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

Q.1. $2^2 + |-5| + 4 \times 5 + 6/5$

Code :- `2^2 + abs(-5) + 4*5 + 6/5`

⇒ 30.2.

Q.2. $x = 20$

$y = 2x$

$z = x + y$

Find $\sqrt{z}$

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

⇒ 7.745967

Q.3. $x = 10, y = 15, z = 5$.

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.

Q.4. A vector in R software is denoted 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

(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

Q.5: Find the sum, product, sqroot of sum & product for the given values.

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

<u>code</u> :-

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

Q.6. Find the sum, prod, max, min. values of x = c(2, 8, 9, 11, 10, 7, 6) 12

Sum (x)	max (x)	prod (x)	min (x)
=> [1] 53	[1] 11	[1] 665280	[1] 2

* <u>Matrix</u> :- x <- matrix (nrow = 4, ncol = 2, data = c(1, 2, 3, 4, 5, 6, 7, 8))

38

Q.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 * 2

	[,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

9-12-19

PRACTICAL NO- 2.1

* Binomial Distribution

n : Total number of trials.
p : P (success)
q : P (failure)
r : number 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 :-

Q.1 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
```

PRACTICAL NO :- 3

Q.1 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, 1^{st} cond^n is satisfied.

Also,

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

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

$$= 1$$

$\therefore$ 2^{nd} condition is also satisfied

→ 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
```

3 0.25

ce of 'x

$x^2 P(x)$
0.1
0.60
1.8
4.8
6.25

```
> 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
```

Q.3. Find mean & variance of X.

~~xP(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
> 38.6875
```

Scanned with CamScanner

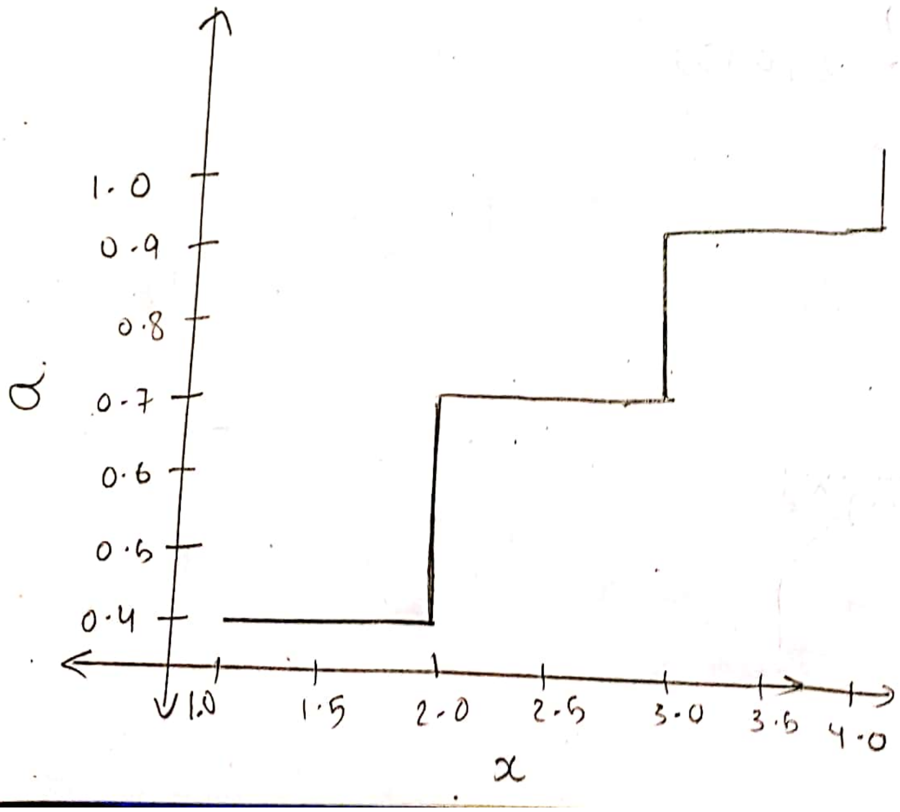
Q.4. i) Find c.d.f of the following p.m.f and draw the graph of c.d.f

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

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

$$
\begin{aligned}
P(x) &= 0 & x < 1 \\
&= 0.4 & 1 \le x < 2 \\
&= 0.7 & 2 \le x < 3 \\
&= 0.9 & 3 \le x < 4 \\
&= 1.0 & x \geq 4
\end{aligned}
$$

Plot(x, a, "s")



Q. (ii)

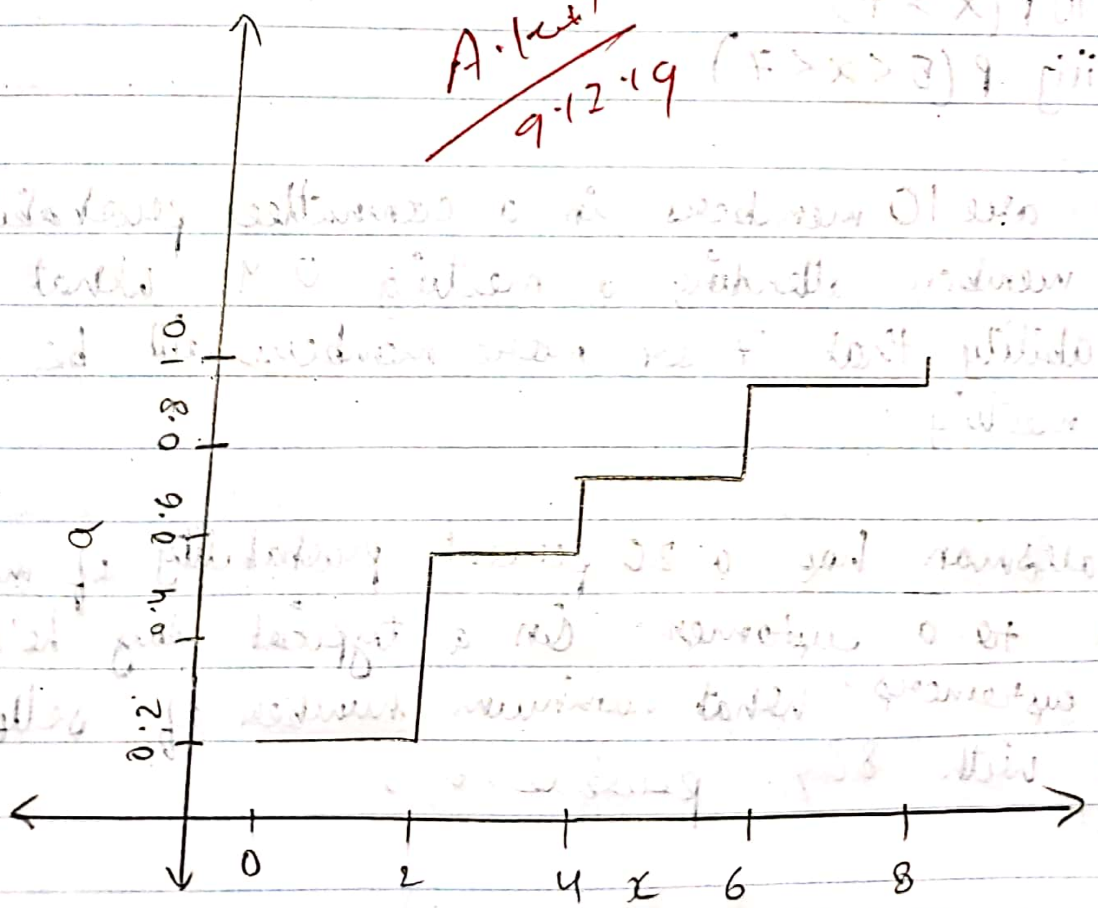
x	0	2	4	6	8
$P(x)$	0.2	0.3	0.2	0.2	0.1

