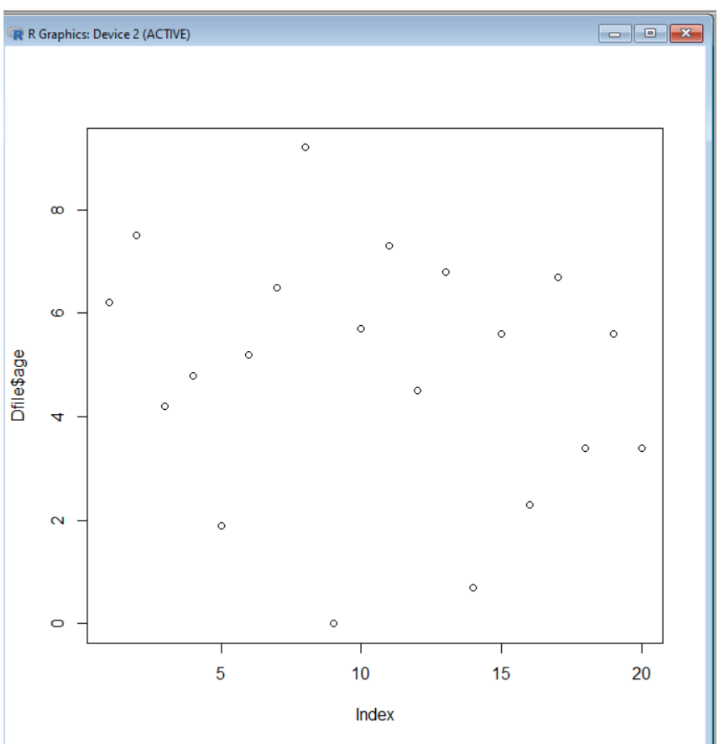
```
>plot(roomsinfo$price)
```




```
>plot(roomsinfo$floorArea)
```

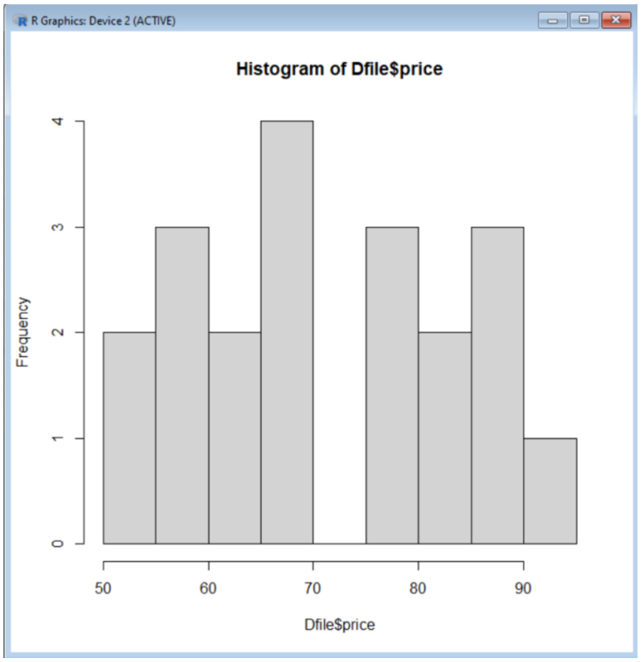


>plot(roomsinfo$age)

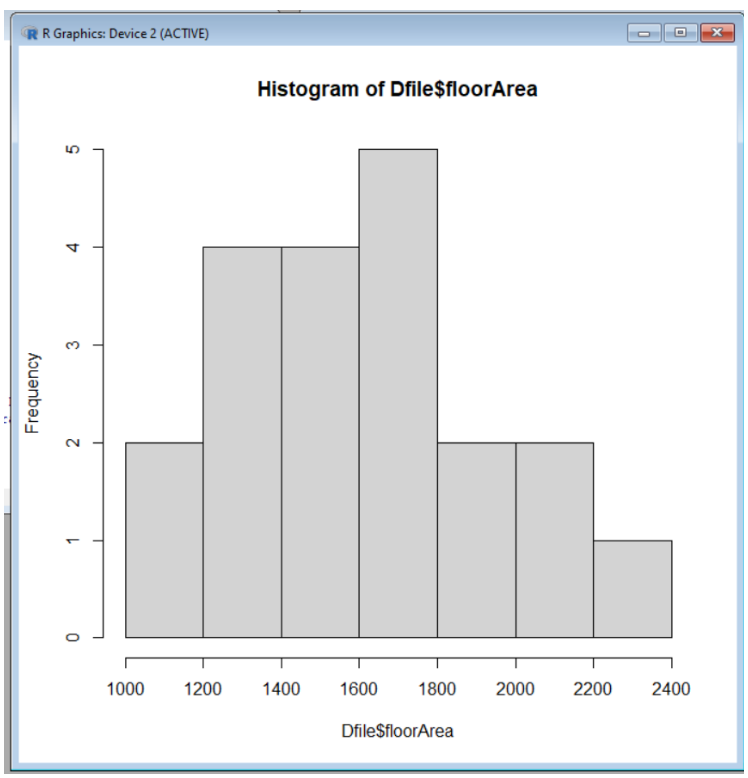


d)Draw histograms of Prices, Floor Area, and Age.

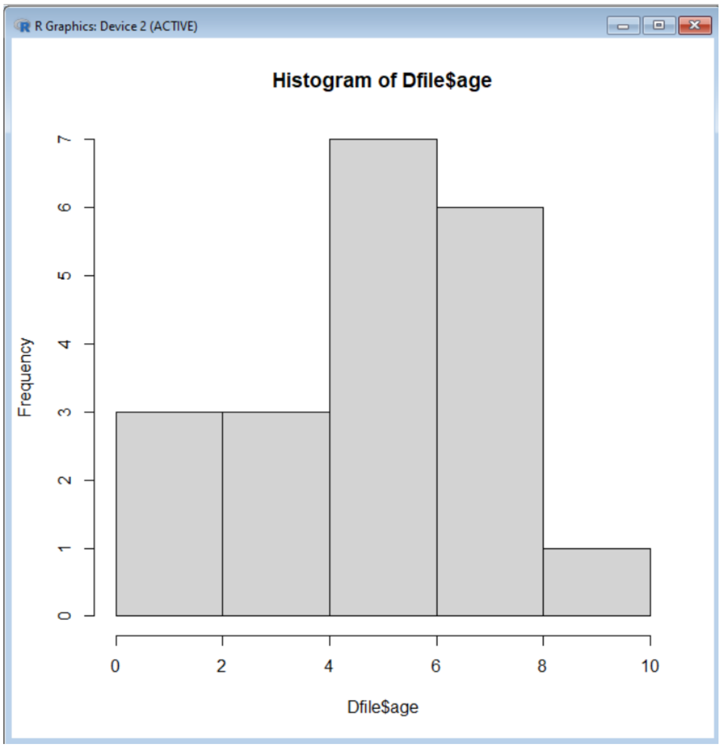
```
>hist(roomsinfo$price)
```




```
>hist(roomsinfo$floorArea)
```

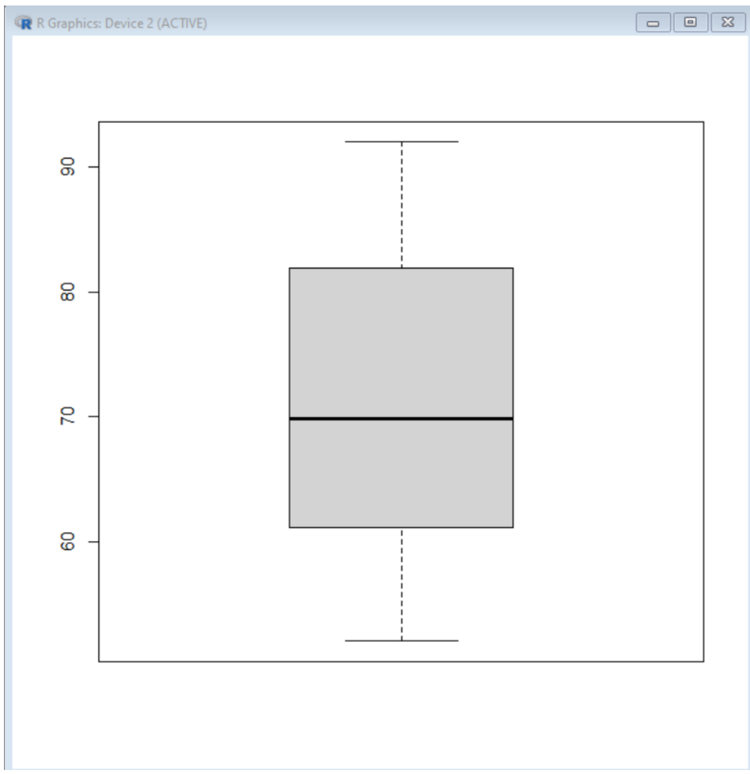


>hist(roomsinfo$age)

