PRACTICAL - 01

- R is a software for statistical analysis and data computing.

- It is an effective data handling software and outcome storage is possible.

- It is capable of graphical display

- R is a free software.

Q1] Solve the following

Code:-

1. $4+6+8\div 2-5$

```
1] 4+6+8/2
  [1] 14
```

2. $2^2+|-3|+\sqrt{45}$

```
2] 2^2 +abs(-3) + sqrt (45)
  [1] 13.7082
```

3. $5^3+7\times5\times8+46/5$

```
3] 5^3 +7*5*8+46/5
  [1] 414.2
```

4. $\sqrt{4^2+5\times3+7/6}$

```
4] sqrt (4^2+5*3+7/6)
  [1] 5.671567
```

5. Round off $46+7+9\times8$

```
5] 46/7+9*8
  [1] 78.57143
```

Scanned with CamScanner

Q2] Solve the following

① c(2,3,5,7) * 2

→ [1] 4 6 10 14

② c(2,3,5,7) * c(2,3)

→ [1] 4 9 10 21

③ c(2,3,5,7) * c(2,3,6,2)

→ [1] 4 9 30 14

④ c(1,6,2,3) * c(-2,-3,-4,-1)

→ [1] -2 -18 -8 -3

⑤ c(2,3,5,7)^2

→ [1] 4 9 25 49

⑥ c(4,6,8,9,4,5)^c(1,2,3)

→ [1] 4 36 512 9 16 125

⑦ c(6,2,7,5)/c(4,5)

Q3] Solve the following

$x = 20$, $y = 30$, $z = 2$.

Find 1) $x^2 + y^3 + z$

2) $\sqrt{x^2 + y}$

3) $x^2 + y^2$

i) > a = x^2 + y^3 + z

> a

[1] 27402

ii) > b = sqrt(x^2 + y)

> b

[1] 20.73644

iii) ~~x~~ = x^2 + y^2

```
> c
[1] 1300
```

Q4] Convert the data in Matrix Form :-

1	5
2	6
3	7
4	8

```
> x = matrix(nrow = 4, ncol = 2, data = c(1,2,3,4,5,6,7,8))
> x
     [,1]  [,2]
[1,]   1     5
[2,]   2     6
[3,]   3     7
[4,]   4     8
```

Q5] Find x+y and 2x+3y where

$$x = \begin{bmatrix} 4 & -2 & 6 \\ 7 & 0 & 7 \\ 9 & -5 & 3 \end{bmatrix} \qquad y = \begin{bmatrix} 10 & -5 & 7 \\ 12 & -4 & 9 \\ 15 & -6 & 5 \end{bmatrix}$$

```
> x = matrix(nrow = 3, ncol = 3, data = c(4,7,9,-2,0,-5,6,7,3))
> x
     [,1]  [,2]  [,3]
[1,]   4    -2     6
[2,]   7     0     7
[3,]   9    -5     3

> y = matrix(nrow=3, ncol=3, data = c(10,12,15,-5,-4,-6,7,9,5))
> y
     [,1]  [,2]  [,3]
[1,]  10    -5     7
[2,]  12    -4     9
[3,]  15    -6     5
```

```
> 2^x + 3^y
     [,1]  [,2]  [,3]
[1,]  38   -19    33
[2,]  50   -12    41
[3,]  63   -28    21
```

```
> x+y
     [,1]  [,2]  [,3]
[1,]  14   -7    13
[2,]  19   -4    16
[3,]  24   -11    8
```

Q6] Marks of Statistics of Computer Science student

59, 20, 35, 24, 46, 56, 55, 45, 27, 22, 27, 58, 54, 40, 50, 32, 36, 29, 35, 39

```
> x = c(59,20,35,24,46,56,55,45,27,22,27,58,54,
        40,50,32,36,29,35,39)
> breaks = seq(20,60,5)
> a = cut(x,breaks, right=FALSE)
> b= table(a)
> c= transform(b)
```

> x

	a	freq
1	[20,25)	3
2	[25,30)	2
3	[30,35)	1
4	[35,40)	4
5	[40,45)	1
6	[45,50)	3
7	[50,55)	2
8	[55,60)	4

AM

PRACTICAL - 2

Aim: Probability Distribution.

Q] Check whether the following are pmf or not.

1)

x	$p(x)$
0	0.1
1	0.2
2	-0.5
3	0.4
4	0.3
5	0.5

$\Rightarrow$ Since, $P(2) = -0.5$, cannot be pmf because $p(x) \geq 0 \ \forall x$.

ii)

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

$\Rightarrow$ >prob=c(0.2,0.2,0.3,0.2,0.2)

> Sum (prob)

[1] 1.1

It cannot be p.mf. as in p.mf., $\Sigma P(x)=1$

iii)

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

>prob=c(0.2,0.2,0.35,0.15,0.1)

> sum (prob)

