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.** 1758 **Year** 2019-20

This is to certify that the work entered in this journal is the work of Mst. / Ms. Himanthu CHAUHAN ·K

who has worked for the year 2020 in the Computer Laboratory.

Teacher In-Charge

Head of Department

Date : ____________

Examiner

Scanned by CamScanner

★★ INDEX ★★

Scanned by CamScanner

PRACTICAL - 1

AIM : Basics of R software.

1) R is a software for stastical 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

Q1 Solve the followings

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

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

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

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

3/ $5^3+7\times 5\times 8+46/5$

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


4/ $\sqrt{4^2+5\times 3+7/6}$

```
>sqrt(4^2+5*3+7/6)
[1] 5.671567
```

5. round off
$46\div 7+9\times 8$

```
>round(46/7+9*8)
[1]79
```

Q2

```
>c(2,3,5,7)*2
[1]4 6 10 14
>c(2,3,5,7)*c(2,3)
[1] 4 9 10 21
>c(2,3,5,7)*c(2,3,6,2)
[1] 4 9 30 14
>c(1,6,2,3)*c(-2,-3,-4,-1)
[1] -2 -18 -8 -3
>c(2,3,5,7)^2
[1] 4 9 25 49
>c(4,6,8,9,4,5)^c(1,2,3)
[1] 4 36 512 9 16 125
>c(6,2,7,5)/c(4,5)
[1] 1.50 0.40 1.75 1.00
```

Q3

```
>x=20 >y=30 >z=2
>x^2 +y^3+z
[1] 27402
>Sqrt(x^2+y)
[1] 20.73644
>x^z +y^z
[1] 1300
```

Q4/
```
>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
```

Q5/ Find $x+y$ and $2x+3y$ where $x = \begin{bmatrix} 4 & -2 & 6 \\ 7 & 0 & 7 \\ 9 & -5 & 3 \end{bmatrix}$

$y = \begin{bmatrix} 10 & -5 & 7 \\ 12 & -4 & 9 \\ 15 & -6 & 5 \end{bmatrix}$

```
>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,-6,7,9,5))
>y   [,1]  [,2]  [,3]
[1,]  10    -5     7
[2,]  12    -4     9
[3,]  15     6     5
>X+Y   [,1]  [,2]  [,3]
[1,]    14    -7    13
[2,]    19    -4    16
[3,]    24   -11     8
```



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

Q6/ Marks of stastics of CS Batch B

```
x = c(58,20,35,24,46,56,55,45,27,22,47,58,
      54,40,50,32,36,29,35,39)

>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 - 2

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 $\Sigma P(x) = 1$

$\therefore P(0) + P(1) + P(2) + P(3) + P(4) + P(5) = P(x)$
$= 0.1 + 0.2 - 0.5 + 0.4 + 0.3 + 0.5$
$= 1.0$

$\because P(2) = -0.5$, it can be a probability mass function
$\because p(x) \geq 0 \quad \forall, x$

2)

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

The condition for p.m.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 the $P(x) \neq 1$

3)

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

The condition for pmf is
1) $P(x) \geq 0 \quad \forall x$ satisfy
2) $\Sigma P(x) = 1$

$\Sigma P(x) = P(10) + P(20) + P(30) + P(40) + P(50)$
$= 0.2 + 0.2 + 0.35 + 0.15 + 0.1$
$= 1$

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

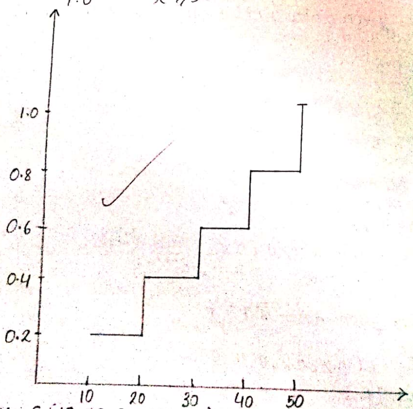
code:

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

Q2 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

$F(x) = 0$	$x < 10$
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	$x \geq 50$