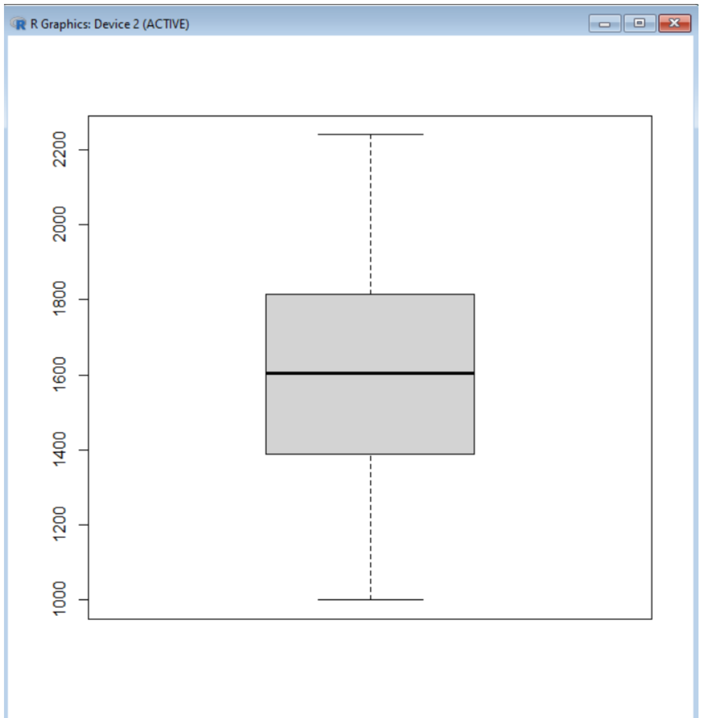


e)Draw box plots of Price, Floor Area, and Age.

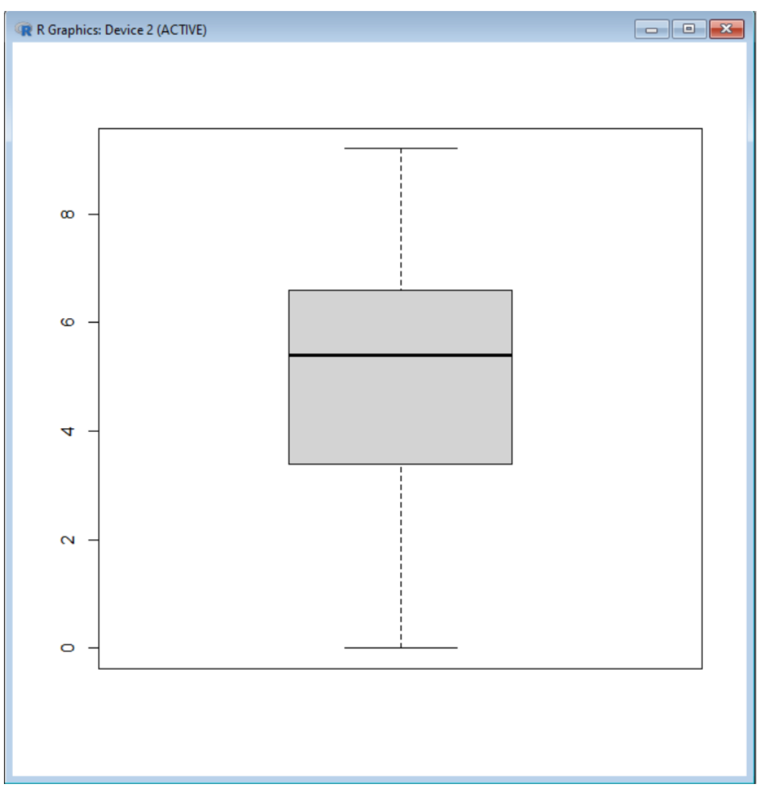
>boxplot(roomsinfo$price)



>boxplot(roomsinfo$floorArea)

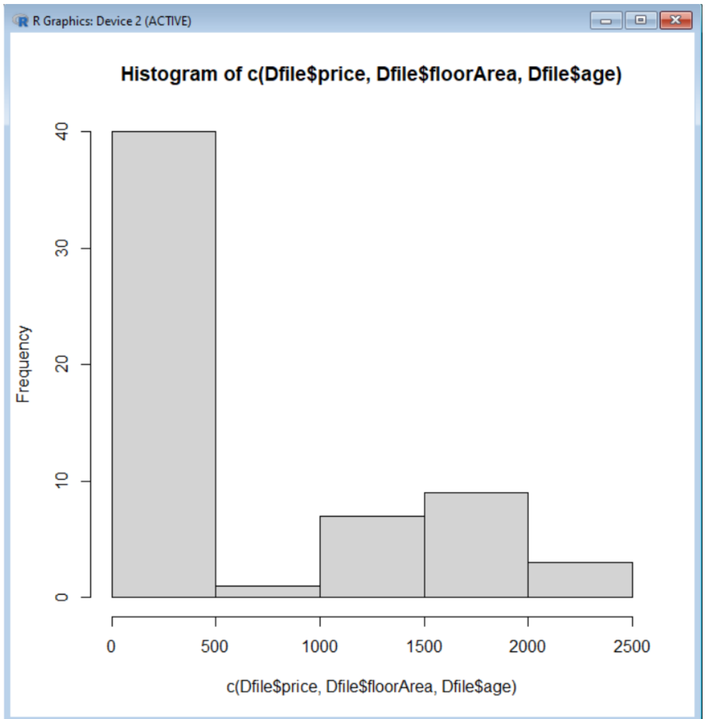


>boxplot(roomsinfo$age)

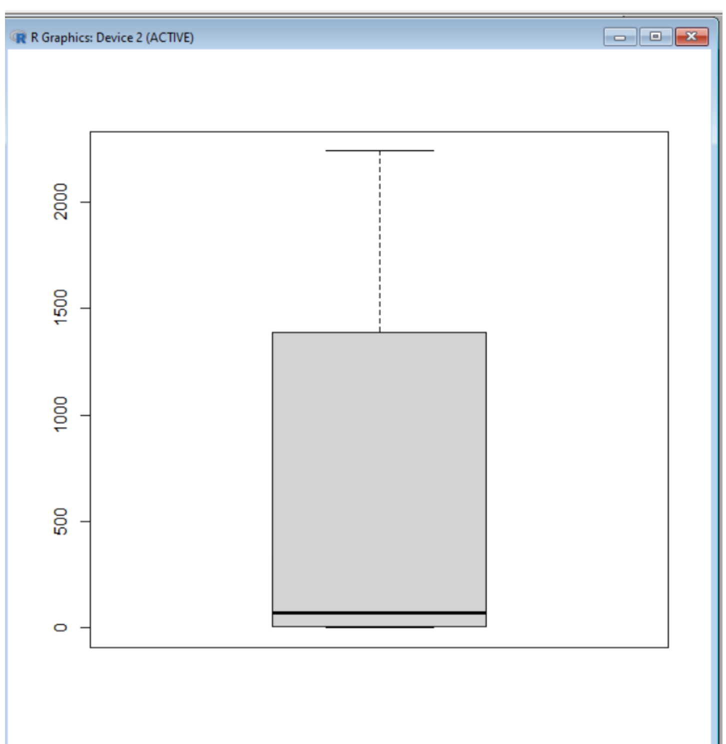


(f.) Draw all the graphs in (c), (d), and (e) in the same graph paper.

