PRACTICAL NO-1

BASICS OF R SOFTWARE

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

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

iii) It is capable of graphical display

iv) It's a free software
+, -, *, /, abs, sqrt

Q1) $2^2+|-5|+4\times5+6/5$.

code:- 2^2 + abs(-5) + 4*5 + 6/5
⇒ 30.2

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

```
code:- x= 20
y=2*x
z= x+y
sqrt(z)
```

⇒ 7.745967

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

i) $x+y+z$
=) $x+y+z$
=30

ii) $x \cdot y \cdot z$
=) x * y * z
=) 750

iii) $\sqrt{xyz}$
=) a = sqrt (x*y*z)
a = 27.38613

iv) round ($\sqrt{xyz}$)
=) round(a)
= 27

Q4) A vector in R software is denied by c

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

2) x = c (2,3,5,7) ^c(2,3)
=) x
=) [1] 4 27 25 543

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

4) c (21, 23, 1, 4) * 3
=) [1] 63 69 3 12

5) c(21, 23, 1, 4) * c(-2, -3, -5, -7)
=) [1] -42, -69, -5, -28

OPO

6) c(2,3,5,7) + c(-2,-3,-1,0)

=) [1] 0 0 4 7.

7) c(2,3,5,7)/2

=) 1.0 1.5 2.5 3.5

Q5) Find the sum, product, sqroot of sum & product for the given value.

4, 9, 2, 5, 7, 8, 3, 6, 15, 12, 10, 9, 8, 13, 14

Code :-

```
i=c(4,9,2,5,7,8,3,6,15,12,10,9,8,13,14)
s= sum(i)
p= prod(i)
s
p
sqrt(s)
sqrt(p)
```

output

```
[1] 125
[1] 8.559323e+12
[1] 11.18034
[1] 2925632
```

Q6) Find the sum, prod, max, min, value of x=c(2,8,9,1, 7,6)^2

sum(x) max(x) prod(x) min(x)

=) [1] 53 [1] 11 [1] 665280 [1] 2

Matrix : x<- matrix(nrow=4, ncol=2, data=c(1,2,3,4,5,6

7) a<- matrix(nrow=3, ncol=3, data=c(4,5,6,7,8,9,4,0,2))
b<-matrix(nrow=3, ncol=3, data=c(6,4,5,11,12,18,9,7,4))

=) a+b

	[,1]	[,2]	[,3]
[1,]	10	18	13
[2,]	9	20	7
[3,]	11	27	6

a*b

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

b*3

	[,1]	[,2]	[,3]
[1,]	18	33	27
[2,]	12	36	21
[3,]	15	54	12

a%*%b

	[,1]	[,2]	[,3]
[1,]	24	77	36
[2,]	20	96	0
[3,]	30	162	8

AM

FDO

PRACTICAL NO - 2.1

* Binomial Distribution

n : Total no of trials
p = P(success)
q = P(failure)
x = no of success out of n
$p[x] = {}^{n}C_{x} \; p^{x} q^{n-x}$
$E[x] = np$
$V(x) = npq$
dbinom (x,n,p)
n*p
n*p*q
pbinom (x,n,p)

* Exercise :-

Q1) 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

output

```
>n = 10
>p = 0.6
>q = 0.4
>a = dbinom (7,n,p)
>a
[1] 0.2149908
>b = dbinom [4,n,8)
>b
[1] 0.1114767
>c=1- pbinom (6,n,8)
>c
[1] 0.3822806
>d= pbinom (4,n,8)
>d
[1] 0.1662386
>e= dbinom (0,n,8)
>e
[1] 0.0001048576
>f=dbinom(10,n,p)
>f
[1] 0.006046618
```

PM

PLO

PRACTICAL NO :- 2.2

* Binomial Distribution

Q2) 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) atmost 4 correct answers.
iii) atleast 8 correct answers.

Q3) Find the complete binomial distribution when n=5 $p = 0.1$

Q4) Find the probability of exactly 10 success out of 100 trials with $p = 0.1$

Q5) 'x' follow binomial distribution with $n=12$ & $p = 0.25$.
Find i) $P(X \leq 5)$
ii) $P(X > 7)$
iii) $P(5 < X < 7)$

Q6) there are 10 members in a committee probability of any member attending a meeting 0.9. What is the probability that 7 or more member will be present meeting

Q7) A salesman has a 2 ...

sell to a customer. On a typical day he'll meet 50 customers. What minimum no of sells will make with 88% probability

Q3) For $n=10, p=0.6$. Find the binomial probability that & plot the graphs of part 2 c d f.

