★★ INDEX ★★

No.	Title	Page No.	Date	Staff Member's Signature
1.	Basics of R-Software		25/11/19	
2.	Binomial Distribution.		2/12/19	
3.	PMF (Probability mass function)	65	9/12/19	AM
4.	Practice	67	23/12/19	
5.	Normal distribution	69	6/12/19	
6.	Z and t distribution summ	71	20/1/2020	
7.	large sample test	73	27/1/2020	AM.
8.	small sample test	75	3/2/2020	
9.	large & small sample test	79	10/2/20	
10.	ANOVA & Coll squared test	81	17/2/20	AM.
11.	Non parametric test	85	24/2/2020	AM

Practical No. 1

i) R is a software for data analysis and statistical computing

ii) It is a software by which effective data handling and outcome storage is possible

iii) It is capable of graphical display

iv) It is a free software

Q1 $2^2 + |-5| + 4 \times 5 + 6/5$

```
> 2^2 + abs (-5) + 4*5 + 6/5
[1] 30.2
```

Q2 $x = 20$
$y = 2x$
$z = x+y$
$\sqrt{z}$

```
> x = 20
> y = 2*x
> z = x+y
> z**0.5
[1] 7.745
```

Q3 $x = 10$
$y = 15$
$z = 5$

a) $x+y+z$
b) xyz
c) $\sqrt{xyz}$
d) round of $\sqrt{xyz}$

```
-> >x=10
>y=15
>z=5
>x+y+z
[1] 30
>x*y*z
[1] 750
>(x*y*z)^0.5
[1] 27.3861
>round((x*y*z)^0.5)
[1] 27
```

* A vector in R software is denoted by the syntax c

```
c (2,3,5,7)^2
>c (2,3,5,7)^2
[1] 4 9 25 49

c(2,3,5,7)^c(2,3)
>c (2,3,5,7)^c(2,3)
[1] 4 27 25 343

c(2,3,5,7,9,11)^ c (2,3)
>c(2,3,5,7,9,11)^ c(2,3)
[1] 4 27 25 343 81 1331

>x=c(1,2,3,4,5,6) ^ c (2,3,4)
[1] 1 8 81 16 125 1296

>c (2,4,6,8) * 3
[1] 6 12 18 24
```

Q9
→ c (2,4,6,8) * c (-2,-3,-5,-7)

[1] -4 -12 -30 -56

Q10
x = c (2,4,6,8)

x+10

[1] 12 14 16 18

Q11. c(2,4,6,8) + c(-2,-3,-1,0)

[1] 0 1 5 8

→ sum product

Q12 Find the sum, product, sqrt of the sum and product for the following values
= (4,9,2,5,7,8,3,6,15,12,10,9,8,13,14)

x = c (4,9 . . . .)

y = sum(x)

z = prod(y)

y^0.5

y

[1] 125

z

[1] 5

```
>y^0.5
[1] 11.18034
[1] 2^0.5
> 2925632
[1]
```

Find the sum, product, maximum, min, values of

```
c(2,8,9,11,10,7,6)^2

x=c(8 - - - - )^2
[1] 4  64  81  121  100  49  36
>sum(x)
[1]455
>Prod(x)
[1] 4 4 2597478400
> max(x)
[1] 121
> min(x)
[1] 4.
```

Matrix

```
x <- matrix (n rows=3, n col=4, c(2,6,7,8,9,4,5,0,2,
                                  1,4,5)
     [1.]  [2.]  [3.]  [4.]
[1.]  2     8     5     1
[2.]  6     9     0     4
[3.]  7     4     2     5
```

Q Addition of matrix

X = matrix (n row = 3, n col = 3, c(4,5,6,7,8,9,4,6,2))
Y = matrix (n row = 3, n col = 3, c(8,9,5,11,12,8,8,9,7))

X

	[,1]	[,2]	[,3]
[1,]	4	7	4
[2,]	5	8	6
[3,]	6	9	2

Y

	[1]	[2]	[3]
[1]	6	11	9
[2]	4	12	7
[3]	5	8	4

X * 2

	[,1]	[,2]	[,3]
[1,]	8	14	8
[2,]	10	16	0
[3,]	12	18	4

Y * 2

	[,1]	[,2]	[,3]
[1,]	12	22	18
[2,]	8	24	14
[3,]	10	10	8

x+y

	[1]	[2]	[3]
[1]	10	18	13
[2]	9	20	7
[3]	11	17	6

x*y

	[,1]	[,2]	[,3]
[1]	29	77	36
[2]	20	96	0
[3]	30	72	8

Ans
2.12.19

Practical No. 2

Binomial distribution

n = Total no. of trials
p = P(success)
q = P(failure)
x = No. of successes out of n

$P(x) = {}^nC_x \, p^x \, q^{n-x}$
$E(x) = np$
$V(x) = npq$

dbinom	$P(x)$
n*p	$E(x)$
n*p*q	$V(x)$
pbinom	$P(x \leq x)$

Exercise :-
Toss a coin 10 times P(Heads) = 0.5. Let x be the no. of heads
find the probability
i) seven heads
ii) four heads
iii) atleast six heads
iv) atmost four heads
v) No heads
vi) All heads
Also find expectation & variance

Output:-

```
>n=10
>p=0.6
>q=0.4
>a = dbinom (7, n, p)
>a
[1] 0.2149908
>b = dbinom (4, n, p)
>b
[1] 0.1114767
>c = 1 - pbinom (6, n, p)
>c
[1] 0.3822806
>d = pbinom (4, n, p)
>d
[1] 0.1662386
>e = dbinom (0, n, p)
>e
[1] 0.0001048576
>f = dbinom (10, n, p)
>f
[1] 0.006046618
```

Q8. for $n = 10$, $p = 0.6$. find the binomial probability that ~~x~~ plot the graphs of pmf & cdf

* formulas
  * probability of atmost x values:
    $P(X \le x) = \text{pbinom}(x, n, p)$
  * Probability of atleast x values:
    $P(X \ge x) = 1 - \text{pbinom}(x, n, p)$
  * if 'x' is unknown & the probability is given as P,, To find x c q binom (P, n, p)

Q2
```
>n=12
>p=1/5
>p
[1] 0.2
>n
[1] 12
i) >dbinom(4,n,p)
[1] 0.1328756
ii) >pbinom(4,n,p)
[1] 0.927445
iii) >1-pbinom(2,n,p)
[1] 0.4416543
```