[1] 1

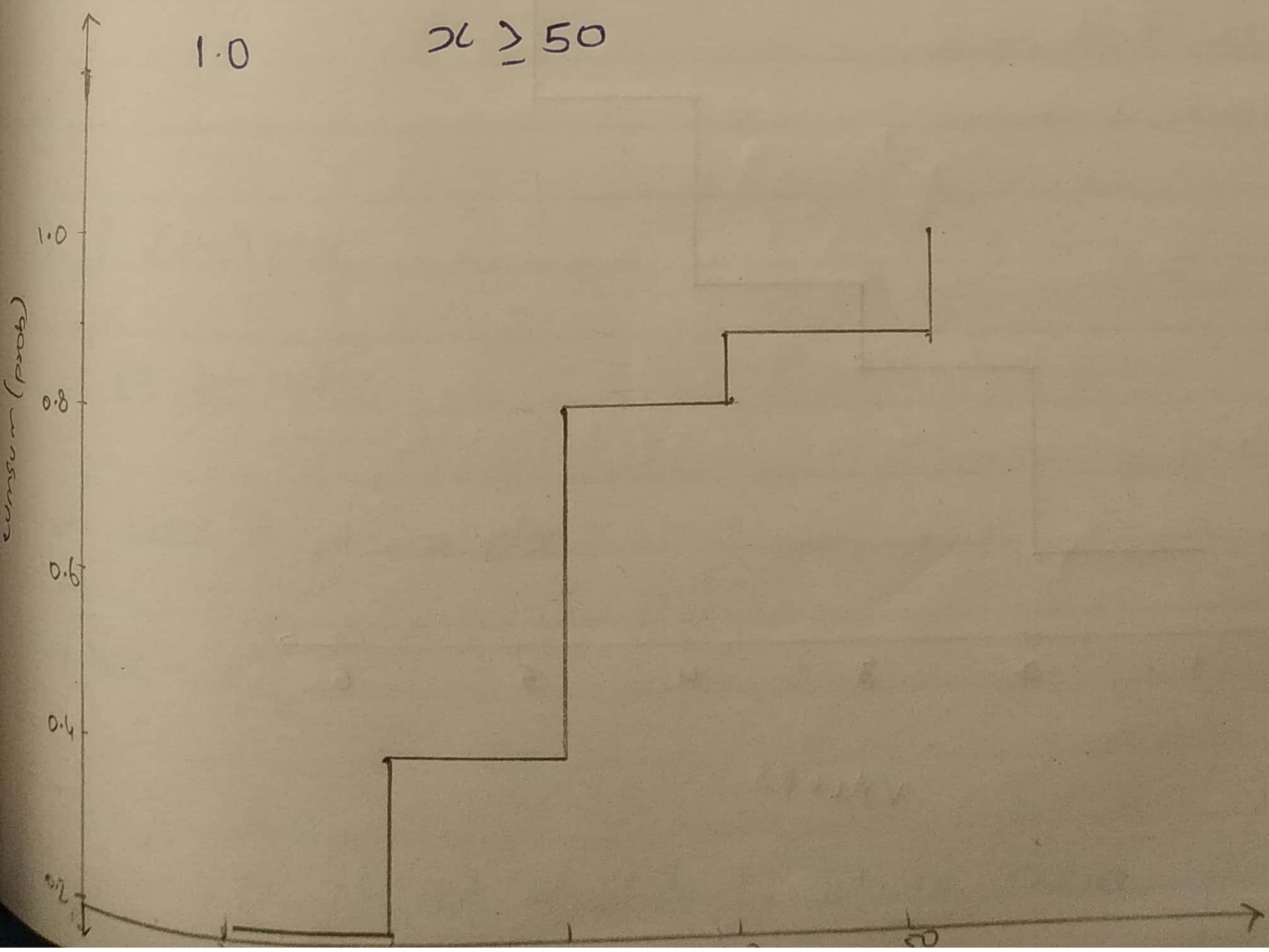
It is a p.m.f as $\Sigma p(x)=1$ & $P(x) > 0 \ \forall x$

Q2]
i] Find c.d.f for the following p.m.f and sketch the graph.

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

```
> prob = c(0.2, 0.2, 0.35, 0.15, 0.1)
> sum(prob)
> cumsum(prob)
[1] 0.20  0.40  0.75  0.90  1
```

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

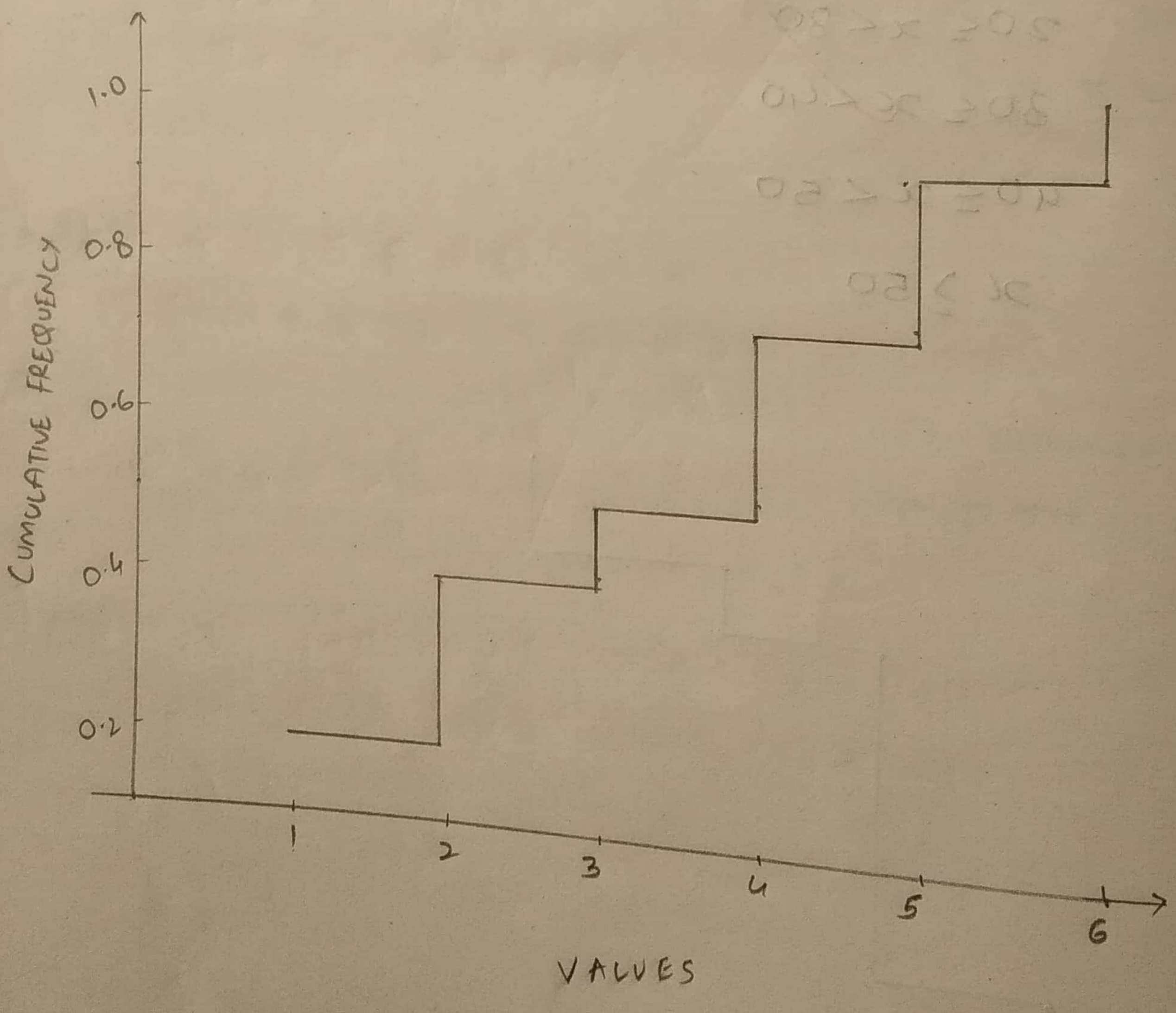


Q2]
ii]

x	1	2	3	4	5	6
$P(x)$	0.15	0.25	0.1	0.20	0.2	0.1

```
> prob=c(0.15,0.25,0.1,0.2,0.2,0.1)
> sum(prob)
> cumsum(prob)
[1] 0.15  0.40  0.50  0.70  0.90  1.0
> x=c(1,2,3,4,5,6)
> plot(x, cumsum(prob), "s", xlab="VALUES", ylab"CUMU
                    FREQ", main="C.D.F")
```

C.D.F.