* formula :-
  * Probability of atmost x values:
    $P(X \le x) = pbinom(x, n, p)$
  * Probability of atleast x values
    $P(X \ge x) = 1 - pbinom(x, n, p)$
  * If 'x' is unknown and the probability is given as p_1. To find x qbinom (p_1, 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] 0.131863
```

Q5)
```
>n =12
>p= 0.25
>pbinom (5,n,p)
[1] 0.9455978
>1-pbinom (7,n,p)
[1] 0.0027815)
>dbinom (6,n,p)
[1] 0.04014945
```

a&c

Q6) > n = 10
> p = 0.9
> 1- pbinom (6,n,p)
[1] 0.9872048

Q7) > n = 30
> p = 0.2
> p1 = 0.88
> qbinom(p1,n,q)
[1] 9

Q8) > n = 10
> p = 0.6
> x = 0:n
> bp = dbinom(x,n,p)
> bp
[1] 0.0001048576 0.0015728640 0.106168320 0.04246737 8
0.11476734

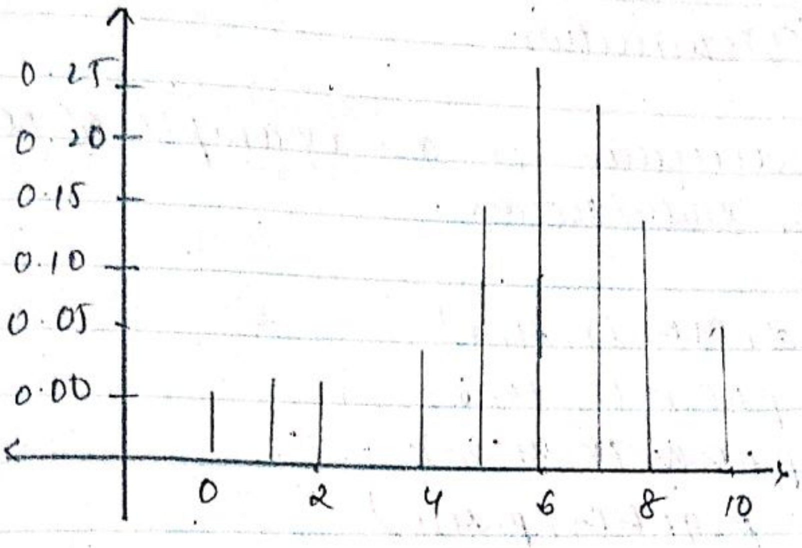
[6] 0.2006551248 0.2508226560 0.2149708480 0.1209373820

[11] 0.0060466176

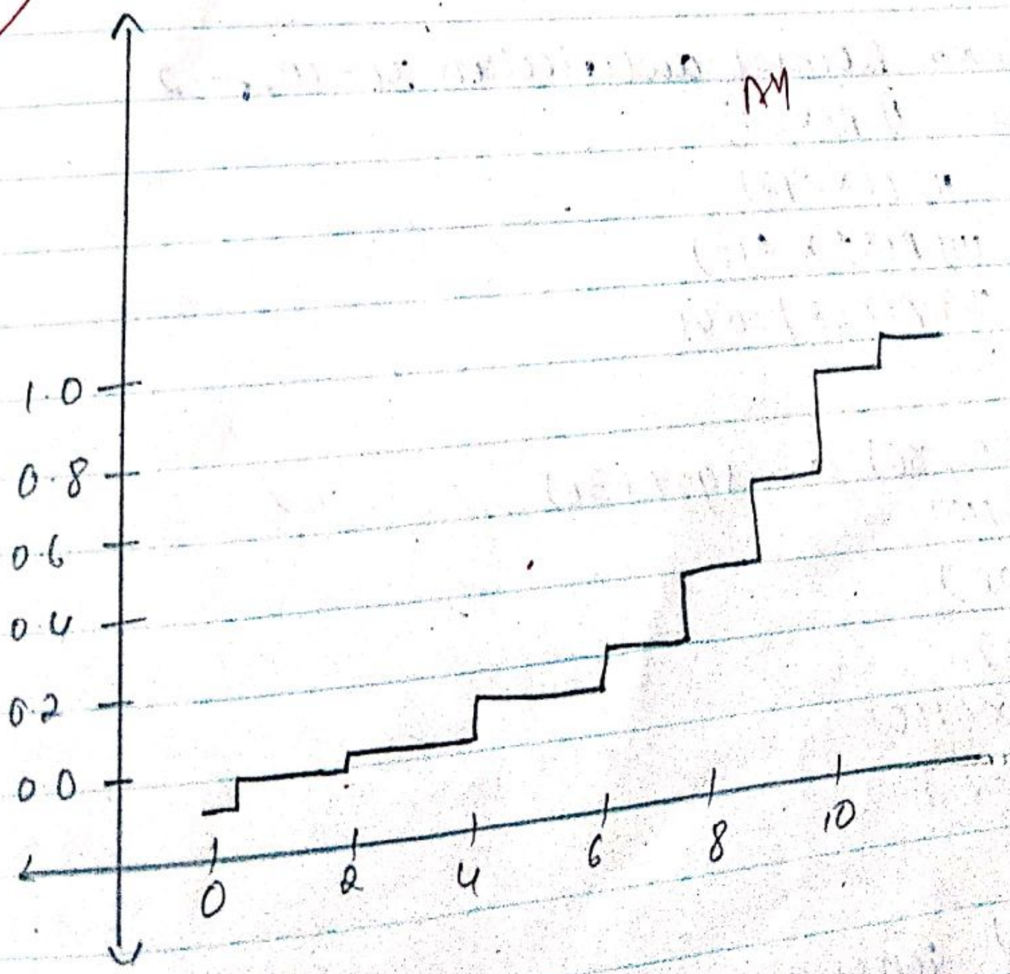
> d = data.frame ("x.value" = x, probability" = bp)
> d

	x value	probability
1	0	0.0001048576
2	1	0.0015728640
3	2	0.0106168320
4	3	0.0424673280
5	4	0.114767360
6	5	0.2006581248
7	6	0.2508226560
8	7	0.2149908480
9	8	0.120932530
10	9	0.0403107840

>plot (x, bp, "h")



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



AM

PRACTICAL NO:-3

Q1) check the following are p.m.f 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 satisfied

Also,

$\Sigma P(X) = P(10) + P(20) + P(30) + P(40) + P(50)$

$= 0.3 + 0.2 + 0.3 + 0.1 + 0.1$

$= 1$

$\therefore$ 2nd condition is also satisfied

→ Here 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 than 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, 2^{nd} 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
		$\Sigma XP(X) = 3.45$	$\Sigma X^2 P(X) = 13.55$

Mean $= E(X) = \Sigma X P(X) = 3.45$

Var $= V(X) = \Sigma X^2 P(X) - [E(X)]^2$
$$= 13.55 - (3.45)^2$$
$$= 1.6475$$

011

