PERFORMANCE

Term	Remarks	Staff Member's Signature
I	Complete	[illegible]
II	Complete	[illegible] 14/3/20

Exam Seat No.________________

THAKUR COLLEGE OF SCIENCE & COMMERCE

NAAC Accredited with Grade "A" (3rd Cycle)

THAKUR ® TRUSTS

ISO 9001 : 2015 Certified

Degree College

Computer Journal

CERTIFICATE

SEMESTER II **UID No.** ________________

Class FYBSC **Roll No.** 1755 **Year** 2019-20

This is to certify that the work entered in this journal is the work of Mst. / Ms. AADARSH YADAV

who has worked for the year 2019-20 in the Computer Laboratory.

Teacher In-Charge

Head of Department

Date : 14/02/20

Examiner

INDEX

Scanned with CamScanner

PRACTICAL - 01

Aim : Basics of R software.

1) R is a software for statical analysis and data computing.

2) It is an effective data handling software and outcome storage is possible.

3) It is capable of graphical display.

4) It is a free software.

Q.1) Solve the followings :

1) $4+6+8\div 2-5$

```
> 4+6+8/2-5
[1] 9
```

2) $2^2+|-3|+\sqrt{45}$

```
> 2^2 + abs (-3)+ sqrt (45)
[1] 13.7082
```

3) $5^3+7\times5\times8+46/5$

```
> 5^3 + 7 * 5 * 8 + 46/5
[1] 414.2
```

4) $\sqrt{4^2+5\times3+7/6}$

> sqrt (4^2 + 5 * 3 + 7/6)

[1] 5.671567

5) round off $46\div7+9\times8$

> round (46/7+9*8)

[1] 79

Q.2)

a) > c(2,3,5,7) * 2

[1] 4 6 10 14

b) > c(2,3,5,7) * c(2,3)

[1] 4 9 10 21

c) > c(2,3,5,7) * c(2,3,6,2)

[1] 4 9 30 14

d) > c(1,6,2,3) * c(-2,-3,-4,-1)

[1] -2 -18 -8 -3

e) > c(2,3,5,7)^2

[1] 4 9 25 49

f) > c(4,6,8,9,4,5)^c(1,2,3)

[1] 4 36 512 9 16 125

g) > c(6,2,7,5) / c(4,5)

[1] 1.50 0.40 1.75 1.00

Q.3) > x=20 > y=30 > z=2

> x^2 + y^3 + z

[1] 27402

> sqrt (x^2+y)

[1] 20.73644

> x^2 + y^2

[1] 1300

Scanned with CamScanner

Q.4) > x <- matrix(nrow=4, ncol=2, data=c(1,2,3,4,5,6,7,8))

> x		[,1]	[,2]
[1]	[1,]	1	5
	[2,]	2	6
	[3,]	3	7
	[4,]	4	8

Q.5) Find x+y and 2x+3y where $x = \begin{bmatrix} 4 & -2 & 6 \\ 7 & 0 & 7 \\ 9 & -5 & 3 \end{bmatrix}$

$y = \begin{bmatrix} 10 & -5 & 1 \\ 12 & -4 & 9 \\ 15 & -6 & 5 \end{bmatrix}$

> x <- matrix(nrow=3, ncol=3, data=c(4,7,9,-2,0,-5,6,7,3))

> x	[,1]	[,2]	[, 3]
[1,]	4	-2	6
[2,]	7	0	7
[3,]	9	-5	3

> y <- matrix(nrow=3, ncol=3, data=c(10,12,15,-5,-4,-6, 1,7,9,5))

> y	[, 1]	[, 2]	[, 3]
[1,]	10	-5	1
[2,]	12	-4	9
[3,]	15	6	5

> x+y

	[, 1]	[, 2]	[, 3]
[1,]	19	-7	13
[2,]	19	-4	16
[3,]	24	-11	8

58

```
> 2 * X + 3 * Y
      [,1]   [,2]   [,3]
[1,]   38    -19     33
[2,]   50    -12     41
[3,]   63    -28     21
```

Q.6) Marks of statistic of CS Batch A

x = c(58, 20, 35, 24, 46, 56, 55, 45, 27, 22, 47, 58, 54, 40, 50, 52, 36, 29, 35, 39)

```
> x = c(data)
> breaks = seq(20, 60, 5)
> a = cut(x, breaks, right = FALSE)
> b = table(a)
> c = transform(b)
> c
        a     Freq
1  [20,25)     3
2  [25,30)     2
3  [30,35)     1
4  [35,40)     4
5  [40,45)     1
6  [45,50)     3
7  [50,55)     2
8  [55,60)     4
```

PRACTICAL - 02

TOPIC: Probability distribution.

1) Check whether the followings are p.m.f or not,

x	$p(x)$
0	0.1
1	0.2
2	-0.5
3	0.4
4	0.3
5	0.5

If the given data is p.m.f then

$\Sigma P(x) = 1$

$\therefore P(0) + P(1) + P(2) + P(3) + P(4) + P(5) = P(x)$

$= 0.1 + 0.2 - 0.5 + 0.4 + 0.3 + 0.5$

$= 1.0$

$\therefore P(2) = -0.5$, it can't be a probability mass function

$\therefore p(x) \geq 0 \quad \forall, x$

2)

x	1	2	3	4	5
$P(x)$	0.2	0.2	0.3	0.2	0.2

The condition for pm.f is $\Sigma P(x) = 1$

So,

$\Sigma P(x) = P(1) + P(2) + P(3) + P(4) + P(5)$

$= 0.2 + 0.2 + 0.3 + 0.2 + 0.2$

$= 1.1$

$\therefore$ The given data is not a pmf because $P(x) \neq 1$.

3)

x	10	20	30	40	50
$P(x)$	0.2	0.2	0.35	0.15	0.1

The condition for p.m.f. is satisfy

1) $P(x) \geq 0 \quad \forall x$
2) $\sum P(x) = 1$

$\sum P(x) = P(10) + P(20) + P(30) + P(40) + P(50)$

$= 0.2 + 0.2 + 0.35 + 0.15 + 0.1$

$= 1$

$\therefore$ The given data is p.m.f.

