PROB& STAT LAB ASSIGNMENT

SUBMITTED BY

AVISEK MANDAL- 102203700(3CO35)

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

SUBMITTED TO - DR.SUMIT DEVI

TIET,PATIALA
AUG- DEC 2024

Assignment 05

Question 1: Consider that X is the time (in minutes) that a person has to wait in order to take a flight. If each flight takes off each hour X ~ U(0, 60). Find the probability that

(a) waiting time is more than 45 minutes, and

Ans:

b1 = punif(45, min = 0, max = 60, lower.tail = FALSE)

print(b1)

```
> b1 = punif(45, min = 0, max = 60, lower.tail = FALSE)
> print(b1)
[1] 0.25
```

(b) waiting time lies between 20 and 30 minutes.

c1 = punif(30, min = 0, max = 60) - punif(20, min = 0, max = 60)

print(c1)

```
> c1 = punif(30, min = 0, max = 60) - punif(20, min = 0, max = 60)
> print(c1)
[1] 0.1666667
```

Question 2: The time (in hours) required to repair a machine is an exponential distributed random variable with parameter λ = 1/2.

a) Find the value of density function at x = 3

Ans.

a2 = dexp(3, rate = 1/2)

print(a2)

```
> a2 = dexp(3, rate = 1/2)
> print(a2)
[1] 0.1115651
```

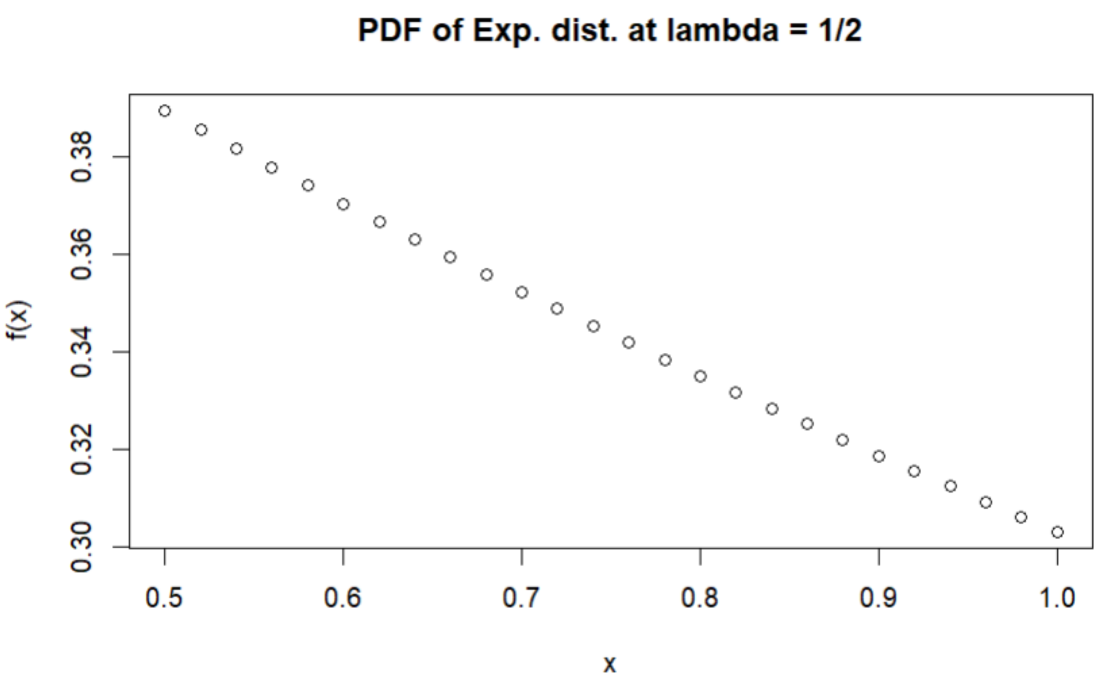
b) Plot the graph of exponential probability distribution for 0 ≤ x ≤ 5.

Ans.

x = seq(0.5, by = 0.02)

fx = dexp(x, rate = 1/2)

plot(x, fx, xlab = "x", ylab = "f(x)", main = "PDF of Exp. dist. at lambda = 1/2")

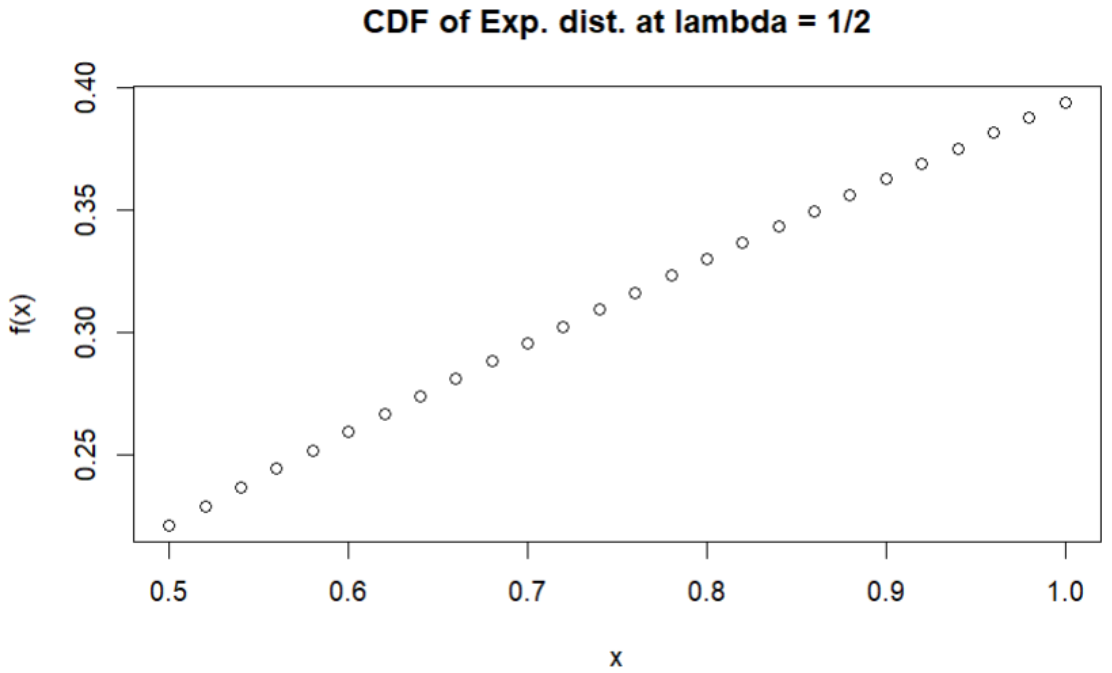


c) Find the probability that a repair time takes at most 3 hours.