```
> 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] 271.25
> var = sum(b) - (sum(a)^2)
> var
```

FAQ...

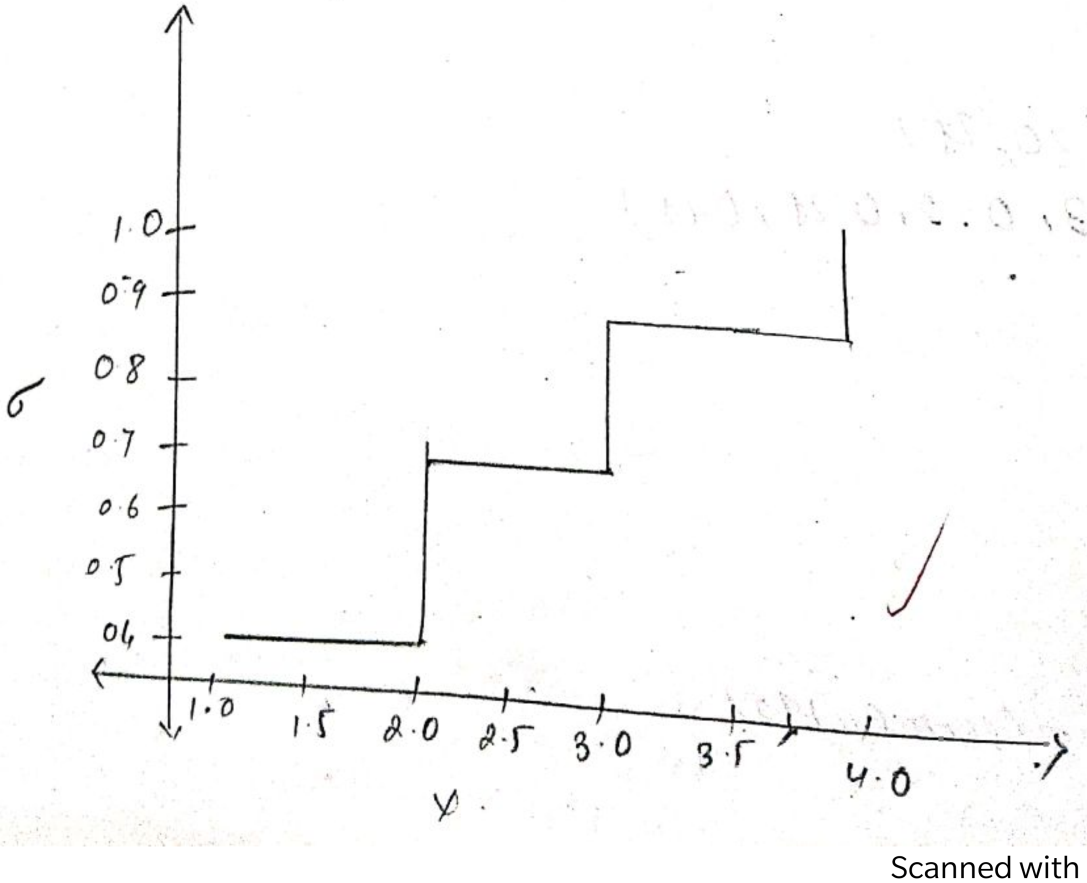
Q4) Find c.d.f of the following p.m.f and draw the group of c.d.f

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

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

F(X) = 0 x < 1
= 0.4 1 ≤ x < 2
= 0.7 2 ≤ x < 3
= 0.9 3 ≤ x < 4
= 1.0 x ≥ 4

Plot (x, a, "s")



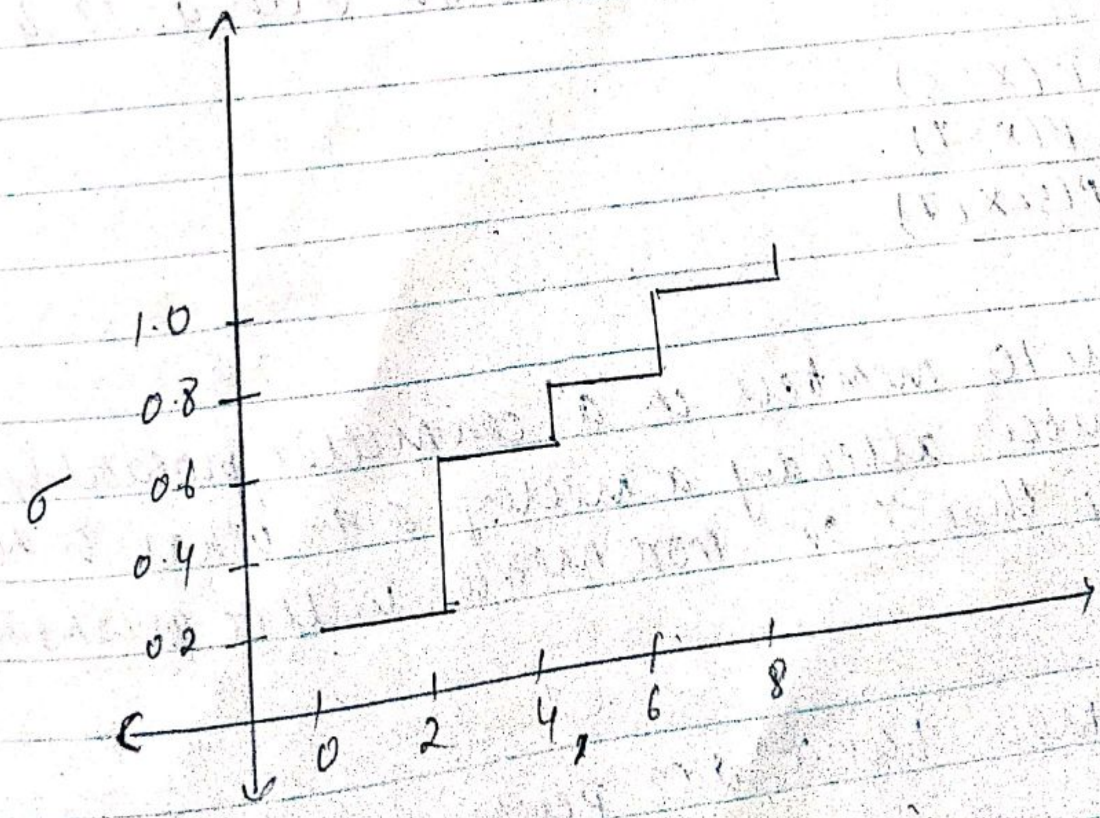
Q4) ii)

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)
> prob = c(0.2,0.3,0.2,0.2,0.1)
> a = cumsum(prob)
> a
[1] 0.2    0.5  0.7  0.9  1.0
```