Q3 ->
```
>n=5
>p=0.1
>dbinom(0,n,p)
[1] 0.59049
>dbinom(1,n,p)
[1] 0.32805
>dbinom(2,n,p)
[1] 0.0729
>dbinom(3,n,p)
[1] 0.0081
>dbinom(4,n,p)
[1] 0.00045
>dbinom(5,n,p)
[1] 1e-05
```

Q4
```
>x=10
>n=100
>p=0.1
>dbinom(x,n,p)
[1] 0.1318653
```

Q5
```
>n=12
>p=0.25
>pbinom(5,n,p)
[1] 0.9455478
>1-pbinom(7,n,p)
[1] 0.0027815
>dbinom(6,n,p)
[1] 0.00401445
```

Pratical No:- 2.2

Binomial distribution :-

Suppose there are 12 MCQ's in an ENGLISH question paper. Each question has 5 answer out of which only one is correct. find the probability of having i) 4 correct answers ii) atleast 4 correct answers iii) atleast 3 correct answers

find the complete binomial distribution when $n=5$ & $p=0.1$

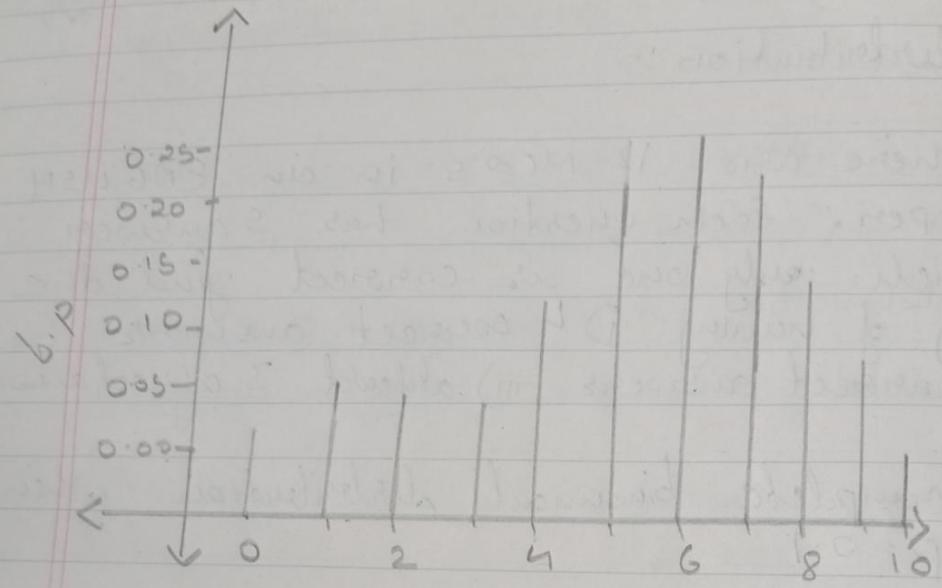
find the probability of exactly 10 successes out of 100 trials with $p=0.1$

'X' follow binomial distribution with $n=12$ & $p=0.25$
find i] $P(x \leq 5)$
ii] $P(x > 7)$
iii] $P(5 < x < 7)$

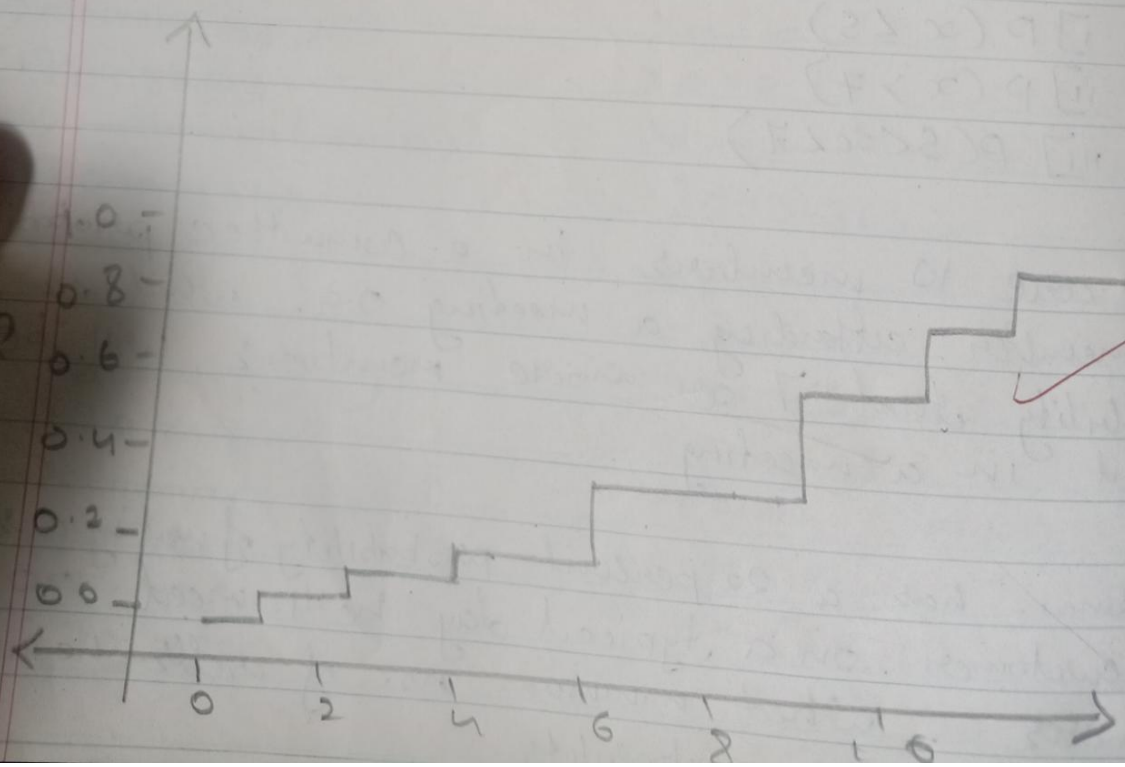
There are 10 members in a committee probability of any member attending a meeting 0.9. what is the probability that 7 or more members will be present in a meeting