```
> X = C (10, 20, 30, 40, 50)
> plot (X, cumsum (Prob), "S")
```


Q2 Find

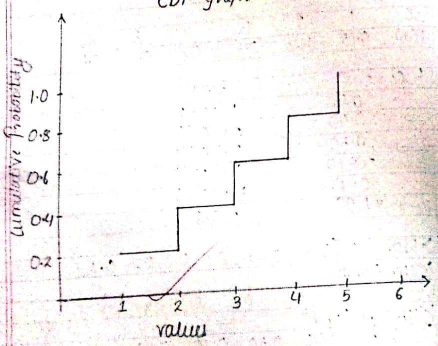
x	1	2	3	4	5	6
$P(x)$	0.15	0.25	0.1	0.2	0.2	0.1

$F(x) = 0$	$x < 1$
$= 0.15$	$1 \leq x < 2$
$= 0.40$	$2 \leq x < 3$
$= 0.50$	$3 \leq x < 4$
$= 0.70$	$4 \leq x < 5$
$= 0.90$	$5 \leq x < 6$
$= 1.00$	$x \geq 6$

```
> prob = c (0.15, 0.25, 0.1, 0.2, 0.2, 0.1)
> sum (prob)
[1] 1
> cumsum (prob)
[1] 0.15 , 0.40, 0.50, 0.10 , 0.90 , 1.00
> x = c (1, 2, 3, 4, 5, 6)
> plot (x, cumsum (prob), "S". xlab = "value",
    ylab = "cumulative probability",
    main = "CDF graph ; "col = "brown")
```

52

CDF Graph



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

(i) $f(x) = 3 - 2x \; ; \; 0 \leq x \leq 1$
(ii) $f(x) = 3x^2 \; ; \; 0 < x < 1$

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

$= [3x - x^2]_0^1 = 2.$

$\because$ The $\int_0^1 f(x) = 1 \quad \therefore$ It is not a pdf

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

$\int_0^1 f(x)$

$= \int_0^1 3x^2$

$= 3 \int_0^1 x^2$

$= \left[3 \frac{x^3}{3}\right]_0^1 \qquad \because x^n = \frac{x^{n+1}}{n+1}$

$= x^3$

$= 1$

The $\int_0^1 f(x) = 1 \quad \therefore$ It is a p.d.f

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

2- i) dbinom (4, 12, 0.2)
[1] 0.1328756
ii) pbinom (4, 12, 0.2)
[1] 0.9274445
iii) 1 - pbinom (5, 12, 0.2)
[1] 0.01940528

3- 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] 0.9455978
3) 1 - pbinom (7, 12, 0.25)
[1] 0.00278151
4) dbinom (6, 12, 0.25)
[1] 0.04014945

8

PRACTICAL - 3

TOPIC : Binomial distribution

$P(X=x)$ = dbinom (x, n, p)
$P(X \leq x)$ = pbinom (x, n, p)
$P(X > x)$ = 1 - pbinom (x, n, p)
If x is unknown
$P1 = P(X \leq x)$ = qbinom (P_1, n, p)

1) Find the probability of exactly 10 success in hundred trials with P=0.4

2) Suppose there are 12 mcq, Each question has 5 options out of which 1 is correct 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) $P(X \leq 5)$
iii) $P(x>7)$
iv) $P(5<x<7)$

8

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 salesman 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

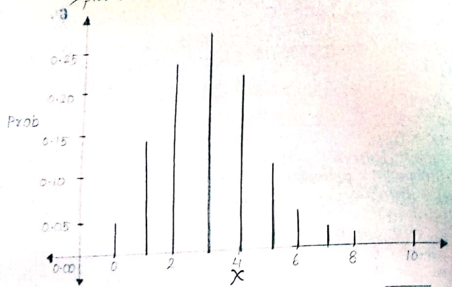
7) X follows binomial distribution with n=10, p=0·3 plot the graph of p.m.f and c.d.f.

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

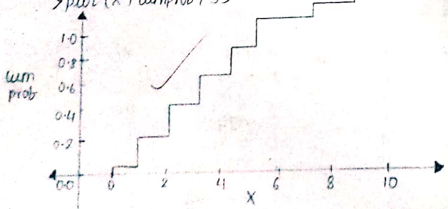
6)
```
qbinom (0.88 ,f30, 0.2)
[1] 9
```

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

>plot (X , prob ,"h")



>plot (X , cumprob , "s")



PRACTICAL - 4

AIM : Normal Distribution.

i.) $P(x=X) = $ dnorm (x, μ, σ)
ii.) $P(X \leq x) = $ pnorm (X, μ, σ)
iii) $P(X > x) = 1 - $ pnorm (x, μ, σ)
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. $P(X \leq 15)$ ii. $P(10 \leq X \leq 13)$ 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
```