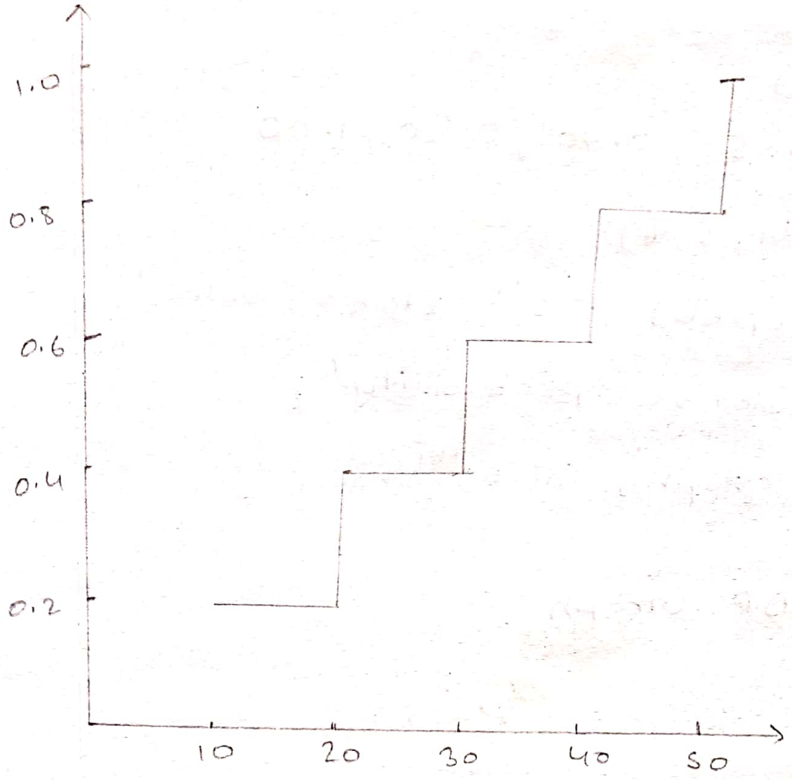
Code :

```
> prob= c(0.2 , 0.2 , 0.35, 0.15, 0.1)
> sum (prob)
[1] 1
```

Q.2) Find the c.d.f. for the following p.m.f. and sketch the graph.

x	10	20	30	40	50
$P(x)$	0.2	0.2	0.35	0.15	0.1

$$
\begin{aligned}
F(x) &= 0 && x < 10 \\
&= 0.2 && 10 \leq x < 20 \\
&= 0.4 && 20 \leq x < 30 \\
&= 0.75 && 30 \leq x < 40 \\
&= 0.90 && 40 \leq x < 50 \\
&= 1.0 && x \geq 50
\end{aligned}
$$




```
> x=c(10,20,30,40,50)
> plot(x, comsum(prob), "s")
```

Q.3) Find :

X	1	2	3	4	5	6
P(X)	0.15	0.25	0.1	0.2	0.2	0.1

$$F(x) = \begin{cases} 0 & x<1 \\ 0.15 & 1\le x<2 \\ 0.40 & 2\le x<3 \\ 0.50 & 3\le x<4 \\ 0.70 & 4\le x<5 \\ 0.90 & 5\le x<6 \\ 1.00 & x\ge 6 \end{cases}$$

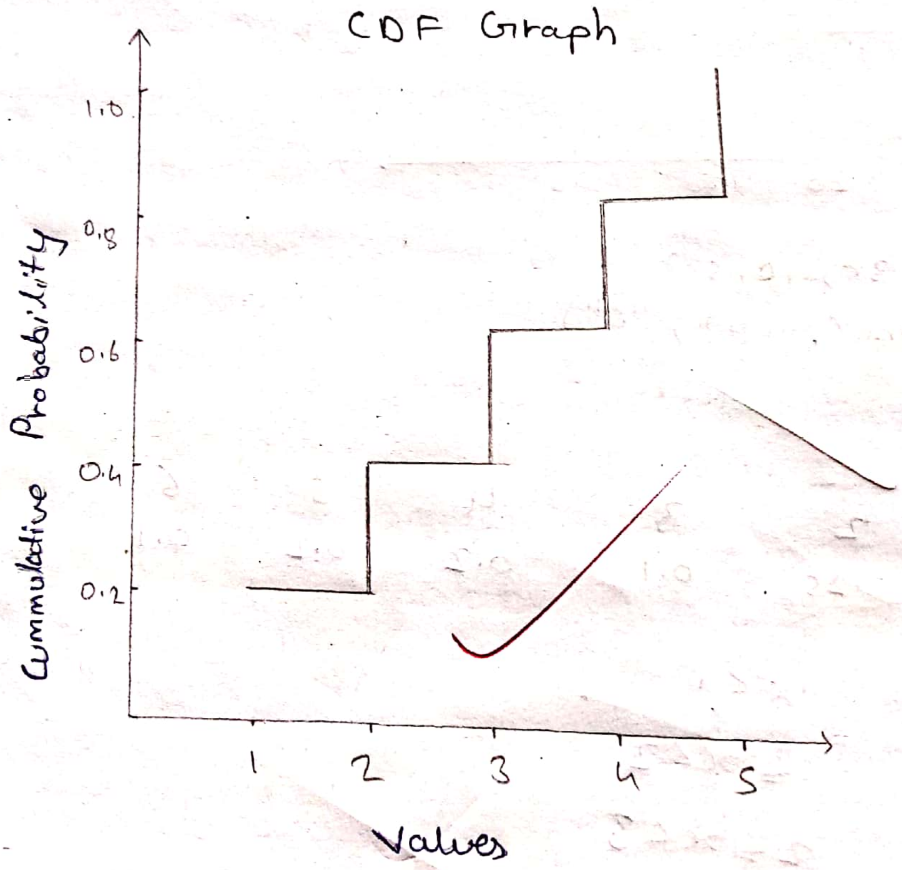
```
> prob=c(0.15, 0.25, 0.1, 0.2, 0.1)
> sum(prob)
[1] 1
```

> cumsum (prob)

[1] 0.15, 0.40 , 0.50, 0.70 , 0.90, 1.00

> x=c(1, 2, 3, 4, 5, 6)

> plot (x, cumsum (prob), "s", xlab = "value",
    ylab = " Commulative probability",
    main = "CDF graph", col = "brown")



Q.4) Check that whether the following is p.d.f. or not.

i) $f(x) = 3-2x \quad ; \quad 0 \le x \le 1$

ii) $f(x) = 3x^2 \quad ; \quad 0 < x < 1$

i) $f(x) = 3-2x$

$= \int_0^1 f(x)$

$= \int_0^1 (3-2x)dx$

$= \Rightarrow \int_0^1 \cancel{3dx} \left[3x - \frac{2x^2}{2}\right]_0^1$

$= 3 - 1$

$= 2 \neq 1$

It is not a pdf.

$\therefore \int_0^1 f(x) \neq 1$

ii) $f(x) = 3x^2$

$\int_0^1 f(x) = 1$

$= \int_0^1 3x^2 \, dx$

$= \left[\frac{3x^3}{3}\right]_0^1 dx$

$= 1$

The $\int_0^1 f(x) = 1$ $\because$ It is a pdf.