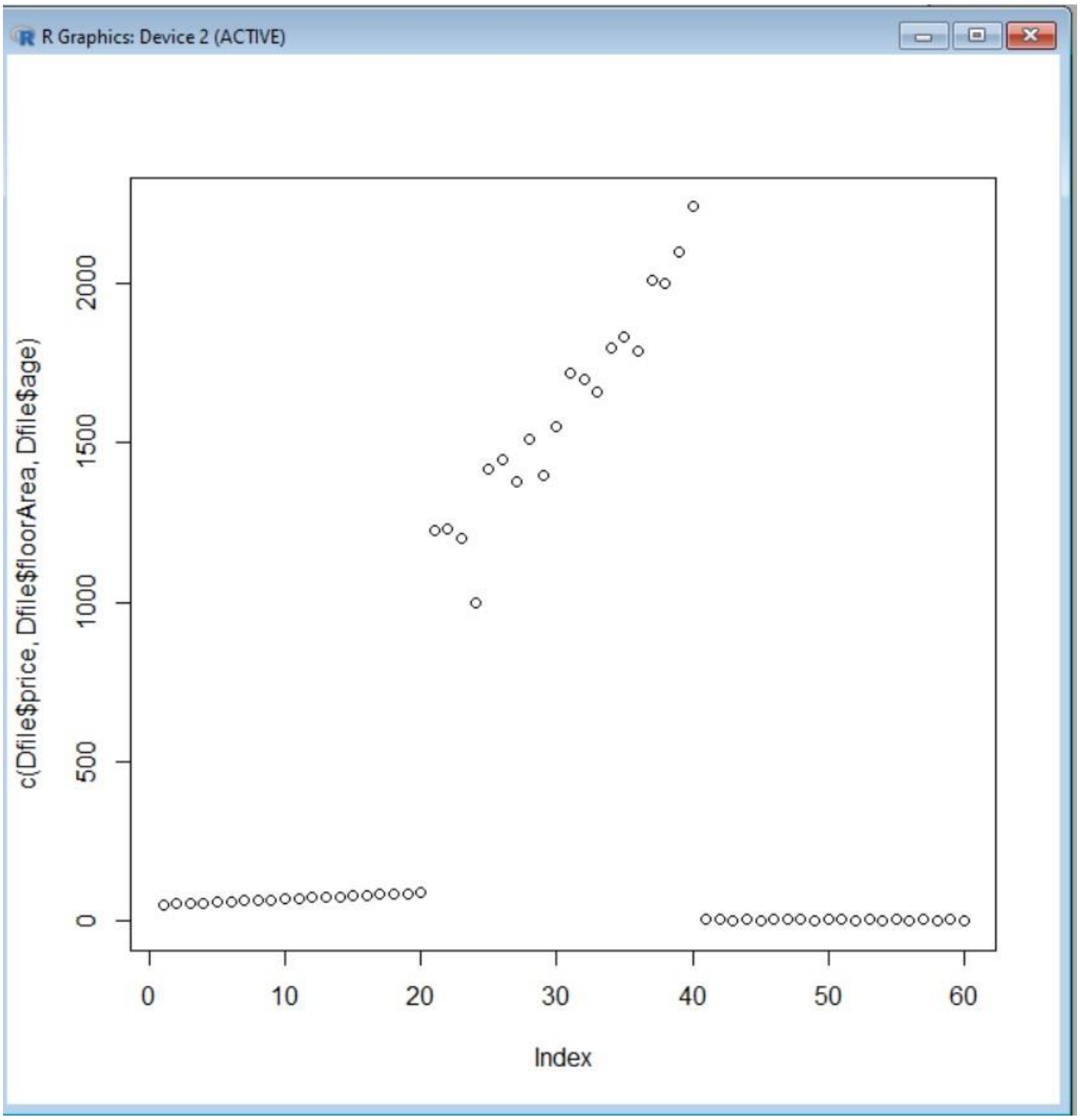
hist(c(roomsinfo$price, roomsinfo$floorArea, roomsinfo$age)



>boxplot(c(roomsinfo$price, roomsinfo$floorArea, roomsinfo$age))



>plot(c(roomsinfo$price, roomsinfo$floorArea, roomsinfo$age))



Name: Yash Daga

Registration Number: 20BCE7323

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

Ans > z<-list(rep(4,3),rep(3,3),rep(5,3))

> z

[[1]]

[1] 4 4 4

[[2]]

[1] 3 3 3

[[3]]

[1] 5 5 5

(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.

Ans > d<-list(rep(4,10),rep(6,10),rep(3,10))

> d

[[1]]

[1] 4 4 4 4 4 4 4 4 4 41

[[2]]

[1] 6 6 6 6 6 6 6 6 6 6

[[3]]

[1] 3 3 3 3 3 3 3 3 3 3

(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

Ans > f<-list (rep(3,4),rep(1:2,2),rep(5,3),rep(4,2),rep(7,6),6,34,21,54)

> f

[[1]]

[1] 3 3 3 3

[[2]]

[1] 1 2 1 2

[[3]]

[1] 5 5 5

[[4]]

[1] 4 4

[[5]]

[1] 7 7 7 7 7 7

[[6]]

[1] 6

[[7]]

[1] 34

[[8]]

[1] 21

[[9]]

[1] 54

(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 9.

Ans > a<-list(2,1,3,4)

> b<-list(5,7,12,6,8)

> c<-append(a,b)

> c

[[1]]

[1] 2

[[2]]

[1] 1

[[3]]

[1] 3

[[4]]

[1] 4

[[5]]

[1] 5

```
[[6]]
[1] 7

[[7]]
[1] 12

[[8]]
[1] 6

[[9]]
[1] 8

> length(c)
[1] 9
```

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.

Ans > x=seq(from=3,by=0.5,length.out=9)

> x

[1] 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5 7.0

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

Ans > a=seq(from=2,to=14)

> a

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

> i=a%%2==0

> a[i]

[1] 2 4 6 8 10 12 14

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

Ans > 2*x

[1] 6 7 8 9 10 11 12 13 14

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

a)Mean Median Standard Deviation for Math

> mean(A$Math)

[1] 20

> median(A$Math)

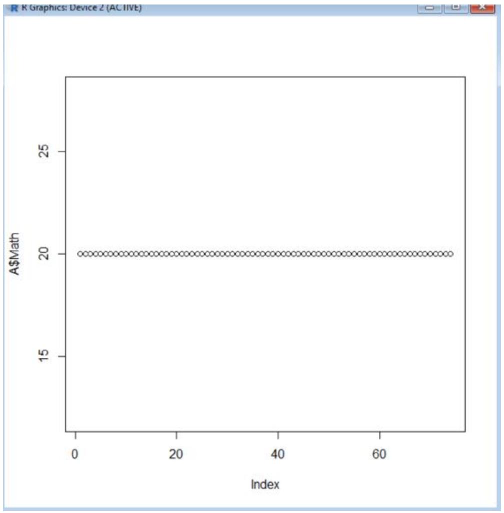
[1] 20

> sd(A$Math)

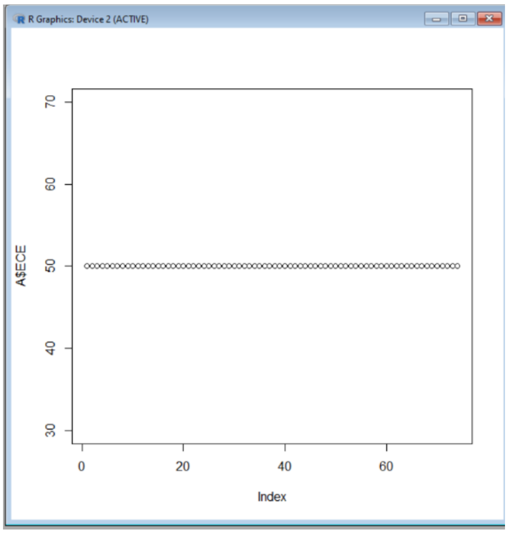
[1] 0

b) Graphs for Math, ECE

Ans > plot(A$Math)

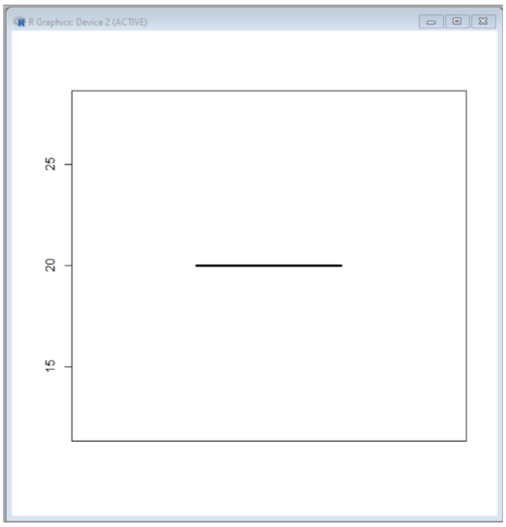


> plot(A$ECE)



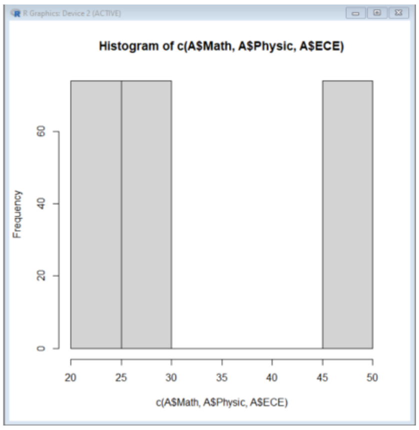
c)Box for Math

> boxplot(A$Math)



d)All graphs in one

> hist(c(A$Math,A$Physic,A$ECE))



Name – Yash Daga   Roll Number – 20BCE7323

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

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

Ans >dbinom(x = 3, size = 5, prob = 0.95)

[1] 0.02143438

(b)Find all the probabilities.

Ans >x <- 0:5
>dbinom(x, size = 5, prob = 0.95)
[1] 0.0000003125
[2]0.0000296875
[3]0.0011281250
[4]0.0214343750
[5]0.2036265625
[6] 0.7737809375

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

>x <- 0:20
>dbinom(x, size = 20, prob = 0.2)
[1] 1.152922e-02
[2]5.764608e-02
[3]1.369094e-01
[4]2.053641e-01
[5]2.181994e-01
[6] 1.745595e-01
[7]1.090997e-01
[8]5.454985e-02
[9]2.216088e-02
[10]7.386959e-03
[11] 2.031414e-03
[12]4.616849e-04

[13]8.656592e-05
[14]1.331783e-05
[15]1.664729e-06
[16] 1.664729e-07
[17]1.300570e-08
[18]7.650410e-10
[19]3.187671e-11
[20]8.388608e-13
[21] 1.048576e-14

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 foundin error each day.?