A salesman has a 20 percent probability of making a sell to a customer. on a typical day he'll meet 30 customers. what minimum no. of sells he will make with 88% probability

```
> plot (x, bp, "h")
```




```
> cp = pbinom (x, n, p)
> plot (x, cp, "s")
```



```
>p=0.9
>1-pbinom(6,n,p)
[1]0.9872048

>n=30
>p=0.2
>p1=0.88
>q binom(p1,n,p)
[1]9

>n=10
>p=0.6
>x=0=n
>bp=dbinom(x,n,p)
>bp
]0.0001048576  0.0015728640  0.1061683200  0.4246732800
0.1114767360

]0.2006581248  0.2508226560  0.2149908480  0.1209323520
0.0431070400  [11] 0.0060466176
d=data frame("x.values" = x, "Probability" = bp)
d

x.values  probability
0         0.0001048576
2         0.0015728640
3         0.0106168320
4         0.0424673280
5         0.1114767360
6         0.2006581248
7         0.2508226560
8         0.2149908480
9         0.1209323520
10        0.0060466176
```

Practical No:- 3

1. check the following are pmf or not?

i]

x	1	2	3	4	5
$P(x)$	0.2	0.5	0.5	0.4	0.4

ii]

x	10	20	30	40	50
$P(x)$	0.3	0.2	0.3	0.1	0.1

iii]

x	0	1	2	3	4
$P(x)$	0.4	0.2	0.3	0.2	0.1

Solutions

i] → It is not a p.m.f since it does not satisfy the first condition

ii] → Since all values of $P(x)$ are more than zero and less than one. 1st condn is satified. Also,

$$\Sigma P(x) = P(10) + P(20) + P(30) + P(40) + P(50)$$
$$= 0.3 + 0.2 + 0.3, 0.1, 0.1$$
$$= 1$$

∴ 2nd condition is also satified

→ Hence, it is a p.m.f

```
>prob=c(0.3, 0.2, 0.3, 0.1,0.1)
>prob
[1] 0.3   0.2   0.3   0.1   0.1
>sum(prob)
[1] 1
```

iii) -> Since all the values of $P(x)$ are more than zero & less than one then first condition is satisfied.

Also,

$$\Sigma P(x) = P(0) + P(1) + P(2) + P(3) + P(4)$$
$$= 0.4 + 0.2 + 0.3 + 0.2 + 0.1$$
$$= 1.2$$

Here, 2nd condition is not satisfied since the value is greater than 1. Hence, it is not a p.m.f

Q2 following is a p.m.f of x

x	1	2	3	4	5
$P(x)$	0.1	0.15	0.2	0.3	0.25

find mean & variance of 'x'

x	$P(x)$	$xP(x)$	$x^2P(x)$
1	0.1	0.1	0.1
2	0.15	0.30	0.60
3	0.2	0.6	1.8
4	0.3	1.2	4.8
5	0.25	1.25	6.25
		$\Sigma xP(x) = 3.45$	$\Sigma x^2P(x) = 13.55$

$$\text{Mean} = E(x) = \Sigma xp(x) = 3.45$$
$$\text{Var} = V(x) = \Sigma x^2p(x) - [E(x)]^2$$
$$= 13.55 - (3.45)^2$$
$$= 1.6475$$

```
>x=c (1,2,3,4,5)
>prob=c (0.1,0.15,0.2,0.3,0.25)
>a= x*prob
>sum (a)
[1] 3.45
>b= x*a
>sum (b)
[1] 13.55
>var = 13.55 - (3.45^2)
[1] 1.6475
```

Q3 find mean & variance of x

x	P(x)
5	0.1
10	0.3
15	0.2
20	0.25
25	0.15

```
>x= c (5,10, 15,20,25)
>prob=c ( 0.1,0.3,0.2 ,0.25,0.15)
> a = x*prob
> sum(a)
[1] 15.25
> b=x * a
> sum (b)
[1] 241.25
>Var = sum(b)-(sum (a) ^ 2)
>Var
> 38.6875
```

Q4. i) find c.d.f of the following p.m.f & draw the graph of c.d.f

x	1	2	3	4
$P(x)$	0.4	0.3	0.2	0.1

```
> x = c(1,2,3,4)
> prob x = c(0.4, 0.3, 0.2, 0.1)
> a = cumsum(prob x)
> a
[1] 0.4 0.7 0.9 1.0
```

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

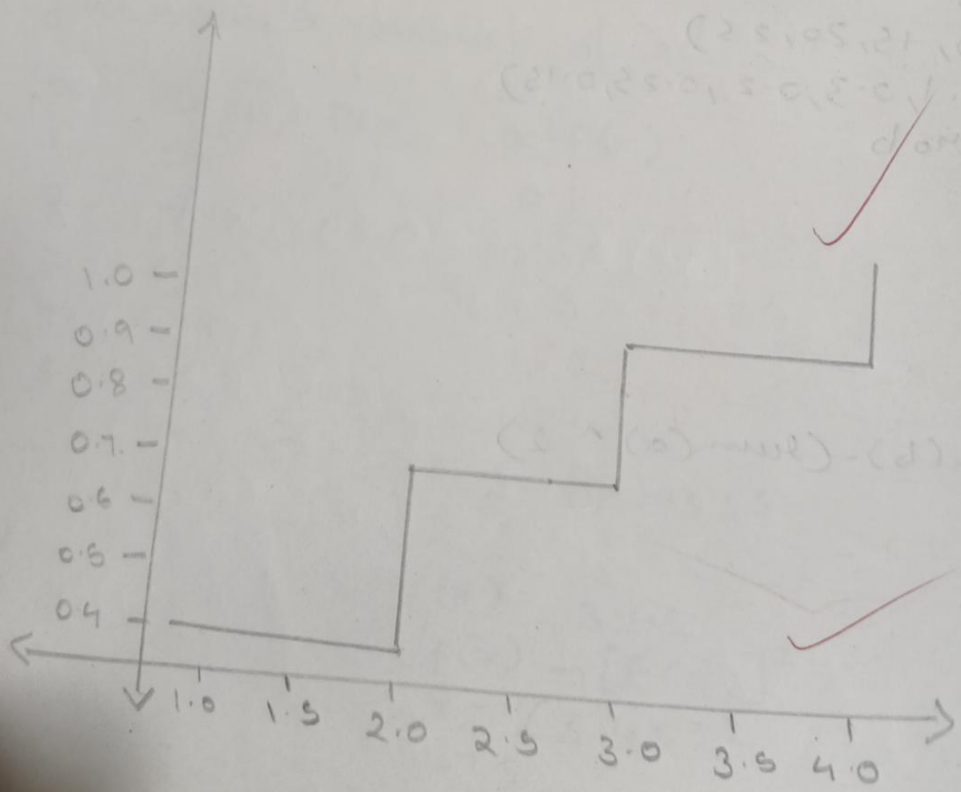
$= 0.4 \quad 1 \le x < 2$

$= 0.7 \quad 2 \le x < 3$

$= 0.9 \quad 3 \le x < 4$

$= 1.0 \quad x \geq 4$

```
plot(x, a, "s")
```



Q.4 ii]

x	0	2	4	6	8
$P(x)$	0.2	0.3	0.2	0.2	0.1

```
>x = c(0,2,4,6,8)
>probx = c(0.2, 0.3, 0.2, 0.2, 0.1)
>a = cumsum(probx)
>a
[1] 0.2  0.5  0.7  0.9  1.0
```