```
> 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
```

$$P(x) = \begin{cases} 0.\cancel{2} & x < 0 \\ 0.2 & 0 \leq x < 2 \\ 0.5 & 2 \leq x < 4 \\ 0.7 & 4 \leq x < 6 \\ 0.9 & 6 \leq x < 8 \\ 1.0 & x \geq 8 \end{cases}$$

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

A.Kt
9.12.19



Scanned with CamScanner

PRACTICAL NO :- 2.2

* Binomial Distribution :-

Q.2 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 3 correct answers :-

Q.3 Find the complete binomial distribution when $n=5$ & $p=0.1$.

Q.4. Find the probability of exactly 10 successes out of 100 trials with $p=0.1$.

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

Q.6. 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.

Q.7. 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 number of sells he'll make with 88% probability.

Q.8. For $n = 10$, $p = 0.6$. Find the binomial probability that and plot the graphs of pmf & cdf.

* Formulas :-

* Probability of atmost x values:

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

* Probability of atleast x values :-

$P(X \geq x) = 1 - pbinom(x, n, p)$

* If 'x' is unknown and the probabity is given as P_1, To find X qbinom (P_1, n, P)

Q.2.
```
→ > 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
```

Q.3 →
```
> 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
```

Q.4.
```
> x = 10
> n = 100
> p = 0.1
> dbinom (x,n,p)
[1] 0.1318653
```

Q.5.
```
> n = 12
> p = 0.25
> pbinom (5,n,p)
[1] 0.9455978
> 1-pbinom (7,n,p)
[1] 0.00278151
> dbinom (6,n,p)
[1] 0.04014945
```

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

Q.7. > n=30
> p=0.2
> p1=0.88
> qbinom(p1,n,p)
[1] 9

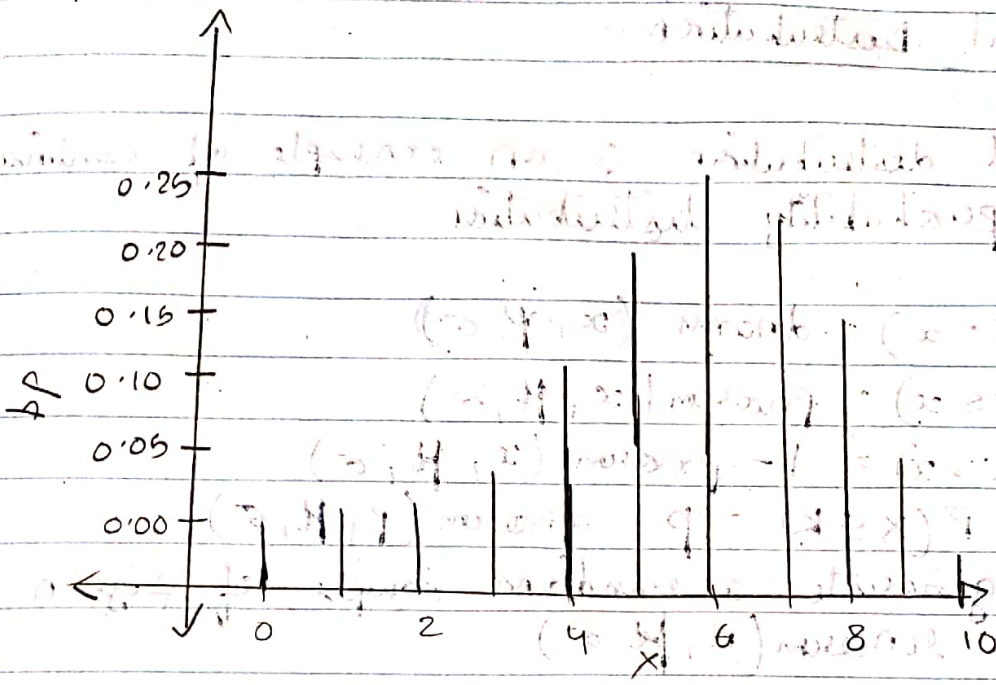
Q.8.> n=10
> p=0.6
> x=0:n
> bp=dbinom(x,n,p)
> bp
[1] 0.0001048576 0.0015728640 0.106168320 0.0424673280 0.111476734
[6] 0.2006551248 0.2508226560 0.2149908480 0.1209323520 0.0403107
[11] 0.0060466176

> d=data.frame("x.values"=x, "Probability"=bp)
> d

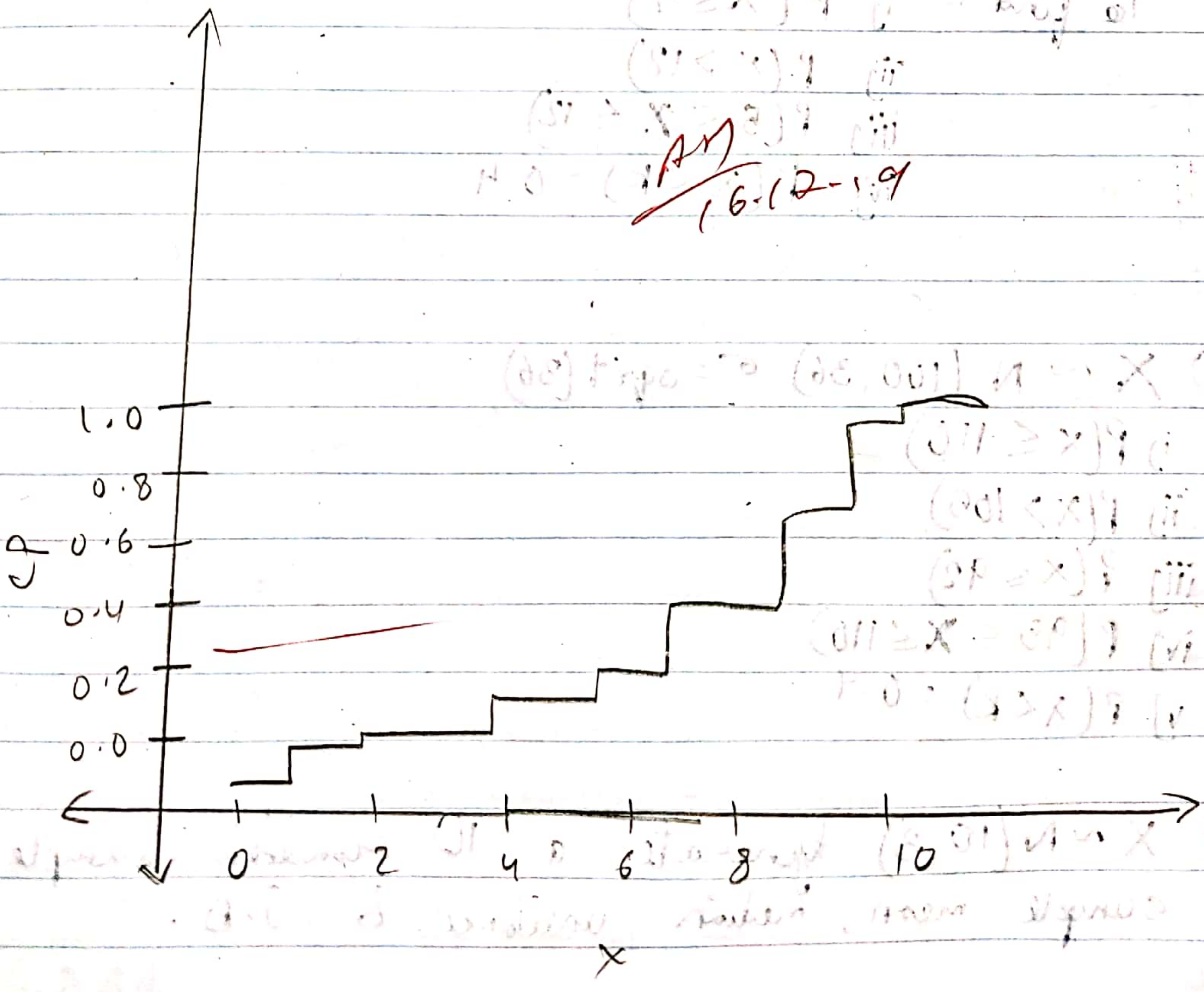
	x.values	Probability
1	0	0.0001048576
2	1	0.0015728640
3	2	0.0106168320
4	3	0.0424673280
5	4	0.1114767360
6	5	0.2006681248
7	6	0.2508226660
8	7	0.2149908480
9	8	0.1209323520
10	9	0.0403107840
11	10	0.0060466176

Q.9 > n=30
> p=0.2
> p1=0.88
...ne.

~~Q.8.~~ Q.9. > plot (x, bp, "h")