$$
P(X) = \begin{cases} 0 & x < 0 \\ 0.2 & 0 \le x < 2 \\ 0.5 & 2 \le x < 4 \\ 0.7 & 4 \le x < 6 \\ 0.9 & 6 \le x < 8 \\ 1.0 & x \ge 8 \end{cases}
$$

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



PRACTICAL-NO 4.

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

$[n=8, p=0.6, q=0.4]$ - given

i) $P(X=7) = \binom{8}{7} (0.6)^7 \cdot (0.4)^{8-7}$
$= {}^8C_7 \times 0.2799 \times 0.4$
$= 8 \times 0.2799 \times 0.4 = 0.08957$

ii) $P(X \le 3) = P(0) + P(1) + P(2) + P(3)$
$= {}^8C_0 (0.6)^0 (0.4)^8 + {}^8C_1 (0.6)^1 (0.4)^{8-1}$
$+ {}^8C_2 (0.6)^2 (0.4)^6 + {}^8C_3 (0.6)^2 (0.4)^5$

$= 1 \times 1 \times 0.00065536 + 8 \times 0.6 \times 0.0016384$
$+ 28 \times 0.36 \times 0.004096$
$+ 56 \times 0.26 \times 0.01024$
$= 0.1736704$

iii) $P(X = 2 \text{ or } 3) = P(2) + P(3)$
$= {}^8C_2 (0.6)^2 (0.4)^2 + {}^8C_3 (0.6)^3 (0.4)^3$
$= 28 \times 0.36 \times 0.004096 + 56 \times 0.216 \times 0.01024$
$= 0.04128768 + 0.12386304$
$= 0.16515012$

FAO -

PRACTICAL NO - 5.

* Normal Distribution

Normal distribution is an example of continuous probability distribution.

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

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

iii] $X \sim N(100, 36)$ $\sigma =$ sqrt (36)
i) $P(X \le 110)$
ii) $P(X > 105)$
iii) $P(X \le 92)$
iv) $P(95 \le X \le 110)$
v) $P(X < k) = 0.9$

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

Solution i] pnorm(7,10,2)

[1] 0.0668072

>cat("P(X<=7) is =",p)

P(X<=7) is= 0.0668072

>p2 = 1-pnorm(12,10,2)

>cat("P(X>12) is=",p2)

P(X>12) is = 0.1586553

>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

iii] x=rnorm(10,10,3)

>x

[1] 9.195412 11.360665 5.331607 17.160876 11.912200
9.372691 12.436043 13.810725 10.908016 13.367803

>am= mean(x)

>am

[1] 11.48346

>me= median(x)

>me

[1] 11.62647

>n=10

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

>variance

[1] 9.018713

>sd= sqrt(variance)

>sd

[1] 5.003...

8AO

11) >p1 = pnorm (110, 100, 6)

>p1

[1] 0.9522096

> cat ("p(X<=110) is =" p1)

p(X <=110) is = 0.95220967

>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 = 0.09121122>

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

> cat ("p(95<=X<110) is =", p4)