a

```
>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  12.27246
```

2. X follows normal distribution with $\mu = 10$, $\sigma = 2$
Find i $P(X<7)$ 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(04, 10, 2)
>a5
[1] 9.493306.
```


Q3 Generate 5 random numbers from a normal distribution $\mu = 15$, $\sigma = 4$ Find sample Mean, median, S.D and print it.

CODE:

```
>rnorm(5,15,4)
[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
```

Q4 $X \sim N(30,100)$, $\sigma = 10$

i. $P(x \leq 40)$

ii. $P(x > 35)$

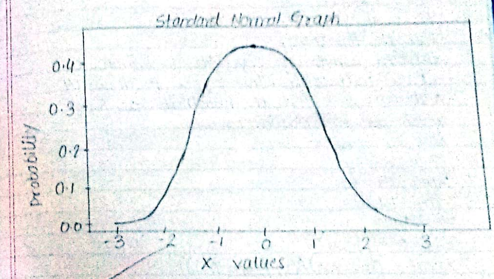
iii. $P(25 < x < 35)$

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
```

Q5 Plot the standard normal graph.

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

PRACTICAL - 5

TOPIC : Normal and t-test

1. Ho : $\mu = 15$ (Null) H1 : $\mu \neq 15$ (allarative)

Test the hypothesis Random sample of size 400 is 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 > accept the value
0.05 < less than Reject

```
>mo=15
>mx=14
>n=400
>sd=3
>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 = μ = 15

2) Test the hypothesis Ho = μ = 10 against H1 : $\mu \neq 10$ A random sample size of 400 is drawn with sample mean = 10.2 and SD = 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: proportional of smokers in 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 900)

```
>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 9 field crops are insect pollutell. 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 0.1 so value is accepted.

5) Test the hypothesis H0: u=12.5 from the following sample at 5% level of significance

```
>x=c(12.25,11.97,12.15,12.08,12.31,12.28,11.94,11.89,
   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-pnorm(abs(t)))
>pvalue
[1] 0
```

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

a

PRACTICAL - 6

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 SD 30. Test the hypothesis that the population mean is 250 or not on 5% level of significance.

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.

1. Solution :

```
>mo = 250
>mx = 275
>sd = 30
>n = 100
>zcal = (mx - mo)/(sd/(sqrt(n)))
>cat ("calculated value 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 Ho = μ = 250.

2. 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 : ", 7 -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 SD 2.5 the sample means are 67.5 and 68 Test the hypothesis μ1 = μ2 at 5% of significance.

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

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

3. Solution :

```
>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(abs(zcal)))
>pvalue
[1] 2.417564e-07 ∴ (Rejected)
```

4.

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



```
[1] 0.2450211
```

∴ The value is greater than 0.01 we accept the value

5. Ho : P1=P2 against H1 : P1 ≠ P2.

```
>n1 = 600
>n2 = 900
>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] 1.75322e-10
```

∵ Value is less than 0.01 the value is rejected

6. $H_0 : P_1 = P_2$ vs $H_1 : P_1 \neq P_2$

```
>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).

PRACTICAL - 7

TOPIC : Small sample test

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

$H_0 : \mu = 66$

```
>x = c (66., 63, 66, 67, 68, 69, 70, 70, 71)
>t.test (x)
one sample t-test
 data : x
 t = 68.319 , df = 9, pvalue = 1.558e-13
 alternative hypothesis
 True mean is not equal to 0
95 percent confidence interval
   65.65171          70.14829
sample estimate
 mean of x
   67.9
```

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

a

2. Two groups of student scorred the following marks Test the hypothesis that there is no significant difference between the 2 groups.

