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 CS **Roll No.** 1753 **Year** 2019-20

This is to certify that the work entered in this journal is the work of Mst. / Ms. Mishra Anushka Abhay

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

Teacher In-Charge

Head of Department

Date : ____________

Examiner

★★ INDEX ★★

PRACTICAL NO. 1

Aim: Basics of R-software

1) R is a software for statiscal 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 following:

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) $5^3+7\times 5\times 8+46/5$
> 5^3 + 7*5*8 + 46/5
[1] 414.2

4) $\sqrt{4^2+5\times3+716}$
> sqrt(4^2+5*3+716)
[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^z + y^z
[1] 1300

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

04

		[,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,
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	-4	8

NO

> 2 * x + 3 * y

	[,1]	[,2]	[,3]
[1,]	38	-19	33
[2,]	50	-12	41
[3,]	63	-28	21

Q.6 Marks of statistics of CS Batch of

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

G

Practical No. 02.

Topic : Probability distribution.

1) Check whether 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 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 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 $p(x) \neq 1$

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

1) $P(x) \geq 0$ $\forall x$ 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
```


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

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

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

Q.3 Find

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

$F(x) = 0 \quad x < 1$

$= 0.15 \quad 1 \le x < 2$

$= 0.40 \quad 2 \le x < 3$

$= 0.50 \quad 3 \le x < 4$

$= 0.70 \quad 4 \le x < 5$

$= 0.9 \quad 5 \le x < 6$

$= 1.00 \quad x \ge 6$

```
> 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 = "Cumulative probability",
  main = "CDF graph", col = "brown")
```

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

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

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

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

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

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

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

$$= 3-1$$

$$= 2 \neq 1$$

$\therefore$ It is not a pdf

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

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

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

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

$$= 1$$

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

PRACTICAL No. 3

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 unknoown

$p_1 = P(X \le x) =$ 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) $p(x=5)$ iii) $p(x>7)$

ii) $p(x \le 5)$ 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 = 0.0000

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] 0.00278157

4)
5)
6)
7)

4) dbinom (6,12, 0.25)

[1] 0.04014945

5) > ~~bd~~dbinom (0,10, 0.15)
[1] 0.1968744
> 1 - pbinom (2,20, 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)

	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

5) The probability of a salesman matching a sale to customer 0.15. Find the probability of.

i) No sales out of 10 customer

ii) More than 3 sales

Practical No :: 4

Aim : Normal Distribution.

i) $P(z=x)$, domain (z, μ, σ)
ii) $P(x \le x)$, pnorm (x, μ, σ)
iii) $P(x > x) = 1 -$ Pnorm (x, μ, σ)
iv) To generate random number from a normal distribution (n random number) the R code is
Pnorm (n, μ, σ)

Q.1 A random variable x follows normal distribution with mean, $\mu - 12$ & $30 = \sigma - 3$, find i) $P(x \le 15)$
ii) $P(10 \le x \le 13)$ iii) $P(x > 14)$
iv) Generate 5 observation (random number)

code :

```
> p1 = Pnorm (15, 12, 13)
> p1
[1] 0.8413447
> cat ("P(x <=15) ." P1)
P (x<=15) =0.8413447
>P2 = Pnorm (13, 12, 73) - Pnorm (10, 10, 3)
>P2
[1] 0.3780611
> cat ("P(10<3xc = 13) = "P2)
P(10< = x <=13) = 0.378061
> P3= 1-Pnorm (14, 12, 3)
```

```
> P3
[1] 0.2524905
> cat ("P(x>14) = " P3)
  P(x>14) = 0.2524925
> P4 = rnorm (5, 12, 3)
> P4
[1] 15.254728    16.548505    11.280515    6.419849
    12.272460
```

2) x follows normal distribution with $\mu = 10$, $\sigma = 2$
find i) $P(X \leq 7)$ ii) $(5 < x < 12)$ iii) $P(X > 12)$
iv) generate 10 observation 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] a4 = rnorm (10, 10, 2)
> a4
```

Scanned with CamScanner

```
[1] 11.600931   9.120417   12.637741   8.073359
    8.72180     9.193724   9.366824    11.707104
    9.537584    10.715006
```



```
> as = qnorm (0.4,10,2)
> as
[1] 9.49330
```

Q.3 Generate 5 random no. from a normal distribution $\mu = 15$, $\sigma = 4$ find sample mean, median, S.D. & print it.