PRACTICAL - 03

TOPIC : Binomial

\# $P(X=x) = dbinom(x,n,p)$
\# $P(X \le x) = pbinom(x,n,p)$
\# $P(x > x) = 1 - pbinom(x,n,p)$
\# If x is unknown
$p_1 = P(x \le x) = qbinom(p_1, n, p)$

1) Find the probability of exactly 10 success is hundred trials with $p = 0.1$

2) Suppose there are 12 mcq, Each question has 5 options out of which 1 is correct.
i) Find the probability of having exactly 4 correct answers.
ii) Atmost 4 correct answers.
iii) More than 5 correct answers

3) Find the complete distribution when $n=5$ and $p=0.1$.

4) $n=12$, $p=0.25$, find the following probabilities
i) $P(X=5)$
ii) $P(X \le 5)$
iii) $P(x>7)$
iv) $P(5<x<7)$

1) > x=dbinom (10, 100, 0.1)
> x
[1] 0.1318653

2) i) > x=dbinom (4, 12, 0.2)
> x
[1] 0.1328756

ii) > x= pbinom (4, 12, 0.2)
> x
[1] 0.9274445

iii) 1-pbinom (5, 12, 0.2)
[1] 0.01940528

3) x=dbinom (0:5, 5, 0.1)
0 - 0.59049
1 - 0.32805
2 - 0.07290
3 - 0.00810
4 - 0.00045
5 - 0.00001

4) 1) dbinom (5, 12, 0.25)
[1] 0.1032414

2) pbinom (5, 12, 0.25)
[1] 0.9455918

3) 1-pbinom (7, 12, 0.25)
[1] 0.00278151

4) dbinom (6, 12, 0.25)
[1] 0.04014945

5) > dbinom (0, 10, 0.15)
[1] 0.1968744
> 1 - pbinom (3, 20, 0.15)
[1] 0.3522748

6) qbinom (0.88, 30, 0.2)
[1] 9

7) > n=10
> p=0.3
> x=0:n
> prob = dbinom (x, n, p)
> comprob = pbinom (x, n, p)
> d = data.frame ("X values" = x, "probability" = prob)
> print (d)

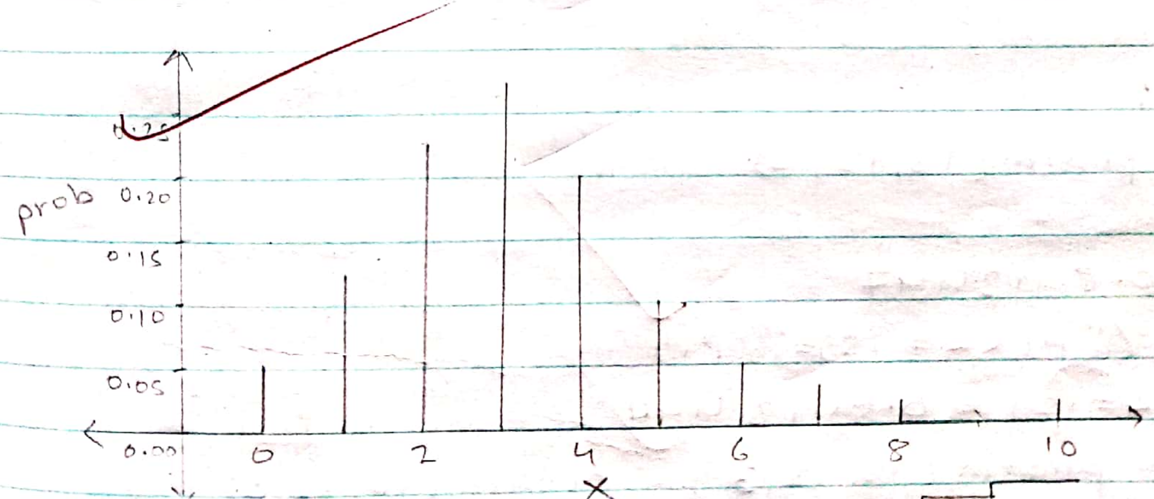
	Xvalues	probability
1	0	0.0282
2	1	0.1210
3	2	0.2334
4	3	0.2668
5	4	0.2001
6	5	0.1029
7	6	0.0367
8	7	0.0090
9	8	0.0014
10	9	0.0001
11	10	0.0000

5) The probability of a salesman making a sale to customer 0.15. Find the probability of
i) No sales out of 10 customer
ii) More than 3 sales out of 20 customer.

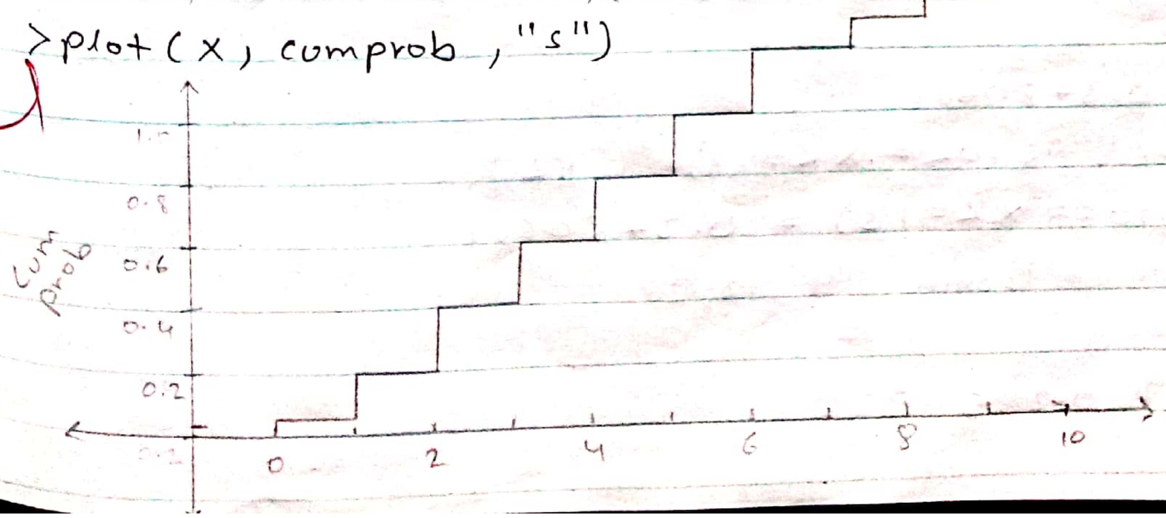
6) A sales man has 20% probability of making a sale to customer. out of 30 customers what minimum number of sales he can make with 88% of probability.

