Practical no.1

Basic of R-software

(1) R is a software of dataanalysis and statistical computing.
(2) It is a software by which, effective data handling and outcome storage is possible.
(3) It is capable of graphical display
(4) It is a free software.

Q.1 $2^2 + |-5| + 4\times5 + 6/5$

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

Q.2 $x = 20$

$y = 2x$

$z = x + y$

sqrt(z)

```
> x = 20
> y = 2x
> z = x + y
> sqrt(7)
[1] 7.745967
```

Q.3 $x = 10$

$y = 15$

$z = 5$

(a) $x + y + z$

(b) xyz

(c) $\sqrt{xyz}$

(d) round $\sqrt{xyz}$

ES :-

```
> x = 10
> y = 15
> z = 5
> x + y + z
[1] 30
> x * y * z
[1] 750
> sqrt (x*y*z)
[1] (27.38613)
```

A vector in R software is denoted by the syntax

```
c(2,3,5,7)^2

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

c (2,3,5,7)^(2,3)
4  27  25  343

c (2,3,5,7,9)^(2,3)
warning message

c(1,2,3,4,5,6)^(2,3,4)
[1] 1  8  9  64  25  216

c(4,6,8,10) * 3
12  18  24  30
```

Scanned by CamScanner

c(4,6,8,10) * c(-1,-2,-3,-4)

[1] -4 -12 -24 -40

c(2,3,5,7) * c(-2,-3)

[1] 0.2500000 0.37037037 0.40

c(2,3,5,7) + 10

12 13 15 17

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

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

[1] 1.0 1.5 2.5 3.5

find the sum of product maximum and minimum for values of

x = c(2,8,4,9,11,10,7,6)^2

> sum(x)

[1] 455

> (x)

[1] 4425974784

> max(x)

[1] 121

> min(x)

[1] 4

4 matrix (nrow=4, 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

$$\rightarrow \underset{x}{\begin{bmatrix} 4 & 7 & 4 \\ 5 & 8 & 0 \\ 6 & 9 & 2 \end{bmatrix}} + \underset{y}{\begin{bmatrix} 6 & 11 & 9 \\ 4 & 12 & 7 \\ 5 & 8 & 4 \end{bmatrix}}$$

x ← matrix (nrow = 3, ncol = 3, data = c(4,5,6,7,8,9,4,0,2)

y ← matrix (nrow = n, col = 3, data = c(6,4,5,11,12,8,7,4)

> x

	[1]	[2]	[3]
[1]	4	7	4
[2]	5	8	0
[3]	6	9	2

> y

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

> ~~z~~x*z

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

y*z

[1]	[1]	[2]	[3]
[2]	12	22	18
[3]	8	24	14
	10	16	8

x+y

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

x*y

	[1]	[2]	[3]
[1]	24	77	36
[2]	20	96	0
[3]	30	92	8

Practical 2

Binomial distribution

n = total no. of trials
p = P(success)
q = p(failure)
x = no. of success out of n

$P(x) = {}^nC_x \cdot p^x \cdot q^{n-x}$; $x = 0, 1, \ldots n$

$E(x) = np$
$V(x) = npq$

dbinom (x, n, p)   P(x)
n * p   E(x)
n * p * q   V(x)
pbinom (x, n, p) $P(X \le x)$, $P(X > x) = 1 - \text{pbinom}(x, n, p)$

Q.1) Toss a coin 10 times with probability (Head = 0.6)
let x be the no. of heads
find the probability of :-
1) 7 Heads
2) 4 Heads
3) Atmost 4 Heads
4) Atleast 6 Heads
v) No Heads
vi) All Heads
vii) expectation and variance.

```
> n=10
> p=0.6
> q=0.4
> dnium(7,10,0.6)
[1] 0.2149908
> dbinom(4,10,0.6)
[1] 0.114767
> pbinom(4,10,0.6)
[1] 0.1662386
> 1-pbinom(6,10,0.6)
[1] 0.3822806
> dbinom(0,10,0.6)
[1] 0.0000d048567
> dbinom(10,10,0.6)
[1] 0.006066618
> exp=10*0.6
> exp
[1] 6
> var= 10*0.6*0.4
> var
> 2.4
```

Q.2 There are 12 MCQ in an english question paper and question has 5 answers and only 1 of them is correct find the probability of having

1. 4 correct answer
2. atmost 4 correct answer
3. atleast 3 correct answer

$n = 12,\ P = 1/4$

(i) $P(x=4)$

```
> dbinom (4,12, 1/5)
[1] 0.1328756
```

(ii) $P(x \le 4)$

```
> pbinom (4,12, 1/5)
[1] 0.927445
```

(iii) $P(x \ge 3) = P(x > 2)$

```
> 1-pbinom(2,12,1/5)
[1] 0.4416543
```

Q.3 find complete binomial distribution where $n = 0.5,\ P = 0.1$

```
> dbinom (0.5,0.1)
[1] 0.59049
> dbinom (1,5,0.1)
[1] 0.32805
> dbinom (2,5,0.1)
[1] 0.0729
> dbinom(3,5,0.1)
[1] 0.0081
> dbinom (4,5,0.1)
[1] 0.0045
dbinom (5,5,0.1)
[1] 1e-05
```