GR1 - 18, 22, ~~27~~, 21, 17, 20, 17, 23, 20, 22, 21
GR2 - 16, 20, 14, 21, 20, 18, 13, 15, 17, 21

Ho: There is not difference b/w the 2 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 = 2.2573  df = 16.376  p-value = 0.03798
alternative hypothesis :
   True difference in means is not equal
to 0 95 percent confindence interval :
 0.1628205     5.0371195
sample estimates :
 mean of x  mean of y
   20.1           17.5

>p-value = 0.03798
>if (pvalue > 0.05) {cat("accept ho")}
else {cat("reject ho")}
reject Ho.
```

(PAIRED - T-test)

3) The sales data of 6 shops before & after a special campaign are given below.

Before : 53, 28, 31, 48, 50, 42
After : 58, 29, 30, 55, 56, 45

Test the hypothesis that the campaign is affective or not.

Ho : There is no significance difference of sales before & after campaign.

```
>x = c(Before)
>y = c(after)
>t.test(x, y, paired = T, alternative = "greater")
     paired t-test
   data: x & y
   t = -2.7815, df = 5, pvalue = 0.9806
   alternative hypothesis:
      True difference in means is greater than 0
   95 percent confidence interval:
      -6.035547   inf
   sample estimates
   mean of the difference
            -3.5
```

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

Scanned by CamScanner

4) Following are the weights before & 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 :- H0 : There is bio significance difference

```
> x = c (Before)
> y = c (after)
> t.test (x, y, paired = T, alternative = "less")
```

paired t-test

data : x & y
t = 4.3458, df = 5, pvalue = 0.9963
alternative hypothesis : true difference in means is less than 0.
95 percent confindence interval :
– inf 29.0295

sample estimates :
mean of the differences
19.83333

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

5. 2 medicines are applied to two groups of patient respectively.

gr1 : 10, 12, 13, 11, 14
gr2 : 8, 9, 12, 14, 15, 10, 9
Is there any significance difference b/w 2 medicines

H0 : There is no significance difference

```
> x = c (grp1)
> y = c (grp2)
> t.test (x, y)
```

data : x & y
t = 0.80384, df = 9.7594, pvalue = 0.4406
alternative hypothesis : true difference in means is not equal to 0
95 percent confidence interval
– 0.9698553 4.2981886.
Sample estimates :
mean of x mean of y.
12.0000 10.33333

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

PRACTICAL - 8

TOPIC : Large and Small Test

#Questions·

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:- H0 : μ = 55   H1 : μ ≠ 55

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

∵ The value is less than 0.05 we will reject the value H0 = μ = 55

Q2. In a big city 350 out of 700 males are found to be smokers. Does the information supports that exactly half of the males in the city are smokers? Test at 1% LOS.

Solⁿ:- p = 0.5   H0 : μ = 0.5

```
>Q = 1-P
>p = 350/700
>n = 700
>zcal = (p-P)/sqrt((P*Q/n))
>zcal
[1] 0
>pvalue = 2 * (1 - pnorm (abs (zcal)))
>pvalue
[1] 1
```

∵ The value of 1 is greater than 0.01 we accept the Hypothesis.

Q3. 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:- H0 : P1 = P2   as H1 : P1 ≠ P2

```
>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))
>zcal
[1] 2.084842
>pvalue=2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.03708364
∵ pvalue is less than 0.05 we reject the hypothesis.

Q4. 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.

Soln:-
H0: 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
∵ The value is less than 0.05 we reject the hypothesis.

Q5. The flower stems are selected and the heights are found to (cm) 63, 63, 68, 69, 71, 71, 72, test the hypothesis that the mean height is 66 or not at 1% LOS.

Soln:- Ho : μ=66
>x = c(63, 63, 68, 69, 71, 71, 72)
>t.test(x)

One sample t-test
data : x
t = 47.94, df = 6, pvalue = 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
∴ p value is 0.05 less than we reject the hypothesis.