Ans.

Fx = pexp(x, rate = 1/2)

plot(x, Fx, xlab = "x", ylab = "f(x)", main = "CDF of Exp. dist. at lambda = 1/2")

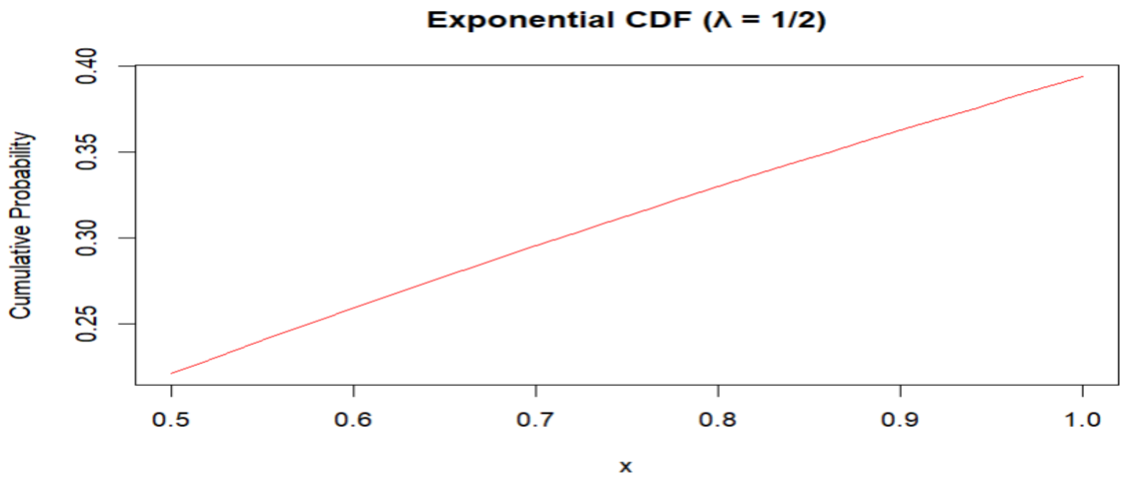


d) Plot the graph of cumulative exponential probabilities for 0 ≤ x ≤ 5.

Ans.

cumulative_vals <- pexp(x_vals, rate = lambda)

plot(x_vals, cumulative_vals, type = "l", main = "Exponential CDF (\u03bb = 1/2)", xlab = "x", ylab = "Cumulative Probability", col = "red")

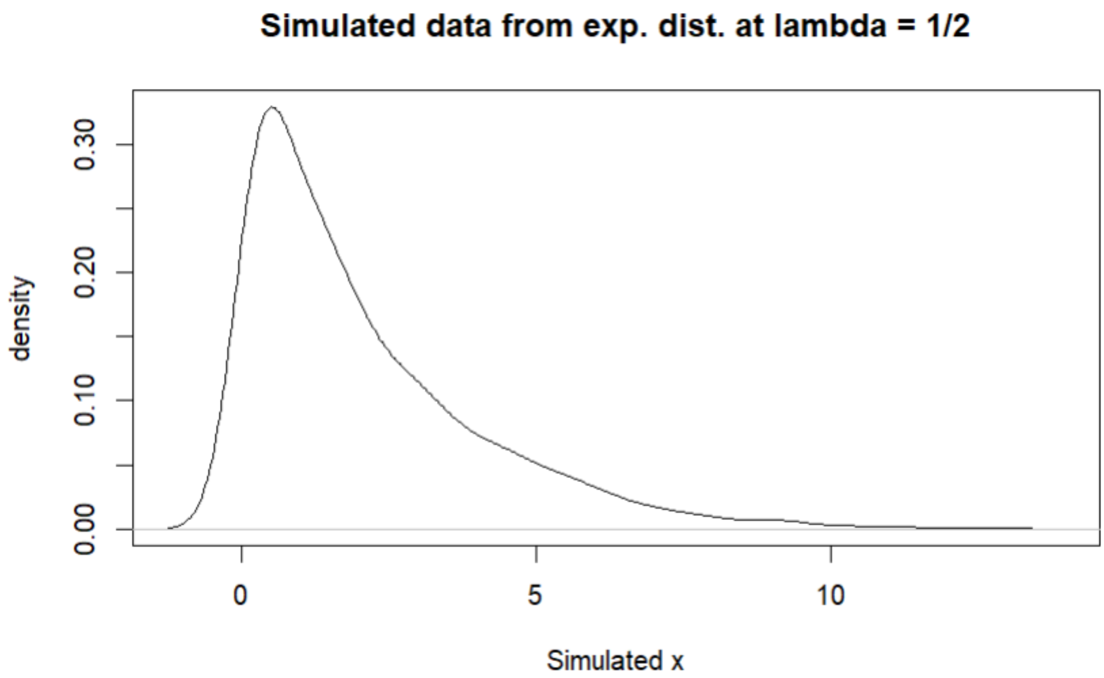


e) Simulate 1000 exponential distributed random numbers with λ = ½ and plot the simulated data.

Ans.

x_sim = rexp(1000, rate = 1/2)

plot(density(x_sim), xlab = "Simulated x", ylab = "density", main = "Simulated data from exp. dist. at lambda = 1/2")



Question 3: The lifetime of certain equipment is described by a random variable X that follows Gamma distribution with parameters α = 2 and β = 1/3

a) **Find the probability that the lifetime of equipment is at least 1 unit of time.**

Ans.

alpha = 2

beta = 1/3

a_1 = 1 - pgamma(1, shape = alpha, scale = beta)

print(a_1)

```
> alpha = 2
> beta = 1/3
> a_1 = 1 - pgamma(1, shape = alpha, scale = beta)
> print(a_1)
[1] 0.1991483
```

b) **What is the value of c, if P(X ≤ c) ≥ 0.70?**

Ans. b3 = qgamma(0.70, shape = alpha, scale = beta)

print(b3)

```
> b3 = qgamma(0.70, shape = alpha, scale = beta)
> print(b3)
[1] 0.8130722
```

Assignment 06

Question 1: he joint probability density of two random variables X and Y is f(x, y) = 2(2x + 3y)/5; 0 ≤ x, y ≤ 1 0; elsewhere Then write a R-code to

(i) check that it is a joint density function or not? (Use integral2())

Ans.

f <- function(x, y){

  2*(2*x + 3*y)/5}

I <- integral2(f, xmin=0, xmax=1, ymin=0, ymax=1)

I$Q

```
> library(pracma)
> f <- function(x, y){
+   2*(2*x + 3*y)/5
+ }
>
> I <- integral2(f, xmin=0, xmax=1, ymin=0, ymax=1)
> I$Q
[1] 1
```

(ii) find marginal distribution g(x) at x = 1.