$P(x) =$

$= 0.2 \quad 0 \le x < 2$

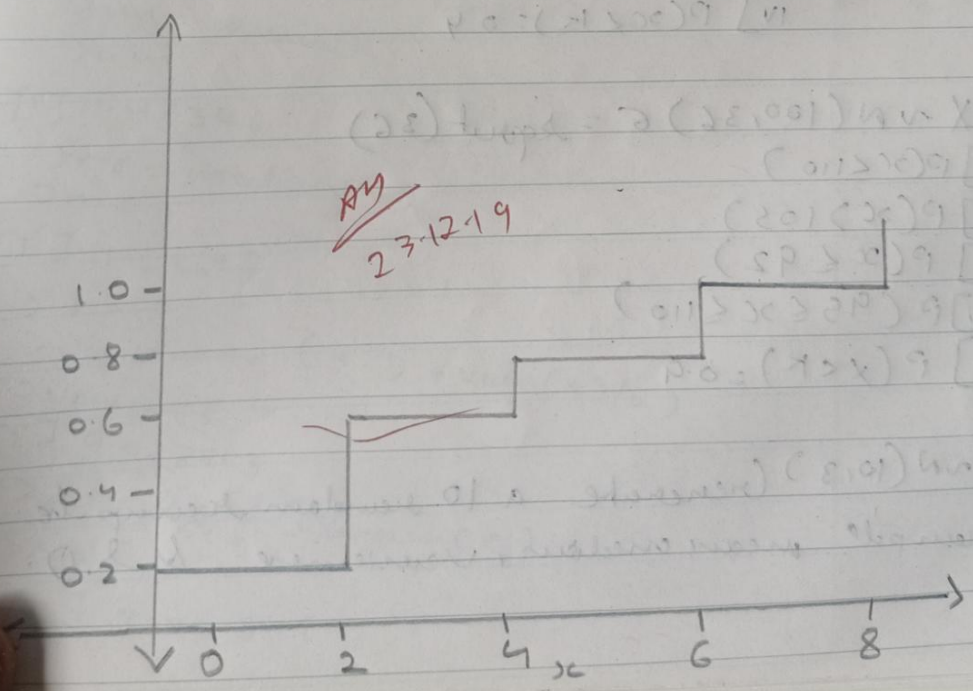
$0.5 \quad 2 \le x < 4$

$0.7 \quad 4 \le x < 6$

$0.9 \quad 6 \le x < 8$

$1.0 \quad x \ge 8$

plot (x, a, "s")



AM
23.12.19

Practical-4

* Normal distribution :-

Normal distribution is an an example of continous probability distribution

i] $P(X=x)$ = dnorm (x, μ, σ)
ii] $P(X \leq x)$ = pnorm (x, μ, σ)
iii] $P(x > x)$ = 1 - p (norm) (x, μ, σ)
iv] $1 - P(X \leq k)$ = p qnorm (P, μ, σ)
v] To generate a random sample of size n
rnorm (n, μ, σ)

i) x follow normal distribution $\mu = 10, \sigma = 2$.
To find:- i] $P(x \leq 7)$
ii] $P(x > 12)$
iii] $P(5 \leq x \leq 12)$
iv] $P(x < k) = 0.4$

ii) $X \sim N(100, 36)$ σ = sqrt(36)
i] $P(x \leq 110)$
ii] $P(x > 105)$
iii] $P(x \leq 92)$
iv] $P(95 \leq x \leq 110)$
v] $P(X < K) = 0.9$

iii) $X \sim N(10, 3)$ Generate a 10 random sample. find sample mean, median, variance & SD.

```
pnorm(4,10,2)
[1] 0.0668072
>cat("P(x<=7) is= ",p1)
P(x<=7) is= 0.0668072
>p2 = 1- pnorm(12,10,2)
>cat("P(x>12) is= ",p2)
P(x>12) is = 0.158655 3
>p3= p2- pnorm(5,10,2)
>cat("P(5<=x<=12) is = ",p3)
P(5<=x<=12) is= 0.1524456
>k= qnorm(0.4,10,2)
>cat("P(x<k)=0.4, k is= ",k)
P(x<k)=0.4, k is= 9.493306
```

```
11] x=rnorm(10,10,3)
>x
[1] 9.193412 11.340685 5.331607 17.160876
11.912250 9.372691 12.436043 13.810725
10.908016 13.367803
>am = mean(x)
>am
[1] 11.48346
>me= median(x)
>me
[1] 11.6267
> n=10
>variance= (n-1)*var(x)/n)
>variance
[1] 9.018713
>sd = sqrt(variance)
>sd
[1] 3.003117
```

ii] >p1-pnorm (110,100,6)
>p1

[1] 0.9522096

>cat("P(x<=110) is= ",p1)

P(x<=110) is= 0.9522096

>p2=1-pnorm(105,100,6)

>cat("P(x>105) is= ",p2)

P(x>105) is= 0.2023284

>p3=pnorm(92,100,6)

>cat("P(x<=92) is= ",p3)

P(x<=92) is= 0.9121122

>p4=pnorm(110,100,6)-pnorm(95,100,6)