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

Q3] Check whether the following is p.d.f. or not.

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

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

Solⁿ :-

i) $\int_0^1 f(x)\,dx$

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

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

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

$= \underline{\underline{2}}$

$\therefore$ It is not equal to 1, It is not a p.d.f.

Scanned with CamScanner

ii) $\int_0^1 f(x) * dx$

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

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

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

$= \underline{\underline{1}}$

Since, integration is equal to 1,
It is a p.d.f.

A.V. 5/7/19

Sol^n :-

1) > dbinom(10,100,0.1)
[1] 0.1318653

2] 1) exactly 4 correct answers
> dbinom(4,12,0.2)
[1] 0.1328756

2) Atmost 4 correct answers
> pbinom(4,12,0.2)
[1] 0.9274445

3) More than 5 correct answers
> 1-pbinom(5,12,0.2)
[1] 0.01940528

3] dbinom(0:5,5,0.1)
[1] 0.59049 0.32805 0.07290 0.00810 0.0004

4]
i) P(x=5) => dbinom(5,12,0.25) [1] 0.1032416
ii) P(x≤5) => pbinom(5,12,0.25) [1] 0.9455978
iii) P(x>7) => 1-pbinom(7,12,0.25) [1] 0.00278151
iv) P(5<x<7) = dbinom(6,12,0.25) [1] 0.04014945

PRACTICAL - 3

TOPIC: BINOMIAL DISTRIBUTION

COMMAND:

i) $P(X=x)$ = dbinom (x, n, p)

ii) $P(X \le x)$ = pbinom (x, n, p) (atleast)

iii) $P(X > x)$ = 1 - pbinom (x, n, p) (atmost)

iv) If x is unknown and $P_1 = P(X \le x)$

→ qbinom (p_1, n, p)

Q1] Find the probability of exactly 10 sucess in 100 ~~trial~~ trials, with $P = 0.1$.

Q2] Suppose there ~~as~~ are 12 MCQ. Each question has 5 options out of which one is correct find the probability of having,

1) exactly 4 correct answer.

2) Atmost 4 correct answer.

3) More than 5 correct answer.

Q3] Find the complete distribution when $n=5$, $p=0.1$.

Q4] $n=12$, $p=0.25$ find the following probability,

i) $P(X=5)$ ii) $P(X \le 5)$ iii) $P(X>7)$ iv) $P(5<X<7)$

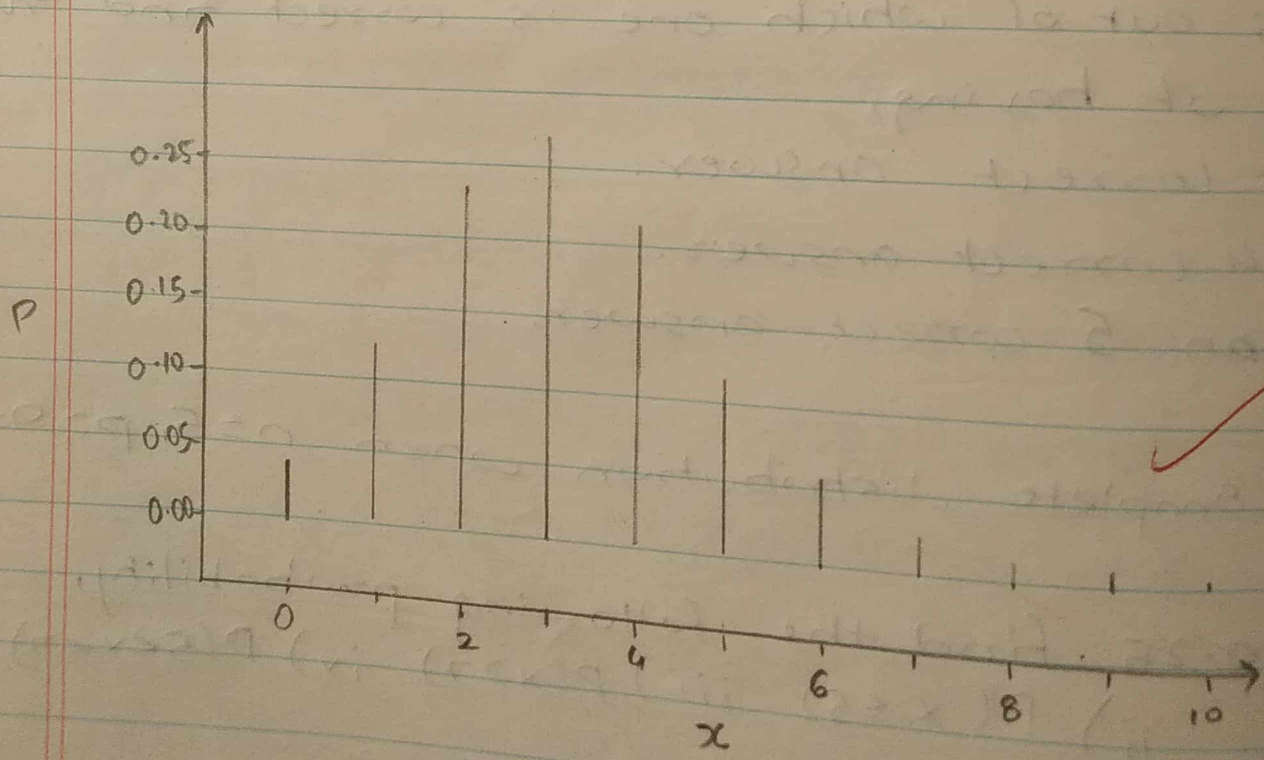
Q5] Probability of a salesman making a sale to a customer 0.15. Find the probability of

i) no sale out of 10 customer.

ii) more than 3 sale out of 20 customer.

Q6] A salesman has a 20% probability of making a sale out of 30 customers what min. no. of sales he can make with 88% probability.

Q7] X follows binomial distribution with $n=10$, $p=0.3$, plot the graph of p.m.f and c.d.f.