Ans.

f2 <- function(y){

  f(1, y)}

integrate(f2, 0, 1)$value

```
> f2 <- function(y){
+   f(1, y)
+ }
> integrate(f2, 0, 1)$value
[1] 1.4
```

(iii) find the marginal distribution h(y) at y = 0.

Ans.

f3 <- function(x){

  f(x, 0)}

integrate(f3, 0, 1)$value

```
> f3 <- function(x){
+   f(x, 0)
+ }
> integrate(f3, 0, 1)$value
[1] 0.4
```

(iv) find the expected value of g(x, y) = xy

Ans.

f4 <- function(x, y){

  f(x, y) * x * y}

exp <- integral2(f4, 0, 1, 0, 1)$Q

exp

```
> f4 <- function(x, y){
+   f(x, y) * x * y
+ }
> exp <- integral2(f4, 0, 1, 0, 1)$Q
> exp
[1] 0.3333333
```

Question 2: The joint probability mass function of two random variables X and Y is f(x, y) = {(x + y)/30; x = 0, 1, 2, 3; y = 0, 1, 2} Then write a R-code to

(i) display the joint mass function in rectangular (matrix) form.
(ii) check that it is joint mass function or not? (use: Sum())
(iii) find the marginal distribution g(x) for x = 0, 1, 2, 3. (Use:apply())
(iv) find the marginal distribution h(y) for y = 0, 1, 2. (Use:apply())
(v) find the conditional probability at x = 0 given y = 1.
(vi) find E(x), E(y), E(xy), V ar(x), V ar(y), Cov(x, y) and its correlation coefficient.

Ans.

```
> f <- function(x, y) (x+y)/30
> x <- c(0:3)
> y <- c(0:2)
> m1 <- matrix(c(f(0, 0:2), f(1, 0:2), f(2, 0:2), f(3, 0:2)), nrow=4, ncol=3, byrow=TRUE)
> m1
           [,1]       [,2]       [,3]
[1,] 0.00000000 0.03333333 0.06666667
[2,] 0.03333333 0.06666667 0.10000000
[3,] 0.06666667 0.10000000 0.13333333
[4,] 0.10000000 0.13333333 0.16666667
> sum(m1)
[1] 1
>
>
> gx <- apply(m1, 1, sum)
> # sum along row axis
>
> hy <- apply(m1, 2, sum)
> # sum along col axis
>
> m1[1, 2] / hy[2]
[1] 0.1
>
> Ex <- sum(x * gx)
> Ex
[1] 2
>
> Ey <- sum(y*hy)
> Ey
[1] 1.266667
>
> Exy <- sum(outer(x, y, f) * m1)
> Exy
[1] 0.1088889
>
```

ASSIGNMENT-7

Question 1: Use the rt(n, df) function in r to investigate the t-distribution for n = 100 and df = n − 1 and plot the histogram for the same.

Ans.

n <- 100

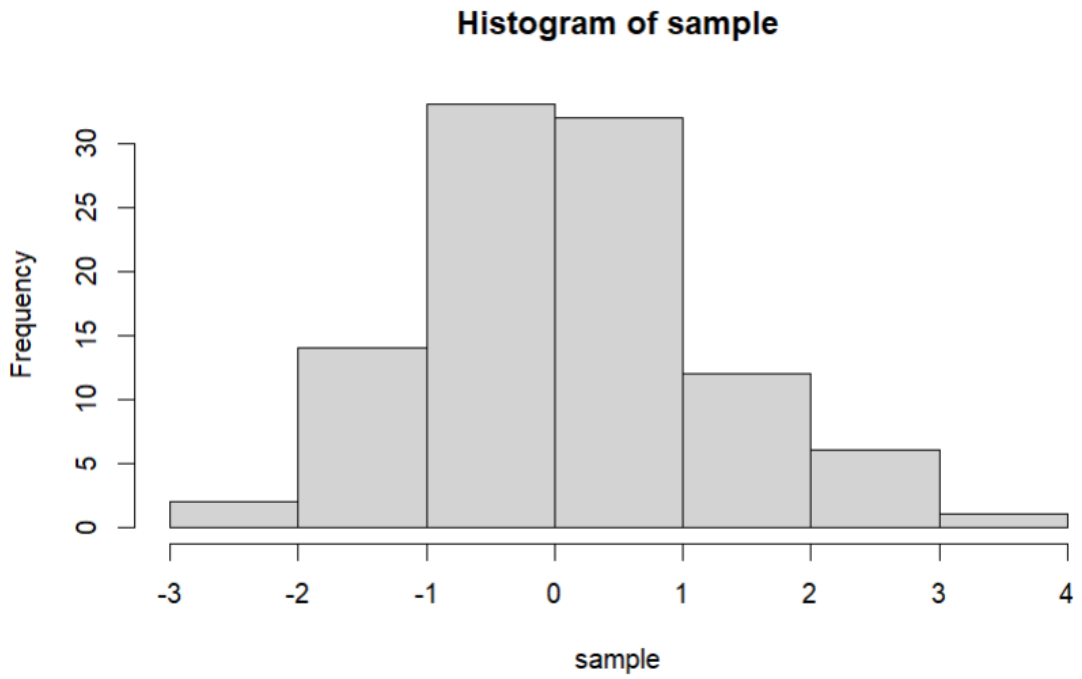
df <- n-1

sample <- rt(n, df)