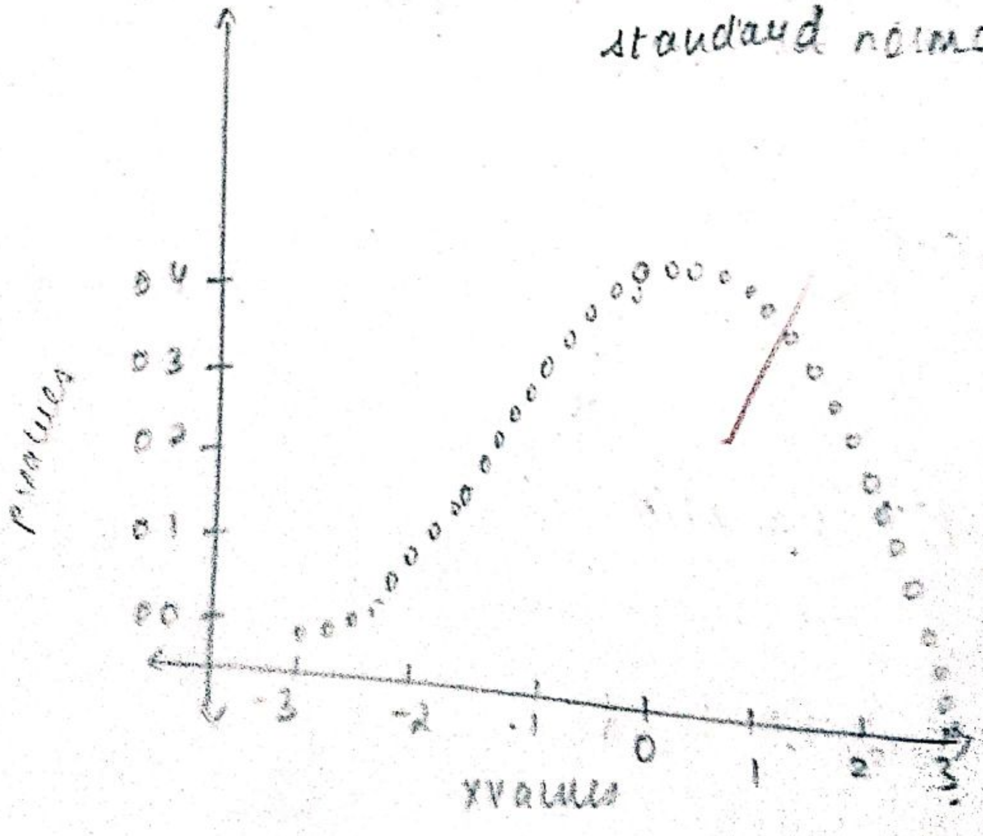
p (95<=X<=110) is = 0.74988137

>k = qnorm (0.9 ;100, 6)

> cat ("p(X<k)=0.9, k is =", k)

p(X<k): 0.9 k is = 107.68937

standard normal curve



v) plot the standard normal curve

x = seq(-3, 3 by=0.1)

y = dnorm(x)

plot(x, y, xlab = "x values", ylab = "probability", main = "standard normal curve")

x ~ N(50, 100)

find i) P(X ≤ 60)

ii) P(X > 65)

iii) P(45 ≤ X ≤ 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 0.668072

>p3 = pnorm(60, 50, 10) - pnorm(45, 50, 10)

>cat("P(45<= X <=60) is", p3)

P(45<= X <=60) is 0.5328072

20/1/

PRACTICAL NO-6

z and t distribution sums

Q1) Test the hypothesis H0: μ=20 against H1: μ≠20
A sample of size 400 is selected and the sample mean is 20.2 and the standard deviation is 2.25
Test at 5% level of significance

```
>n = 400
>mo = 20
>mx = 20.2
>sd = 2.25
>z cal = (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)
pvalue is 0.0754
```

∵ 0.0754 > 0.05
∴ H0: μ≠20 is accepted

Q2) Test the hypothesis H0: μ=250 against H1: μ≠250
A sample of size 100 has a mean of 275 & standard deviation 30. check the hypothesis at 5% level of significance.

```
>n=100
>mo=250
>mx=275
```

> sd=30
> zcal=(mx-mo)/(sd/sqrt(n))
> cat ("z calculated is", zcal)
z calculated is 8.333337
> pvalue= 2* (1-pnorm (abs(zcal)))
> cat ("PVALUE is", pvalue)
PVALUE is 0 >
0 < 0.075
∴ H0: μ= 250 is rejected

Q3) We want to test the hypothesis H0, P=0.2 against H1: P ≠ 0.2 (P= population proportion). A sample of 400 is selected and the sample proportion is calculated 0.125. Test the hypothesis at 1% level of significance.

> P = 0.2
> Q= 1-P
> p = 0.125
> n = 400.
> zcal= (p-P)/sqrt((P*Q/n))
> zcal
[1] -3.75
> pvalue=2* (1-pnorm(abs(zcal)))
> cat ("PVALUE is", pvalue)
PVALUE is 0.000176846 >
∴ P value < 0.01
∴ H0: μ= 0.2 is rejected

Q4) In a big city 325 men out of 600 men are found to be self-employed. Does this informat support the conclusion that exactly half of the men in the city are self employed

```
> p = 325/600
> P = 0.5
> n = 600
> Q = 1 - P
> zcal = (p - P)/(sqrt(P*Q/n))
> zcal
[1] 2.041241
> pvalue = 2*(1-pnorm(abs(zcal)))
> cat("P VALUE IS", pvalue)
PVALUE IS 0.04122683 >
```

∴ 0.0412 = 0.05

∴ It is rejected

Q5) Test the hypothesis $H_0 : \mu = 50$ against $H_1 : \mu \neq 50$

A sample of 30 is collected.