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



AM 16-12-19

PRACTICAL - 4

* Normal Distribution :-

Normal distribution is an example of continuous ~~non~~ 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 size n.
rnorm (n, μ, σ)

(i) 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) $P(X < K) = 0.4$

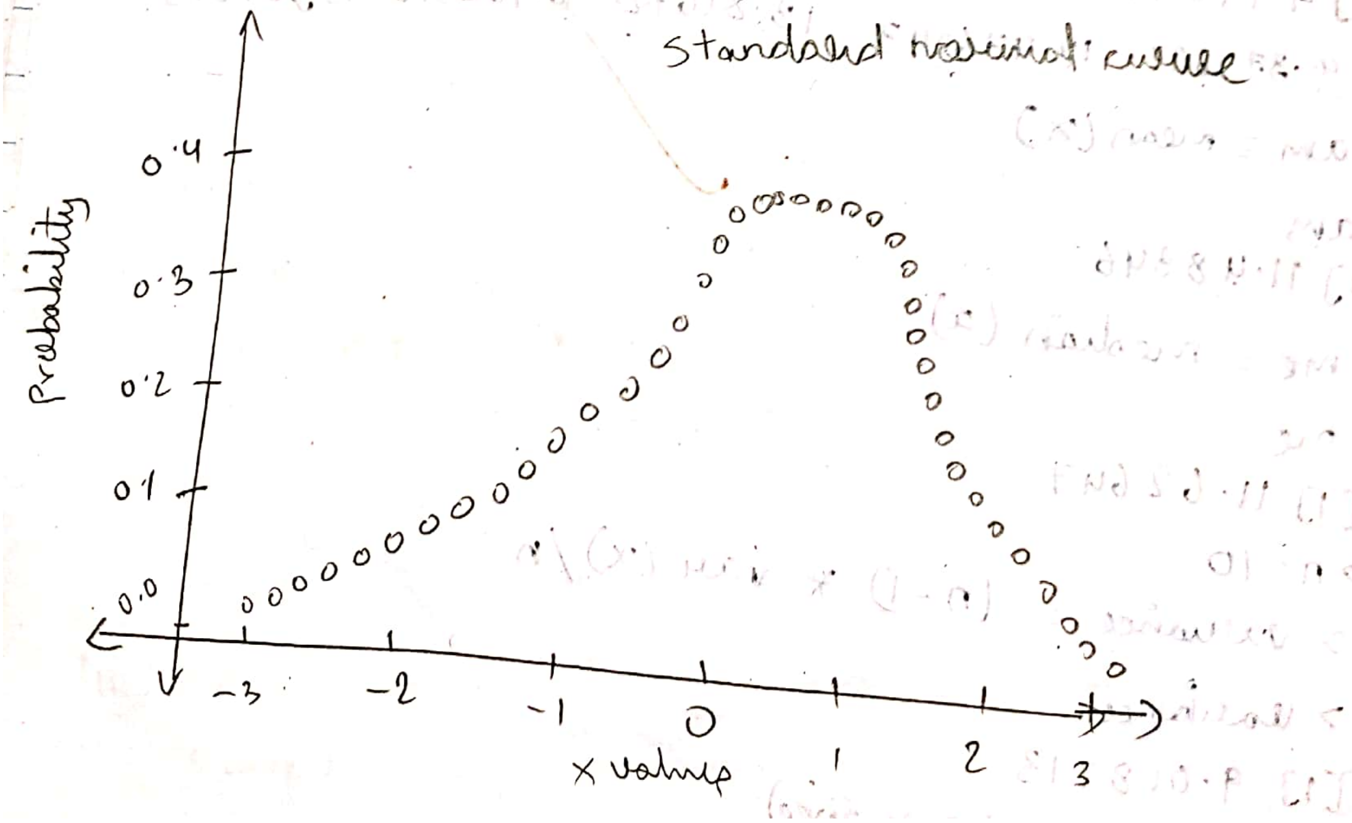
(ii) $X \sim N(100, 36)$ σ = sqrt(36)
i) $P(X \le 110)$
ii) $P(X > 105)$
iii) $P(X \le 92)$
iv) $P(95 \le X \le 110)$
v) $P(X < K) = 0.9$

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

Solution :- i) > p1=pnorm (7,10,2)
[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 = 0.1586553
> 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] 9.193412 11.340685 5.331607 17.160876 11.912250
9.372691 12.436043 13.810725 10.908816 15.357803
> 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] 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 = 0.09121122 >
> 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 sandard normal curve

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

(v) 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
```