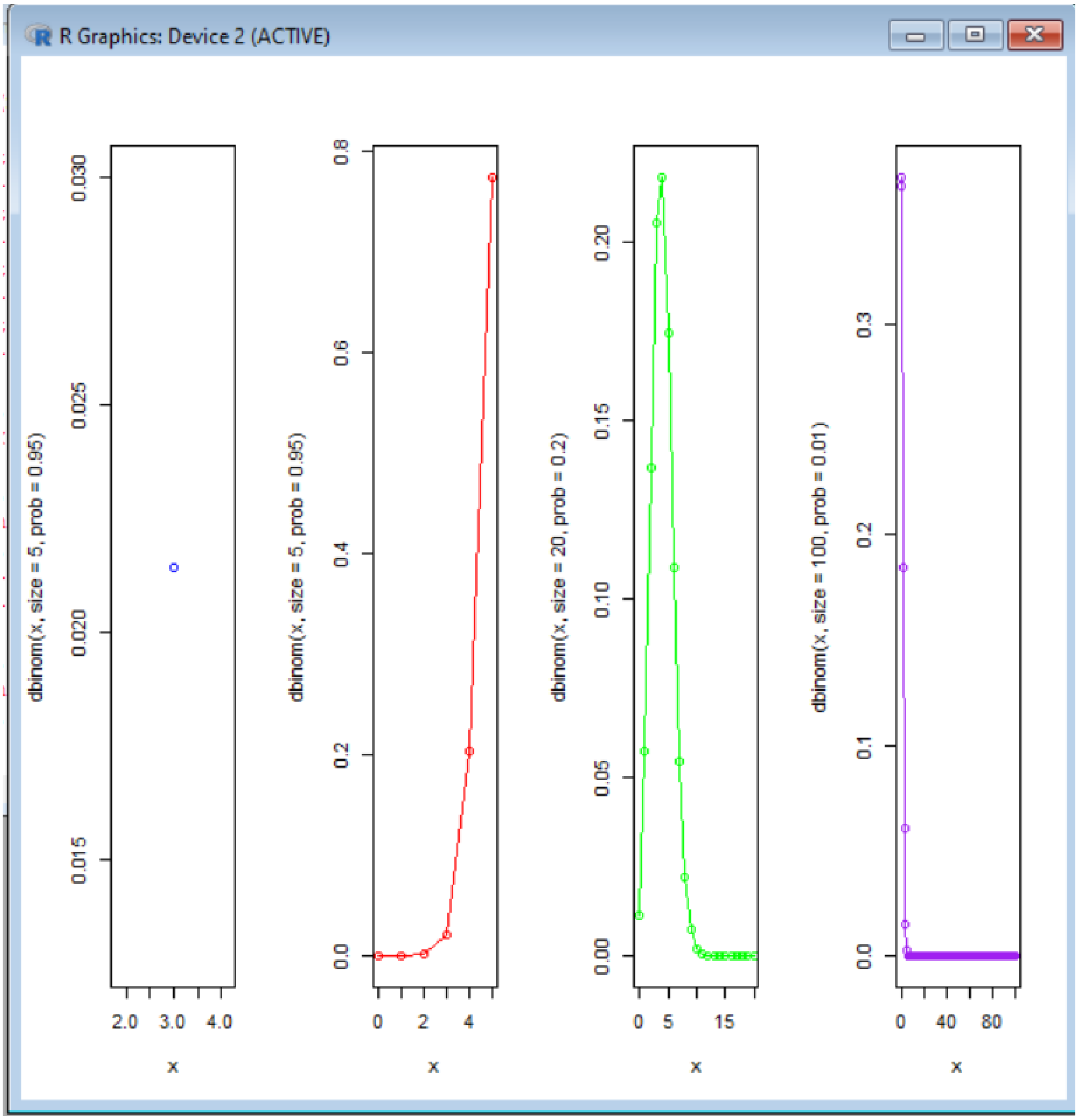
Ans

```
>x <- 0:100
>dbinom(x, size = 100, prob = 0.01)
[1]  3.660323e-01  3.697296e-01  1.848648e-01  6.099917e-02
1.494171e-02
 [6]  2.897787e-03  4.634508e-04  6.286346e-05  7.381694e-06
7.621951e-07
 [11]  7.006036e-08  5.790112e-09  4.337710e-10  2.965956e-11
1.861747e-12
 [16]  1.078184e-13  5.785707e-15  2.887697e-16  1.344999e-17
5.863367e-19
 [21]  2.398650e-20  9.230014e-22  3.347893e-23  1.146841e-24
3.716614e-26
 [26]  1.141263e-27  3.325359e-29  9.206008e-31  2.424382e-32
6.079954e-34
 [31]  1.453457e-35  3.315151e-37  7.220499e-39  1.502889e-40
2.991491e-42
 [36]  5.698078e-44  1.039212e-45  1.815713e-47  3.040667e-49
4.882708e-51
 [41]  7.521343e-53  1.111802e-54  1.577594e-56  2.149411e-58
2.812590e-60
 [46]  3.535467e-62  4.269888e-64  4.955382e-66  5.526836e-68
5.924459e-70
 [51]  6.103988e-72  6.044749e-74  5.753549e-76  5.263395e-78
4.627377e-80
 [56]  3.909263e-82  3.173103e-84  2.474154e-86  1.852815e-88
1.332276e-90
```

[61] 9.195842e-93 6.090970e-95 3.870118e-97 2.357936e-99
1.376951e-101
[66] 7.703225e-104 4.126306e-106 2.115098e-108 1.036813e-110
4.856976e-113
[71] 2.172673e-115 9.273039e-118 3.772701e-120 1.461680e-122
5.387028e-125
[76] 1.886367e-127 6.267832e-130 1.973343e-132 5.877609e-135
1.653336e-137
[81] 4.383845e-140 1.093365e-142 2.558996e-145 5.605686e-148
1.145943e-150
[86] 2.178859e-153 3.838722e-156 6.239651e-159 9.310773e-162
1.268066e-164
[91] 1.565513e-167 1.737722e-170 1.717116e-173 1.492009e-176
1.122294e-179
[96] 7.159768e-183 3.766713e-186 1.568973e-189 4.851495e-193
9.900000e-197
[101] 1.000000e-200

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

```
>par(mfrow = c(1,4))
>x = 3
>plot(x,dbinom(x, size = 5, prob = 0.95),type = 'o', col = 'blue')
>x <- 0:5
>plot(x,dbinom(x, size = 5, prob = 0.95),type = 'o', col = 'red')
>x <- 0:20
>plot(x,dbinom(x, size = 20, prob = 0.2),type = 'o', col = 'green')
>x <- 0:100
>plot(x,dbinom(x, size = 100, prob = 0.01),type = 'o', col = 'purple')
```


R Graphics: Device 2 (ACTIVE)
dbinom(x, size = 5, prob = 0.95)
0.015
0.020
0.025
0.030
2.0
3.0
4.0
x
dbinom(x, size = 5, prob = 0.95)
0.0
0.2
0.4
0.6
0.8
0
2
4
x
dbinom(x, size = 20, prob = 0.2)
0.00
0.05
0.10
0.15
0.20
0
5
15
x
dbinom(x, size = 100, prob = 0.01)
0.0
0.1
0.2
0.3
0
40
80
x

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.

```
>y <- 1-pbinom(3, 5, 0.95)
>y
[1] 0.9774075
 >a <- dbinom(x = 4, size = 20, prob = 0.2)
>a
[1] 0.2181994
>round(a,4)
[1] 0.2182
>g <- 4
>q <- 1- pbinom(g, size = 20, prob = 0.2)
>q
[1] 0.3703517
>round(q,4)
[1] 0.3704
```

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

(a)at least 10 survive,

```
>f = 9
>c <- 1 - pbinom(f, size = 15, prob = 0.4)
>c
[1] 0.0338333
```

(b)from 3 to 8 survive,

```
>j = 8

>g = 3

 >n <- pbinom(j, size = 15, prob = 0.4)

>n

[1] 0.9049526

>m <- n - pbinom(g, size = 15, prob = 0.4)
>m
[1] 0.8144507
```

(c)exactly 5 survive?

```
>i=5

>c <- 1 - pbinom(i, size = 15, prob = 0.4)
>c

[1] 0.5967844
```

Name – Yash Daga Roll Number – 20BCE7323

Problem Set-5

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?

Ans > dpois(6,4)

[1] 0.1041956

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

Ans > dpois(1,2)

[1] 0.2706706

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

Ans > dpois(3,2)

[1] 0.180447

Q.3) In a manufacturing process where glass products are made, defects or bubbles occur, occasionallyrendering the piece undesirable for marketing. It is known that, on average, 1 in every 1000 ofthese items produced has one or more bubbles. What is the probability that a random sample of8000 will yield fewer than 7 items possessing bubbles?

Ans > ppois(6,8)

[1] 0.3133743

Q.4) On average, 3 traffic accidents per month occur at a certain intersection. What is theprobability that in any given month at this intersection

(a)exactly 5 accidents will occur?

Ans > dpois(5,3)

[1] 0.1008188

(b)fewer than 3 accidents will occur?

Ans > ppois(2,3)

[1] 0.4231901

(c)at least 2 accidents will occur?