5]
i) > dbinom (0, 10, 0.15)

[1] 0.1968744

ii] > 1-pbinom (3, 20, 0.15)

[1] 0.3522748

Q6] > qbinom (0.88, 30, 0.2)

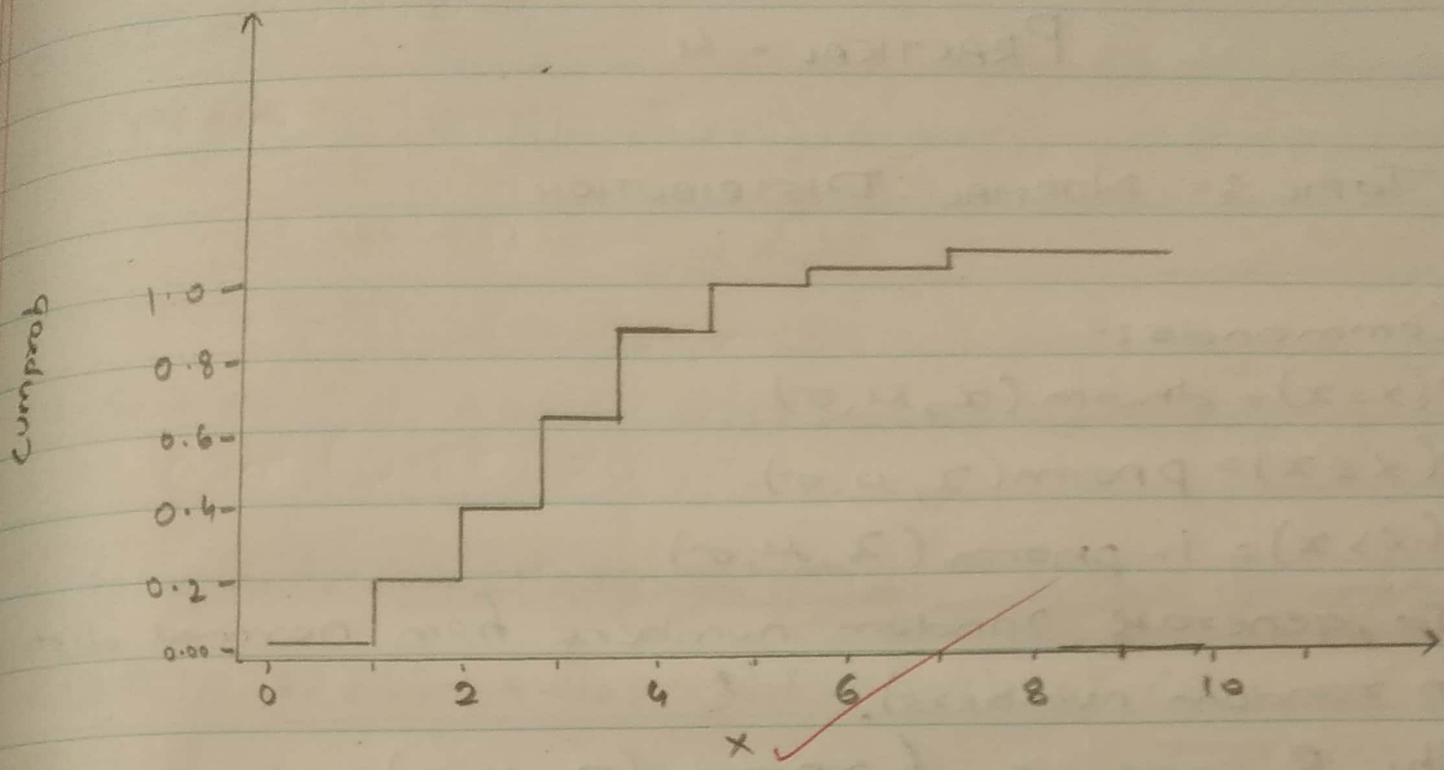
[1] 9.

Q7]

```
>n=10
>p=0.3
>x=0:n
>prob= dbinom(x,n,p)
>cumprob=pbinom( x,n,p)
>d= data.frame("x values"= x, "Probability"= prob)
>print (d)
```

X values	Probability
0	0.02824
1	0.121060
2	0.23334
3	0.26682
4	0.20012
5	0.10291
6	0.03675
7	0.009001
8	0.0014467
9	0.0001377
10	0.0000

plot (x, prob, "h")
plot (x, cumprob, "s")



An
9/12/19

1/20

PRACTICAL - 4

TOPIC :- NORMAL DISTRIBUTION

Commands:-

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) To generate random numbers from normal distribution (n random numbers),
the R code is, { rnorm (n, μ, σ)

Q1] A random variable x follows normal distribution with mean $= \mu = 12$ and standard deviation $= \sigma = 3$, Find, 1) $P(X \le 15)$ 2) $P(10 \le X \le 13)$ 3) $P(X > 14)$
4) Generate 5 observations (random numbers).

⟹

1) $\mu = 12, \sigma = 3$

```
> p1 = pnorm (15,12,3)
> cat ("P(x≤15) = ", p1)
[1] 0.8413447
```

2)

```
> p2= pnorm (13,12,3) - pnorm (10,12,3)
> p2
[1] 0.3780661
> cat ("P(10≤x≤13) = ", p2)
```

> 1
[1]
> c
4)
>
[1]
Q2]
1
5
⟹
>
> p
[1
2)
>
>
[1
3)
>
>
[1]

3)
```
>p3 = 1-pnorm(14,12,3)
[1] 0.2524925
>cat("P(x>14) =", p3)
```

4)
```
> rnorm(5, 12, 3)
[1] 10.512180   11.944169   15.115499   10.366201   8.408162
```

Q2] x follows normal distribution with $\mu = 10, \sigma = 2$, find
1) $P(x \leq 7)$ 2) $P(5 < x < 12)$ 3) $P(x > 12)$ 4) Generate 10 observations
5) Find k such that $P(x < k) = 0.4$

⇒

1) $P(x \leq 7)$
```
>p1= pnorm(7,10,2)
>p1
[1] 0.0668072
```

2) $P(5 < x < 12)$
```
>p2= pnorm(12,10,2) - pnorm(5,10,2)
>p2
[1] 0.8351351
```

3) $P(x > 12)$
```
>p3= 1-pnorm(12, 10, 2)
>p3
[1] 0.1586553
```

8

4] Generate 10 observations

>rnorm(10, 10,2)

[1] 7.324489 8.308746 8.394823 11.790873 9.487554

10.366571 12.905423

8.835328 10.756449 13.338475

5] qnorm(0.4,10,2)

[1] 9.493306

Q3] Generate 5 random no's from normal distribution and mean = μ = 15 , S.D = 4 . Find sample mean, median, S.D and print it.