AM 6.1.20

Practice Practical

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

- $P(x=7) = \binom{8}{7} (0.6)^7 (0.4)^{8-7}$ .... $(n=8, p=0.6, q=0.4)$

$= {}^8C_7 \times 0.2799 \times 0.4$

$= 8 \times 0.2799 \times 0.4$

$= 0.08957$

2) $P(X \le 3) = P(0) + P(1) + P(2) + P(3)$

$= {}^8C_0 (0.6)^0 (0.4)^8 + {}^8C_1 (0.6)^1 (0.4)^8 + {}^8C_2 (0.6)^2 (0.4)^6$

$+ {}^8C_3 (0.6)^2 (0.4)^5$

$= 1 \times 1 \times 0.00065536 + 8 \times 0.6 \times 0.0016384 + 28 \times$

$0.36 \times 0.004096 + 56 \times 0.216 \times 0.01024$

$= 0.1736704$

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

$= {}^8C_2 (0.6)^2 (0.4)^6 + {}^8C_3 (0.6)^3 (0.4)^5$

$= 28 \times 0.36 \times 0.004096 + 56 \times 0.216 \times 0.01024$

$= 0.04128768 + 0.12386304$

$= 0.16515072$

Scanned with CamScanner

PRACTICAL – 6

* Z & t distribution sums :-

Q.1. 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 and 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("P value is ", pvalue)

P value is 0.07544036>
```
∵ 0.0754 > 0.05

∴ $\mu_0 : \mu = 20$ is acceptable.

Q.2 Test the hypothesis $\mu_0 : \mu = 250$ against $\mu_1 : \mu \neq 250$. A sample of size 100 has a mean of 275 and standard deviation 30. Check the hypothesis at 5% level of significance

```
> n = 100
> mo = 250
```

```
> mx = 275
> sd = 30
> zcal = (mx-m0) / (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>
```

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

Q.3. We want to test the hypothesis $\mu_0 : P = 0.2$ 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% 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
> pval = 2*(1-pnorm(abs(zcal)))
> cat ("P value is ", pvalue)
P value is 0.0001768346>
```

$\because$ Pvalue < 0.01

$\therefore$ $\mu_0 : P = 0.2$ is rejected.

Q.4. In a big city 325 men out of 600 men are found to be self-employed. Does this information support 47 the conclusion that exactly half of the men in the city are self employed

P = 0.5

p = 325/600

n = 600

```
> P = 0.5
> Q = 1-P
> n = 600
> p = 325/600
> 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.

Q.5. Test the hypothesis $\mu_0 : \mu = 50$ against $\mu_1 : \mu \neq 50$

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]