Ans > ppois(1,3)

[1] 0.1991483

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

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

Ans > dbinom(10,10,0.99)

[1] 0.9043821

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

Ans > dbinom(12,12,0.99)

[1] 0.8863849

Q.6) The acceptance scheme for purchasing lots containing a large number of batteries is to testno more than 75 randomly selected batteries and to reject a lot if a single battery fails. Supposethe probability of a failure is 0.001.

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

Ans > pbinom(0,75,0.001)

[1] 0.9277087

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

Ans > dpois(20,0.075)

[1] 1.209285e-41

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

Ans > x=1:75

> round(sum(dpois(x,0.001)),4)*10

[1] 0.01

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

Ans > par(mfrow = c(1,4))

> x=2

> plot(x,dpois(1,2),type='o',col='blue')

> x<-3

> plot(x,dpois(5,3),type='o',col='red')

> plot(x,ppois(2,3),type='o',col='red')

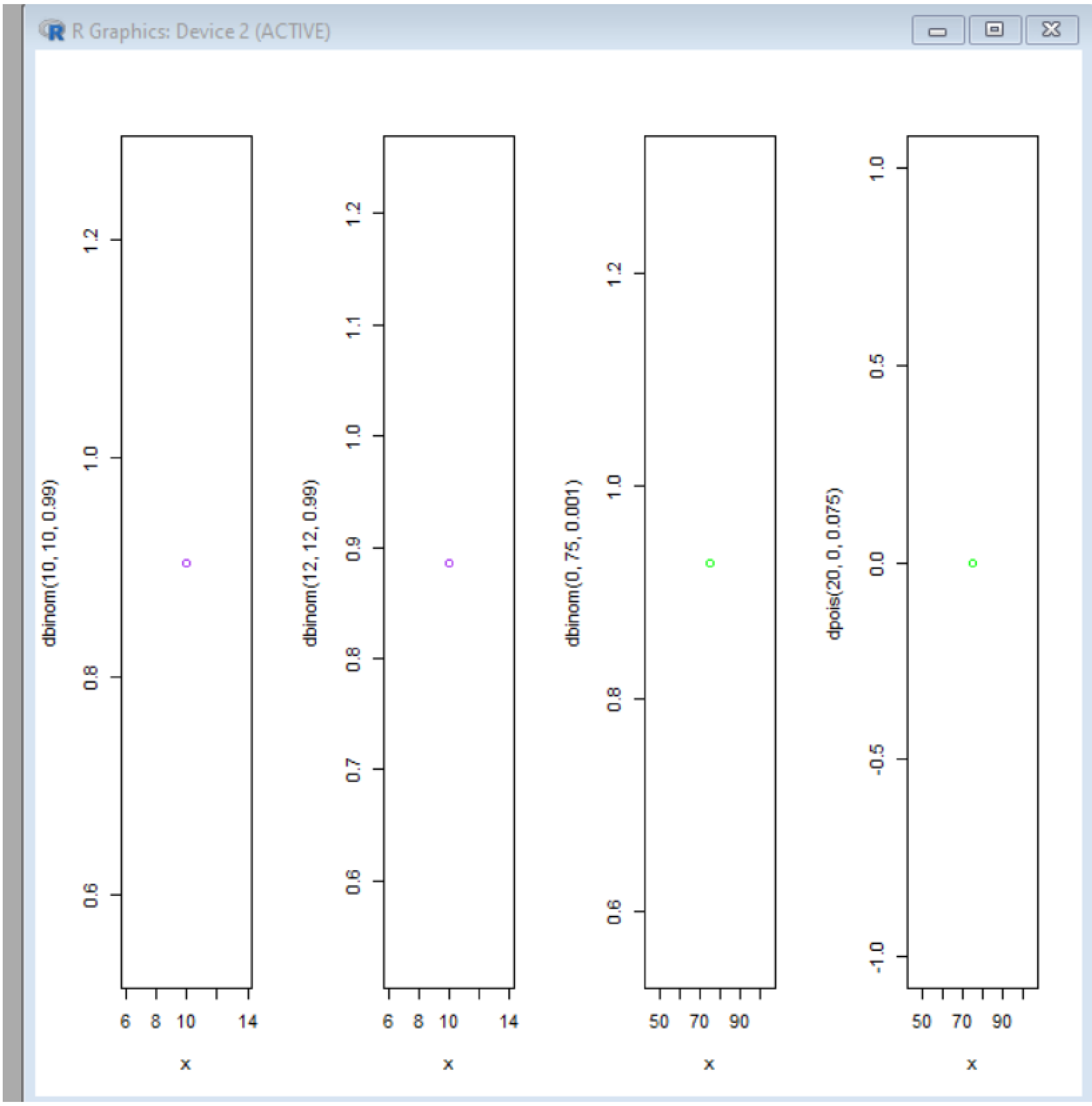
> plot(x,ppois(1,3),type='o',col='red')

> x<-10

> plot(x,dbinom(10,10,0.99),type='o',col='purple')

> plot(x,dbinom(12,12,0.99),type='o',col='purple')

```
> x<-75
> plot(x,dbinom(0,75,0.001),type='o',col='green')
> plot(x,dpois(20,0,0.075),type='o',col='green')
```



Name – Yash Daga Roll Number – 20BCE7323

Problem Set=6 (Answers)

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?
Ans > x=pnorm(125,100,15,lower.tail=TRUE)
> x
[1] 0.9522096
> x*100
[1] 95.22096

(b) What percentage of people have an IQ > 110?
Ans > x=pnorm(110,100,15,lower.tail=FALSE)
> x
[1] 0.2524925
> x*100
[1] 25.24925

(c) What percentage of people have 110 < IQ < 125?
Ans > x=pnorm(110,100,15,lower.tail=TRUE)
> x
[1] 0.7475075
> x*100
[1] 74.75075

(d) Find 25% for standard normal distribution.
Ans > qnorm(0.25,mean=0,sd=1,lower.tail=TRUE)
[1] -0.6744898

(e) Find 25% normal distribution with mean and standard deviation 2&3.
Ans > qnorm(0.25,mean=2,sd=2,lower.tail=TRUE)
[1] 0.6510205

(f) What IQ separates the lower 25% from the others?
Ans > qnorm(0.25,mean=100,sd=15,lower.tail=TRUE)
[1] 89.88265

(g) What IQ separates the top 25% from the others?
Ans > qnorm(0.25,mean=100,sd=15,lower.tail=FALSE)
[1] 110.1173

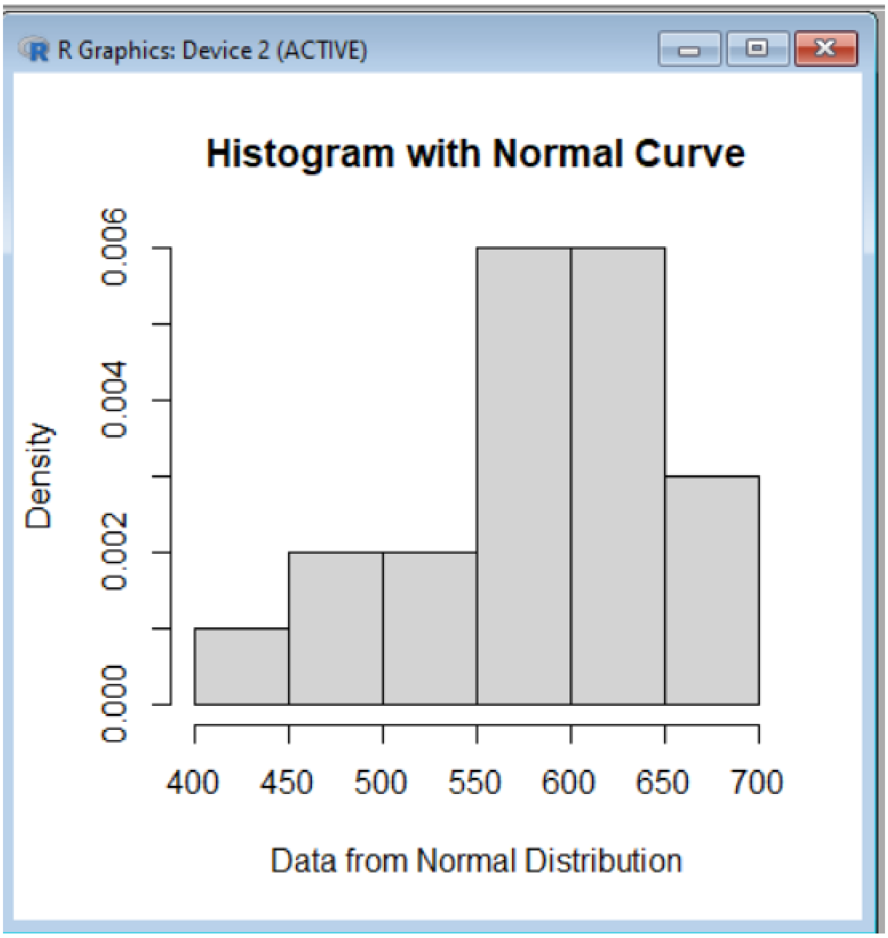
(h) Find 25 percentile for mean 100 and SD 15.
Ans > qnorm(0.25,mean=100,sd=15,lower.tail=FALSE)
[1] 110.1173

