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 2 **UID No.** ______

Class FYBSc(CS) A **Roll No.** 1742 **Year** 2019-2020

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

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

Teacher In-Charge

Head of Department

Date : 14/03/20

Examiner

★★ INDEX ★★

Scanned with CamScanner

08 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.

Q.1 Solve the followings

1. $4+6+8\div2-5$
```
> 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

$46 \div 7 + 9 \times 8$

> round (46/7 + 9*8)

[1] 79

Q2

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

→ 4 6 10 14

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

→ 4 9 30 14

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

→ 4 9 25 49

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

→ 1.50 0.40 1.75 1.00

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

→ 4 9 10 21

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

→ -2 -18 -8 -3

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

→ 4 36 512 9 16 125

Q3

> 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

Q4. >x <- matrix(nrow = 4, ncol=2 data = c(1,2,3,
4,5,6,7,8))

> x

	[,1]	[,2]
[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(n row = 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 (n row = 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	6
[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 statistic 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

G

Practical - 2

Topic - Probability distribution

1) Check wether the following 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 pmf then $\Sigma P(x) = 1$

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

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

$= 1$

$\because P(2) = -0.5$, it can be a probability mass function.

$\because P(x) \geq 0 \ \forall, x$

2)

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

The Condition for PmP 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

i) $p(x) \geq 0 \ \forall x$ satisfy

ii) $\sum p(x) = 1$

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

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

$= 1$

$\therefore$ The given data is pmf

Code:

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

Q2 Find the cdf for the following PMF and sketch the graph

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

$F(x) = 0 \quad x < 10$

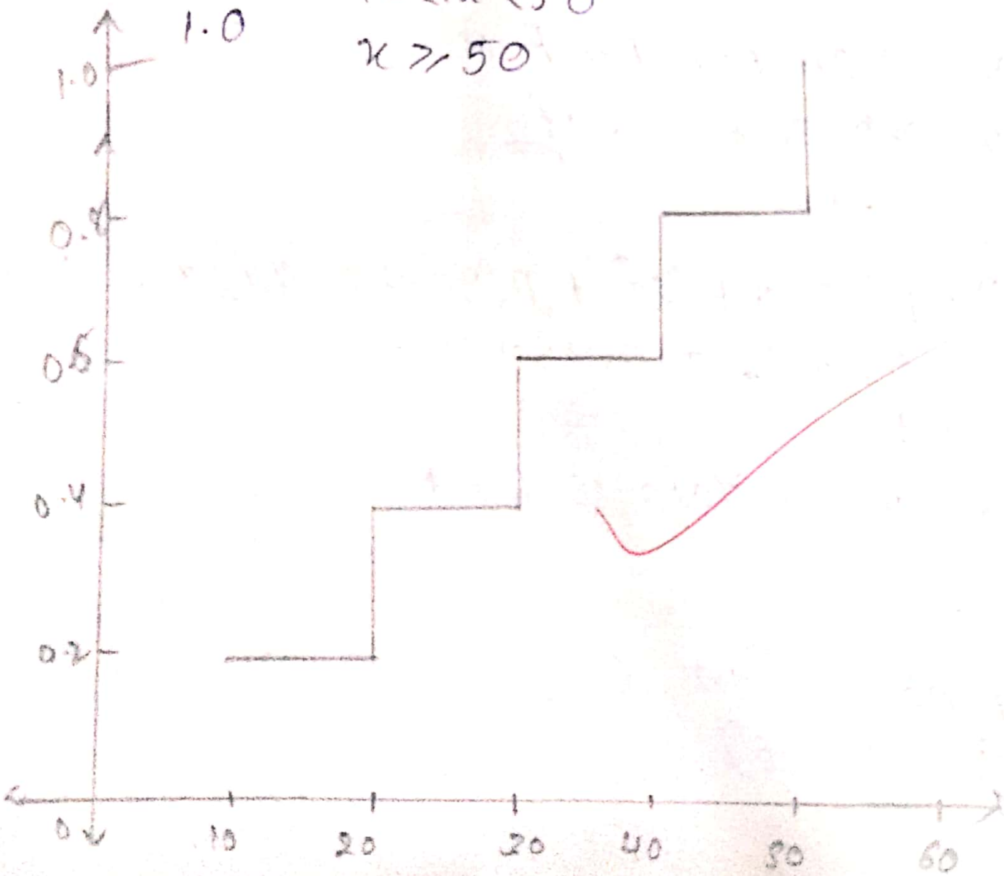
$0.2 \quad 10 \le x < 20$

$0.4 \quad 20 \le x < 30$

$0.75 \quad 30 \le x < 40$

$0.90 \quad 40 \le x < 50$

$1.0 \quad x \geq 50$