```
> mo = 50
> x = c(50, 49, 52, 44, 45, 48, 46, 45, 49, 45, 40, 47, 55, 54, 4, 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
[1] 49.3333
> variance = (n-1) * var(x) / n
> variance
[1] 30.955
> sd = sqrt(variance)
> sd
[1] 5.56
> zcal = (mx-mo) / (sd / sqrt(n))
> zcal
[1] -0.6562
> pvalue = 2 * (1-pnorm (abs(zcal)))
> cat ("P value is ", pvalue)
P value is 0.511633
```

$\because 0.511 > 0.05$

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

PRACTICAL-7

Aim :- Large Sample Test..

Q.1 Two random sample of size 1,000 & 2,000 are drawn from two population with the standard deviation 2 & 3 respectively. Test the hypothesis that two population means are equal or not at 5% level of significance. Sample means are 67 & 68 respectively.

→ $H_0 : \mu_1 = \mu_2$ against $\mu_1 : \mu_1 \neq \mu_2$

n1 = 1000
n2 = 2000
mx1 = 67
mx2 = 68
sd1 = 2
sd2 = 3

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

0 < 0.05
$\mu_1 : \mu_1 \neq \mu_2$
∴ It is rejected

Q.2. A study of noise level in two hospitals is done following data is calculated. First sample size 84 first sample mean 61.2 first standard 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.

→ >n1 = 84
>n2 = 34
>mx1 = 61.2
>mx2 = 59.4
>sd1 = 7.9
>sd2 = 7.8
>zcal = (mx1 - mx2) / sqrt((sd1^2/n1) + (sd^2/n2))
>cat ("Zcalculated is ", zcal)
Zcalculated is 1.13117 >
>pvalue = 2*(1 - pnorm(abs(zcal)))
>cat ("Pvalue is ", pvalue)
Pvalue is 0.258006 >

0.258 > 0.01
Hence it is accepted

Q.3. 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 = 30. Test: $p_1 = p_2$ against $p_1 \neq p_2$

→ >n1 = 250
>n2 = 200
>p1 = 44/250
>p2 = 30/200
>p = (n1*p1) + (n2*p2) / (n1+n2)

```
>p
[1] 0.1644444
>q=1-p
>q
[1] 0.8355556
>zcal = (p1-p2)/sqrt(p*q*(1/n1+1/n2))
>cat("Z calculated is ", zcal)
Z calculated is  0.7393881>
>pvalue = 2*(1-pnorm(abs(zcal)))
pvalue >cat("Pvalue is ", pvalue)
Pvalue is  0.4596896>
```

0.459 > 0.05
Hence it is accepted

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

H_0, $\mu_1 = \mu_2$ at 5% level of significance

```
-> >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
[1] 0.41
>zcal= (p1-p2)/sqrt(p*q*(1/n1+1/n2))
>cat("Z calculated is ", zcal)
Z calculated is -4.724751>
>pvalue = 2*(1-pnorm(abs(zcal)))
>cat("Pvalue is ", pvalue)
Pvalue is 2.303972 x 10^-6 >
```

0.00000230 < 0.05
Hence it is rejected.

Q.5. Following are the two independent sample from the two populations. Test equality of the two population means at 5% level of significance. $H_0 = \mu_1 : \mu_2$ at 5%

Samples $S_1 = [74, 77, 74, 73, 79, 76, 82, 72, 75, 78, 77, 78, 76, 76]$
$S_2 = [72, 76, 74, 70, 70, 78, 70, 72, 75, 79, 79, 74, 75, 78, 72], 74, 80]$

```
> x1 = c(74,77,74,73,79,76,82,72,75,78,77,78,76,76) 80]
> l1 = length(x1)
> l1
[1] 14
> x2 = c(72,76,74,70,70,78,70,72,75,79,79,74,75,78,72,74,80)
> 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, pvalue = 0.1387
```

> Pvalue is 0.1387 | 0.1387 > 0.05
Hence it is accepted.

27.4.20

PRACTICAL - 8

Aim :- Small Sample Test.

Q.1. The random sample of 15 observations are given by 80, 100, 110, 105, 122, 70, 120, 110, 101, 88, 83, 95, 89, 107, 125. Do this data support the assumption that population mean is 100?