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.

```
> RandomData=rnorm(20,mean=572,sd=51)
> mean(RandomData)
[1] 570.8001
> sd(RandomData)
[1] 39.81884
> RandomData<-rnorm(20,mean=572,sd=51)
> mean(RandomData)
[1] 578.7784
> RandomData<-rnorm(20,mean=572,sd=51)
> mean(RandomData)
[1] 584.0696
> sd(RandomData)
[1] 68.1673
```

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

```
> std=sd(RandomData)

> m=mean(RandomData)

> hist(RandomData,xlab="Data from Normal Distribution",
freq=FALSE,main="Histogram with Normal Curve")
```


R Graphics: Device 2 (ACTIVE)
Histogram with Normal Curve
Density
0.000
0.002
0.004
0.006
400
450
500
550
600
650
700
Data from Normal Distribution

Name – Yash Daga Roll Number – 20BCE7323

Assignment-8

Q.1) Test the hypothesis that the mean systolic blood pressure in a certain population equals 140mmHg. The standard deviation has a known value of 20 and a data set of 55 patients is available.BBb

Ans

```
> No<-seq(1:55)

> Status <- c(rep(0,25),rep(1,30))

> M<-
c(120,115,94,118,111,102,102,131,105,107,115,139,115,113,114,105,115,134,
109,109,93,118,109,106,125,150,142,119,127,141,149,144,142,149,161,143,1
40,148,149,141,146,159,152,135,134,161,130,125,141,148,153,145,137,147,1
75)

> BP<-data.frame(No,Status,M)

> BP
```

R Console

```
> BP
   No Status   M
1   1      0 120
2   2      0 115
3   3      0  94
4   4      0 118
5   5      0 111
6   6      0 102
7   7      0 102
8   8      0 131
9   9      0 105
10 10      0 107
11 11      0 115
12 12      0 139
13 13      0 115
14 14      0 113
15 15      0 114
16 16      0 105
17 17      0 115
18 18      0 134
19 19      0 109
20 20      0 109
21 21      0  93
22 22      0 118
23 23      0 109
24 24      0 106
25 25      0 125
26 26      1 150
27 27      1 142
28 28      1 119
29 29      1 127
30 30      1 141
31 31      1 149
32 32      1 144
33 33      1 142
34 34      1 149
35 35      1 161
36 36      1 143
37 37      1 140
38 38      1 148
```

```
34 34      1 149
35 35      1 161
36 36      1 143
37 37      1 140
38 38      1 148
39 39      1 149
40 40      1 141
41 41      1 146
42 42      1 159
43 43      1 152
44 44      1 135
45 45      1 134
46 46      1 161
47 47      1 130
48 48      1 125
49 49      1 141
50 50      1 148
51 51      1 153
52 52      1 145
53 53      1 137
54 54      1 147
55 55      1 175
```

```
> MU=140
> XB=mean(BP$M)
> Sigma=20
> N=55
> ##Z-Value
> Z=(XB-MU)/(Sigma/sqrt(N))
> Z
[1] -3.660905
> ##P-Value
> P=2*pnorm(-abs(Z))
> P
[1] 0.0002513257
> if(P<0.5) {
+ print("The Null Hypothesis is Rejected")
+ } else{
+ print("The Null Hypothesis is Accepted")
+ }
[1] "The Null Hypothesis is Rejected"
```

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.

Ans

```
> a=100
> b=43
> c=b/a
> d=0.5
> e=1-d
> #Z – value
> f=(c-d)/(sqrt((d*e)/a))
> f
[1] -1.4
> ## P - value
> g=2*pnorm(-abs(f),lower.tail=FALSE)
> g
[1] 1.838487
> if(f<g){
+ print("This Coin is Fair")
+ } else{
+ print("This Coin is Not Fair")
+ }
[1] "This Coin is Fair"
```

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 0.5kilogram. Test the hypothesis that μ=8 kilograms against the alternative that μ is not equal to 8kilograms if a random sample of 50 lines is tested and found to have a mean breaking strength of7.8 kilograms. Use a 0.01 level of significance.

Ans

```
> a=7.8
> b=8
> c=0.5
> d=50
> ##Z-value
> e=(a-b)/(c/sqrt(d))
> e
[1] -2.828427
> ##P-Value
> f=2*pnorm(-abs(Z))
> f
[1] 0.0002513257
> if(e<f){
+ print("The Null Hypothesis is Rejected")
+ }else{
+ print("The Null Hypothesis is Accepted")
+ }
[1] "The Null Hypothesis is Rejected"
```

Name – Yash Daga   Roll Number – 20BCE7323

Problem Set -8

Q.1) An outbreak of salmonella-related illness was attributed to ice produced at certain factory. Scientists measured the level of salmonella in 9 randomly sampled batches ice cream. The levels (in MPN/g) were:0.593,0.142,0.3290,691,0.231,0.793,0.519,0.392,0.418Is there evidence that the mean level of salmonella in ice cream greater than 0.3 MPN/g.

Ans > 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)

```
	One Sample t-test

data:  x
t = 2.2051, df = 8, p-value = 0.02927
alternative hypothesis: true mean is greater than 0.3
95 percent confidence interval:
 0.3245133       Inf
sample estimates:
mean of x
0.4564444
```

```
> 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)

        One Sample t-test

data:  x
t = 2.2051, df = 8, p-value = 0.02927
alternative hypothesis: true mean is greater than 0.3
95 percent confidence interval:
 0.3245133       Inf
sample estimates:
mean of x
0.4564444
```

Q.2) Suppose that 10 volunteers have taken an intelligence 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.

Ans a = c(65, 78, 88, 55, 48, 95, 66, 57, 79, 81)

```
> t.test (a, mu=75)

        One Sample t-test

data:  a

t = -0.78303, df = 9, p-value = 0.4537

alternative hypothesis: true mean is not equal to 75

95 percent confidence interval:

 60.22187 82.17813

sample estimates:

mean of x

    71.2
```

> qt(0.975, 9)

[1] 2.262157

```
> a = c(65, 78, 88, 55, 48, 95, 66, 57, 79, 81)
> t.test (a, mu=75)

        One Sample t-test

data:  a
t = -0.78303, df = 9, p-value = 0.4537
alternative hypothesis: true mean is not equal to 75
95 percent confidence interval:
 60.22187 82.17813
sample estimates:
mean of x
     71.2

> qt(0.975, 9)
[1] 2.262157
> |
```

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

Ans> a = c(175, 168, 168, 190, 156, 181, 182, 175, 174, 179)

> b = c(185, 169, 173, 173, 188, 186, 175, 174, 179, 180)

> t.test(a,b, var.equal=TRUE, paired=FALSE)

Two Sample t-test

data: a and b

t = -0.94737, df = 18, p-value = 0.356

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-10.93994   4.13994

sample estimates:

mean of x mean of y

174.8   178.2

> qt(0.975, 18)

[1] 2.100922

```
> a = c(175, 168, 168, 190, 156, 181, 182, 175, 174, 179)
> b = c(185, 169, 173, 173, 188, 186, 175, 174, 179, 180)
> t.test(a,b, var.equal=TRUE, paired=FALSE)

        Two Sample t-test

data:  a and b
t = -0.94737, df = 18, p-value = 0.356
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -10.93994   4.13994
sample estimates:
mean of x mean of y
    174.8     178.2

> qt(0.975, 18)
[1] 2.100922
>
```

Name – Yash Daga Roll Number – 20BCE7323

Problem Set-9

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 Taneda L.) and Cottonwood (Populus deltoids 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 ingrams/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		

Ans: > 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(28186, 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)

> cor(x,y)

[1] 0.9434695

```
> 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
> y = c(28186, 29266, 26215, 30162, 38867, 37831, 44576, 46097, 59698, 67705, 66088, 78486, 89869, 77369, 67095, 85156, 69571, 84160
> cor(x,y)
[1] 0.9434695
>
```

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

Ans

> x2 = c(70, 92, 80, 74, 65, 83)

> y2 = c(74, 84, 63, 87, 78, 90)

> cor(x2,y2)

[1] 0.2396639

```
> x2 = c(70, 92, 80, 74, 65, 83)
> y2 = c(74, 84, 63, 87, 78, 90)
> cor(x2,y2)
[1] 0.2396639
>
```

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

Individual	Strength, x	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

Ans

```
>  x3 = 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)

>  y3 = 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)

> cor(x3,y3)

[1] 0.3916965
```

```
> x3 = 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,
> y3 = 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
> cor(x3,y3)
[1] 0.3916965
> |
```

Name – Yash Daga Roll Number – 20BCE7323

Lab-10

Q.1) In a certain type of metal test specimen, the normal stress on a specimen is known to befunctionally related to the shear resistance. The following is a set of coded experimental data onthe 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?