Q4] X follows N(30,100)

$\mu = 30$, $\sigma = 10$

find 1) $P(X \leq 40)$ 2) $P(X>35)$ 3) $P(25<X<35)$ 4) Find K such that $P(X<k)$=0.

=> 1) $P(X \leq 40)$

>P1 = pnorm(40, 30,10)

>P1

[1] 0.8413447

2) $P(X>35)$

>P2=1-Pnorm(35,30,10)

>P2

[1] 0.3085375

3) P(25 < x < 35)
```
>p3 = pnorm(35, 30, 10) - pnorm(25, 30, 10)
>p3
[1] 0.3829249
```

4]
```
qnorm(0.6, 30, 10)
[1] 32.53347
```

Q3] Sol^n:-

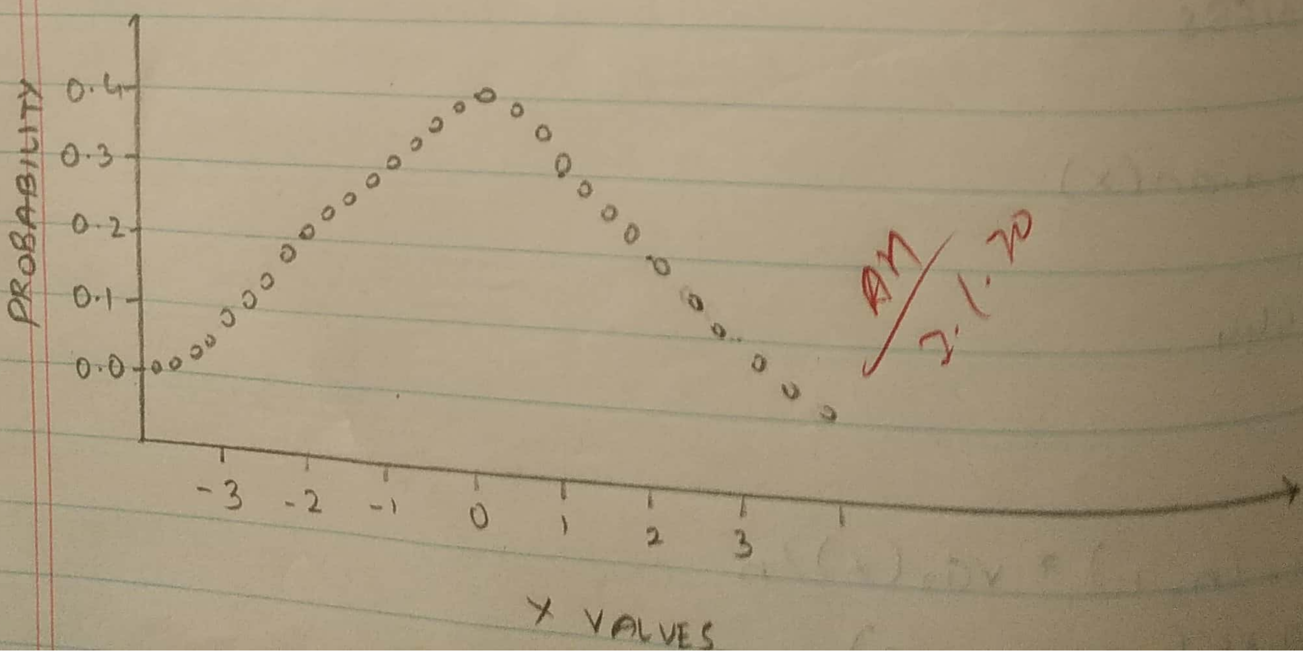
$n = 5, \mu = 15, \sigma = 4$
```
x=rnorm(5, 15, 4)
```
~~[1] 17.154528 12.79657 12.168378 15.146658 21.97~~
```
>x
[1] 15.27357 20.67309 16.30102 15.58444 13.74727

>am=mean(x)
>am
[1] 16.31588

>me=median(x)
>me
[1] 15.58444

>n=5
>variance=(n-1)*var(x)/n
>sd = sqrt(variance)
>sd
[1] 2.332582
```

>cat("samplemean =", am)
>cat("sample median =", me)
>cat("sample s.d =", sd)
[1] Sample mean = 16.31588
[1] sample median= 15.58444
[1] sample variance = 2.332583

Q5] Plot the standard normal graph.

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

STANDARD NORMAL GRAPH



AM
7.1.20

PRACTICAL - 05

AIM :- Normal and T-test

Q1] Test the Hypothesis

$H_0 : \mu = 15$

$H_1 : \mu \neq 15$

Random sample of 400 size is drawn & it is calculated. The sample mean is 14 & standard deviation is 3. ~~Hypothesis~~ at 5% level of significance.

```
-> > mo = 15, mx = 14 , sd = 3, n = 400
> zcal = (mx-mo) / (sd) (sqrt (n)))
> cat ("calculated value at z is = ", zcal)
Calculated value at z is -6.66667
> pvalue = 2* (1-pnorm (abs(zcal)))
> pvalue
[1] 2.616796 e-11
```

$\because$ value of p is less than 0.05,

we reject $H_0 : \mu = 15$.

Q] Test the Hypothesis

$H_0 : \mu = 10$

$H_1 : \mu \neq 10$

Random sample of size 400 is drawn with sample mean 10.2 and s.d = 2.5. ~~These~~ Test the hypothesis at 5% level of significance.

$\rightarrow$ > mo = 10 , mx = ~~14~~ 10.2, s.d = 2.5, n = 400

> zcal = (mx - mo) / (sd / sqrt(n)))

> cat(" calculated value at z is ", zcal)

calculated value at z is = 1.7778

> pvalue

[1] 0.07544

∴ value of p is greater then 0.01

we accept $H_0 : \mu = 10$.

Q] Test the Hypothesis of proportion of smokers in our college is collected & proportion is calculated as 0.125. Test the Hypothesis at 5% level of significance. (sample size is 400)

$\rightarrow$ > P = 0.2 , p = 0.125 , n = 400 , Q = 1 - P

> zcal = (p - P) / (sqrt (P*Q/n))

> zcal

[1] -3.75

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

> pvalue

[1] 0.00017683

∴ value of p is less than 0.05 we reject $H_0 = 0.2$.