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

```
> M0 = 50
> x = c(50,49,52,44,45,48,46,45,49,45,40,47,55,54,46,58,
  47,44,59,60,61,41,52,44,55,56,46,45,48,49)
> n = length(x)
> n
[1] 30
> mx = mean(x)
> mx
```

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

> variance

[1] 30.955

> sd = sqrt (variance)

> sd

[1] 5.563772

> zcal

[1] -0.6562965

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

> cat ("pvalue is", pvalue)

PVALUE is 0.5116334>

$\because$ 0.511 > 0.05

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

20-4-20

PRACTICAL NO. 7

LARGE SAMPLE TEST

Q1) Two randoms sample of size 1000 and 2000 are drawn from 2 population with standard deviation 2, 3 test the hypothesis that the two population means are equal to 5% level of significance sample means are 67 & 68 resp.

Ho: $\mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$

```
>n1 = 1000
> n2 = 2000
> mx1 = 67
> mx2 = 68
> sd1 = 2
> sd2 = 3
> zcal = (mx1 - mx2) / sqrt((sd1^2 / n1) + (sd2^2 / n2))
> cat ("z calculated is", zcal)
calculated is -10.84652
> pvalue = 2* (1- pnorm(abs(zcal)))
> cat ("Pvalue is", pvalue)
Pvalue is 0
```

∵ 0 < 0.05

∴ It is rejected

Q2) A study of noise level is done in two hospitals following data is calculated first sample size = 84, mean = 61.2, sd = 7.9. second sample size, 34, mean = 59.4, sd = 7.8 at 1 level of significance. Test H0 $\mu_1 = \mu_2$ at 1 level of significance.

H0: $\mu_0 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$

```
> n1=84
> n2=34
> mx1=61.2
> mx2=59.4
> sd1=7.9
> sd2=7.8
> zcal=((mx1-mx2)/(sqrt((sd1^2/n1)+(sd2^2/n2))
> cat ("z calculated is", zcal)
> calculated is 1.131117
> pvalue= 2* (1-pnorm(abs(zcal)))
> cat ("pvalue is ", pvalue)
pvalue is 0.258006
```

∴ 0.25 > 0.01

∴ H is accepted

pvalue is greater than 0.01

Q3) From each of the population of oranges the following data is collected check whether the proportion of that oranges are equal or not. n1=250, bad oranges = 44, n2=200, bad oranges = ...

Test: p1 = p2 against p1 ≠ p2

H0: P1=P2 against
H1: P1≠P2

n1=250
n2=200
p1=46/250
p2=30/200
p=(n1*p1)+(n2*p2)/(n1+n2)
>p
[1] 0.1688889
>q=1-p
>q
[1] 0.8311111
>zcal=(p1-p2)/sqrt(p*q*(1/n1+1/n2))
>cat("zcalculated is",zcal)
zcalculated is 0.7393817
>pvalue=2*(1-pnorm(abs(zcal)))
>cat("Pvalue is",pvalue)
Pvalue is 0.4596896
We accept the H0: P1=P2

Q4) Random sample of 400 men & 600 females were asked whether they want an ATM nearby. 200 male & 390 female were in favour of the proposal. Test the hypothesis that the population of male and female favour in the proportion are equal or not at 5% level of significance

```
>H0: P1 = P2  against ...
>n1 = 400
>n2 = 600
>p1 = 200/400
>p2 = 390/600
>p = (n1*p1+n2*p2)/(n1+n2)
>p
[1] 0.59
>q = 1-p
>zcal = (p1-p2)/(sqrt(p*q*(1/n1+1/n2)))
>zcal
[1] -4.7247
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 2.303972e-06
```

∴ pvalue is less than 0.05, we reject $H_0: P_1 = P_2$.

Q5) Following are the two independent sample from the two population. Test equality of two population mean at 5% level of significance.

for S1 = [74, 77, 74, 73, 79, 76, 82, 72, 75, 78, 77, 78, 76, 71]

S2 = [72, 76, 74, 70, 70, 78, 70, 72, 75, 79, 79, 74, 76, 16]

) H0: $\mu_1 = \mu_2$ against H1: $\mu_1 \neq \mu_2$

>x1= c(74, 77, 74, ---, 76, 70)

>n1= length(x1)

>mx1= mean(x1)

>variance = (n1-1)* var(x1)/n1

> variance

[1] 0.4508

>sd1= sqrt(variance)

> sd1

[1] 0.6714

>x2 = c(72, 76, 74, ---, 74, 80)

> n2= length(x2)

> mx2 = mean(x2)

>variance = (n2-1)* var(x2)/n2

> variance

[1] 0.6163

>sd2= sqrt(variance)

> sd2

[1] 0.7850

> tcal= (mx1-mx2)/sqrt((sd1^2/n1)+(sd2^2/n2))

> tcal

[1] 1.4851

> cat ("t calculated is:" tcal)

[1] t calculated is: 1.4851

> t.test (x1, x2)

p value = 0.1387

H0 is accepted

AM 27-01-20

730

PRACTICAL - 8

Aim :- small sample test

Q1) The random sample of 15 observations are given by 80, 100, 110, 105, 122, 70, 110, 100, 88, 83, 95, 89, 107, 125. Do this data support the assumption that population mean is 100?