Ans
```
> x<-c(26.8,25.4,28.9,23.6,27.7,23.9,24.7,28.2,26.9,27.4,22.6,25.6)
> 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)
> lm(y~x)

Call:
lm(formula = y ~ x)
```

Coefficients:

(Intercept)   x

   42.5443   -0.6844

> sxy<-sum(x*y)-((sum(x)*sum(y))/length(x))

> sxx<-sum(x*x)-((sum(x)*sum(x))/length(x))

> syy<-sum(y*y)-((sum(y)*sum(y))/length(x))

> slope<-sxy/sxx

> const<-(sum(y)-(slope*sum(x)))/length(x)

> slope

[1] -0.6844133

> const

[1] 42.5443

> plot(x,y,col="blue")

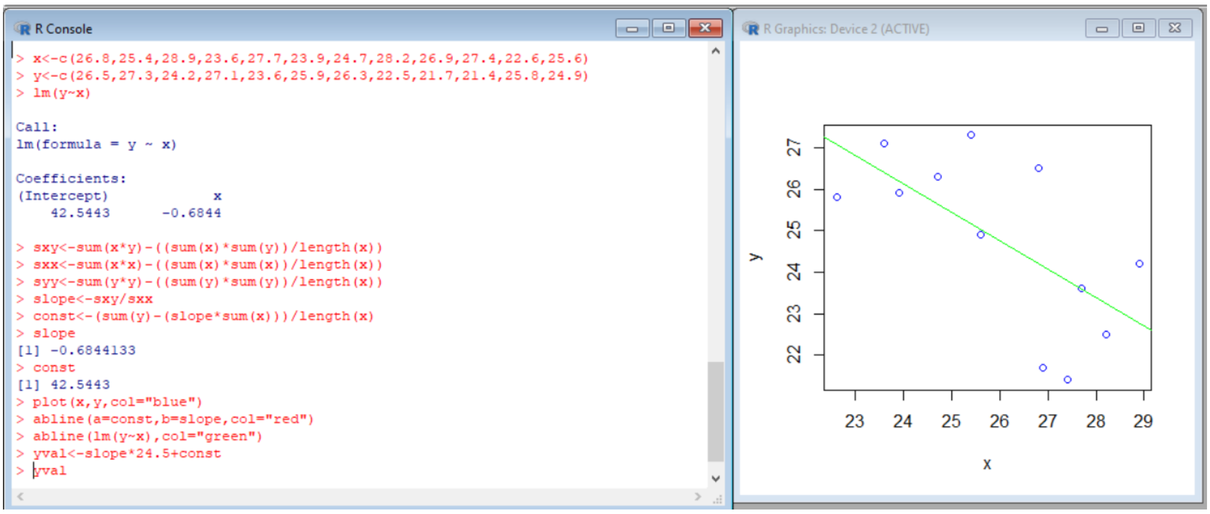
> abline(a=const,b=slope,col="red")

> abline(lm(y~x),col="green")

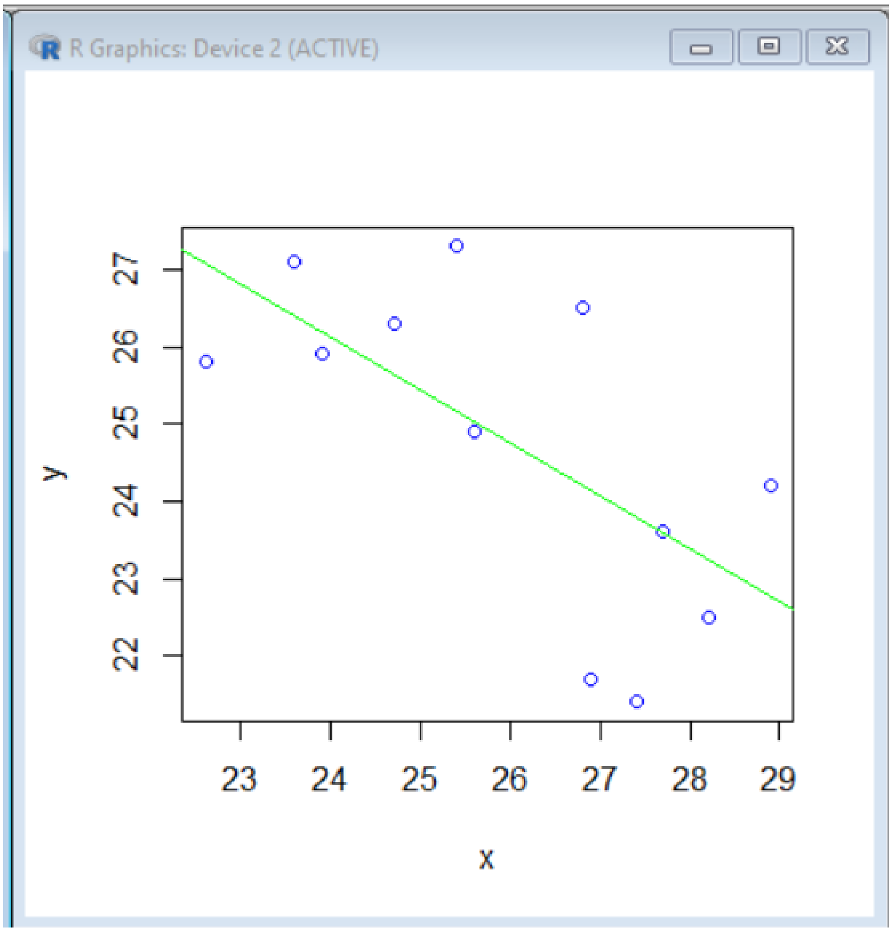
> yval<-slope*24.5+const

> yval

[1] 25.77618



```
R Console
> x<-c(26.8,25.4,28.9,23.6,27.7,23.9,24.7,28.2,26.9,27.4,22.6,25.6)
> 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)
> lm(y~x)

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
    42.5443      -0.6844

> sxy<-sum(x*y)-((sum(x)*sum(y))/length(x))
> sxx<-sum(x*x)-((sum(x)*sum(x))/length(x))
> syy<-sum(y*y)-((sum(y)*sum(y))/length(x))
> slope<-sxy/sxx
> const<-(sum(y)-(slope*sum(x)))/length(x)
> slope
[1] -0.6844133
> const
[1] 42.5443
> plot(x,y,col="blue")
> abline(a=const,b=slope,col="red")
> abline(lm(y~x),col="green")
> yval<-slope*24.5+const
> yval
```

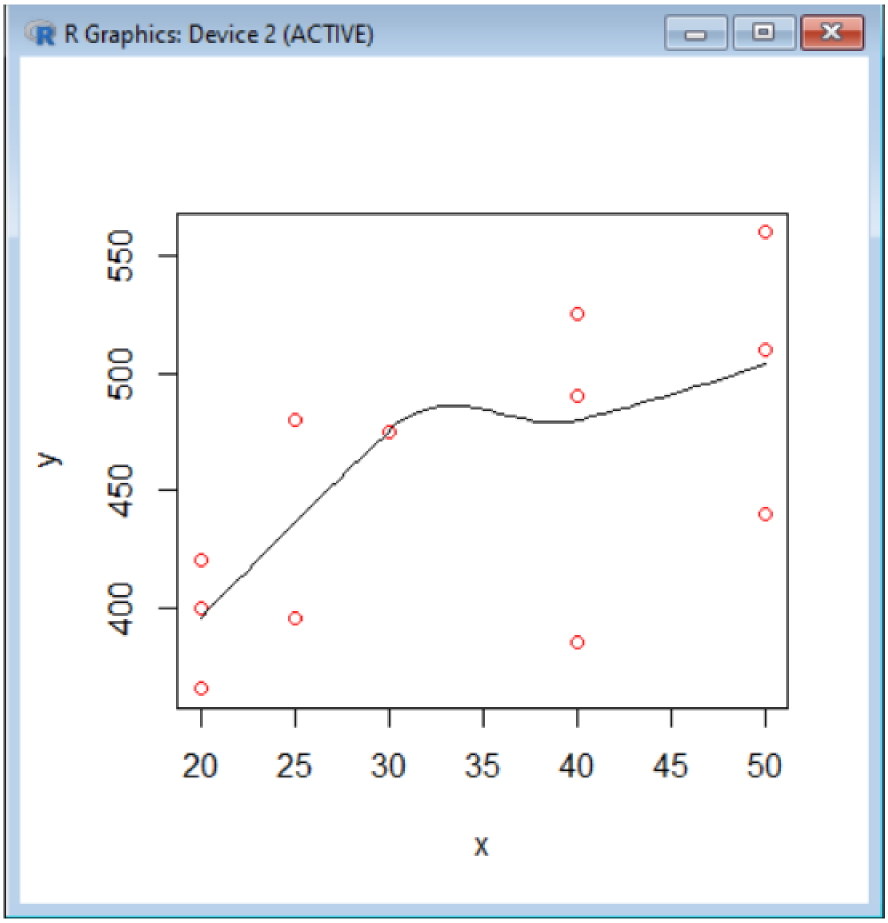


Q.2) A study was made by a retail merchant to determine the relation between weekly advertisingexpenditures and sales.(a)Plot a scatter diagram.(b)Find the equation of the regression line to predict weekly sales from advertising expenditures.