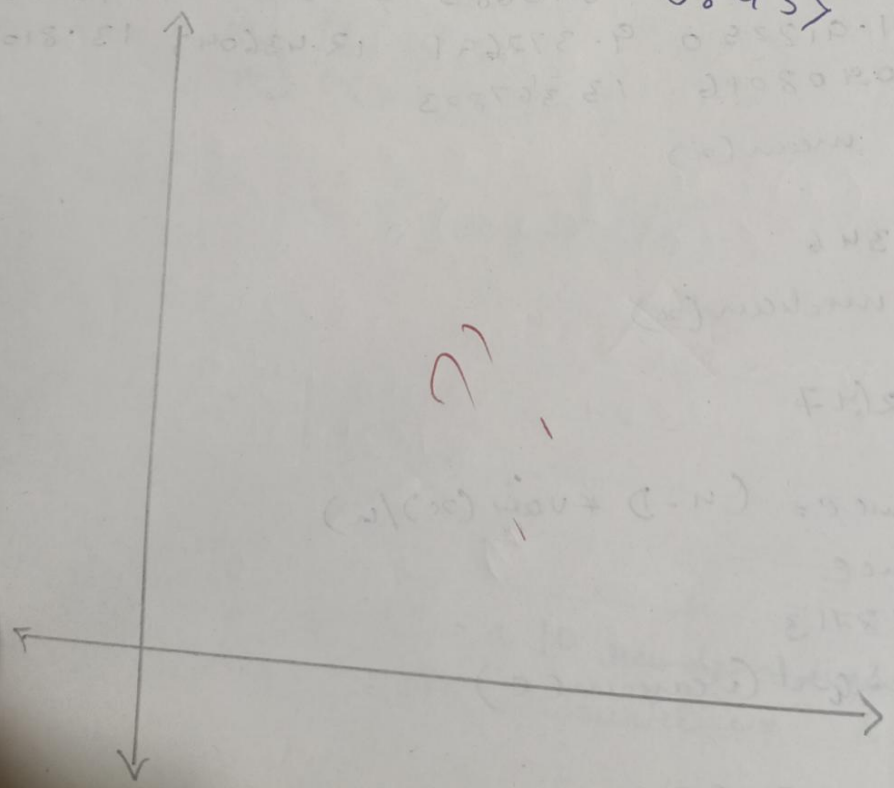
>cat("P(95<=x<=110) is= ",p4)

P(95<=x<=110) is= 0.7498813

>k=qnorm(0.9,100,6)

>cat("P(x<k)=0.9, k is= ",k)

P(x<k)=0.9, k is= 107.6893



iv) Plot the standard normal curve

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

v) $x \sim N(50, 100)$

Find i) $P(x \leq 60)$

ii) $P(x > 60)$

iii) $P(45 \leq x \leq 60)$

```
> p1 = pnorm(60, 50, 10)
> cat("P(x<=60) is ", p1
P(x<=60) is 0.8413447
> p2 = 1-pnorm(65, 50, 10)
> cat("P(x>65) is ", p2)
P(x>65) is 0.6680072
> p3 = pnorm(60, 50, 10) - pnorm(45, 50,
> cat("P(45<=x<=60) is ", p3)
P(45<=x<=60) is 0.5328072
```

Practice practical

1] $P(x=x) = \binom{n}{x} p^{x} q^{n-x}$

$P(x=7) = \binom{8}{7} (0.6) \quad (0.4)^{8-7} \quad (n=8$

$= {}^{8}C_{7} \times 0.2799 \times 0.4$

$= 8 \times 0.2799 \times 0.4$

$= 0.08954$

2] $P(x \leq 3) = P(0) + P(1) + P(2) + P(3)$

$= {}^{8}C_{0} (0.6)(0.4)^{8} + {}^{8}C_{1} (0.6)^{1}(0.4$

(0.6)

$+ {}^{8}C_{3} (0.6$

$= 1 \times 1 \times 0.00065536 + 8 \times 0.6 \times 0.0016$

$0.36 \times 0.004096 + 56 \times 0.216 \times 0.0$

$= 0.1736704$

3] $P(x=2 \text{ or } x=3) = P(2) + P(3)$

$= {}^{8}C_{2} (0.6)^{2} (0.4)^{0} + {}^{8}C_{3} (0.6$

$= 28 \times 0.36 \times 0.004096 + 56 \times 0.216 \times 0.01$

$= 0.04128768 + 0.123363 = 4$

$= 0.16515012$

z & t distribution

Test the hypothesis $\mu_0 = \mu = 20$ against $\mu_1 : \mu \neq 20$.

A sample of size 400 is selected and the sample mean is 20.2 & the standard deviation is 2.25. Test at 5% level of significance

```
>n=400
>mo=20
>mx=20.2
>sd=2.25
>zcal=(mx-mo)/(sd/sqrt(n))
>cat("z calculated is", zcal)
```

z calculated is 1.777778>

```
>pvalue = 2*(1-pnorm(abs(zcal)))
>cat("Pvalue is", pvalue)
```

P value is 0.07544036>

$\therefore$ 0.0754 > 0.05

$\therefore$ $\mu_0 : \mu = 20$ is acceptable.

Test the hypothesis $\mu_0 : \mu = 250$ against $\mu_1 : \mu$
A sample of size 100 has a mean of 2
check the hypothesis a 5 %

```
>n = 100
>mo = 250
>mbc = 275
>sd = 30
>zcal = (mbc-mo) /(sd/sqrt(n))
>cat ("z calculated is", zcal)
z calculated is 8.333333>
>pvalue = 2*(1-pnorm(abs(zcal)))
>cat ("p value is", pvalue)
p value is 0>
```

$\therefore 0 < 0.05 \therefore \mu_0 = \mu = 250$ is unacceptable

Q3 We meant to test the hypothesis μ_0, $P = 0.2$ against $\mu_1: P \neq 0.2$ (P = Population Proportion). A sample of 400 is selected and the sample proportion is calculated 0.125. Test the hypothesis at 1% of significance.

→

```
> p = 0.2
> q = 1-p
> p = 0.125
> n = 400
> zcal = (p-p)/sqrt(p*q/n))
> zcal
[1] -3.75
> Pval = 2*(1-pnorm(abs(zcal)))
> cat ("p value is", pvalue)
p value is 0.00017683346>
```

$\therefore$ Pvalue < 0.01

$\therefore \mu_0 = P = 0.2$ is

In the city 325 men out of 600 men are found to be self-employed. Does this 272 information support the conclusion that exactly half of the men in the city are self employed.

p=0.5

p=325/600

n = 600

>p0=0.5

q = 1-p0

>n=600

>p=325/600

>zcal= (p-p0) /sqrt (p* q/n)

>zcal

[1] 2.041241

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

>cat ("Pvalue is ", pvalue)

Pvalue is 0.04122683

>∴ 0.0412 < 0.05

∴ It is rejected

. Test the hypothesis H0: μ = 50 against H1: μ ≠ 50 A sample of 30 is collected

[50, 49, 52, 44, 45, 48, 46, 45, 49, 45, 40, 47, 55, 54, 56, 58, 47, 44, 59, 60, 61, 41, 52, 44, 55, 56, 46, 45, 49]

>m0=50

> x = c(50, 49, 52, 44, 45, 48, 46, 45, 49, 45, 40, 47, 55, 54, 49, 48, 47, 44, 59, 60, 61, 41, 52, 44, 55, 56, 46, 45, 48, 49)

>n=length(x)

```
>mx = mean(x)
>mx
[1] 49.3333
>variance = (n-1)*var(x)/n
>variance
[1] 30.455
>sd=sqrt(variance)
>sd
[1] 5.56
>zcal= (mx-mo) / (sd/sqrt(n))
>zcal
[1] -0.6562
>pvalue= 2*(1-pnorm(abs(zcal)))
>cat("pvalue is", pvalue)
Pvalue is 0.517633
```

$\therefore$ 0.511 > 0.05

$\therefore$ $H_0 : \mu = 50$ is accepted.