Q] Last year farmers lost 20% of their crop. A random sample of 60 field ie are collected and it is found that 9 field crops are insect polluted. Test the Hypothesis at 1% level of significance.

P=0.2 , p = 9/60 , n=60 , Q=1-P

```
> zcal = (p-P)/(sqrt(P*Q/n))
> pvalue = ((p-P))/sqrt(P*Q/n))
> pvalue
[1] 0.3329
```

Ans → since pvalue > 0.01, we accept $H_0 = 0.2$

Q] Test the hypothesis $H_0 : \mu = 12.5$, from the following sample and 5% level of significance.

```
x=c(1.25,11.97,12.15, 12.08, 12.31,12.28,11.94,11.89,12.16,12.04)
n=length(x)
n
[1] 10
> mx = mean(x)
mx
[1] 11.007
variance = (n-1)*var(x)/n
variance
[1] 10.594912
sd=sqrt(variance)
sd
```

```
> mo =12.5
> t=(mx-mo)/(sd/sqrt(n))
> pvalue = 12*(1-pnorm(abs(t)))
> po
```

PRACTICAL - 06

Aim : T-test

1) Let the population mean (the amount spent per customer in a restaurant) is 250 a sample of 100 customers selected the sample mean is calculated as 275 and S.D. 30, test the hypothesis that population mean is 250 or not at 5% level of significance.

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

Solution :-

1) mx = 275 , mo = 250, SD = 30 , n = 100

→ $H_0 : \mu = 275$ against $H_1 : \mu \neq 275$

>zcal = (275-250)/(30/sqrt(100))

>zcal

[1] 8.3333

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

>pvalue

[1] 0

∵ pvalue = 0 < 0.05

we reject H_0 at 5% level of significance.

Sol 2]

→ $H_0 : P = 0.8$ against $H_1 : P \neq 0.8$

```
> P=0.8
> Q=1-P
> p= 750/1000
> n=1000
> zcal = (p-P)/sqrt(P*Q/n))
> zcal
[1] -3.952847
> pvalue = 2 * (1-pnorm(abs(zcal))
> pvalue
[1] 7.72268e-05
```

∵ pvalue < 0.1

We reject value of H_0 at 1% level of significance.

Q3] Two random sample of size 1000 and 2000 are drawn from two population with the same S.D = 2.5. The sample means are 67.5 and 68 respectively. Test the hypothesis $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$ at 5% level of significance.

Q4] A study of noise level in two hospital is given below: Test the claim that the two hospital have same level of noise at 1% level of significance

	Hospital A	Hospital B
size	84	34
Means	61.2	59.4
S.D	7.9	7.5

Q5] In a sample of 600 students in our college, 400 use blue ink. In another college from sample of 900 students, 450 use blue ink. Test the Hypothesis that the proportion of students using blue ink in two colleges are equal or not at 1% level of significance.

Sol^n:-

Q3] >n1 = 1000
>n2 = 2000
>mx1 = 67.5
>mx2 = 68
>s.d1 = 2.5
>s.d2 = 2.5
>zcal = (mx1 - mx2)/sqrt(sd1^2/n1) + (sd2^2/n2)

>zcal
[1] -5.1639
>cat("zcalculated is = ", zcal)

[1] zcalculated is = -5.1639

>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 2.4175e-7
∵ pvalue < 0.05
we reject Ho at 5% level of significance.

Q4] >n1 = 84
>n2 = 34
>mx1 = 61.2
>mx2 = 59.4
>s.d1 = 7.9
>s.d2 = 7.5
>zcal = (mx1 - mx2)/sqrt(sd1^2/n1) + (sd2^2/n2)
>zcal
[1] 1.162

>pvalue = 2* (1-pnorm(abs(zcal)))
>pvalue
[1] 0.245

∵ pvalue > 0.01
~~we~~ we accept Ho at 1% level of significance.

Q5] Ho : $P_1 = P_2$ ag $H_1 : \mu_1 \neq \mu_2$
>n1=600
>n2=900
> p1 =400/600
> p2 = 450/900
>p=(n1*p1+n2*p2)/(n1+n2)
>q=1-p
>zcal = (p1-p2)/sqrt (p*q*(1/n1 + 1/n2))
>zcal
[1] 6.3814
>pvalue = 2* (1-pnorm(abs(zcal)))
>pvalue
[1] 1.75322e-10

∵ pvalue < 0.01
we reject Ho at 1% level of significance

$H_0: P_1 = P_2$ or $H_1: P_1 \neq P_2$

Q6] For sample size, $n_1 \in 200$

```
> n1 = 200
> n2 = 200
> p1 = 44/200
> p2 = 30/200
> p = (n1 * p1 + n2 * p2) / (n1+n2)
> q = 1-p
> zcal = (p1-p2)/sqrt (p*q*(1/n1+1/n2))
> zcal
[1] 1.8027
> pvalue = 2* (1-pnorm (abs(zcal)))
> pvalue
[1] 0.07142
```

$\because$ pvalue > 0.05 $\therefore$ It accepts H_0 at 5% level of significance

AM
27-01-20

PRACTICAL-7

Topic: Small Sample Test

1) The marks of 10 students are given by 63, 63, 66, 67, 68, 69, 70, 70, 71, 72. Test the hypothesis that the sample comes from opposition that the sample with average 66.

$H_0 . \mu = 66$

```
> x = c(63,63,66,67,68,69,70,70,71,72)
> t = test (x)
    One sample t-test
data = x
t = 68.319, df = 9, p-value = 1.558e-13
alternative hypothesis: mean is not equal to 0.
95 percent confidence interval :
65.65171      70.14829
Sample estimates:
mean of x
   67.9
```

As $1.558e^{-13} < 0.05$, we reject it.

```
> pvalue = 1.558e-13
> if (pvalue > 0.05) { cat ("accept Ho") } else { cat
                                      ("reject Ho") }
reject H.>
```

2] Two groups of students score the following marks. Test the hypothesis that there is no significant difference between two groups.

Group 1 - 18, 22, 21, 17, 20, 17, 23, 20, 22, 21

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

Ho : There is no difference between two groups.

```
> x=c(18,22,21,17,20,17,23,20,22,21)
> y=c(16,20,14,21,20,18,13,15,17,21)
> t.test(x,y)
        Welch Two Sample t-test
data : x and y
t = 2.2573 , df = 16.376 , p-value = 0.03798
alterative 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
> pvalue = 0.03798
> if(pvalue >0.05) { cat(" accept Ho")}else{cat("reject Ho")}
> reject Ho
```

3] Sales data of 6 shops before and after one special campain are given below.