sample

hist(sample)

```
> n <- 100
> df <- n-1
> sample <- rt(n, df)
> sample
  [1] -0.476063055 -1.087727480  0.741159585  2.522273110 -0.497780315  0.932718189
  [7] -0.303034728  0.761018891 -0.140329635  0.386384741  0.659048757 -0.838092400
 [13] -1.519592589  2.477168668  0.311075979  2.003757635 -1.773633804  0.034769872
 [19] -0.850747113  0.932690224 -0.344677968 -0.938903119  1.195648730  0.391955463
 [25] -0.277794289  2.551351536  0.935972793  0.101151614 -0.841076989  3.730750157
 [31]  1.099402481 -0.578995688  0.100842832  0.317954186 -2.129088731  1.171311705
 [37] -0.928788802 -0.968073158 -0.378471073 -1.061415022 -0.735495899  0.151700956
 [43] -1.178135037  0.068008440  0.076758969 -0.679510894  2.946259780 -0.751817695
 [49]  0.517030980 -0.554361497 -0.333291016  1.481079193 -0.347558595  0.517322493
 [55] -1.204204222  1.732936122  0.901469879 -0.741219000 -0.863108521  0.267227391
 [61]  0.231505288  1.436683122 -0.295454992 -0.009788239  1.159050492  0.476308065
 [67]  0.446900388 -1.528036793 -1.256858228 -0.877105897 -1.653097421  0.710286596
 [73] -0.256398130 -1.848330646  0.841566117  1.078919938  0.371826943 -0.740087568
 [79]  0.437319041  1.863863115  1.114268399  1.170331748 -1.156724298 -0.485232561
 [85] -0.319397401 -1.769179603 -1.761859355 -0.001138658 -0.969902211  1.385323573
 [91] -0.075674345  0.419535335 -0.380965886  0.976957587 -1.081569715  0.718341162
 [97]  2.548275599 -2.096780435  0.952017809  0.467389842
> hist(sample)
```



Question 2: Use the rchisq(n, df) function in r to investigate the chi-square distribution with n = 100 and df = 2, 10, 25.

Ans. n <- 100

df <- c(2,10,25)

s1 <- rchisq(n, df[1])

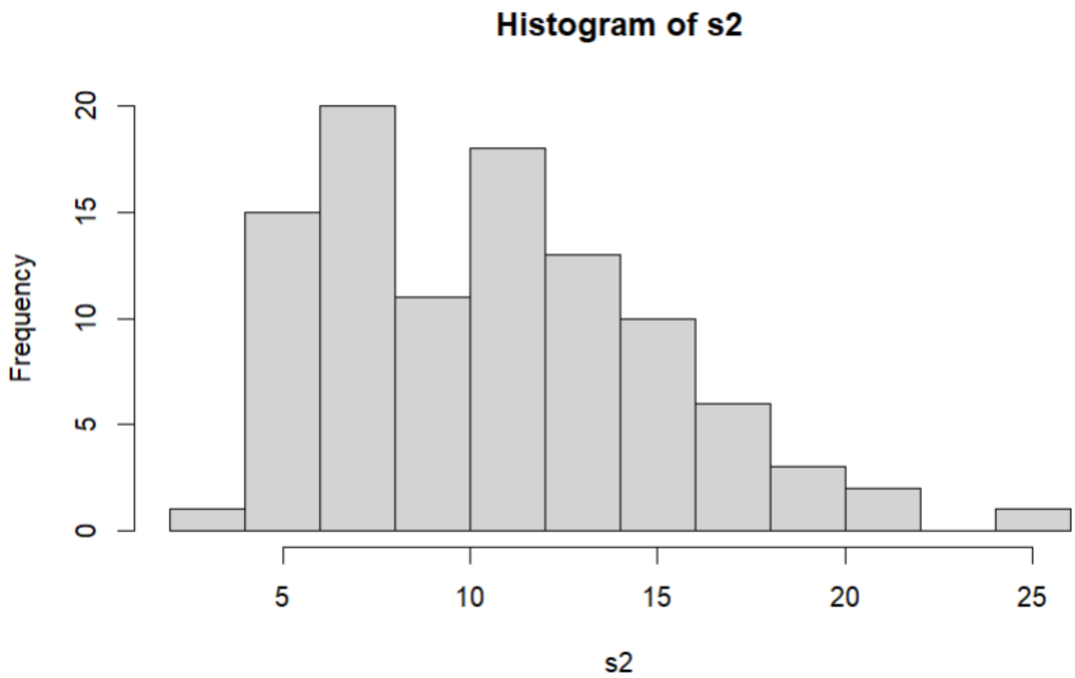
s2 <- rchisq(n, df[2])

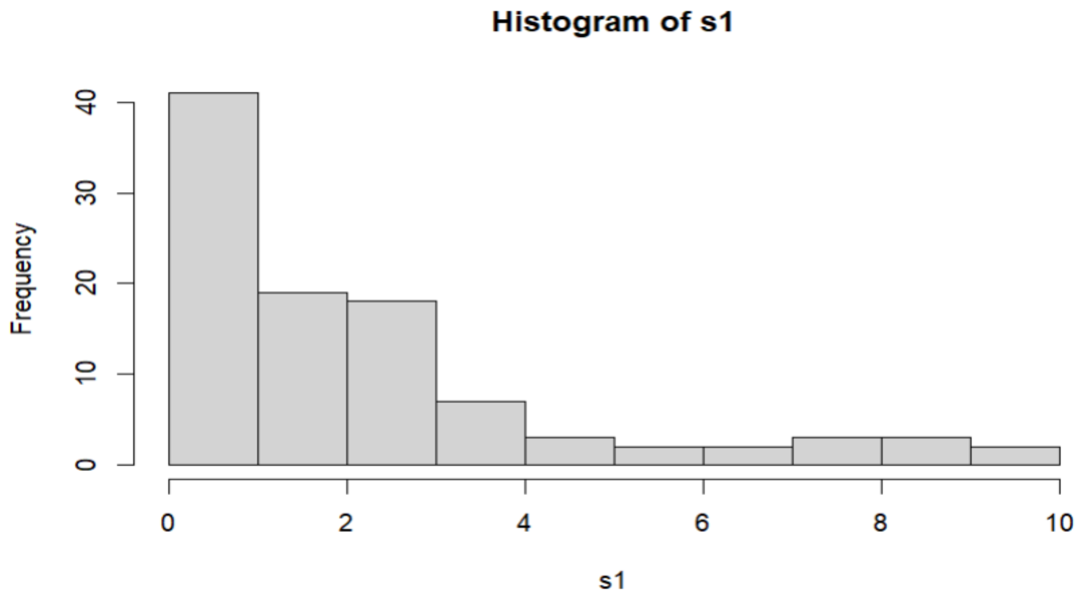
s3 <- rchisq(n, df[3])