Aim:- large sample test

Two random sample of size 1000 & 2000 are drawn from two population with the standard deviation 2 & 3 respectively. Test the hypothesis that two population mean are equal or not at 5% level of significance: sample means are 67 & 68 respectively

→ $\mu_0 = \mu_1 = \mu_2$ against $\mu_1 : \mu_1 \neq \mu_2$

$n_1 = 1000$
$n_2 = 2000$
$mxc_1 = 67$
$mxc_2 = 68$
$sd1 = 2$
$sd2 = 3$

```
>n1 = 1000
>n2 = 2000
>mxc1 = 67
>mxc2 = 68
> sd1 = 2
> sd2 = 3
> zcal = (mxc1 - mxc2) / sqrt((sd1^2/n1) + (sd
> cat (" zcalculated is ", zcal)
zcalculated is  -10.8465
> pvalue = 2 * (1 - pnorm(abs(zcal)))
```

>cat ("P value is ", pvalue)
Pvalue is 0>

Q2 A study of noise level in two hospitals is done following data is calculated. first sample size 34 first sample mean 61.2 first std. deviation 7.9 second sample size 34, second sample mean 59.4 second sd = 7.8. Test H_0, $\mu_1 = \mu_2$ at 1% level of significance

H_0 ?

->n1 = 34
>n2 = 34
>m1 = 61.2
>m2 = 59.4
>sd1 = 7.9
>sd2 = 7.8
>zcal=(m1 - m2) sqrt ((sd1^2/n1) + (sd^2/n2))
>cat ("z calculated is ", zcal)
z calculated is 1.13117>
>pvalue = 2* (1-pnorm (abs (zcal)))
>cat ("P value is ", pvalue)
>Pvalue is 0.258006>

3 From each of the population of orange trees following data is collected check whether the proportion of that orange are equal or not n1 = 250, bad orange = 44, n2 = 200 bad orange = 30. Test - p1 = p2 against p1 ≠ p2

```
> n1=250
  n2=200
  p1=44/250
  p2=30/200
  p=(n1*p1)+(n2*p2)/(n1+n2)
> p
[1] 0.164444
>q=1-p
>q
[1] 0.835556
>zcal=(p1-p2)/sqrt(p*q*(1/n1+1/n2))
> cat("z calculated is ",zcal)
z calculated is 0.7393581
>pvalue=2*(1-pnorm(abs(zcal)))
>cat("Pvalue is ",pvalue)
Pvalue is 0.4596896
```

4) Random sample of 400 men & 600 females were asked whether they went on ATM. 200 male & 390 female were in favour of the proposal. Test the hypothesis that the population of male and female favour in the proposal are equal or not at 5% level

```
>n1=400
>n2=600
>p1=200/400
>p2=390/600
>p=(n1*p1+n2*p2)/(n1+n2)
>p
[1] 0.59
>q=1-p
>q
```

```
>zcal = (p1-p2)/sqrt(p*q*(1/n1+1/n2))
>cat ("z calculated is", zcal)
zcalculated is -4.72445
>pvalue = 2* (1-pnorm (abs(zcal)))
>cat ("Pvalue is, ", pvalue)
Pvalue is 2.803 x e^-6
```

$\therefore$ 0.02 < 0.05

$\therefore$ H_0 is rejected

Q5 following are the two independent samples from the two population. Test equality of the two population means at 5% level of significance $H_0 = \mu_1 : \mu_2$ at 5%.

samples $S_1 = [74, 77, 74, 73, 79, 76, 82, 72, 75, 78, 77, 78, 76, 76]$

$S_2 = [72, 76, 74, 70, 70, 78, 70, 72, 75, 79, 79, 74, 75, 78, 72]$

```
>x1 = c [74, 77, 74, 73, 79, 76, 82, 72, 75, 78, 77, 78, 76, 76, 80]
>l1=length(x1)
>l1
[1] 134
>x2 = c (72, 76, 74, 70, 70, 78, 72, 75, 79, 79, 74, 75, 78, 72, 74, 80]
>l2=length(x2)
>l2
[1] 17
>mx1 = mean(x1)
>mx2=mean(x2)
>mx1
[1] 76.21429
>mx2
[1] 74.58824
> t.test (x1,x2)
```

Welch two sample t-test

data: x1 & x2

t = 1.5227, df = 28.941, p value = 0.1387

>p value is 0.1387 | 0.1387 > 0.05

Hence it is accepted

Practical- 8

Topic:-

a) The random sample of is observation is given by

80, 100, 110, 105, 122, 70, 120, 110, 101, 88, 83, 95, 89, 107, 125.

Do this data support the assumption that the population mean is 100.

```
>x=c(80,100, 110, 105,122,70, 120, 110, 101, 88, 83, 95, 89, 107,
     125)
>t.test (x)
```

One sample t-test

data: x

t = 24.022, df = 14, p-value = 8.819e-13

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

91.37775 109.28892

sample estimates:

mean of x

100.3333

$\therefore 8.819 \times 10^{-13} < 0.05$

$\therefore H_0$ at 5% level of significance

Q2 Two groups of 10 students score the following marks:-

group1 — 18, 22, 21, 14, 20, 17, 23, 20, 22, 21
group2 — 16, 20, 14, 21, 20, 18, 13, 15, 17, 21

Test the ~~by~~ hypothesis, there is no significant between score and 1% level of significance

```
>g1=c (18,22,21,17,20,17,23,20,22,21)
>g2=c (16,20,14,21,20,18,13,15,17,21)
>t.test (g1,g2)
        welch two sample t-test
```

data: g1 & g2
t = 2.2573, df=16.376, p-value = 0.3798
alternative hypothesis: true difference in means 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

3 Two types of medicines are used for five patients for reducing their weights. The decrease in the weight after medicine is given below

medicine A = (10, 12, 13, 11, 14)
Medicine B = (8, 9, 12, 14, 15, 10, 9)

```
sa = c (66, 67, 75, 76, 82, 84, 89, 90, 92)
sb = c (64, 66, 74, 78, 82, 85, 87, 92, 93, 95, 97)
t.test (sa, sb)
```

Welch two sample t-test

data: sa & sb

t = -0.61273, df = 17.956, p-value = 0.5477

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-12.796000 7.018223

sample estimates:

mean of x & y

80.11111 83.00000

0.5477 > 0.05

∴ Ho at 5% level of significance.