7) X follows binomial distribution with $n=10, p=0.3$ plot the graph of p.m.f. and c.d.f.

> plot (x, prob, "h")



>plot (x, comprob, "s")



PRACTICAL-04

Aim: Normal Distribution

i) $P(x = x) = dnorm(x, \mu, \sigma)$
ii) $P(X \le x) = pnorm(x, \mu, \sigma)$
iii) $P(X > x) = 1 - pnorm(x, \mu, \sigma)$
iv) To generate random numbers from a normal distribution (n random numbers) the R code is
rnorm(n, μ, σ)

Q.1) A random variable X follows normal distribution with mean $= \mu = 12$ and S.D. $= \sigma = 3$. Find
i) $P(X \le 15)$ ii) $P(10 \le X \le 13)$ iii) $P(X > 14)$
iv) generate 5 observations (random numbers)

Code:

```
> P1 = pnorm(15, 12, 3)
> P1
[1] 0.8413447
> cat("P(X<=15)=", P1)
P(X<=15) = 0.8413447
> P2 = pnorm(13, 12, 3) - pnorm(10, 12, 3)
> P2
[1] 0.3780661
> cat("P(10<=X<=13)=", P2)
P(10<=X<=13) = 0.3780661
> P3 = 1 - Pnorm(14, 12, 3)
> P3
[1] 0.2524925
```

```
> cat("P(X>14) =", P3)
P(X>14) = 0.2524925
> P4 = rnorm(5,12,3)
> P4
[1] 15.254723   16.548505   11.280515   6.419944 12.272460
```

Q.2> X follows normal distribution with $\mu=10, \sigma=2$

Find i> $P(X\leq 7)$ ii> $P(5<X<12)$ iii> $P(X>12)$

iv> Generate 10 observations

v> Find K such that $P(X<K)=0.4$

Code:

```
> a1 = pnorm(7,10,2)
> a1
[1] 0.668072
> a2 = pnorm(5,10,2) - pnorm(12,10,2)
> a2
[1] 0.8351351
> a3 = 1 - Pnorm(12,10,2)
> a3
[1] 0.1586553
> a4 = rnorm(10,10,2)
> a4
[1] 11.608931   9.920417   12.637741   8.073354
     8.721380   9.193726   9.366824    11.707106
     9.537584   10.715006
> a5 = qnorm(0.4,10,2)
> a5
[1] 9.493306
```

Q.3) Generate 5 random numbers from a normal distribution $\mu = 15$, $\sigma = 4$. Find sample mean, median, S.D and print it.

Code:

```
> rnorm (15, 15, 4)
[1] 10.7649   7.793249   9.953444   13.345904
    17.509668
> am = mean(x)
> am
[1] 11.87345
> cat(" sample mean is = ", am)
sample mean is = 11.87345
> me = median (x)
> me
[1] 10.76499
> cat(" Median is =", me)
Median is = 10.76499.
> n=5
> v =(n-1) * var (x) /n
> v
[1] 11.09965
> SD= sqrt (v)
> SD
[1] 3.33163
> cat(" S.D. is = ", SD)
S.D is = 3.331613
```

Q.4) $X \sim N(30, 100)$, $\sigma = 10$

i) $P(x \le 40)$

ii) $P(x > 35)$

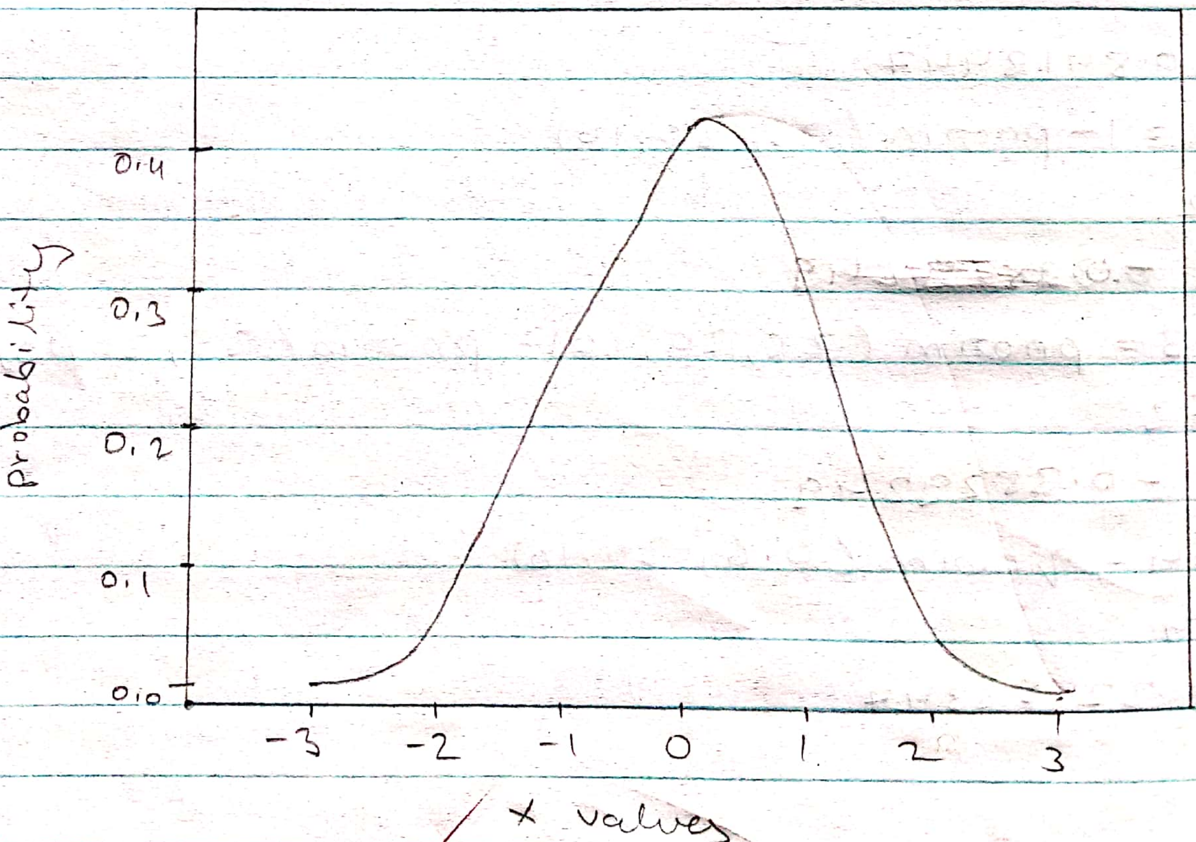
iii) $P(25 < X < 35)$

iv) Find K such that $P(X < K) = 0.6$

```
> f1 = pnorm (40,30, 10)
>f1
[1] 0.8413447
>f2=1-pnorm (35, 30,10)
>f2
[1] 0.3085375
>f3= pnorm (25,30,10)-pnorm(35,30,10)
> f3
[1] -0.3829249
> f4= qnorm (0.6, 30,10)
>f4
[1] 32.53347
```

Q.5) Plot the standard normal graph.

```
> x = seq (-3, 3, by = 0.1)
> y = dnorm (x)
> plot (x, y, xlab = "x values", ylab = "probability",
  main = "standard normal graph")
```



PRACTICAL – 05

TOPIC: Normal and t–test.