```
→ >x=c(80,100,110,105,122,70,120,110,101,88,83,95,89,107,125)
>a= length(x)
>a
[1] 15
>t.test(x)
```

One Sample t-test

data : x

t = 24.029, df = 14, p-value = 8.819 x e^-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

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

$\therefore H_0 : \mu = 100$ is rejected.

Q.2. Two groups of 10 students scored the following marks.

Group 1 : 18, 22, 21, 17, 20, 17, 23, 20, 22, 21.

Group 2 : 16, 20, 14, 21, 20, 18, 13, 15, 17, 21

Test the hypothesis that there is no significant difference between the marks at 1% level of significance.

$H_0: \mu_1 : \mu_2$ at 10% ~~level of signitance~~

→ >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)

~~One Sample t-test~~

Welch Two sample t-test

data : x & y

t = 2.2573, df = 16.376, p-value = 0.03798

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

∴ 0.03798 > 0.01

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

Q.3. 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.

$H_0: \mu_1 : \mu_2$ at 5% LOS

A = (10, 12, 13, 11, 14)

B = (8, 9, 12, 14, 15, 10, 9)

Is there a significant difference in the efficiency of medicine.

$H_0 : \mu_1 = \mu_2$

→ > m1 = c (10, 12, 13, 14, 11)
> m2 = c (8, 9, 12, 14, 13, 10, 9)
> t.test (m1, m2)

Welch Two Sample t-test

data : m1 & m2
t = 0.80384, df = 9.7694, 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.

Q.4. The weight reducing diet program is conducted & the observation is 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 significant difference in weight against reduced weight.
H_1 : The diet program

> 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 9.416556

sample estimates :

mean of the difference

8.5

∴ 1 > 0.05

∴ Ho is accepted at 5% level of significance.

Q.5. 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.

$H_0: \mu_1 = \mu_2$ at 5% los

> a = c (66, 67, 75, 76, 82, 84, 89, 90, 92)

> b = c (64, 66, 74, 78, 82, 85, 87, 92, 93, 95, 97)

> t.test (a, b)

Welch Two Sample t-test

data : a & b

t = -1.05, df = ~~10.051~~ 17.956, p-value = ~~0.3183~~ 0.5477

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

95 percent confidence interval :

-12.79600 ~~75.66312~~ 7.018223

sample estimates :

mean of x & mean of y

80.11111 ~~147.63636~~ 83.00000

∴ 0.5477 > 0.05

∴ Ho at 5% level of significance is accepted.

Scanned with CamScanner

Q.6. Following are the marks before & after a training program. Test whether the program is effective or not. $H_0: \mu_1 = \mu_2$ at 5% level of significance

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 difference in means is less than 0.

-Inf -2.123361

sample estimates:

mean of the differences

-3.6

∵ 0.0007784 < 0.05

∴ H_0 is rejected ~~at~~ 5% level of significance.

AM
03.02.20

PRACTICAL - 9

Aim :- Large & Small Sample Test.

i) The arithmetic mean of a sample of 100 items from a large population is 52 & if the s.d. is 7. Test the hypothesis that the population mean is 55. Against the alternative more than 55 at 5% L.O.S.

ii) In a big city 350/700 males are found to be smokers. Thus this info. supports that exactly half of the males in the city are smokers? Test at 1% L.O.S.

iii) Thousand articles from a factory A are found to have 2% defectives 1500 articles from a factory B are found to have 1% defective. Test at 5% L.O.S. that the two factories are similar or not.

iv) A sample of size 400 was drawn & a sample mean is 99. Test at 5% L.O.S. that the sample comes from a population with mean 100 & variance 64?

v) The flower stems are selected and the heights are found to be (in cms) 63, 63, 68, 69, 71, 71, 72, Test the hypothesis that the mean height is 66 or not at 1% L.O.S.

vi) Two random samples were drawn from 2 normal population & their values are A = 66, 67, 75, 76, 82, 84, 88, 90, 92
B = 64, 66, 74, 78, 82, 85, 87, 92, 93, 95, 97. Test whether the population have the same variance at 5% L.O.S

$H_0 : \mu = 55$

i) >n = 100
>mx = 52
>m0 = 55
>sd = 7
> zcal = (mx-m0) /(sd /sqrt(n))
>cat ("Z calculated is ",zcal)
Z calculated is -4.285714 >
>pval = 2 * (1- pnorm (abs (zcal)))
>pval
[1] 1.82153e-05

Since, pvalue is 0.00001821
∴ 0.00001821 < 0.05
∴ $H_0 : \mu = 55$ is rejected at 5% L.O.S.

ii) > P=0.5
>p=350/700
>n = 700
>Q = 1-P
>zcal = (p-P)/sqrt (P *Q/n)
>cat (" Z calculated is ", zcal)
Z calculated is 0 >
>pval = 2 * (1- pnorm (abs (zcal)))
>pval
[1] 1

Since, pvalue is 1
∴ 1 > 0.01
∴ Hypothesis is accepted at 1% L.O.S.
$H_0 : \mu :$

iii) > 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))
>zcal
[1] 2.084842
>pval = 2 * (1-pnorm(abs(zcal)))
>pval
[1] 0.03708364
```

∵ 0.03708364 < 0.05

∴ It is rejected at 5% L.O.S.

iv)

```
>n = 400
```

Ho: μ = 100

```
>mx = 99
>m0 = 100
>var = 64
>sd = sqrt(var)
>zcal = (mx-m0)/(sd/sqrt(n))
>cat("Z calculated is ", zcal)
Zcalculated is -2.5 >
>pval = 2 * (1-pnorm(abs(zcal)))
>cat("P value is ", pval)
P value is 0.01241933>
```

∵ 0.01241 < 0.05

∴ It is rejected at 5% L.O.S.
Ho: 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 = 6 , pvalue = 5.522e-09

alternative hypothesis : true mean is not equal to 0

95 percent confidence interval :

64.66479   71.62092

sample estimates :

mean of x

68.14286

$\because$ 0.00000000552 < 0.05

$\therefore$ $H_0 : \mu = 66$ is accepted.

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

AN 10/2/20

F test to compare two variances

data : a & b

F = 0.70686 , num df = 8 , denom df = 10 , p-value : 0.6359

alternative hypothesis : true ratio of variances is not equal to 1

95 per cent confidence interval : 0.188362   3.0360393

sample estimates .

ratio of variances 0.7068567

$\because$ 0.6359 > 0.05 $\therefore$ It is accepted at 5% L.O.S.

3

PRACTICAL NO 1-10

Aim :- ANOBA & CHI square test.