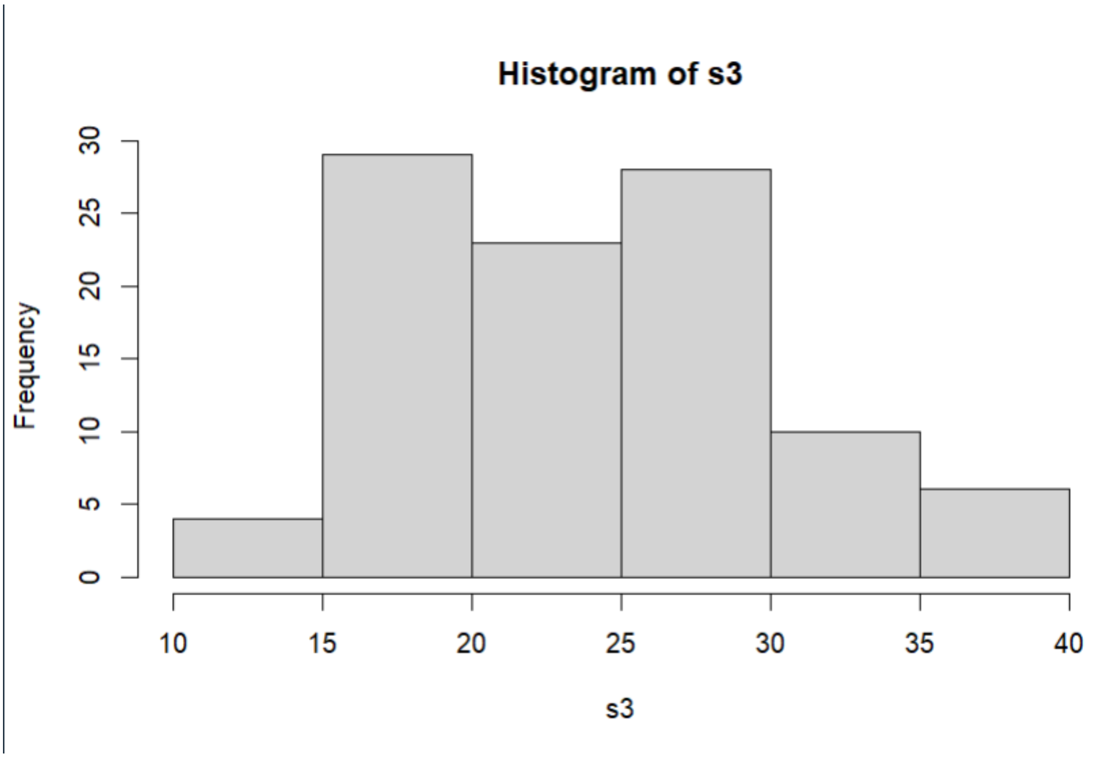
hist(s1)

hist(s2)

hist(s3)


Histogram of s2
Frequency
0
5
10
15
20
5
10
15
20
25
s2


Histogram of s1
Frequency
0
10
20
30
40
0
2
4
6
8
10
s1

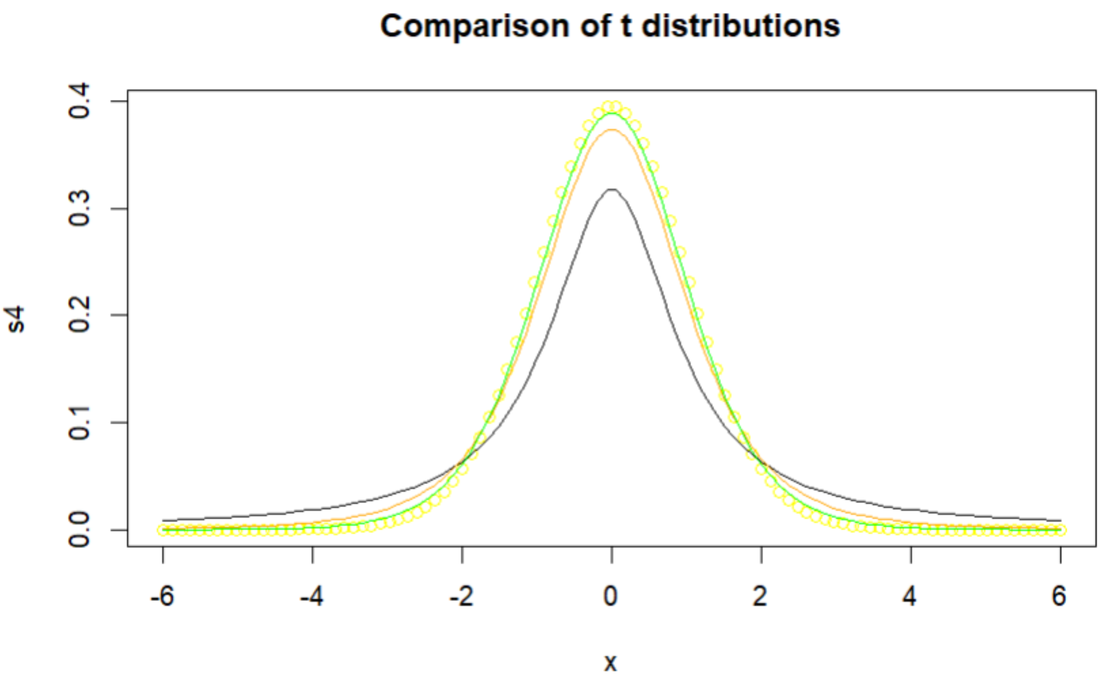


Question 3: Generate a vector of 100 values between -6 and 6. Use the dt() function in r to find the values of a t-distribution given a random variable x and degrees of freedom 1,4,10,30. Using these values plot the density function for students t-distribution with degrees of freedom 30. Also shows a comparison of probability density functions having different degrees of freedom (1,4,10,30).

Ans.

```
x <- seq(-6, 6, length=100)

df <- c(1, 4, 10, 30)

color <- c("black", "orange", "green", "yellow")

s1 <- dt(x, df[1])

s2 <- dt(x, df[2])

s3 <- dt(x, df[3])

s4 <- dt(x, df[4])
```

```
plot(x, s4, main = "Comparison of t distributions", col=color[4])
for(i in 1:3){
  lines(x, dt(x, df[i]), type="l", col=color[i])}
```



Question 4: Write a r-code

(i) **To find the 95th percentile of the F-distribution with (10, 20) degrees of freedom.**

(ii) **To calculate the area under the curve for the interval [0, 1.5] and the interval [1.5, +∞) of a F-curve with v1 = 10 and v2 = 20 (USE pf()).**

(iii) **To calculate the quantile for a given area (= probability) under the curve for a F-curve with v1 = 10 and v2 = 20 that corresponds to q = 0.25, 0.5, 0.75 and 0.999. (use the qf())**

(iv) **To generate 1000 random values from the F-distribution with v1 = 10 and v2 = 20 (use rf())and plot a histogram.**

Ans.