Q6

```
>b=c(71,72,74,60,70,74,76,70,73,75)
>a=c(74,77,74,73,79,76,82,72,75,78)
>t.test(b,a, paired=T, alternative="less")
```

Paired t-test

data: b & a

t = -4.4641, df = 9, p-value = 0.0007784

alternative hypothesis: true difference in means less than 0 ~~95 percent~~

95 percent confidence interval:

-Inf -2.123361

sample estimates:

mean of the differences

-3.6

0.0007784 < 0.05

∴ H_0 is rejected at 5% level of significance

is there a significant difference in the efficiency of medicine

$H_0 = \mu_1 = \mu_2$

```
>med1 = c (10,12,13, 11,14)
>med2 = c(8,9,12,14,15,10,9)
>t.test (med1, med2)
```

Welch two sample t-test

data: med1 & med2

t = 0.80384, df = 9.7594, p-value = 0.4406

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-1.781171 3.781171

sample estimates:

mean of x mean of y

12 11

0.4406 > 0.05

∴ H_0 : $\mu_1 = \mu_2$ is accepted at 5% level of significance

4 The weight reducing diet prog. is conducted & the observation are noted for 10 observation. Test whether the program is effective or not

Before- 120, 125, 115, 130, 123, 119, 122, 127, 128, 118

After- 111, 114, 107, 120, 115, 112, 112, 120, 119, 112

H_0 : there is no significant diff. in weight

H_1 : The diet prog. reduced weight

```
>b = c(120, 125, 115, 130, 123, 119, 122, 127, 128, 118)
>a = c(111, 114, 107, 120, 115, 112, 112, 120, 119, 112)
>t.test(b, a, paired=T, alternative = "less")
```

Paired t-test

data: b & a
t = 17, df = 9, p-value = 1
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
-Inf 9.416555 6
sample estimates:
mean of the differences
8.5

= 1 > 0.5
∴ H_0 is accepted at 5% level of significance

5 sample A: 66, 67, 75, 76, 82, 84, 89, 90, 92
sample B: 64, 66, 72, 78, 82, 85, 87, 92, 96, 95, 9

Test the population mean is equal or not.

The following marks are before & after training program. Test whether the prog. is effective or not.

Before: 71, 72, 74, 69, 70, 74, 76, 70, 73, 75
After: 74, 77, 74, 73, 79, 76, 82, 72, 75, 78

Practical – 9

large and small sample test

Q1 The arithmatic mean of a sample of 100 items from a large population is 52. if standard diviation is 7, test the hypothesis that the population mean is 55 against the alternative more than 55% at 5% LOS

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

Q3 Thousand articles from a factory are found to have 2% defectives 1500 articles from a factory B are found to have 1% defective test at 5% LOS that the two factories are similar or not

Q4 A sample of size 400 was drawn and sample mean is 99 at 5% LOS that the sample comes from a population of mean 100 & variance 64

5 The following items are selected and heights are found to be (in cms) 63, 63, 68, 69, 71, 71, 72. Test the hypothesis that mean hight is 66 is o.k or not at 1% level of significance

Q6 Two random samples were drawn from 2 normal population & these values are A - 66, 67, 75, 76, 82, 84, 88, 90, 42
B - 64, 66, 74, 78, 82, 85, 87, 42, 43, 45, 44,

Test whether the populations have same variance at 5% los

1.

```
> n=100
> mx=52
> mo = 55
> sd = 7
> zcal = (mx-mo) / (sd/sqrt(n))
> cat ("zcalculated is", zcal)
-4.28714
> pvalue = 2 * (1-pnorm(abs(zcal)))
> cat ("Pvalue is ", pvalue)
```

Pvalue is 1.82153×10^{-5}

$\therefore 1.82153 \times 10^{-5} < 0.05$

$H_0 = \mu = 55$ is rejected

2.

```
> p = 0.5
> Q = 1-p
> P=350/700
> n = 700
> zcal = (p-P)/sqrt(p*Q/n)
> zcat ("zcalculated is", zcal)
```

z calculated is 0

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

```
>cat ("Pvalue is",pvalue)
Pvalue is 1
```

$1 > 0.01$

$\therefore$ It is accepted

```
>n1 = 1000
>n2 = 1500
>p1 = 20/1000
>p2 = 15/1500
>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))
>cat ("Zcalculated is", zcal )
Zcalculated is 2.084842
>pvalue = 2*(1-pnorm (abs (zcal)))
>cat ("PVALUE is", pvalue)
PVALUE is 0.03708364
```

$\therefore 0.037 < 0.05$

$\therefore$ $H_0: \mu$ is rejected

4.

```
>n=100
>mxc=dlap
>m0 =100
>var =64
>sd =sqrt(var)
>zcal =(mxc-m0)/sq/sqrt(n))
>cat ("Zcalculated is -2.5>
>pval = 2* (1-pnorm(abs(zcal)))
>cat ("Pvalue is", Pval)
Pvalue is 0.01241435>
```

$\therefore$ 0.1241 < 0.05

$\therefore$ It is rejected at 5% L.O.S

Ho : μ is rejected

5.

```
>x=c(63,63,68,69,71,71,72)
>t.test
        One sample test
data: x
t= 47.94 ,df=6, pvalue = 5.522e-09
alternative hypothesis : true mean is not equal to zero
95 percent confidence interval:
 64.6679        71.62092
sample estimate:
mean of x :
68.14286
```

$\therefore$ $5.522 \times 10^{-9} < 0.01$

$\therefore$ Ho : μ is rejected

H_0 : population is same
H_1 : Variance in population

```
>x = c(66, 67, 75, 76, 82, 84, 88, 90, 92)
>y = c(64, 66, 74, 78, 82, 85, 87, 92, 93, 95, 97)
> f = var.test(x,y)
> f
```

F = 0.7068, num df = 8, denom df = 10,
P.value = 0.635
95% confidence interval
0.1833 3.0360
ratio of variance
0.7068

∴ Pvalue is greater than 0.05
we accept that the population has same variance

A1

Practical No.:-10

Aim:- ANOVA & CHI square test

Q1 Use the following data to test the cleanliness of home & child is independent or not.