1) Ho: $\mu = 15$ $H_1 : \mu \neq 15$
↳ null ↳ alternative

Test the hypothesis

Random sample of size 400 is ~~drawn~~ drawn and it is calculated. The sample mean is 14 and S.D. is 3. Test the hypothesis at 5% level of significance.

0.05 > accep the value

0.05 < Less then Reject

```
> mo=15
> mx=14
> sd=3
> n=400
> zcal= (mx-mo)/(sd(sqrt(n))
> zcal
[1] - 6.666667
> cat ("calculated value of z is =", zcal)
Calculated value of z is = -6.666667
> pvalue = 2* (1-pnorm(abs(zcal)))
> pvalue
[1] 2.616796e-11
```

∵ The value is less than 0.05 we will reject the value of Ho $= \mu = 15$.

2) Test the hypothesis Ho: μ = 10 against H1: μ ≠ 10
A random sample size of 400 is drawn with sample mean = 10.2 and S.D. = 2.25.
Test the hypothesis at

```
> mo = 10
> n = 400
> mx = 10.2
> sd = 2.25
> zcal = (mx-mo)/(sd/(sqrt(n)))
> zcal
[1] 1.77778
> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 0.07544036
```

∵ The value pvalue is greater than 0.05
∴ The value is accepted.

3) Test the hypothesis Ho: preparation of smokers is college is 0.2. A sample is collected and calculated the sample proportional as 0.125. Test the hypothesis at 5% level of significance (sample size is 400)

```
> P = 0.2
> p = 0.125
> n = 400
> Q = 1-P
> zcal = (p-P)/(sqrt(P*Q/n))
> cat("calculated value of z is =", zcal)
[1] Calculated value of z is = -3.75
> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 0.0001768346 (Reject)
```

4) Last year farmer's lost 20% of their crops. A random sample of 60 fields are collected. And it is found that a field crops are insect polluted. Test the hypothesis at 1% level of significance.

```
>P=0.2
>p=9/60
>n=60
>zcal=(p-P)/(sqrt(P*Q/n))
>zcal
[1] -0.9682458
>pvalue = 2* (1-pnorm (abs(zcal)))
>pvalue
[1] 0.3329216
```

∵ The value is 0.01. so value is accepted.

5) Test the hypothesis Ho : u = 12.5 from the following sample at 5% level of significance.

```
>x=c(12.25, 11.97, 12.15, 12.08, 12.31, 11.94, 11.89, 12.16,
                                                      12.04)
>n= length (x)
>n
[1] 10
>mx = mean (x)
>mx
[1] 12.107
```

> variance = (n-1) * var(x)/n

> variance

[1] 0.019521

> sd = sqrt(variance)

> sd

[1] 0.1397176

> mo = 12.5

> t = (mx - mo) / (sd/ sqrt(n))

> t

[1] 8.894909

> pvalue = 2* (1- pnom (abs(t)))

> pvalue

[1] 0

∴ The value is less than 0.05 the value is accepted.

PRACTICAL - 06

Aim: Large Sample test

1) Let the population mean (the amount spent per customer in a restaurant) is 250. A sample of 100 customers selected the sample mean is calculated as 275 and S.D 30. Test the hypothesis that the population mean is 250 or not on 5% level of significance.

Soln:
```
> mo = 250
> mx = 275
> sd = 30
> n = 100
> zcal = (mx-mo) / (sd/(sqrt(n)))
> cat(" Calculated values of Z is = ", zcal)
[1] Calculated value of Z is = 8.333333
> pvalue = 2 * (1-pnorm(abs(zcal)))
> pvalue
[1] 0
```
∵ The value is less than 0.05 we will reject the value of the Ho = μ = 250.

2) In a random sample of 1000 students it is found that 750 use blue pen. test the hypothesis that the population proportion is 0.8 at 1% level of significance.

Soln:
```
> P = 0.8
> Q = 1-P
> p = 750/1000
> n = 1000
> zcal (p-P) / (sqrt (P*Q/n))
> cat (" Calculated value of Z is : ", zcal)
[1] calculated value of Z is : -3.952847
```

```
> pvalue = 2* (1-pnorm (abs ( zcal)))
> pvalue
[1]  7.72268e - 05
```

∴ The value is less 0.01. we reject!

3) To random sample of size 1000 & 2000 are drawn from two population with same s.d 2.5. The sample means are 67.5 and 68. Test the hypothesis. $\mu_1 = \mu_2$ at 5% level of significance.

```
> n1=1000
> n2=2000
> mx1 = 67.5
> mx2 = 68
> sd1 = 2.5
> sd2 = 2.5
> zcal=(mx1 - mx2) / sqrt((sd1^2/n1) + (sd2^2/n2))
> zcal
[1] -5.163978
> pvalue = 2* (1-pnorm (ab(zcal)))
> pvalue
[1] 2.417564e -07
```

∴ (Rejected)

4) A study of noise level in 2 hospital is given below test the claim that 2 hospital have same level of noise at 1% level of significance.

	Hos. A	Hos B
	84	34
	61.2	59.4
	7.9	7.5

```
>n1=84
>n2=34
>mx1=61.2
>mx2= 59.4
>sd1=7.9
>sd2 =7.5
>zcal=(mx1-mx2)/sqrt((sd1^2/n1)+(sd2^2))
>zcal
[1] 1.162528
>pvalue
[1] 0.2450211
```

∴ The value is greater than 0.01 we accept the value.

5) In a sample of 600 students is ~~clg~~ clg 400 used Blue ink. In another clg from a sample of 900 students 450 use blue ink. Test the hypothesis that the preparation of students using blue ink in two colleges are equal or not at 1% level of significance.

$H_0: P_1 = P_2$ against $H_1: P_1 \neq P_2$

```
>n1= 600
>n2 = 400
>p1= 400/600
>p2= 450/900
>p=(n1*p1+n2*p2)/(n1+n2)
>p
[1] 0.566067
```

```
>q=1-p
>q
[1] 0.4333333
>zcal = (p1-p2) /sqrt (p* q* (1/n1 + 1/n2))
>zcal
[1] 6.381534
>pvalue = 2* (1-pnorm (abs(zcal)))
>pvalue
[1] 1.753222e-10
```

∵ value is less than 0.01 the value is ~~Ha~~ rejected.