```
> x = c(c(10,20,30,40,50)
> plot = (x, cumsum(Prob)."s")
```

Q2

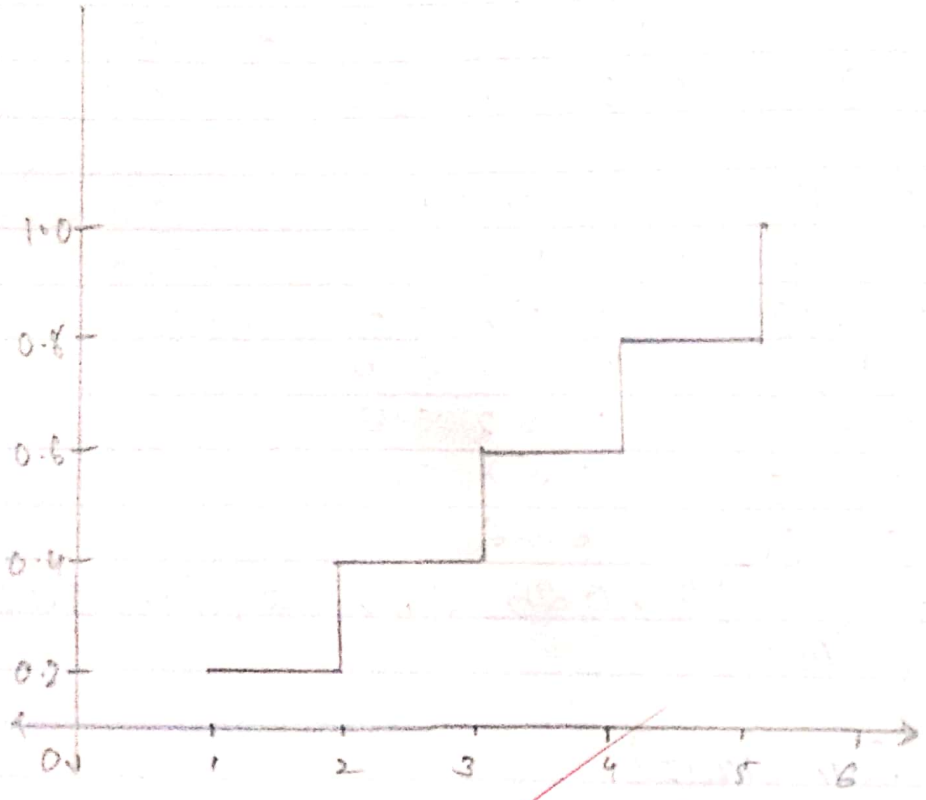
Q. Find

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

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