Q6 Two random samples were drawn from 2 normal population and there values are A - 66, 67, 75, 76, 82, 84, 88, 90, 92 B - 64, 66, 674, 78, 82, 85, 87, 92, 93, 95, 97. Test whether the population have the same variance at 5% LOS.

Solⁿ:-

```
>x = c(66, 67, 75, 76, 82, 84, 88, 90, 92)
>y = c(64, 66, 74, 78, 82, 92, 93, 95, 97)
```

Ho : $\sigma 1 = \sigma 2$

```
>var.test(x, y)
```

F test to compare the two variance

data : x & y.

F = 0.63833

num df = 8

denom df = 9.

pvalue = 0.5383

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

95 percent confidence interval :

0.1556172   2.781378

sample estimates :

ratio of variances

0.6383349.

∴ pvalue is 0.05 we accept the value.

Q7. A company producing light bulbs finds that 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 : μ : 1200

```
>mx = 1150
>m0 = 1200
>sd = 125
>n = 100
>zcal (mx-m0)/(sd/sqrt(n))
>zcal
[1] -4
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 6334288e-05
```

∴ pvalue is less than we reject the hypothesis.

Q8. From each of two consigments 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.

Sol^n:- H0 : u : P1 ≠ P2

```
>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.
>pvalue=2*(1-pnorm(abs(zcal))
>pvalue
[1]0.3613104
```

∵ pvalue is less than 0.05 we reject the hypothesis.

PRACTICAL - 9

TOPIC : Non-parametric Testing of Hypothesis using R - enviroment

1. The following data represent earnings (in dollars) for a random sample of five common stocks listed on the New york stock exchange. Test whether median earnings is 4 dollars

Data :- 1.68 , 3.35 , 2.50 , 6.23 , 3.24.

```
>x<- c(1.68, 3.35, 2.50, 6.23, 3.24);
>n<- length(x);
>n
>[1]5
>x>4;
[1]FALSE FALSE FALSE TRUE FALSE
>s<- sum(x>4); s;
[1] 1
> binom.test(s,n,p=0.5, alternative ="greater");
        Exact binomial test
data: s and n.
number of successes = 1, number of trials =5, p-value = 0.9688
alternative hypothesis: true probability of success
is greater than 0.5; 95
95 percent confidence interval:
 0.01020622 1.00000000
  sample estinates:
  probality of success
        0.2.
```

2. The scores of 8 students in reading before and after lesson are as follows : Test whether there is effect of reading

Student NO :	1	2	3	4	5	6	7	8
Score before :	10	15	16	12	09	07	11	12
Score After :	13	16	15	13	09	10	13	10

CODE :

```
>b <- c(10,15,16,12,09,07,11,12);
>a <- c(13,16,15,13,08,10,13,10);
>D <- b-a ;
> wilcox -test (D, alternative = "greater");
```

Wilcoxon signed rank test with continuity correction data : D

V = 10.5 , p- value = 0.8722

alternative hypothesis : true location is greater than 0

warning message :

In wilcox.test.default(D, alternative = "greater");

cannot compute exact p-value with ties

∴ p-value is greater than 0.05 we accept it.

3. The diameter of a ball bearing was measured by 6 inspectors, each using two different kinds of calipers. The results were, Test whether average ball bearing for.

Inspector :	1	2	3	4	5	6
Caliper 1 :	0.265	0.268	0.266	0.267	0.269	0.264
Caliper 2	0.263	0.262	0.270	0.261	0.271	0.260

Caliper 1 and Caliper 2 are same.

CODE :

```
>x <- c (0.265,0.268,0.266,0.267,0.269,0.264);
>y <-6 (0.263,0.262,0.270,0.261,0.271,0.260);
>wilcox.test (x,y, alternative ="greater")
```

Wilcoxon rank sum test

data : X and Y

W = 24, P = 0.197

alternative hypothesis : true location shift is greater than 0.