Code :

```
> rnorm (5,15,4)
[1] 10.7649   7.793249   9.953444   13.345304
    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 ("SD is = ", SD)
SD ~~SI~~ = 3.33163

Q.4 $x \sim n(30, 100)$, $\sigma = 10$

i) $P(x \leq 40)$

2) $P(x > 35)$

3) $P(25 < x < 35)$

4) 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] 0.3085375
>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 = "xvalues", ylab = "probability",
main ="standard normal graph")
```

Scanned with CamScanner

Practical: 5

Topic : Normal & t-test

1) $H_0 : \mu = 15$ $H_1 : \mu \neq 15$
↳ null ↳ alternative

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 then reject.

```
>mo=15
>mx=14
>sd=3
>n=400
>zcal=(mx-mo)/(sd/(sqrt(n)))
>zcal
[1] -6.6667
>cat("calculated value of z is =",zcal)
calculated value of z is = -6.6667
>pvalue=2*(1-pnorm(abs(zcal)))
>pvalue
[1] 2.616796e-11
```

∴ The value is less than 0.05 we will reject the value of $H_0 = \mu = 15$.

2) Test the hypothesis $H_0: \mu = 10$ against $H_1: \mu \neq 10$
A random sample size of 400 is drawn with sample mean = 10.2 & 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] 1.7778
>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 H_0: preparation of smokers is college is 0.2. A sample is collected & calculated the sample proportion 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 9 field crops are insect polluted. Test the hypothesis at 1% level of significance.

```
>p = 0.2
> p=9/60
> n= 60
> z cal
[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 $H_0: \mu = 12.5$ from the following sample at 5% level of significance

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

Scanned with CamScanner

```
>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]  8.894509
>pvalue = 2* (1-pnorm(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 & s.d. 30.
Test the hypothesis that the population mean is 250 or not, on 5% level of significance

→
```
> 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 is = 8.33333.
> pvalue = 2 * (1 - pnorm (abs (zcal)))
> pvalue
[1] 0
```

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

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

Sol^n. > 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] 7.72268 -05
∴ The value is less 0.01 we reject!

3) Two random sample of size 1000 & 2000 are drawn from two population with same s.d 2.5. The sample means are 67.5 & 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]2.417514e-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.

H.S.A	H.S B
87	34
61.2	59.4
7.9	7.5

```
>n1=87
>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]1.162828
>pvalue
[1]0.2450211
```

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

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 preparation of students using blue ink in two colleges are equal or not at 10% 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] 0.4333
> zcal = (p1 - p2) (sqrt (p * q * (1/n1 + 1/n2))
> zcal
[1] 6.381534
> p value = 2 * (1 - pnorm (ab (zcal)))
> pvalue
[1] 1.75322e-10
```

∴ value is less than 0.01 the value is rejected.

6) $H_0 : p_1 = p_2$ against $H_1 : p_1 \neq p_2$

```
> 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

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 = 68.319, 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 H0")}
else {cat ("reject H0")}

reject H0

Q.2 Two groups of students score the following marks. 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.

Sol^n: H_0: 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 & 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 H0")}
  else {cat ("reject H0")}
reject H0>
```

Scanned with CamScanner

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

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

Before : 53, 28, 31, 48, 50, 42
After : 57, 29, 30, 55, 56, 45

Test the hypothesis that the campaign is effective or not.

Sol^n: H_0 : There is no significant difference of sales before & 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 & 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.x
```

PRACTICAL No. 8

Large & small sample test

1) $H_0 : \mu = 55$

$H_1 : \mu \neq 55$

```
>n=100
>mx=52
>m0=55
>sd=7
>zcal=(mx-m0)/(sd/(sqrt(n)))
>zcal
[1] -4.285714
>cat("calculated z value is ="zcal)
calculated z value is = -4.285714
>pvalue=2*(1-pnorm(abs(zcal)))
>pvalue
[1] 1.82153e-05
```

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

2) H_0 : μ

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

Since pvalue is greater than 0.05 we ~~so~~ accept hypothesis at 1% level of signifiance

3) $H_0 : P_1 = P_2$ 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)
> 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 significance.