```
> 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")
```

ϕ



3. Check that wether the following is P.d.f or not

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

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

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

$= \int_0^1 f(x)$

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

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

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

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

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

$\int_0^1 f(x)$

$= \int_0^1 3x^2$

$= 3\int_0^1 x^2$

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

$= x^3$

$= 1$

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

.d.f

1) > x = dbinom (10,100,0.1)

> x

[1] 0.13

2) i) dbinom (4,12,0.2)

[1] 0.1328756

(ii) pbinom(4,12,0.2)

[1] 0.427445

(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] 0.08 298157

(4) dbinom (6,12,0.25)

[1] 0.04014945

Scanned with CamScanner

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

$p_1 = P(X \leq x)$ = qbinom (p_1, n, p)

1) Find the probability of exactly 10 success in hundred trials with ~~p0.1~~ $p = 0.1$

2) Suppose there are 12 mcq, each question has 5 options out of which one is correct find the probability of having exactly 4 correct answers.
ii) atmost 4 correct answer.
iii) more than 5 correct answer.

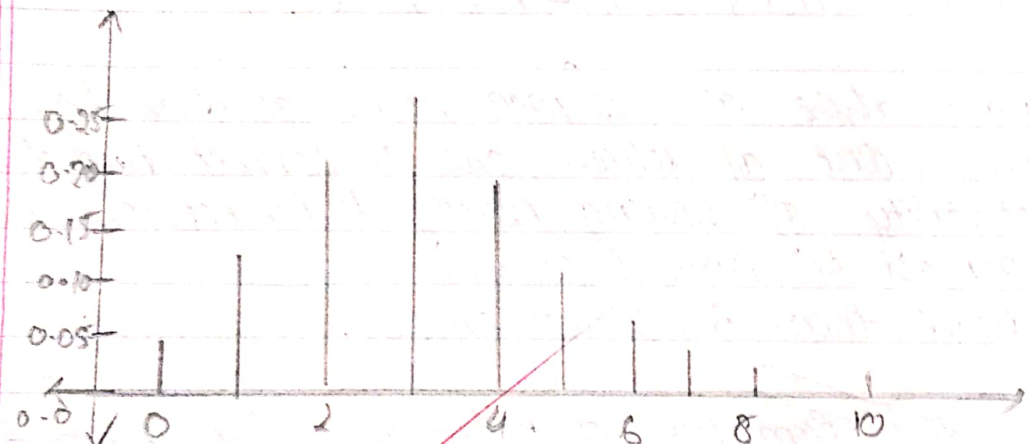
3) Find the complete distribution when $n = 5$ & $p = 0.1$

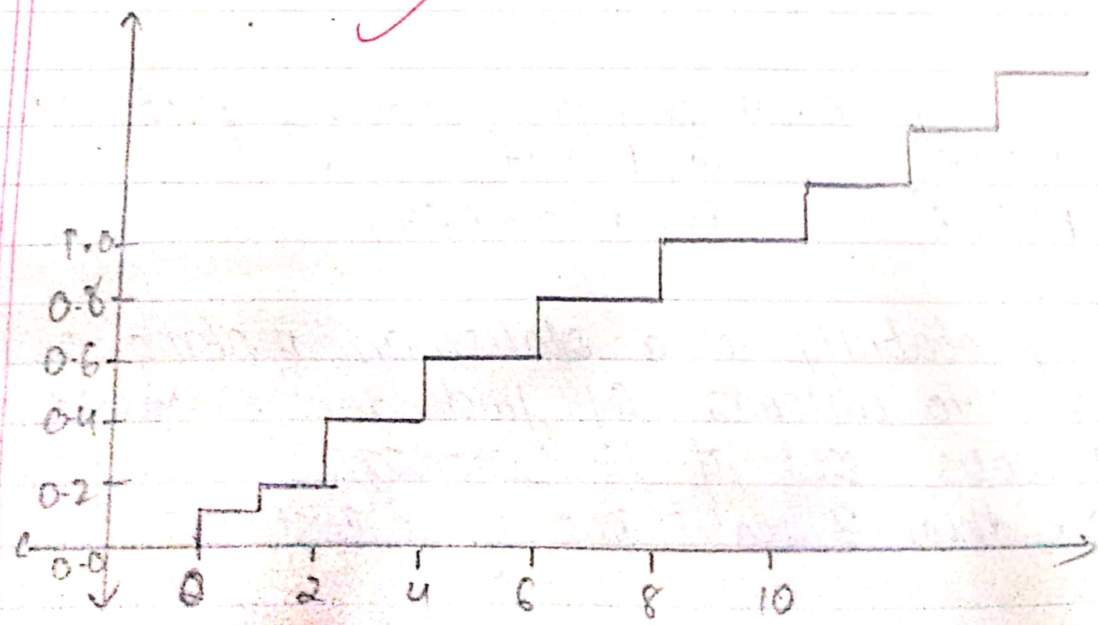
4) $n = 12$, $p = 0.25$ find the following probabilities.
i) $P(X = 5)$
ii) $P(X \leq 5)$
iii) $P(X > 7)$
iv) $P(5 < X < 7)$

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 80% 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) > ~~d~~ 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)
> 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

Practical-4

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) A random variable X follows normal distribution with mean = μ = 12 and S.D = σ = 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
```

FA.

Q3

```
> Cat ("P(x>14)= ", P3)
P (x>14) = 0.2524925
> P4 = rnorm (5,12,3)
> P4
[1] 15.2547283    16.548505  11.280515  6.419944
    12.272460
```

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) Generates 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    10.915006
> a5 = qnorm (0.4,10,2)
> 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. 48

Code:

```
>rnorm (5,15,4)
[1] 10.7649   7.793249   9.953444   13.345904
    17.509668
>am = mean (x)
>am
[1] 11.87845
> 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)
  SD is = 3.331613
```

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

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
```

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. H0 : M = 15   H1 : M ≠ 15

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 > accepthe the value

0.05 < less than Reject.