Q.1. Use the following data to test the cleanliness of home & child is independent or not.

Cleanliness of Home

		Clean	Dirty
Cleanliness of Child	Clean	70	50
	Fairly clean	80	20
	Dirty	35	45

Sol^n :- Ho : 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

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

∵ Ho : 0.00000269 < 0.05 ∴ It is rejected

Q.2. Use the following data to find vaccination and a particular disease is independent or not.

55

		Disease	
		Affected	Not affected
vaccination	given	20	30
	Not given	25	35

Solⁿ:- H0 : Disease & 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
X-squared = 0, df = 1, p-value = 1
```

H0 : 1 > 0.05

∴ It is accepted.

Q-3. Perform ANOVA for the following data:

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

H_0: 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))
> 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=9, p-value=0.00183

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

Call:
   aov(formula = values ~ ind, data = d)
Terms:
                        ind       Residuals
Sum of squares       71.06410     18.16667
Deg. of Freedom         3             9
Residual standard error: 1.420746
Estimated effects maybe unbalanced
```

here H_0 : $0.00183 < 0.05$

$\therefore$ It is rejected.

Q.4. The following data gives the life of tyre 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 average life of tyre is same or not.

```
>x1=c(20,23,18,17,18,22,24)
>x2=c(20,23,18,17(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)
[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
>anova=aov(values~ind, data=d)
>anova
Call:
    aov(formula=values~ind, data=d)
Terms:
                      ind        Residuals
Sum of Squares     91.4381       89.0619
Deg. of freedom       3             20
```

Residual standard error : 2.116236
Estimated effects maybe unbalanced.

Here Ho : 0.002349 < 0.05
∴ It is rejected.

Q.5. 1000 students of a college are graded according to the IQ & their economic condⁿ of their home check that is there any association b/w IQ & the economic condⁿ of their home.

		IQ	
		High	Low
Economic Condⁿ.	High	460	140
	Med	330	200
	Low	240	160

Solⁿ. Ho : Association b/w IQ & economic condⁿ.

```
>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-09

$\therefore$ H_0 : $0.00000000236 < 0.05$

$\therefore$ It is rejected.

AM
17-2-20

PRACTICAL NO :- 11

Aim :- Non-parametric test

Q.1. Following are the amounts of sulphur oxide emitted by industries in 20 days. Apply sign Test to test the hypothesis to check if population is 21.5 (17, 15, 20, 29, 19, 18, 22, 25, 27, 9, 24, 20, 17, 6, 24, 14, 15, 23, 24, 26)

Ho: population median is 21.5

```
> 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
> pv = pbinom(sp,n,0.5)
> pv
[1] 0.4119015
```

∵ 0.4119015 > 0.05

∴ Ho: population median is 21.5 is accepted.

Q.2. Follow

612, 619, 631, 628, 643, 640, 655, 649, 670, 663. Apply sign to test the hypothesis to check if population median is 625 against alternative i.e. its greater than 625 at 1% l.o.s.

Note: If alternative is greater, pv = pbinom(sn, n, 0.5)

Ho: population median is 625.

```
>x=c(612,619,631,628,643,640,666,649,670,663)
>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] 0.9546875
```

$\because$ $0.9546875 > 0.01$

$\therefore$ Ho: population median at 625 is accepted.

Q.3. Ten observations are (36, 32, 21, 30, 24, 25, 20, 22, 20, 18). Using sign test check the hypothesis that population median is 25 against it is less than 25 at 5% L.O.S.

Ho: population median is 25.

```
>x=c(36,32,21,30,24,25,20,22,20,18)
>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
```

$\because$ $0.2539063 > 0.05$

$\therefore$ Ho: population median at 25 is accepted.

Q.4. 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 population median is 60 against the alternative, it is greater than 60 at 5% L.O.S.

→ Ho : population median is 60.

```
> x = c(63,65,60,89,61,71,58,51,69,62,63,39,72,65)
> 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

∵ 0.06186 > 0.05

∴ Ho : population median at 60 is accepted

Q.5. 15, 17, 24, 25, 20, 21, 32, 28, 12, 25, 24, 26 use WSRT to test the Hypothesis that population median is 20 against the alternative, is less than ~~20~~ at 5% L.O.S.

→ Ho : population median is 20.

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

Wilcoxon signed rank-test with continuity correction

data : x

V = 48.5, p-value = 0.9232

Alternative hypothesis : true location is less than 20

∵ 0.9232 > 0.05

∴ Ho : population median at 20 is accepted.

8·6, 20, 25, 27, 30, 18. Test the hypothesis that the population median is 25 against the alternative is not 25 at 5% L.O.S. Ho : population median is 25.

```
->> x = c(20,25,27,30,18)
> wilcox.test(x, at = "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 25

∵ 0.7127 > 0.05

∴ Ho : population median at 25 is accepted.

Abh
9.3.20

Scanned with CamScanner

INDEX

Scanned with CamScanner