-) > x = c(180, 100, 110, 125, 122, 70, 120, 110, 101, 88, 83, 95, 87, 107, 125)

> a = length(x)

> a

[1] 15

> t.test(x)

One sample t-test

data : x

t = 24.029 , df = 14 p-value = 8.819 xe-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

∴ $8.819 \times 10^{-13} < 0.05$

∴ $H_0: \mu = 100$ is rejected at 5% level of significance

two groups of 10 students score the following marks 056
grp1: 18, 22, 21, 17, 20, 17, 23, 20, 22, 21
grp2: 16, 20, 14, 21, 20, 18, 13, 15, 17, 21
test the hypothesis that there is no significant difference
in the score at 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, 18, 17, 21)
>t.test(g1, g2)

Welch two sample t-test

data: g1 and g2

t = 2.2573, df = 16.376, p-value = 0.03798
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence intervals
0.1628205 5.0371795

sample estimates:
mean of x mean of y
20.1 17.5

∴ 0.03798 > 0.01

∴ $H_0: \mu_k = 20$ is accepted at 1% level of significance

) Two types of medicines ~~are used on~~ 5 & 7 patients for reducing their weight. The decrease in the weight after using the medicines are given below

A: (10, 12, 13, 11, 14)
B = (8, 9, 12, 14, 15, 10, 9)
To the ... ...ant difference in the efficiency of

DSC

$H_0: \mu_1 = \mu_2$

-> m1 = c(10, 12, 13, 14, 11)

> m2 = c(8, 9, 12, 14, 15, 10, 9)

> t-test (m1, m2)

Welch Two sample t-test

data: m1 & m2

t = 0.80884, df = 9.7594, p-value = 0.4406

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

95 percent confidence interval:

-1.781171 3.781171

sample estimate

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.

Q4> The weight reducing diet program is conducted the observations are noted for 10 participants. 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 significance difference in weight ...

H_1: the diet program 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 ...

t = 17, df = 9, p-value = 1
alternative hypothesis: true difference in means
is less than 0
95 percent confidence interval:
-Inf 9.416556
sample estimates:
mean of the difference
8.5

$\therefore$ 1 > 0.05
$\therefore$ H_0 is accepted at 5% level of significance.

(ii) sample A: 66, 67, 75, 76, 82, 84, 89, 90, 92
sample B: 64, 66, 74, 78, 82, 85, 87, 92, 93, 95, 97
test the population mean are equal or not.
H_0: ??
x_a = c (66, 67, 75, 76, 82, 84, 89, 90, 92)
x_b = c (64, 66, 74, 78, 82, 85, 87, 92, 93, 95, 97)
x_t = test (9, 8)
Welch two sample t-test
data = a and 21
t = -1.05, df = 17.956 p value = 0.5477
alternative hypothesis true difference in means is
not equal to 0
95 percent confi ~~dence~~ interval:
-12.79 7.018223
sample estimates
80.1111 83.0000
$\therefore$ 0.5477 > 0.05

∴ H0 at 5% level of significance is accepted

Q6 Following are the marks before & after a training program test whether the program is effective or not

Sol:- before - 71, 72, 74, 69, 70, 74, 76, 70, 73, 75
after - 74, 77, 74, 73, 79, 76, 82, 72, 75, 78

→ b = c(71, 72, 74, 69, 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 = -4.4691, df = 9, p-value = 0.0007784
alternative hypothesis: true diff in means is less than

-Inf - 2.123361
sample estimates:
mean of the differences
-3.6

∴ 0.0007784 < 0.05
∴ H0 is rejected at 5% level of significance

03.02.20

PRACTICAL NO-9

LARGE AND SMALL SAMPLE TEST

Q1) The arithmatic mean of a sample of no of items from a large population is 57. If standard deviation is 7, test the hypothesis that the population mean is 55 against the alternative more than 55% @ 5% LOS

Q2) In a big city 850 out of 700 males are found to be smokers. Does this info support that exactly half of males in the city of smokers? Test at 1% LOS.

Q3) Thousand articles from a factory A are found to have 2% defectives 1500 articles from a factory B are found to have 1% defectives. 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 that comes from a population of mean 100 & variance 64?

Q5) The flower stems are selected and heights are found to be (in cms) 63, 63, 18, 69, 71, 72, Test the hypothesis that mean height is 66 or not at 1% level of significance

Q6) Two random samples were drawn from 2 normal population & their values are A - 66, 67, 75, 76, 82, 84, 89, 90, 92
A - 64, 66, 74, 78, 82, 85, 87, 92, 93, 95, 97
B - 64, 66, 74, 78, 82, 85, 87 ... have same variance at 5% LOS
... populations

830

1)

```
>n = 100
>mx = 52
>mo = 55
>sd = 7
>zcal = (mx-mo)/(sd/sqrt(n))
>cat("z calculated is", zcal)
-4.285714
>pvalue = 2*(1-pnorm(abs(zcal)))
>cat("Pvalue is", pvalue)
```

Pvalue is 1.82153 $\times 10^{-5}$

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

$H_0: \mu = 55$ is rejected.

2.

```
>p = 0.5
>q = 1-p
>p = 350/700
>n = 700
>zcal = (p-P)/sqrt(P*Q/n)
>cat("z calculated 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("z calculated is",zcal)
>z calculated is 2.084842
>pvalue=2*(1-pnorm(abs(zcal)))
>cat("pvalue is",pvalue)
pvalue is 0.03708364

∵ 0.037<0.05

∴ H0: μ is rejected

(v) >n=400
>mx=99
>m0=100
>var=64
>sd=sqrt(var)
>zcal=(mx-m0)/(sd/sqrt(n))
>cat("z calculated is",zcal)
z calculated is -2.5
>cat("pvalue is",pval")

030

pvalue is 0.01241933?

∴ 0.01241 < 0.05

It is rejected at 5% LOS

H0: μ

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

>a = length(x)

> a

[1] 7

>t-test(x)

One sample t test

data: x

t = 47.94, df = 26, pvalue = 5.522e-05

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval

64.66477 71.62092

sample estimates

mean of x

68.14286

∴ 0.0000000552 < 0.05

∴ H0: μ = 66 is accepted

Scanned with CamScanner

PRACTICAL-10

ANOVA & Chi-square test

1) Use the following data to test whether the cleanliness of home & cleanliness of the child independent or not

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

-) Ho: CC & CH

```
>x = c(70, 80, 35, 50, 20, 45)
>m = 3 (rows)
>n= 2 (columns)
>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 - sq test

data : y
X- squared = 25.646, df = 2, p-value = 2.698e-06
∴ pvalue is less than 0.05, we reject
Ho: CC & CH.