```
> m0 = 15
> mx = 14
> n = 400
> sd = 3
> zcal = (mx - m0) / (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 H0 = M = 15

2) Test the hypothesis H0 = M = 10 against H1 : M ≠ 10

A random sample size of 400 is drawn with sample mean = 10.2 and SD = 2.25

Test the hypothesis at

```
> m0 = 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 p value is greater than 0.05
∴ The value is accepted.

3. Test the hypothesis H0 : 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 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 farmers lost 20% of their crops. A random sample of 60 fields are collected and it is found that 9 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))
[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 $H_0: \mu = 12.5$ 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
> m0 = 12.5
> t = (mx-m0)/(sd/sqrt(n))
> t
[1] - 8.8949049
> pvalue = 2* (1-pnorm(abs(t)))
> pvalue
[1] 0
∴ The value is less than 0.05 the value is ~~to~~ accepted.

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

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

1. Solution:

> mo = 250

> mx = 275

> sd = 30

> n = 100

> z cal = (mx - mo) / (sd / (sqrt (n)))

> cat ("Calculated value of z is =", z cal)

[1] Calculated value of z is = 8.3333333

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

> Pvalue

[1] 0

∴ The value is less than 0.05 we will reject the value of $H_0 = \mu = 250$.

2

```
> 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 SD 2.5 the sample means are 67.5 and 68 Test the hypothesis $M_1 = M_2$ 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 is clg 400 used blue ink. In other clg from a sample of 900 students 450 use blue ink Test the hypothesis that the ~~pattern~~ porportion of students using blue ink in two colleges are equal or not at 1% level of significance.

3. Soln

```
> n1= 1000
>n2 = 2000
>m x1 = 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= 7.0
>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.

$H_0: P1 = P2$ against $H_1: P1 \neq P2$

```
> n1 = 600
> n2 = 900
> p1 = 400/600
> p2 = 450/900
> p = (n1 * p1 + n2 * p2)/(n1+n2)
> p
[1] 0.5666667
> 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] 1.753226e-10
```

∴ value is less than 0.01 the value is rejected.

6 $H_0 : p_1 = p_2$ as $H_1 : p_1 \neq p_2$

```
> n1 = 200
> n2 = 200
> p1 = 44/200
> p2 = 30/200
> p = (n1*p1 + n2*p2) / (n1+n2)
> p
[1] 0.185
> 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

AIM : Small Sample test.

1. The marks of 10 students are given by 66, 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 - test

data : x

t= 68.319 , df = 9 , pvalue = 1.558e-13

alternative hypothesis.

True mean is not equal to 0 95% confidence interval

65.65171 70.14829

sample estimates

mean of x

67.9

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

2. Two groups of students scored the following marks. Test the hypothesis that there is no significant difference b/w the 2 groups.

GR1 - 18, 22, 21, 17, 20, 17, 23, 20, 22, 21
GR2 - 16, 20, 14, 21, 20, 18, 13, 15, 17, 21
Ho : There is no 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 as present confidence interval:
    0.1628205        5.0371795
  sample estimate :
   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)
```

) 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 effective 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

∴ P value is greater than 0.05, we accept the hypothesis at 5% level of significance.

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 5.

Soln: H_0 : There is no 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 confidence 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 medicians 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?

H_0: 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

Aim: 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 :- $H_0 : \mu = 55$ $H_1 : \mu \neq 55$

```
> n=100
> mx = 52
> m0 = 55
> sd = 7
> zcal - ((mx-m0)/(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 $H_0 = \mu = 55$

Q.2 In a
to be
mach
The

Soln :

```
> Q = 1
> p = 3
> n = 7
> zcal
> zcal
[1] 0
> pvalue
> pvalue
[1] 1
```

∵ The
accep

Q3 Thous
2%
foun
the

Soln.

```
> n1 =
> n2 =
> p1 =
> p2 =
> p =
```

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 H_0 : 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:- H_0 : $P_1 = P_2$ 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)
[1] p 0.014
```

> q = 1-p

> q

[1] 0.986

> zcal= (p1-p2)/sqrt(p*q*(1/n1 + 1/n2))

> zcal

[1] 2.084842

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

> pvalue

[1] 0.03708364

∵ pvalues is less than 0.05 we reject the hypothesis

Q4 A sample of size 400 was drawn at a sample mean 99. Test at 5% LOS that the sample comes from a population with mean 100 and variable 64.

→

$H_0 : \mu = 100$

> 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 (Reject).

5)

Scanned with CamScanner

1) The flower item 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.

→ $H_0 : \mu = 66$

> 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-04

alternative hypothesis : true mean is not equal to 0

95 percent confidence interval :

64.66479 71.62092

sample estimates :

means of x

68.14286

∴ Pvalue ~~is~~ '0.05' less than we reject the hypothesis.

6 Two random samples were drawn from a normal population and there valuues are.
A- 66, 67, 75, 76, 82, 84, 88, 90, 92 ~~B- 84, 66, 74, 78, 82, 85, 87, 92, 93~~, 88.
B- 64, 66, 74, 78, 82, 85, 87, 92, 93, 95, 97.
Test wether the population have the same variance at 5% los

→

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

$H_0 : \sigma_1 = \sigma_2$

> var-test (x,y)

F-test to compare 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 2:

95 percent confidence interval:

0.1558172 2.781878

sample estimates:

ratio of variance.

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 SD 125. A sample of 100 bulbs have mean 1150 hours. Test wether the difference between population and sample mean is significantly different?

Sol^n $H_0 : \mu = 1200$