∴ p-value is ~~less~~ ~~the~~ greater than 0.05 we accept it

4. An officer has three elective typewriters A, B, and C. In a study of machine usage, firm has kept records of machine usage rate of seven weeks. Machine A was out for repairs for two weeks. It is of interest to find out which machine has better usage rate. Analyze the following data on usage rates and determine if there is a significant difference in average rate.

A	B	C
12.3	15.7	32.4
15.4	10.8	41.2
10.3	45.0	35.1
08.0	12.3	25.0
14.6	08.2	08.2
–	20.1	18.4
–	26.3	32.5

CODE:

```
>x <- c(12.3, 15.4, 10.3, 8.0, 14.6);
>n1 <- length(x);
>n1
[1] 5
>y <- c(15.7, 10.8, 45.0, 12.3, 8.2, 20.1, 26.3);
>n2 <- length(y);
>n2
[1] 7
>z <- c(32.4, 41.2, 35.1, 25.0, 8.2, 18.4, 32.5);
>n3 <- length(z);
>n3
[1] 7
```



```
>X <- c(x, y, z);
>g <- c(rep(1, n1), rep(2, n2); rep(3, n3));
>kruskal.test(X, g)

        Krushkal-Wallis rank sum test
data: X and g.
kruskal-wallis chi-squared = 5.217, df = 2,
                        p-value = 0.07365.
```

∴ p-value is greater than 0.05 we accept it

PRACTICAL - 10

AIM : Chi-square tests & ANOVA (Analysis of variance)

Q1. Use the following data to test whether the condition of home & condition of child are independent or not.

Cond. child \ Cond. Home	Clean	Dirty
Clean	70	50
Fairly clean	80	20
Dirty	35	45

H0 : condition of home & child are independent

```
>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-squared test

Data : y
X-squared = 25.646
df = 2
p-value = 2.698e-06
// They are dependent

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

Q2. Test the hypothesis that vaccination & disease are independent or not.

Disease \ vaccine	Affected	Not Affected
Affect	70	46
Non-affected	35	37

H0 : Disease & vaccine are independent

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

```
> pv = chisq.test(Y)
> pv
pearson's chi squared test with
Yates' continuity correlation
data: Y
X - square = 2.0275
df = 1
p- value = 0.1545
```

∵ p-value is more than 0.05 we accept the hypothesis at 5% level of significance

// They are Independent.

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

TYPE	OBSERVATIONS
A	50, 52.
B	53, 55, 53
C	60, 58, 57, 56
D	52, 54, 54, 55

H0 : 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"
> one way.test(values ~ ind, data = d, var.equal = T)
        one-way analysis of means.
data : values and ind.
F = 11.735    df = 3  ,  denom df = 9 ,
Pvalue = 0.00183
```

p-value is less than 0.05 we reject the hypothesis

```
> anova = aov(value ~ ind, data = d)
> summary (anova)
            Df    Sum     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
```

Q4. Following data gives a life of tires of 4 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

H0 : The mean's of A, B, C, D are equal

```
>X1 = C(20, 23, 18, 17, 18, 22, 24)
>X2 = C(B)
>X3 = C(C)
>X4 = C(D)
>d = stalk (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 & ind
F = 6.8445, num df = 3    denom df = 20
```

p-value is less than 0.05 we reject the hypothesis

```
>anova = aov (values ~ ind, data = d)
>summary (anova)
            Df    sumsq    meansq   Fvalue   Pr(>F)
ind          3    91.94    30.479   6.845    0.0025 **
Residuals   20    89.06    4.453
Signif codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
>x = read.csv("C:/users/admin/Desktop/marks.csv")
>x
     stats   maths
1     40      60
2     45      48
3     92      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
>am1 = mean (x $ maths stat)
>am1
[1] 39.4
>m1 = median (x $ stats)
>m1
[1] 38.5
>m2 = median (x $ maths)
[1] 37
>n = length (x $ stats)
>n
[1] 10
```

```
>sd = sqrt((n-1)*var(x$stats)/n)
>sd
[1] 12.6491
>n1 = length(x$maths)
>n1
[1] 10
>sd1 = sqrt((n-1)*var(x$maths)/n)
>sd1
[1] 15.2
>cor(x$stats, x$maths)
[1] 0.830618
```