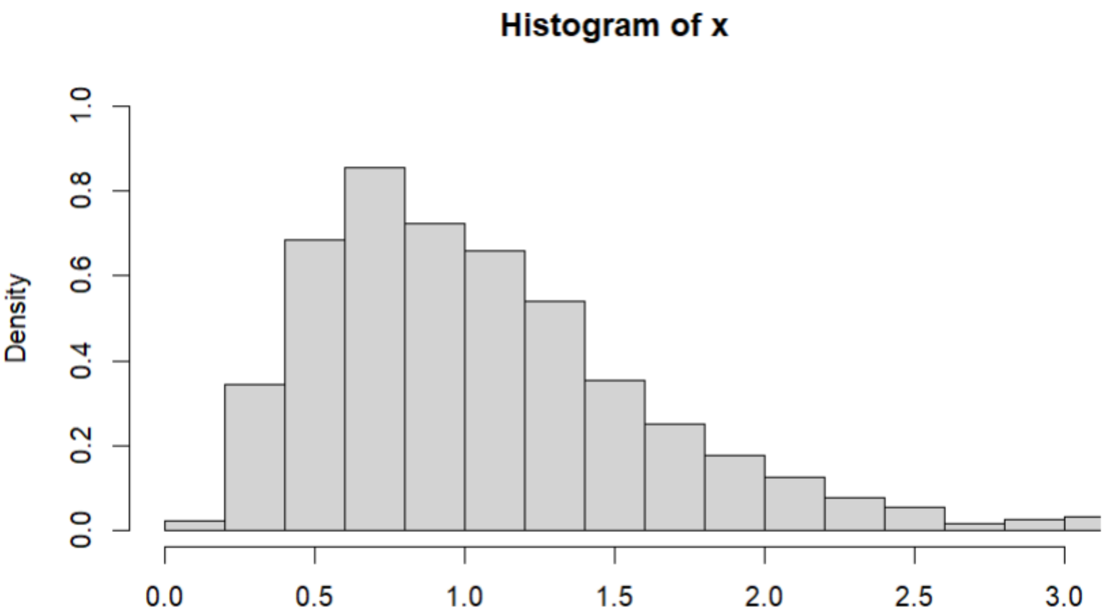
```
qf(0.95, df1 = 10, df2 = 20)
```

```
x = 1.5
v1 = 10
v2 = 20
pf(x, df1 = v1, df2 = v2, lower.tail = TRUE)
pf(x, df1 = v1, df2 = v2, lower.tail = FALSE)
q = c(0.25, 0.5, 0.75, 0.999)
v1 = 10
v2 = 20
qf(q[1], df1 = v1, df2 = v2, lower.tail = TRUE)
qf(q[2], df1 = v1, df2 = v2, lower.tail = TRUE)
qf(q[3], df1 = v1, df2 = v2, lower.tail = TRUE)
qf(q[4], df1 = v1, df2 = v2, lower.tail = TRUE)
x = rf(1000, df1 = 10, df2 = 20)
hist(x,
    breaks = "Scott",
    freq = FALSE,
    xlim = c(0, 3),
    ylim = c(0, 1),
    xlab = "")
```

```
> qf(0.95, df1 = 10, df2 = 20)
[1] 2.347878
```

```
> x = 1.5
> v1 = 10
> v2 = 20
>
> pf(x, df1 = v1, df2 = v2, lower.tail = TRUE)
[1] 0.7890535
>
> pf(x, df1 = v1, df2 = v2, lower.tail = FALSE)
[1] 0.2109465
```

```
> q = c(0.25, 0.5, 0.75, 0.999)
> v1 = 10
> v2 = 20
>
> qf(q[1], df1 = v1, df2 = v2, lower.tail = TRUE)
[1] 0.6563936
> qf(q[2], df1 = v1, df2 = v2, lower.tail = TRUE)
[1] 0.9662639
> qf(q[3], df1 = v1, df2 = v2, lower.tail = TRUE)
[1] 1.399487
> qf(q[4], df1 = v1, df2 = v2, lower.tail = TRUE)
[1] 5.075246
```



Assignment 08

Question 1: A pipe manufacturing organization produces different kinds of pipes. We are given the monthly data of the wall thickness of certain types of pipes (data is available on LMS Clt-data.csv)

The organization has an analysis to perform and one of the basic assumption of that analysis is that the data should be normally distributed.

You have the following tasks to do:

(a) Import the csv data file in R.

Ans. data <- read.csv("C:\\\\Users\\\\admin\\\\Downloads\\\\Clt-data.csv")

data

```
2         12.61742
3         12.36972
4         13.22335
5         13.15919
6         12.67549
7         12.36131
8         12.44468
9         12.62977
10        12.90381
11        12.76122
12        12.80959
13        13.02825
14        12.62958
15        13.17570
16        13.16581
17        12.39500
18        12.82689
19        12.49288
20        13.07618
21        13.07141
22        13.05835
23        12.64817
24        13.16075
25        12.68849
26        12.63552
27        12.48344
28        13.27466
```

(b) Validate data for correctness by counting number of rows and viewing the top ten rows of the dataset.

dim(data)

head(data, 10)

```
> dim(data)
[1] 9000    1
>
> head(data, 10)
   Wall.Thickness
1        12.35487
2        12.61742
3        12.36972
4        13.22335
5        13.15919
6        12.67549
7        12.36131
8        12.44468
9        12.62977
10       12.90381
```

(c) Calculate the population mean and plot the observations by making a histogram.