4) $H_0 : \mu = 100$

```
> mx = 99
> m0 = 100
> sd = 8
> n = 400
> zcal = (mx - m0) / (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% level of significance

5) $H_0: \mu = 66$

> x = c(63, 63, 68, 69, 71, 71, 72)

> t.test(x)

One sample t-test

data 8x

t = 47.94, df = 6, p-value = 5.522e-09

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

67.64479 71.62092

sample estimate:

mean of x

68.14286

Since pvalue is less than 0.05 we reject the hypothesis at 1% level of significance.

7) $H_0 : \mu_0 = 1200$

```
> m0 =1200
> mx =1150
> n =100
> sd =125
> zcal (mx - m0) / (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

8) $H_0 : u = P_1 \neq P_2$

```
> n1 = 200
> n2 = 300
> p1 = 44/200
> p2 = 56/300
> p = (n1 * p1 + n2 * p2) / (n1 * n2)
> p
```

Scanned with CamScanner

[1] 0.1
> q = 1-p
> zcal = (p1 - p2) / sqrt(p * q * (1/n1 + 1/n2))

> zcal
[1] 0.9128709

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

Since p value is greater than 0.05 we accept the hypothesis.

PRACTICAL No. 9

Topic : chi-square Test & ANOVA

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

	Clean	dirty
Clean	70	50
fairly clean	80	20
dirty	35	45

Sol^n: 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 25.644, df = 2, p-value = 2.698

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

Q.2 Test the hypothesis that that vaccination & disease are independent or not.

	Aff	Not Aff
~~Differ~~ Disease Aff	70	46
N.A.	35	37

Sol^n: H_0: Disease & vaccination are independent

> 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
Pearson's chi-squared test
with Yates continuity correction

data : y
x-squared = 2.0275 , df = 1 , pvalue
= 0.01545

∴ 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	Obs
A	50, 52
B	53, 55, 53
C	60, 58, 75, 57
D	52, 54, 54, 55

Scanned with CamScanner

Sol^n: 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 = 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, num df = 3, denom df = 5
pvalue = 0.00183

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

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

Q.7 The following data given the life of tyres of four brand

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: H_0: The average life of type is equal for (A, B, C, D)

> x_1 = c(20, 23, 18, 17, 18, 22, 24)
> x_2 = c(19, 15, 17, 20, 16, 17)
> x_3 = c(21, 19, 22, 17, 20)
> x_4 = c(15, 17, 16, 18, 14, 16)
> d = stack(list(b1 = x_1, b2 = x_2, b3 = 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 and ind

F = 6.8445 , num df = 3, denom df = 20,

p-value = 0.002349

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

PRACTICAL - 10.

26

TOPIC : Non Parametric Test

Q.1 Following are the amounts of sulphur oxid emitted by a industry in 20 days. Apply sign test, to test the hypothesis that the population median is 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.

H_0 : Population median is 21.5

Solⁿ: 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
```

Scanned with CamScanner

```
> 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

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

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

Sol. H_0: ~~Pop~~ Population median is 625

```
> x = c(612, 619, 631, 628, 643, 640, 655, 649,
  670, 663)
> 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] 0.0541875

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

Q.3 Following are value 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.

Sol:- H_0: population median = 60
H_1: 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)
```

Scanned with CamScanner

wilcox signed Rank Test
with continuity correction

data : x
V = 145 , p-value = 0.02298
alternative hypothesis : true location
is greater than 6 .

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, 45, 20, 21, 32, 28, 12, 25, 24, 26.

Sol^n : H_0 : population median is 12
H_o : 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 connection

data : x
v = 66 , p-value = 0.996
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.8 The weights of students before & after they stop smoking are given below. Using WSRT test that there is no significant change.

Before

61, 75, 75, 62,
72

After

71, 74, 72, 60,
73.

soP: H_0: Before & after there is no significant change

H_1: There is a change.

```
> x = c(65, 75, 75, 62, 72)
> y = c(71, 74, 72, 66, 73)
> d = x-y
> wilcox.test(x, alter = "two.sided", mu=0)
```

wilcox signed RANK Test with continuity connection

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 accepts the hypothesis at 5% level of significance