6) $H_0 : P_1 = P_2$ agains $H_1 : P_1 \neq P_2$

```
>n1=200
>n2= 200
>p1=44/200
>p2= 30/200
>p=(n1*p1+n2*p2)/(n1+n2)
>p
[1] 0.185
>q= 1-p
>q
[1] 0.815
>zcal =(p1-p2) /sqrt (p*q * (1/n1+1/n2))
>zcal
[1] 1.802741
>pvalue = 2* (1-pnorm (abs(zcal)))
>pvalue
[1] 0.0714288
```

∴ Accept ∵ greater than 0.05.

9

5

PRACTICAL - 07

TOPIC : Small Sample Test

Q.1) The marks of 10 students are given by 63, 63, 66, 67, 68, 69, 70, 70, 71, 72. Test the hypothesis that the sample comes from a population with average marks 66.

Sol^n: $H_0 : \mu = 66$

```
> x=c(63, 63, 66, 67, 68, 69, 70, 70, 71, 72)
> t.test (x)
        one sample t-test

data: x
t= 69.319 , df= 9, p-value =1.558e-13
alternative hypothesis: true mean is not
equal to 95 percent confidence interval:
 65.65171  70.14829
 Sample estimates:
 mean of x
      67.9
```

Since p-value is less than 0.05 we reject hypothesis at 5% level of significance

```
> los = 0.05
> pvalue =1.558e-13
> if (pvalue > 0.05) {cat("accept Ho")}
else {cat("reject Ho")}
reject Ho>
```

Q.2) Two groups of students score the following marks : ~~[illegible]~~ test the hypothesis that there is no significance difference between the two groups.
group 1 = 18, 22, 21, 17, 20, 17, 23, 20, 22, 21
group 2 = 16, 20, 14, 21, 20, 18, 13, 15, 17, 21

Solⁿ: Ho : There is no difference between two groups.

```
> x=c(18,22,21,17, 20, 17, 23, 20, 22,21)
> y = c(16,20, 14, 21, 20, 18, 13, 15, 17, 21)
> t.test (x,y)
        Welch   Two Sample t-test
data: x and y
t= 2.2513 , df =16.316 , p-value = 0.03798
alternative hypothesis: true difference means
 it is not equal to 0.
95 percent confidence interval:
 0.1628205  5.0371795
Sample estimates :
mean of x  mean of y :
   20.1       17.5
> pvalue = 0.03798
> if (pvalue >0.05) { cat("accept Ho")}
  else { cat ("reject Ho")}
reject Ho>
```

Since p-value is less than 0.05 we reject hypothesis at 5% level of significance.

Q3) The sales data of 6 shops before and after a special campaign are given ~~[illegible]~~ below :

Before : 53, 28, 31, 48, 50, 42

After : 58, 29, 30, 55, 56, 45

Test the hypothesis that the compaign is effective or not.

Soln, Ho : There is no significant difference of sales before and after the campaign.

```
>x= c(53, 28, 31, 48, 50, 42)
> y= c(58, 29, 30, 55, 56, 45)
> t.test (x, y, paired = T, alternative = "greater")
   Paired t-test
 data : x and y
 t= -2.7815 , df = 5 , p-value = 0.9806
alternative hypothesis : true difference in mean
is greater than 0.
95 percent confidence interval
-6.035547   in
   Sample estimates :
   mean of the difference
               -3.5

> pvalue = 0.9806
> if (pvalue > 0.05) { cat ("accept Ho") }
  else { cat (" reject Ho") }
accept Ho>
```

Since p-value is greater than 0.05 we accept hypothesis at 5% level of significance.

Q.4) Two medicines are applied to the two groups of patient respectively.

group 1: 10, 12, 13, 11, 14

group 2: 8, 9, 12, 14, 15, 10, 9.

Is there any significant difference between two medicines.

Soln: Ho: There is no significant difference between two medicines of two groups

```
> x = c(10, 12, 13, 11, 14)
> y = c( 8, 9, 12, 14, 15, 10, 9)
> t.test (x, y)

        welch Two Sample t-test

data: x and y
t = 0.80384 , df = 9.7594 , p-value = 0.4406
alternative hypothesis: true difference in
mean is not equal to 0.
95 percent confidence interval:
  -1.781171        3.781171
Sample estimates:
Mean of x and y:
       12          11
> pvalue = 0.4406
> if (pvalue > 0.05) { cat("accept Ho") }
else { cat ("reject Ho") }
accept Ho>
```

Since the p-value is greater than 0.05 we accept the hypothesis at 5% level of significance.

Q.5) The following are ~~tears~~ the weight before and after the diet program. Is the diet program effective,

Before : 120, 125, 115, 130, 123, 119

After : 100, 114, 95, 90, 115, 99.

Sol^n: Ho : There is no ~~diff~~ significant difference

```
>x=c(120,125,115,130,123,119)
>y=c(100,114,95,90,115,99)
>t.test(x,y, paired =T, alternative ="less")
```

Paired t-test

data: x and y

t= 4.3458 , df= 5 , p-value= 0.9963

alternative hypothesis: true difference in means is less than 0.

95 percent confidence interval:

Inf 29.0295:

sample estimates:

mean of the differences

19.83333

```
>pvalue =0.9963
>if (pvalue >0.05){ cat(" accept Ho")}
else{ cat(" reject Ho")}
```

accept H_0 >

Since the p-value is greater than 0.05 we accept the hypothesis at 5% level of significance.

PRACTICAL - 08

Topic : Large and Small Sample tests

Q.1) The arithmetic mean of a sample of 100 items from a large population is 52. If the standard is 7, test the hypothesis that the population mean is 55 against the alternative it is more than 55 at 5% LOS.

Soln:
$H_0 : \mu = 55$
$H_1 : \mu \neq 55$

```
> n = 100
> mx = 52
> mo = 55
> sd = 7
> zcal = (mx-mo) / (sd/(sqrt(n)))
> zcal
[1] -4.285714
> cat(" Calculated z value is =" zcal)
  calculated z value is = -4.285714
> pvalue = 2 * (1 - pnorm (abs (zcal)))
> pvalue
[1] 1.82153e-05
```

Since p value is less than 0.05 we reject the hypothesis at 5% LOS.

Q.2) In a big city 350 out of 700 males are found to be smokers. Ddoes the information supports that exactly half of the males in the city are smokers? Test at 1% LOS.