popmean <- mean(data$Wall.Thickness)

popmean

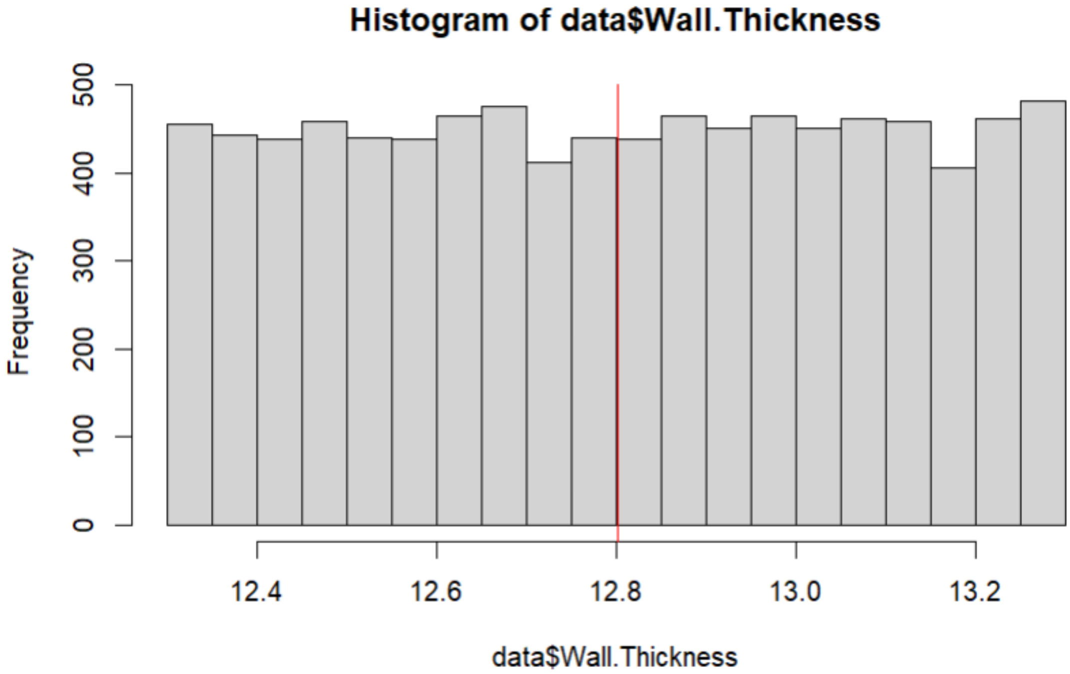
```
> popmean <- mean(data$Wall.Thickness)
> popmean
[1] 12.80205
>
> hist(data$Wall.Thickness)
> abline(v=popmean, col='red')
>
```

(d) Mark the mean computed in last step by using the function abline.

ow perform the following tasks:

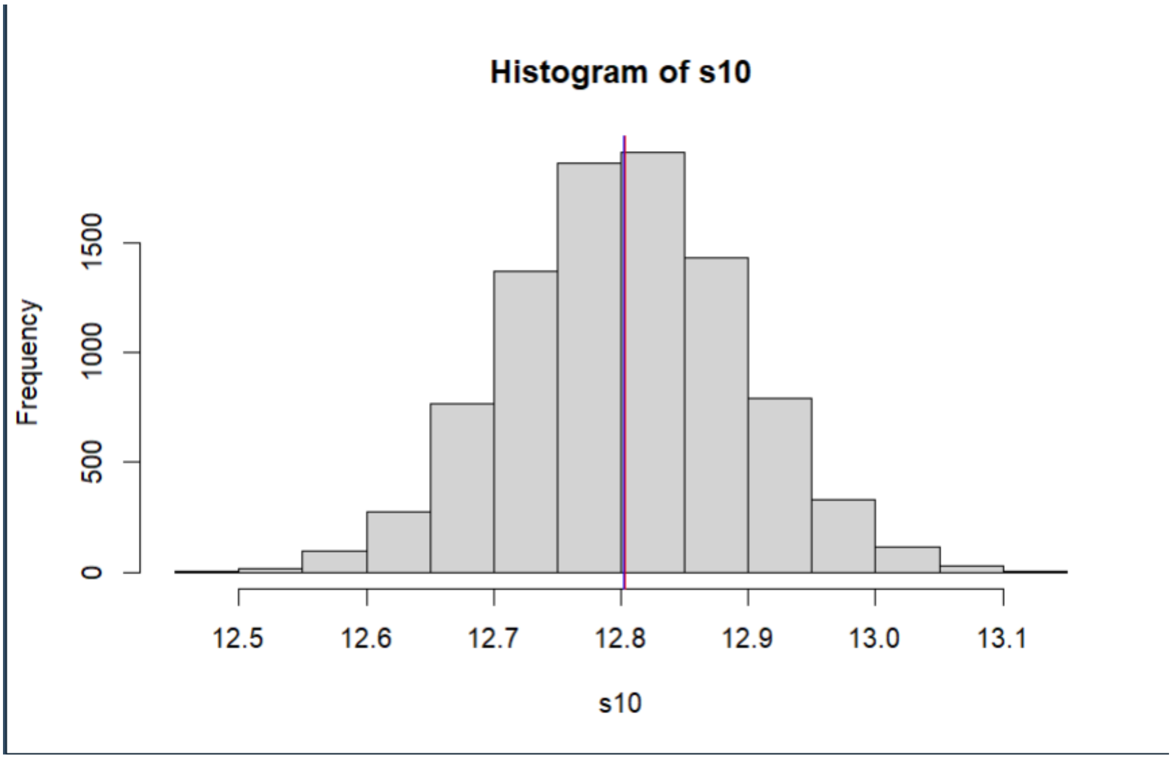
hist(data$Wall.Thickness)

abline(v=popmean, col='red')



(a) Draw sufficient samples of size 10, calculate their means, and plot them in R by making histogram. Do you get a normal distribution.