```
> mx = 1150
> mo = 1200
> sd = 125
> n = 100
> zcal (mx-mo)/(sd/sqrt(n))
> zcal
[1] -11
> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue

[1] 63.34288e-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 rotten apples in two assignments are significantly different at 1% LOS?

	samplesize	No. of badapples
Consignment1	200	44
Consignment2	300	56

Soln H_0 : H_1: $p_1 \neq p_2$

```
>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
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.3513104
```

∴ pvalue is less than 0.05 we reject the hypothesis.

[3] Practical-9

Aim: Non-parametric testing of Hypothesis using R-Environment.

1. The following data represent earnings (in dollars) for a random sample of five common stock listed on the new york stack exchange. Test wether median earnings in 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 TRUE
>S <- Sum (x>4) ;S;
[1] 1
> binom . test (s,n,p=0.5, alternative = "greater");
            Exact binomal test
data : S and n
number of success = 1 , number of trials =5,
p-value alternative hypothesis : true probability of
success = 0.9688. is greater than 0.5,
  95 percent Confidence interval:
     0.01020622     1.0000000
   sample estimates :
   probability of success
              0.2
```

1. The scores of 8 students in reading before and after lesson are as follows:
Test wether 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

Sol:

```
> 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 correct 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 wether 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 caplier 2 are same

Code:

```
>x <- c(0.265, 0.268, 0.266, 0.267, 0.269, 00.64)
>y <- c(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 greater than 0.05 we accept it.

10. office has three electric type writers A, B and C. In a study of machine usage firm has kept records of machine usage rate of seven weeks machine C. It was out of or repairs for two weeks. It is of interest to find out which machine has better usage rate. Analyze the following data on usage rate and determine if there is a significant difference in average.

A	B	C
12.3	15.7	32.4
15.4	10.8	41.2
10.3	45.0	35.1
08.0	19.3	25.0
14.6	8.2	08.2
-	20.1	15.4
-	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 test & ANOVA (Analysis of Variance)

Q. Use the following data to test wether the condition of child are independent or not.

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

H_0: 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 independent

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

2) Test the hypothesis that vaction & disease are independent or not.

Vacine

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

H_0 : Disease & Vaccines 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
```

Scanned with CamScanner

> pv = chisq.test(y)
> pv
Pearson's chi squared test with Yate's continuity correction
data: y
X-squared = 2.0275
df = 1
p-value = 0.1545
∴ p-value is more than 0.05 we accept the hypothesis at 5% level of significance.

3 Perform a ANOVA for the following data

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

H_0 : 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)
        One-way analysis of means
data: values and ind
F = 11.735   df = 3, denom df = 9,
```

pvalue = 0.00183

∴ pvalue is less than 0.05 we reject the hypothesis.

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

> summary (anova)

	DF	Sum	Meansq	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

4 Following data gives a life of the 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

H_0 : The means of A, B, C, D are equal

> x_1 = c(20, 23, 18, 17, 18, 22, 24)

> x_2 = c(B)

> x_3 = c(C)

> x_4 = c(D)

> d = stack (list (b1 = x_1 , b2 = x_2 , b_3 = x_3, b4 = x_4)

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-values = 6.002849

∴ pvalue is less than 0.05 we reject the hypothesis

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

	DF	sum sq	mean sq	F value	p (>F)
ind	3	91.44	30.479	6.845	6.00235 **
residual	20	89.06	4.453		

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

> x = read.csv ("C: /user /admin /Desktop /marks csv")

> x

	Stats	Maths
1	40	60
2	45	68
3	42	47
4	15	20
5	37	25
6	36	27
7	49	57
8	59	58
9	20	25
10	27	27

22

> am = mean (x $stats)

> am

[1] 37

> am1 = mean (x $maths)

> am1

[1] 39.4

> m1 = median (x $maths)

> m1

[1] 38.5

> m2 = median (x $maths)

[1] 37

> n = length(x $stats)

> n

[1] 10