Soln: Ho: U

```
>P=0.5
>Q=1-P
>p=350/700
>n=700
>zcal=(p-P)/(sqrt(P*Q/n))
>cat("Calculated z value is =",zcal)
Calculated z value is =0
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 1
```

Since pvalue is greater than 0.05 we accept hypothesis at 1% LOS.

Q.3) Thousand article from a factory. A are found to have 2% defectives, 1500 articles from a 2nd factory: B are found to have 1% defective. Test at 5% LOS that the two factory are similar are not.

Soln: Ho: $P_1 = P_2$ as $H_1: P_1 \neq P_2$

```
>n1=1000
>n2=1500
>p1=0.02
>p2=0.01
```

```
> p=(n1 * p1+n2 *p2) / (n1+n2)
> p
[1]  0.014
> q = 1-p
> q
[1] 0.986
> zcal = (p1-p2) / sqrt (p* q* (1/n1 + 1/n2))
> p
[1]  0.014
> q = 1-p
> q
[1] 0.986
> zcal = (p1-p2) / sqrt (p*q *(1/n1 + 1/n2))
> zcal
[1]  2.084842
> cat ("calculated z value is =", zcal)
calculated z value is = 2.084842
> pvalue = 2* (1-pnorm (abs (zcal)))
> pvalue
[1] 0.03708364
```

Since pvalue is less than 0.05 we reject the hypothesis at 5% level of signifance.

Q.4) A sample of size 400 was drawn at a sample mean is 99. Test at 5% LOS that the sample comes from a population with mean 100 and variance 64.

Sol^n: Ho : μ = 100

```
> mx=99
>mo=100
>sd= 8
>n=400
>zcal=(mx-mo)/(sd/(sqrt(n)))
>zcal
[1] -2.5
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.01241933
```

Since pvalue is less than 0.05 we reject the hypothesis at 5% LOS.

Q.5) The flower stems are selected and the heights are found to be (cm) 63, 63, 68, 69, 71, 72 test the hypothesis that the mean height is 66 or not at 1% LOS.

Sol^n: Ho : μ = 66

```
>x=c(63, 63, 68, 69, 71, 72)
>t.test(x)
        One sample t-test
data : x
```

t = 47.94 , df = 6, p-value = 5.522e-09
alternative hypothesis : true mean is not equal to 0.
95 percent confidence interval :
64.66479 71.62092
Sample estimates :
mean of x
68.14286
Since pvalue is less than 0.05 we reject hypothesis at 1% level of significance.

Q.6) Two random samples were drawn from 2 normal populations and there values are
A - 66, 67, 75, 76, 82, 84, 88, 90, 92
B - 64, 74, 78, 82, 85, 87, 92, 93, 95, 97.
Test whether the population have the same variance of 5% Los.

Soln: $H_0: \mu = \sigma_1 = \sigma_2$
>x = c(66, 67, 75, 76, 82, 84, 88, 90, 92)
> y = c(64, 66, 74, 78, 82, 85, 87, 92, 93, 95, 97)
>var. test (x, y)
F test to compare two variance
data : x and y
F = 0.70686 , num df = 8, denom df = 10,
p value = 0.6359

Scanned with CamScanner

alternative hypothesis : true ratio of variances is not equal to 1.

95 percent confidence interval :

0.1833662 3.0360393

Sample of variance

0.7068567

```
> pvalue = 0.6359
> if (pvalue > 0.05) { cat ("accept Ho") }
   else { cat ("reject Ho") }
accept Ho >
```

Since p-value is greater than 0.05 we accept hypothesis at 5% LOS.

Q.77) A company producing light bulbs finds than mean life span of the population of bulbs is 1200 hours with s.d. 125. A sample of 100 bulbs have mean 1150 hours. Test whether the difference between population and sample mean is significantly different?

Soln: Ho : $\mu = 1200$

```
> mo = 1200
> mx = 1150
> n = 100
> sd = 125
> zcat (mx - mo) / (sd / (sqrt (n)))
> cat ("Calculated z value is =", zcal)
Calculated z value is = -4.
```

>pvalue = 2 * (1- pnorm (abs (zcal)))

[1] 6.334248e-05

Since pvalue is less than 0.05 we reject hypothesis.

Q.8) From each of two consignments of apples, a sample of size 200 is drawn and number of bad apples are counted. Test whether the proportion of rotton apples in two assignments are significantly different at 1% LOS ?

	Sample Size	No. of bad apples
Consignment 1	200	44
Consignment 2	300	56

Soln: Ho : u = $P_1 \neq P_2$

```
>n1=200
>n2=300
>p1=44/200
>p2=56/300
>p=(n1*p1+n2*p2)/(n1+n2)
>p
[1] 0.2
>q= 1-p
>zcal = (p1-p2)/sqrt(p*q*(1/n1+1/n2))
>zcal
[1] 0.9128709
```

```
> cat("Calculated z value is =", zcal)
Calculate z value is = 0.9128709
> pvalue = 2*(1 - pnorm(abs(zcal)))
> pvalue
[1] 0.3613104
```

Since pvalue is greater than 0.05 we accept the hypothesis.

Practical - 09

TOPIC: Chi-square Test & ANOVA

Q.1) Use the following data to test whether the condition of the home and condition of child are independent or not.

	Clean	Dirty
Clean	70	50
fairly Clean	80	20
Dirty	35	45

Soln: H_0: Condition of home and child are independents.

```
> x=c(70,80,35,50,20,45)
> m=3
> n=2
> y = matrix(x, nrow=m, ncol=n)
> y

      [,1]  [,2]
[1,]   70    50
[2,]   80    20
[3,]   35    45

> pv= chisq.test(y)
> pv
```

Pearson's Chi-square tests.

data : y

X-squared = 25.646 , df = 2 , p-value = 2.698e-

Pvalue is less than 0.05 we reject the hypothesis at 5% level of significance.

Q.2) Test the hypothesis that ~~the~~ the vaccination and the disease are independent or not.

	Aff.	Not aff.
Disease Aff.	70	46
N.A.	35	37

Soln: Ho: Disease and vaccination are independents.

```
> x = c(70,35,46,37)
> m = 2
> n = 2
> y = matrix(x, nrow=m, ncol=n)
        [,1]   [,2]
[1,]    70     46
[2,]    35     37

> pv = chisq.test(y)
> pv
```

data: y

X-squared = 2.0275, df = 1, p-value = 0.1545

∴ Pvalue is more than 0.05 we accept the hypothesis at 5% level of significance.

Q.3) Perform a ANOVA for the following data.