Ans:

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

```
> scatter.smooth(x,y,col="red")
> 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)
> sx<- sum(x)
> sy<-sum(y)
> sxy<- sum(x*y)
> sx2<-sum(x^2)
> lx<-length(x)
> ly<-length(y)
> lxy<-length(x*y)
> b<-(lx*sxy-sx*sy)/(lx*sx2-sx^2)
> b
[1] 3.220812
> a<-(sy-b*sx)/lx
> a
[1] 343.7056
> scatter.smooth(x,y,col="red")
> |
```



Code:

```
> 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)

> sx<- sum(x)

> sy<-sum(y)

> sxy<- sum(x*y)

> sx2<-sum(x^2)

> lx<-length(x)

> ly<-length(y)

> lxy<-length(x*y)

> b<-(lx*sxy-sx*sy)/(lx*sx2-sx^2)

> b

[1] 3.220812
```

```
> a<-(sy-b*sx)/lx
> a
[1] 343.7056
> scatter.smooth(x,y,col="red")
```

Definitions

1) Random Sampling and Probability: Random sampling, or probability sampling, is a sampling method that allows for the randomization of sample selection, i.e., each sample has the same probability as other samples to be selected to serve as a representation of an entire population.
2) Binomial distribution: A binomial distribution can be thought of as simply the probability of a SUCCESS or FAILURE outcome in an experiment or survey that is repeated multiple times.
3) Poisson Distribution: Poisson distribution is a probability distribution that can be used to show how many times an event is likely to occur within a specified period of time.
4) Normal Distribution: probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve.
5) Test of Hypothesis-z test: Z-test is a statistical test to determine whether two population means are different when the variances are known and the sample size is large. Z-test is a hypothesis test in which the **z**-statistic follows a normal distribution. A **z**-statistic, or **z**-score, is a number representing the result from the z-test.
6) Test of Hypothesis - T test: A **t**-test is a statistical test that compares the means of two samples. It is used in hypothesis testing, with a null hypothesis that the difference in group means is zero and an alternate hypothesis that the difference in group means is different from zero.
7) Correlation and Regression: Correlation quantifies the strength of the linear relationship between a pair of variables, whereas regression expresses the relationship in the form of an equation.

Application Oriented Problem

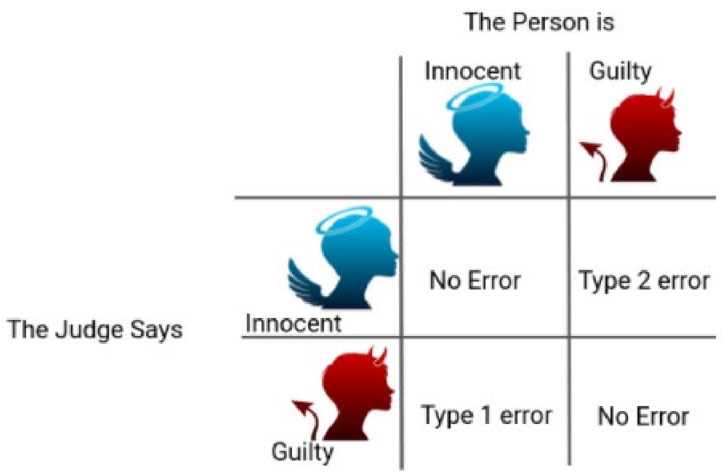
Applications of z-test and t-test problems

- **Fundamentals of Hypothesis Testing**
    - Basic Concepts – Null Hypothesis, Alternative Hypothesis, Type 1 Error, Type 2 Error, and Significance Level
    - Steps to Perform Hypothesis Testing
    - Directional Hypothesis
    - Non-Directional Hypothesis Test
- **What is the Z Test?**
    - One-Sample Z Test
    - Two Sample Z Test
- **What is the t-Test?**
    - One-Sample t-Test
    - Two-Sample t-Test
- **Deciding between the Z Test and t-Test**

Concept behind Hypothesis:

Example: A person is on trial for a criminal offense and the judge needs to provide a verdict on his case. Now, there are four possible combinations in such a case:

- First Case: The person is innocent and the judge identifies the person as innocent
- Second Case: The person is innocent and the judge identifies the person as guilty
- Third Case: The person is guilty and the judge identifies the person as innocent
- Fourth Case: The person is guilty and the judge identifies the person as guilty



Its is just an act where the analysts test their judgement.

Sample for z-test

If girls on average score higher than 600 in the exam. We have the information that the standard deviation for girls' scores is 100. So, we collect the data of 20 girls by using random samples and record their marks. Finally, we also set our α value (significance level) to be 0.05.

Score
650
730
510
670
480
800
690
530
590
620
710
670
640
780
650
490
800
600
510
700

- Mean Score for Girls is 641
- The size of the sample is 20
- The population mean is 600
- Standard Deviation for Population is 100

$$\text{z score} = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}}$$

$$= \frac{641 - 600}{100/\sqrt{20}}$$

$$= 1.8336$$

$$\text{p value} = .033357.$$

Critical Value = 1.645

Z score > Critical Value

P value < 0.05

```
> y=c(650,730,510,670,480,800,690,530,590,620,710,670,640,780,650,490,800,600,5:
> z.test(y,sigma.y=1)
```

```
        One Sample  -test

data:  y
t = 24.17, df = 19, p-value = 4.965e-16
alternative hypothesis: true mean is greater than 100
95 percent confidence interval:
 602.2968      Inf
sample estimates:
mean of x
      641

> |
```

H1: mean> 600 is true

Reject Null-hypothesis cause p-value is less than 0.05

Sample for t-test

Let's say we want to determine if on average girls score more than 600 in the exam. We do not have the information related to variance (or standard deviation) for girls' scores. To a perform t-test, we randomly collect the data of 10 girls with their marks and choose our α value (significance level) to be 0.05 for Hypothesis Testing

Girls_Score
587
602
627
610
619
622
605
608
596
592

- Mean Score for Girls is 606.8
- The size of the sample is 10
- The population mean is 600
- Standard Deviation for the sample is 13.14

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$= \frac{606.8 - 600}{13.14/\sqrt{10}}$$

$$= 1.64$$

Critical Value = 1.833

t score < Critical Value

P value = 0.0678

P value > 0.05

$H_0 : \mu \le 600$

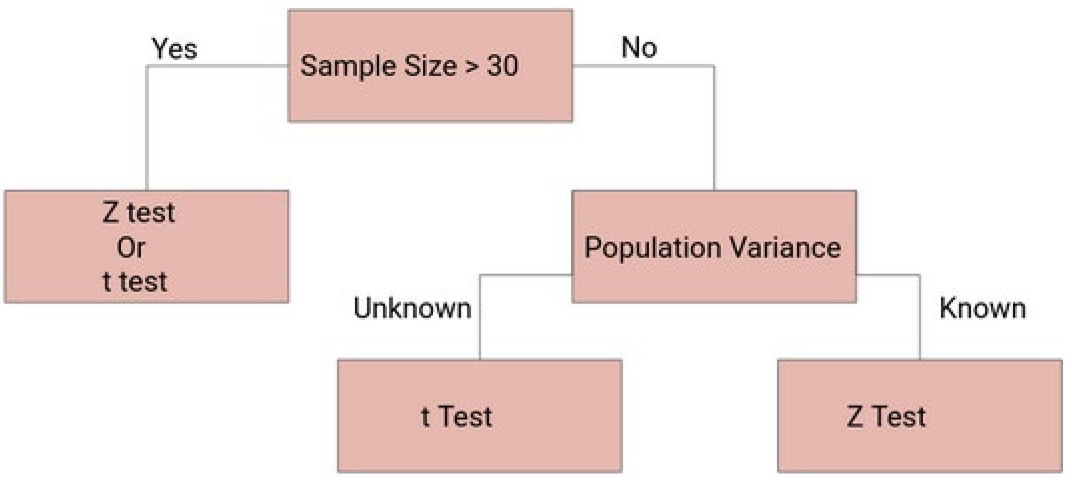
$\boxed{H_1 : \mu > 600}$

P-value is greater than 0.05 thus fail to reject the null hypothesis

```
> x=c(587,602,627,610,619,622,605,608,596,592)
> t.test(x,alternative="greater",mu=13.4)

        One Sample t-test

data:  x
t = 142.82, df = 9, p-value < 2.2e-16
alternative hypothesis: true mean is greater than 13.4
95 percent confidence interval:
 599.1838      Inf
sample estimates:
mean of x
    606.8

> |
```

Deciding between Z Test and T-Test

So when we should perform the Z test and when we should perform t-Test? It's a key question we need to answer if we want to master statistics.



If the sample size is large enough, then the Z test and t-Test will conclude with the same results. For a **large sample size**, **Sample Variance will be a better estimate** of Population variance so even if population variance is unknown, we can **use the Z test using sample variance.**

Similarly, for a **Large Sample**, we have a high degree of freedom. And since **t-distribution approaches the normal distribution**, **t**he difference between the z score and t score is negligible.

Link :[Application of my Problem]