Before : 53, 28, 31, 48, 50, 42

After : 58, 29, 30, 55, 56, 45

Test the hypothesis that campaign in effective or not.

Ho: There is no significant difference of sales before & after the campaign.

Paired t-test

```
>x=c(53,28,31,48,50,42)
>y=c(58,29,30,55,56,45)
>t.test(x,y,paired=T,alternative="greater")
```

Paired t-test

data : x and y

t = -2.7815, df = 5, p-value = 0.9806

alternative hypothesis : true difference in means is greater than 0.

95 percent confidence interval :

-6.035547 Inf

Sample estimates :

mean of the differences

-3.5

```
>Pvalue = 0.9806
>if(pvalue>0.05){cat("accept Ho")}else{cat("reject Ho")}
accept Ho
```

4] Two medicines are applied to ~~typh~~ ... patient

Group1 : 10, 12, 13, 11, 14

Group2 : 8, 9, 12, 14, 15, 10, 9

Is there any significant difference betn two medicine

Ans: ~~H₀~~: There is no significant difference.

H₀:

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

Welch Two sample t-test

data : x and y

t=0.80384 , df = 9.7594, pvalue = 0.4406

alternative hypothesis : true differences 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

~~pvalue = 0.9806~~

```
> pvalue = 0.4406
> if (pvalue > 0.05) {Cat ("accept H0")} else {"(reject H0")}
accept H0
```

5) The followings are the weight before and after a diet program. Is the diet program effective?

Before : 120, 125, 115, 130, 123, 119
After : 100, 114, 95, 90, 115, 99

Paired t-test

H_0 : There is not significant difference.

```
> x=c(120,115,125,130,123,119)
> y=c(100,114,95,90,115,99)
> t.test(x, y, paired = T, alternative = "less")
```

Paired t-test

data : x and y
t = 4.3458 , df = 5, p-value = 0.9963
alternative hypothesis : true difference in means is less than 0.
95 percent confidence interval:
-Inf 29.0295.
sample estimates:
mean of differences
19.8333

AM 6.2.20

```
> pvalue = 0.9963
> if(pvalue>0.05){cat("accept H0")} else{"(reject H0")
}
```

accept H_0

PRACTICAL - 8

Topic : Large and Small Sample Tests.

Q1] $H_0 : \mu = 55$ $\qquad H_1 : \mu \neq 55$

```
> mx = 55
> mo = 52
> n = 100
> sd = 7
> zcal = (mx-mo)/ (sd/sqrt(n))
> zcal
[1] 4.285714
> pval = 2*(1-pnorm(abs(zcal)))
> pval
[1] 1.82153e-05
```

(pval is ~~accepted~~ rejected < 0.05)

Q2] $H_0 : P = 0.5$ $\qquad H_1 : P \neq 0.5$

```
> P = 0.5
> p = 350/700
> n = 700
> q = 1-p
> zcal = (p-P)/sqrt(P*(q/n))
> zcal
[1] 0
> pval = 2* (1-pnorm(abs(zcal)))
> pval
[1] 1
```

(pval is

Ho : $P_1 = P_2$ $H_1 : P_1 \neq P_2$

83]

```
>n1=1000
>n2=1500
>p1=0.02
>p2= 0.01
>p= (n1* p1+ n2 *p2)/(n1+n2)
>q= 1-p
>p
[1] 0.5
>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
```

(pval is rejected <0.05)

Ho : $\mu = 100$ against $H_1 : \mu \neq 100$

```
>mx=100
>mo= 99
>n = 400
>sd =8
>zcal=(mx-mo)/(sd/sqrt(n))
>zcal
[1] 2.5
> pval = 2 * (1-pnorm(abs(zcal)))
>pval
[1] ...
```

... <0.05)

Q5] $H_0 : \mu = 66$ against $H_1 : \mu \neq 66$

```
> x = c(63, 63, 68, 69, 71, 71, 72)
> t.test (x)
        one Sample t-test

data : x
t = 47.94, df = 6, pvalue = 5.2522e-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.
```

> pval < 0.01 , Hence reject H_0.

Q6] $H_0 : \sigma_1 = \sigma_2$ against $H_1 : \sigma_1 \neq \sigma_2$

```
> x = c(66, 67, 75, 76, 82, 84, 88, 90, 92)
> y = c(64, 66, 74, 78, 82, 85, 87, 92, 93, 95, 97)
> var.test (x, y)
        F test to compare two variances
data : x and y
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 percent confidence interval:
   0.1833662      3.0360393
sample estimates:
ratio of variances : 0.7068567
```

Q7] Ho : μ_1 = 1200 against H1 : $\mu_1 \neq$ 1200

```
>n = 100
>mx = ~~1150~~ 1200
>mo = ~~1264~~ 1150
>sd = 125
>zcal = (mx - mo) / (sd / (sqrt (n))
>pval = (1 - pnorm(abs (zcal))) *2
>zcal
[1] 4
>pvalue
[1] 6.334248e-05
```

We reject the pvalue at Ho : μ_1 = 100

Q8] Ho : p1 = p2 against H1 : P1 $\neq$ P2

```
>n1 = 200
>n2 = 300
>P1 = 44/200
>P2 = 56/300
>P = (n1* p1 + n2*p2)/(n1+n2)
>q = 1-p
>zcal = (p1 - p2) / sqrt(p*q *(1/n1+1/n2))
>pval = (1-pnorm (abs(zcal))) *2
>zcal
[1] 0.9128709
>pvalue
[1] 0.3613104
>
```

We accept the pvalue at 1% level of significance.

Practical - 9

TOPIC : Chi square test & ANOVA

Q1] Use the following data to test whether the condition of home & condition of child are independent or not.

cond. child	cond. home clean	dirty
clean	70	50
fairly clean	80	20
Dirty	35	45

H_0: Condition of Home & child are independent

```
>x = c(70,80,35, 50,20,45)
>m = 3
>n = 2
> y = matrix (x, nrow = m, ncol = n)
>y
      [,1]   [,2]
[1,]   70     50
[2,]   80     20
[3,]   35     45
```

> pv = chisq. test(y)
> pv
pearson's chi squared test

data : y
x-squared = 25.646
df = 2
p-value = 2.698e-06
They are dependent.
∴ pvalue is less than 0.05 we reject the hypothesis at 5% level of significance.