Type	Observation
A	50, 52
B	53, 55, 53
C	60, 58, 57, 56
D	52, 54, 54, 55

Solⁿ: Ho: The mean's are equal for A, B, C, D

```
> x1 = c(50, 52)
> x2 = c(53, 55, 53)
> x3 = c(60, 58, 57, 56)
> x4 = c(52, 54, 54, 55)
> d = stack(list(b1=x1, b2=x2, b3=x3,
              b4=x4))
> names(d)
[1] "values" "ind"
> oneway.test(values ~ ind, data=d,
              var.equal=T)
```

~~Oneway stests & values~~

One-way analysis of means

data: values and ind
F=11.735, num df = 3, denom df = 9,
pvalue = 0.00183

```
>anova = aov(values ~ ind, data = d)
>summary(anova)
```

	Df	Sum sq	Mean sq	F value	Pr(>F)
ind	3	71.06	23.688	11.73	0.00183 **
Residuals	9	18.17	2.019		

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

Since p values is less than 0.05 we reject the hypothesis at 5% level of significance.

Q.4) The following data gives the life of tyres of four brands.

Type	life
A	20, 23, 18, 17, 18, 22, 24
B	19, 15, 17, 20, 16, 17
C	21, 19, 22, 17, 20
D	15, 14, 16, 18, 14, 16

Solⁿ: Ho: The average life of tyre is equal for (A, B, C, D)

```
>x1 = c(20,23,18,17,18,22,24)
>x2 = c(19,15,17,20,16,17)
>x3 = c(21,19,22,17,20)
>x4 = c(15,14,16,18,14,16)
>d = stack(list(b1=x1, b2=x2,
  b3=x3, b4=x4))
>names(d)
[1] "values" "ind"
> oneway.test(values ~ ind, data=d,
               var.equal= T)

  One-way analysis of means

data: values and ind
F=6.8445, num df=3, denom df=20,
p-value =0.002349.
```

Since p value is less than 0.05 we reject the ~~importance~~ hypothesis at 5% level of significance.

Q.5) Import excel file in R software

Sol^n: >x = read.csv ("C:/users/admin/Desktop/marks.csv")
>print(x)

	Stats	Maths
1	40	60
2	45	48
3	42	47
4	15	20
5	37	25
6	36	27
7	49	57
8	59	58
9	20	25
10	27	27

```
>am = mean(x$stats)
>am
[1] 37
>me = median (x$stat)
>me
[1] 38.5
>n = length (x$stat)
>sd = sqrt((n- 1) * var(x$stats)/n)
>sd
[1] 12.64911
```

```
>am = mean (x $ maths)
>am
[1] 39.4
> me = median ( x $ maths)
>me
[1] 37
>n= length (x$Maths)
>sd = sqrt ((n-1) * var( x $ Maths)/n)
>sd
[1] 15.2
>cor( x $stats, x$ maths)
[1] 0.830618
```

Scanned with CamScanner

Practical -10

TOPIC: Non - Parametric Test

Q.1) Following are the amount of solphur oxide emitted by a industry in 20 days. Apply sign test, to test the hypothesis that the population median Rs 21.5 at 5% level of significance.
17, 15, 20, 29, 19, 18, 22, 25, 27, 9, 24, 20, 17, 6, 24, 14, 15, 23, 24, 26.

Solⁿ: Ho: Population median is 21.5

```
>x=c(17,15, 20, 29, 19, 18, 22, 25, 27, 9, 24,
     20, 17, 6, 24, 14, 15, 23, 24, 26)
>me = 21.5
>sp=length(x[x>me])
>sn = length(x[x<me])
>n = sp+sn
>n
[1] 20
>pv= pbinom (sp, n, 0.5)
>pv
[1] 0.4119015
```

Since pvalue is more than 0.05 we accept the hypothesis at 5% level of significance.

If the alternative (H_1) not equal to ($H_1 \neq ve$) (<) then pv=pbinom (sp, n, 0.5) and if H_1> me then pv=binom (sn, n, 0.5).

Q.2) Following is data obtained observations apply sign test to the hypothesis that the population median is ~~625~~ 625 against the alternative it is more than 625.

612, 619, 631, 628, 643, 640, 655, 649, 670, 663.

Solⁿ: Ho: population median is 625.

```
> x = c(612, 619, 631, 628, 643)
> me = 625
> sp = length(x[x > me])
> sn = length(x[x < me])
> n = sp + sn
> n
[1] 10
> pv = pbinom(sn, n, 0.5)
> pv
[1] 0.0546875
```

Since pvalue is greater than 0.05 we accept the hypothesis at 5% level of significance.

Q.3) Following are values of a sample test the hypothesis that the population median is 60 against the alternative it is more than 60. at 5% level of significance using wilcoxon signed Rank Test.

63, 65, 60, 89, 61, 71, 58, 51, 69, 62,
63, 39, 72, 69, 48, 66, 72, 63, 87, 69.

Soln: H0 : population median 60
H1 : population median > 60

```
> x = c(63, 65, 60, 89, 61, 71, 58, 51, 69,
        62, 63, 39, 72, 69, 48, 66, 72,
        63, 87, 69)
> wilcox.test(x, alter = "greater", mu = 60)
```

wilcoxon signed Rank Test
with continuity correction

data: x

V = 145, p-value = 0.02298
alternative hypothesis: true location is greater than 60.

Note: If the alternative is less then (alter = "less") and if alternative is not equal then (alter = "two.sided").

Since pvalue is less than 0.05 we reject the hypothesis at 5% level of significance.

Q.4) using WSR test
population median is 12 or less than 12.
15, 17, 24, 25, 20, 21, 32, 28, 12, 25, 24, 26

Soln: H_0: population median is 12.
H_1: population median <12

```
> x=c(15,17,24,25,20,21,32,28,12,25,
      24,26)
> wilcox.test(x, alter="less", mu=12)
```

Wilcoxon signed Rank Test
with continuity correction

data: x
V = 66, p-value = 0.9986
alternative hypothesis: true location is less than 12.

Since pvalue is greater than 0.05 we accept the hypothesis at 5% level of significance.

Q.5) The weights of students before and after they stop smoking are given below. Using WSRT test that there is no significant change.

Before	After.
65, 75, 75, 62, 72	72, 74, 72, 66, 73

Solⁿ: Ho : Before and after there is no significant change

H₁ : There is a change.

```
>x=c(65,75,75,62,72)
>y=c(72,74,72,66,73)
>d=x-y
>wilcox.test(x,alter="two.sided".mu=0)
```

wilcoxon Signed Rank Test with continuity correction

data : x

V=15, p-value = 0.05791

alternative hypothesis: true location is not equal to 0.

Since pvalue is greater than 0.05 we accept the hypothesis at 5% level of significance.