Cleanliness of home

	Clean	Dirty
Clean	70	50
Fairly clean	80	20
Dirty	35	45

Sol:- H_0 = Cleanliness of child & cleanliness of home

```
>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

χ-squared = 25.646, df = 2, p-value = 2.698e-06

$\therefore$ H_0 = 0.0000026 < 0.05 $\therefore$ It is rejected

82

22 Use the following data to find vaccination and a particular discease is indepent or not

Discease

	Affected	Not Affected
Given	20	30
Not given	25	35

Soln:- H_0: Discease & vaccination are independent

```
>x=c(20,25,30,35)
>m=2
>n=2
>y=matrix(x,nrow=m,ncol=n)
>y
     [,1] [,2]
[1,]   20   30
[2,]   25   35
>pv=chisq.test(y)
>pv
```

Pearson's Chi-squared test with Yates continuity correction

data = y

χ-squared = 0, df = 1, p-value = 1

H_0 = 1 > 0.05

$\therefore$ It is accepted

Q3 Perform ANOVA of the following data

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

Ho: The mean of the varieties are equal

```
>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 & ind
F = 11.735, num df = 3, denom df = 9, p-value = 0.00183
>anova = aov(values ~ ind, data = d)
>anova
Call:
    aov(formula = values ~ ind, data = d)
```

The following data gives the life of tires of 4 brands

Type	observation
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

Ho: Test the hypothesis that the average life is same or not.

Soln:-

Ho: The average life of tires is same

```
>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)
>data.frame(list(b1 = x1, b2 = x2, b3 = x3, b4 = x4))
>remove(d)
[1] "values" "ind"
>oneway.test(values~ind, data = d, var.equal=
        one-way analysis of means
data: values & ind
F = 6.8445, num df = 3, denom df = 20, p-value =
0.002349

> anova1 = aov(values~ind, data = d
> anova
```

Call:

aov(formula = values ~ ind, data = d)

Terms:

	ind	Residuals
Sum of squares	91.4381	89.0614
Deg of freedom	3	20

Residual effect may be unbalanced

Q5 1000 students of a college are graded according to their IQ and their economic condition of their home, check that is there any association b/w IQ & the economic condn of their home

		High	Low
Economic Condn	High	460	140
	Mod	330	200
	Low	240	160

Ho :- Association b/w IQ & economic condn

```
>x = c(460, 330, 240, 140, 200, 160)
>m = 3
>n = 2
>y = matrix(x, nrow = nrow = n)
>y
```

	[,1]	[,2]
[1,]	460	140
[2,]	330	200
[3,]	240	160

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

Pearson's Chi-squared test

data = y

X-squared = 39.726, df = 2, p-value = 2.3642-09

∴ Ho = 0.00000000236 < 0.05

∴ It is rejected

Ay 17.2.20

Topic:- Non-parametric test

Q1 Following are the amounts of sulphur oxide emitted by industries in 20 days. Apply sign test to test the hypothesis that the population median 21.5

17, 15, 20, 29, 19, 18, 22, 25, 27, 9, 24, 20, 17, 6, 24, 14, 15, 23, 24, 26

Ho: pop median is 21.5

```
> x = c (17,15,20 . . . . . 24, 26)
> length (x)
[1] 20
> me = 21.5
> sp = length (x[x > me])
> sn = length (x[x < me])
> n = sp + sn
> n
> pv = pbinom (sp, n, 0.5)
> pv
```

Since p-value more than 0.05 we accept no. at 5% level of significance

Q2 Following are the 10 observations
612, 618, 631, 628, 643, 640, 655, 649, 670, 663.

Apply sign test to test the hypothesis that the population median is 625 against the alternative it is greater than 625 at 1% L.O.S

Ho : pop median is 625

Note:- if a alternative is greater than
pv = pbinom (sn, n, 0.5)

```
>> x=c(612, 619, 631, 628, 643, 640, 655, 649
    670, 663)
> length(x)
[1] 10
> 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 more than 0.05 we accept Ho at 5% LOS

3 Ten observations are:-
36, 32, 21, 30, 24, 25, 20, 22, 20, 18. Using sign test to test hypothesis that population median is 25 against the alternative it is less than at 5% LOS

```
>x=c(56,52...,20,18)
>length(x)
[1] 10
>me=25
>sp=length(x[x>me])
>sn=length(x[x<me])
>n=sp+sn
>n
[1]9
>pv=pbinom(sp,n,0.5)
>pv
[1] 0.2539063
```

Since p value is

$\therefore$ 0.2539 > 0.05

$\therefore$ H_0 : population median is 25 is accepted

04 The following are some measurements 63, 65, 60, 89, 61, 71, 58, 51, 69, 62, 63, 39, 72, 65 using wilcoxon signed rank test. Test the hypothesis that the population median is 60 against alternative it is greater than 60 at 5% LOS.

H_0 : pop median is 60

```
->>x=c(63,65,....39,72,65)
>wilcox.test(x,alt="greater",mu=60)
        Wilcoxon signed rank test with
        continuity correction

data: x
V = 68, p-value = 0.06186
alternative hypothesis: true location is
        greater than 60 is accepted
```

$\therefore 0.0618 > 0.05$

H_0: population median is 60 is accepted

Q5 15, 17, 24, 25, 20, 21, 32, 28, 12, 25, 26. Use wilcox.test to test the hypothesis that population median is 20, against the alternative, is less than 20. LOS.

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

Wilcoxon signed Rank test with cointinuity correction

data: x

V = 48.5, p-value = 0.9232

alternative hypothesis: true location is less than 20

$\therefore 0.9232 > 0.05$

$\therefore$ H_0: population median is 50 is accepted

Q6 (20, 29, 27, 30, 18). Test the hypothesis that the population median is 25, against the alternative, is not 25

```
> x = c(20,25,27,30,18)
> wilcox.test (x, alt = "two.side", mu = 25)
```

Wilcoxon signed Rank test with continuity correction

data: x

V = 3.5, p-value = 0.7127

alternative hypothesis: True location is not equal to 25

$\therefore 0.7127 > 0.05$

$\therefore$ H_0 is accepted.