Q2) Use the following data to find a vaccination and a particular class are independent or not

	Dis	
	Aff	Not Aff
Vaccination given	20	30
Not given	25	35

→ Ho: Vaccination & disease 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

... 4
squared = 0, df = 1, pvalue = 1
pvalue is greater than 0.05 we accept
H0: vaccination & disease are independent 062

Perform ANOVA for the following data

variety	observation
A	50, 52
B	53, 55, 53
C	60, 58, 57, 56
D	52, 54, 54, 55

H0: the means 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))
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 = 0.001835
anova = aov(values ~ ind, data = d)
anova
call:
aov(formula = values ~ ind, data = d)

terms	ind	residuals
sum of sq's	71.06410	18.16667
degrees of freedom	3	9

Residual standard error: 1.420746
estimated effects may be unbalanced.

∴ since pvalue is less than 0.05, we reject
H0: The means of the varieties are equal.

Q4) The following data gives the life of type of 4 brand

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.

- Test the hypothesis that the average life for the four brand are same

-) H0: Avg life of 4 brands are 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).
> d = stack(list(b1 = x1, b2 = x2, b3 = x3, b4 = x4))
> names(d)
> oneway.test(values ~ ind, data = d, var.equal=T)
```

One-way analysis of means

data: values & ind
F = 6.945, num df = 3 denom ...

> anova = aq(values ~ ind, data = a)
> anova

call:
a ov (formula = values ~ ind, data = d)

Terms

	ind	Residuals
sum of sq's	91.4381	89.0619
Deg of freedom	3	20

Residual standard error: 2.110237
Estimated effects may be unbalanced

∴ the p-value is less than 0.05, we reject the

H0: life of four type brand are same.

15) one thousand students of a college are graded according to their IQ and the economic condition of their home. check that is there any association between IQ & their economic condition of their home

		IQ	
		high	low
EC	high	460	140
	mid	330	200
	low	240	160

630

→ Ho: EC & IQ are independent
> x = c (460, 330, 240, 140, 200, 160)
> m = 3
> n = 2
> y = matrix (x, nrow = m, ncol = 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 = 2.364e

∴ pvalue is less than 0.05, we reject
Ho: EC & IQ are indipendent.

AM

PRACTICAL-11.

NON-Parametric test

1) following are the amounts of sulphuric acid emmited by industries in 20 days Apply sign test to test the hypothesis that the population median is 21.5

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

H0: population median is 21.5

> x = c(Data)

> median = 21.5

> n = length(x[x<me])

> n = sp+sn

> sp = length(x[x>me])

> n

[1] 20

> pv = pbinom(sp, n, 0.5)

> pv

[1] 0.419015

since p-value is more than 0.05 we accept H0: population median is 21.5

2) following are the 10 observation 612, 619, 631, 628, 643, 640, 655, 640, 670, 663

Apply sign test the hypothesis the population median is 625 against the alternative H1 is greater than 625 at 1% LOS

Only if the alternative is greater than pv = pbinom(sn, n, 0.5)

median is 625

```
> x = c(Data)
> med = 625
> sn = length(x(x<med))
> sp = length(x(x>med))
> n = sp+sn
[1] 10
> pv = pbinom(sn,n,0.5)
> pv
[1] 0.0546875
```

∴ since p-value is more than 0.01 ∴ we accept the H0 population median is 625.

* 10 observations are

36, 32, 21, 30, 24, 25, 20, 22, 20, 18

using sign test. Test the hypothesis is 25 against the alternative it is less than 25 at 5% loss

soln:-

H0: population median is 25

```
> x = c(Data)
> med = 25
> sp = length(x[x>med])
> sn = length(x[x<med])
> n = sp+sn
> n
[1] 9
> pv = pbinom(sp,n,0.5)
> pv
```

[1] 0.25 39063

∴ pvalue is more than 0.05 we accept H0:

population median is 25

Q4) The following are same means moments -

63, 65, 60, 89, 61, 71, 58, 51, 69, 62, 63, 39, 72, 5

using wilcoxon signed rank test. Test the hypothesis that the population median is 60 against the alternative it is greater than 60 at 5% LOS

H0: population median is 60

> x<-c(data)

> 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

since p-value is greater than 0.05, we accept H0: population median is 60

) 15, 17, 29, 25, 20, 21, 32, 28, 12, 25, 24, 21

wilcoxon signed rank test. Test the hypothesis that the population median 20 against the alternative it is less than 20 at 5% LOS.

230

Soln:-

> Ho:: population median is 20

> x = c(Data)

> wilcoxon test (x, alt = "less", mu = 20)

wilcoxon signed rank test with continuity correction

data: x

V = 48.5   p-value = 0.932

alternative hypothesis true location is less than 20.

∴ since, p-value is greater than 0.05, we accept Ho: population median is 20.

6) (20, 25, 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.sided", 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

∵ 0.7127 > 0.05

∴ Ho is accepted

AD 9.7.22