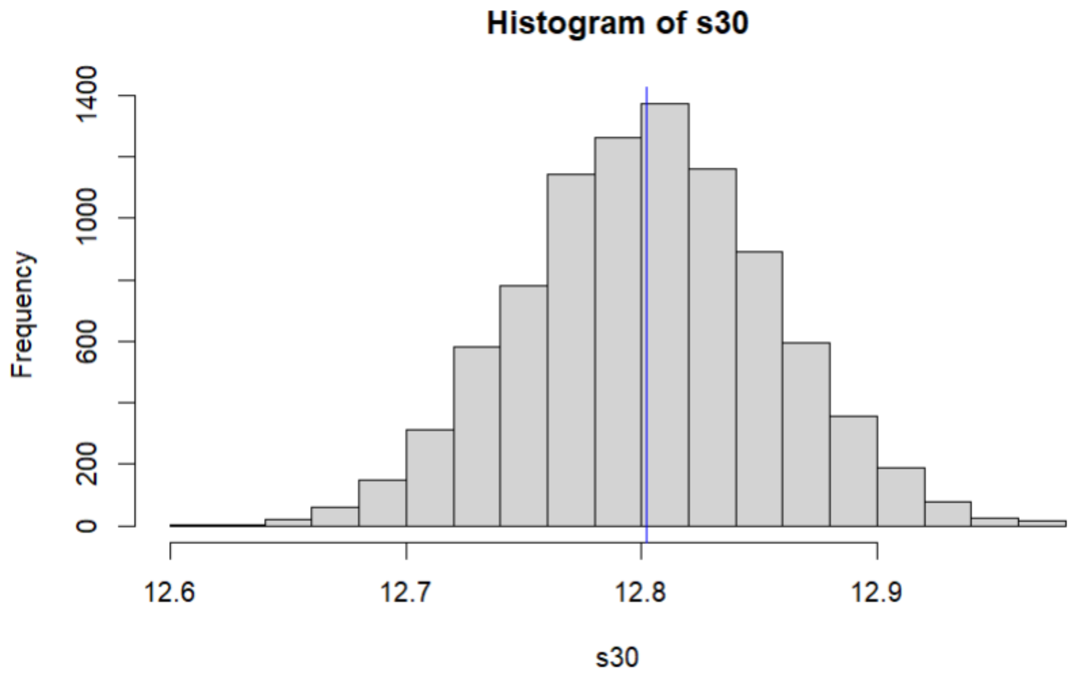
```
s10<-c()
n<-9000
for (i in 1:n) {
  s10[i] = mean(sample(data$Wall.Thickness, 10, replace=TRUE))
}
hist(s10)
abline(v=mean(s10), col='red')
abline(v=popmean, col='blue')
```



(b) Now repeat the same with sample size 50, 500 and 9000. Can you comment on what you observe.

```
s30 <- c()

s50 <- c()

s500 <- c()

n<-9000

for (i in 1:n) {

  s30[i] = mean(sample(data$Wall.Thickness, 30, replace=TRUE))

  s50[i] = mean(sample(data$Wall.Thickness, 50, replace=TRUE))

  s500[i] = mean(sample(data$Wall.Thickness, 500, replace=TRUE))}

hist(s30)

abline(v=mean(s30), col='red')

abline(v=popmean, col='blue')
```

par(mfrow=c(1,3))



Question 2: The following table gives information on ages and cholesterol levels for a random sample of 10

Age	58	69	43	39	63	52	47	31	74	36
Cholesterol	189	235	193	177	154	191	213	165	198	181

Plot the scatter diagram and a regression line that will enable us to predict Cholesterol level on age. Further, estimate the cholesterol level of a 60 year-old man.

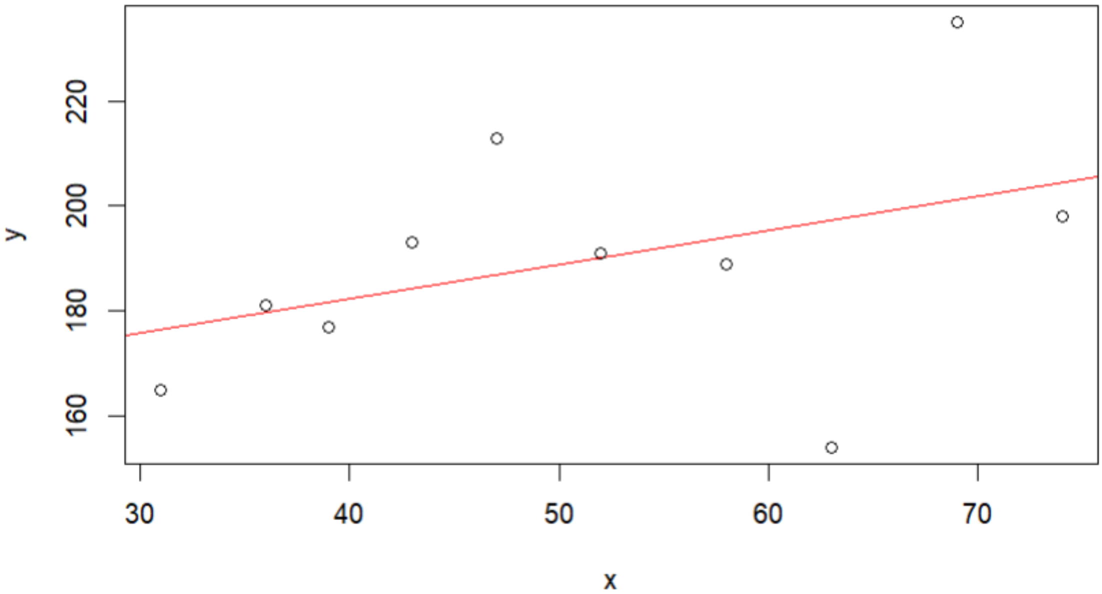
Ans :par(mfrow=c(1,1))

x <- c(58, 69, 43, 39, 63, 52, 47, 31, 74, 36)

y <- c(189, 235, 193, 177, 154, 191, 213, 165, 198, 181)

relation <- lm(y~x)

plot(x, y, abline(lm(y~x), col='red'))



a <- data.frame(x=60)

result <- predict(relation, a)

result

```
> a <- data.frame(x=60)
> result <- predict(relation, a)
> result
       1
195.3184
```

Question 3 : A research methodology course has recently been added to the PhD curriculum at the Thapar Institute of Engineering and Technology, Patiala. To evaluate its effectiveness, students take a test on formulating research problems and writing research papers both before and after completing the course. Below are the marks for a random sample of ten students

Before the test	145	173	158	141	167	159	154	167	145	153
After the test	155	167	156	149	168	162	158	169	157	161

Assume that the differences between the pre-course and post-course test scores are normally distributed, and a high score on the test indicates a strong level of assertiveness. Do the collected data, at 5% level of significance, provide enough evidence to conclude that research scholars become more assertive after completing the course?

Ans.

x <- c(145, 173, 158, 141, 167, 159, 154, 167, 145, 153)

y <- c(155, 167, 156, 149, 168, 162, 158, 169, 157, 161)

t.test(x, y, alternative = "greater", mu=0, paired=FALSE, var.equal=FALSE, conf.level=0.95)

```
> x <- c(145, 173, 158, 141, 167, 159, 154, 167, 145, 153)
> y <- c(155, 167, 156, 149, 168, 162, 158, 169, 157, 161)
>
> t.test(x, y, alternative = "greater", mu=0, paired=FALSE, var.equal=FALSE, conf.level=0.95)

        Welch Two Sample t-test

data:  x and y
t = -1.0152, df = 14.797, p-value = 0.8368
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -10.91342       Inf
sample estimates:
mean of x mean of y
    156.2     160.2
```

---THANK YOU----