Q2] Test the hypothesis that vaccination & disease are independent or not.

Vacine

Disease	Affected	Not Affected
Affected	70	46
Non Affected	35	~~35~~ 37

> pv = chisq. test(y)
> pv

Pearson's chi squared test with continuity

data : y
x-square = 2.0275
df = 1
P-value = 0.1545
∴ pvalue is ~~more~~ than 0.05 we accept the hypothesis at 5% level of Significance.

Q3] Perform a ANOVA for the following data.

Type	Observations
A	50, 52
B	53, 55, 53
C	60, 58, 57, 56
D	52, 54, 54, 55

H_0: The mean's are equal for A,B,C,D.

```
> x1 = c(50,52)
> x2 = c(53,55,53)
> x3 = c(60,58,57,56)
> x4 = c(52,54,54,55)
> d = stack (list(b1=x1, b2=x2, b3=x3, b4=x4))
> names(d)
[1] "values" "ind"
> one way.test (values ~ ind, data=d, var.equal=T)
```

one-way analysis of means

data : values and ind

F = 11.735 df = 3 , denom df = 9 ,

Pvalue = 0.00183

Pvalue is less than 0.05 we reject the hypothesis

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

	Df	Sum	Meansq	F value	Pr(>F)
ind	3	71.06	23.688	11.73	0.00183
Res	9	18.17	2.019		

Signif. codes: 0 '***' 0.001 '**' ...

0.0[illegible]

following data gives a life of) to 4 brands.

Types	Life
A	20, 23, 18, 17, 18, 22, 24
B	19, 15, 17, 20, 16, 17
C	21, 19, 22, 17, 20
D	15, 14, 16, 18, 14, 16.

Ho: The mean's of A, B, C, D, are equal.

```
> x1 = c(20,23,18,17,18,22,24)
> x2 = c(B)
> x3 = c(C)
> x4 = c(D)
> d = stack(list(b1=x1, b2=x2, b3=x3, b4=x4))
> names(d)
[1] values "ind"
> one way.test (values ~ ind, data=d, var.equal=T)
        one-way analysis of means
data : values & ind
f = 6.8445, num dt = 3    denom dt = 20
p-value = 6.002349
```

∵ p-value is less than 0.05 we reject the hypothesis.

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

	Df	Sumsq	meansq	Fvalue	Pr (>F)
ind	3	91.44	30.479	6.845	6.00235**
Residuals	20	89.06	4.453		

'.' 0.05 '.' 0.1 ' ' 1

```
> x = read.csv("C:/users/admin/Desktop/marks...
>x
     Stats   Maths
1     40      60
2     45      48
3     42      47
4     15      20
5     37      25
6     36      27
7     49      57
8     59      58
9     20      25
10    27      27
>am = mean(x $ stats)
>am
[1] 37
> am1 = mean (x $maths)
> am1
[1] 39.4
> m1 = median (x $stats)
>m1
[1] 38.5
>m2 = median(x $maths)
[1] 37
>n = length (x $ stats)
>n
[1] 10
```

```
> sd=sqrt((n-1) * var(x$stats)/n)
> sd
[1] 12.64911
> n1=length (x $maths)
> n1
[1] 10
> sd1 = sqrt ((n-1) * var (x $maths)/n)
> sd1
[1] 15.2
> cor (x$stats, x $maths)
[1] 0.830618.
```

A-V

Practical - 10

TOPIC : No Parametric Test

Following are amongs amount of sulphuric acid emitated by industries in date. apply sign test. Test hypothesis that the population median is 21.5 at 5% level of significance.

(17, 15, 20, 29, 19, 18, 22, 25, 27, 9, 24, 20, 17, 6, 24, 15, ~~229, 17, 6, 26~~ 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
] 20
pv=pbinom(sp,n,0.5)
pv
] 0.4119015
```

Since, pvalue is greater, so accepted.

If alternative me ≠ H1 : mos as me < pv = pbinom (sn,n,0)

Q2] Following is the data of 10 observations apply Sign test to test the hypothesis that the population median is 625 against the alternative it is more than 625.

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

Ho - population median 625

```
> b=c(612,619,631,628,643,640,655,649,670,
      655)
> me = 625
> Sp=length(b[b>me])
> Sn=length(b[b<me])
> n=Sp+Sn
> pv=pbinom(sn,n,0.5)
> pv
[1] 0.989
[1] 0.0546875
> n
[1] 10
```

Since pvalue is greater the 0.05, so we accept

1) The following are the values of a sample. Test the hypothesis that the population median is \$60 against the alternative. It is more than \$60 at 5% LOS using Wilcoxon Signed Rank Test.

63, 65, 60, 89, 61, 71, 58, 51, 69, 62, 63, 39, 72, 69, 48, 66, 72, 63, 87, 69.

H_0: population median = 60

H_1: " " > 60

```
x = c(63, 65, 60, 89, 61, 71, 58, 51, 69, 62, 63, 39, 72,
      69, 48, 66, 72, 63, 87, 69)
wilcox.test(x, alter="greater", mu=60)
```

Wilcoxon signed rank test with continuity correction.

data : w

V = 148.5, p-value = 0.01631

alternative hypothesis : true location is greater than 60.

Since, to pvalue is less than, since accepted.

When alternative is less : alter = "less"

When not eqaul to : alter = "greater"

WSRT
Q4] Population median is
15, 17, 24, 25, 20, 21, 32, 28, 12, 25, 24, 26.
~~Ho: before and after no change~~
~~H1: change~~
>x=c(15,17,24,25,20,21,32,28,12,25,24,26)
>wilcox.test(x, alter = "Less", mu=12)

Wilcoxon signed rank test with continuity correction

data : x
V = 66, pvalue = 0.9986
alternative hypothesis : true location is less than 12.

∵ pvalue is greater, then accepted.

Q5] The weights of students before and after they stop smoking are given below. Using WSRT test that there is no significant chang

Weights Before : 65, 75, 75, 62, 72
Weights After : 72, 74, 72, 66, 73
Ho: befor and after no change
H1 : change.
>x=c(65,75,75,62,72)
>y=c(72,74,72,66,73)

> d = x-y   > d [1] -7 1   3 -4 -1

> wilcox.test(d, alter = "two.sided", mu=0)

Wilcox signed rank test with continuity correction.

data : d

V = 4.5 , Pvalue = 0.4982

alternative hypothesis : true location is not equal to 0.

We accept at 0.05%. level of significance.

A.W
27.03.20

Scanned with CamScanner