Q4 find probability of exactly 10 success outcome 100 trias with p=0.1

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

5) x follows binomial distribution with n=12, p=0.25
find i) $p(x \le 5)$ (ii) $p(x>7)$ (iii) $p(5<x<7)$

i) $p(x \le 5)$
> pbinom(5,12,0.25)
[1] 0.9455978

(ii) 1-pbinom(7,12,0.25)
[1] 0.0027815 1

(iii) $p(5<x<7)$
[1] 0.04014945

Q6 there are 10 numbers in a commitee prob. of any number attending a mating 0.9 words probability that 7 or more member will present in meeting

$n=10,\ x=7,\ p=0.9$
> 1-pbinom(6,10,0.9)
[1] 0.9872048

Scanned by CamScanner

Q.7 A salesman has 90% probability of making a sell of a typical day he will meet 50 customer what minimum no. of he will make with 88% probability

=> n=50 p=0.88
> qbinom(p,n,p)
> q binom (0.88,30,0.2
[1] 9

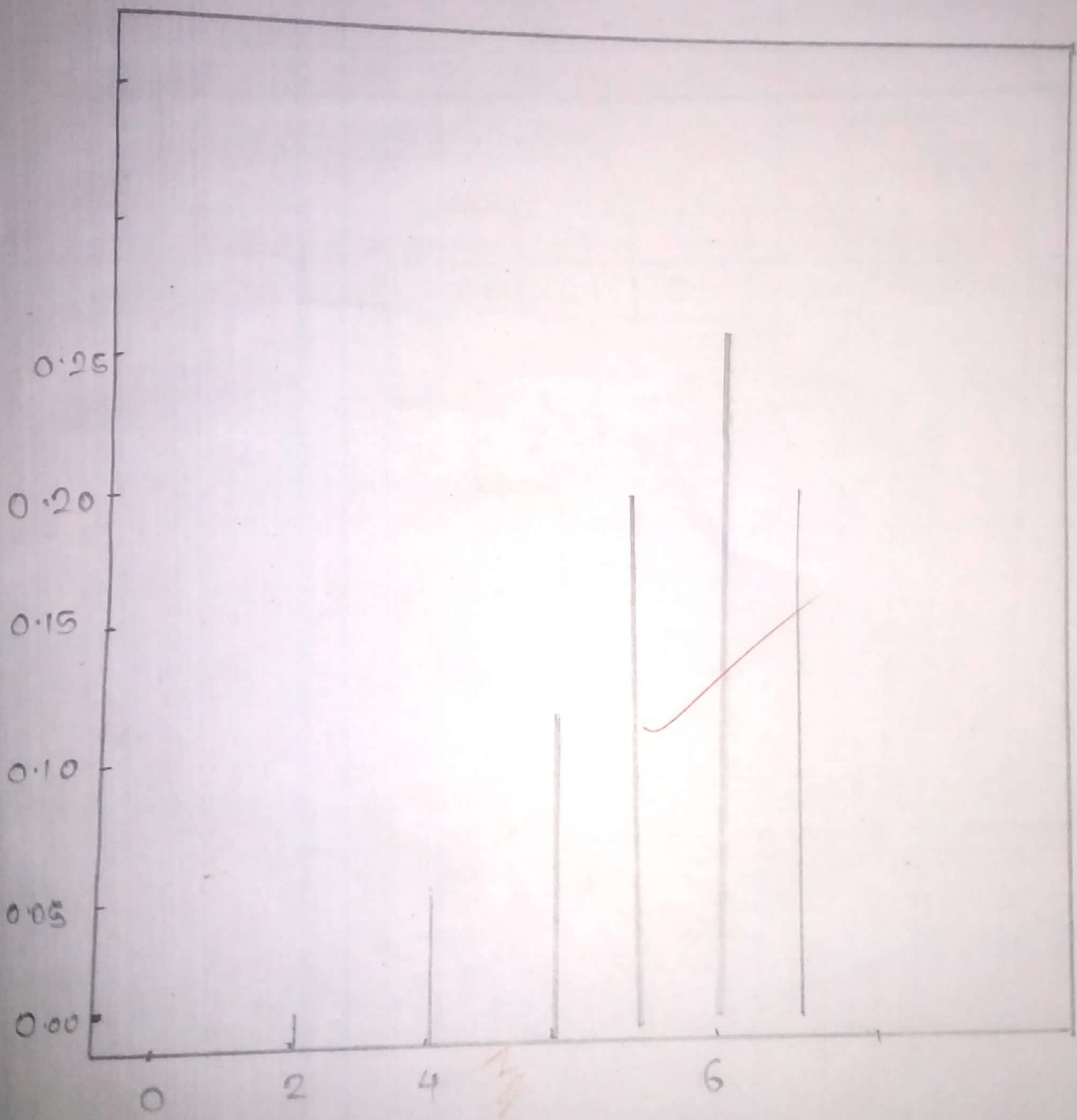
Q.8 n=10, P=0.6 find the binomial distribution plot the graph of p.m.f & c.d.f

>n=10 ; p=0.6
>x=0:n
> bp=dbinom (x,n,p)
> d= data.frame ("x-values"=x, "probability"= bp)
>print (d)

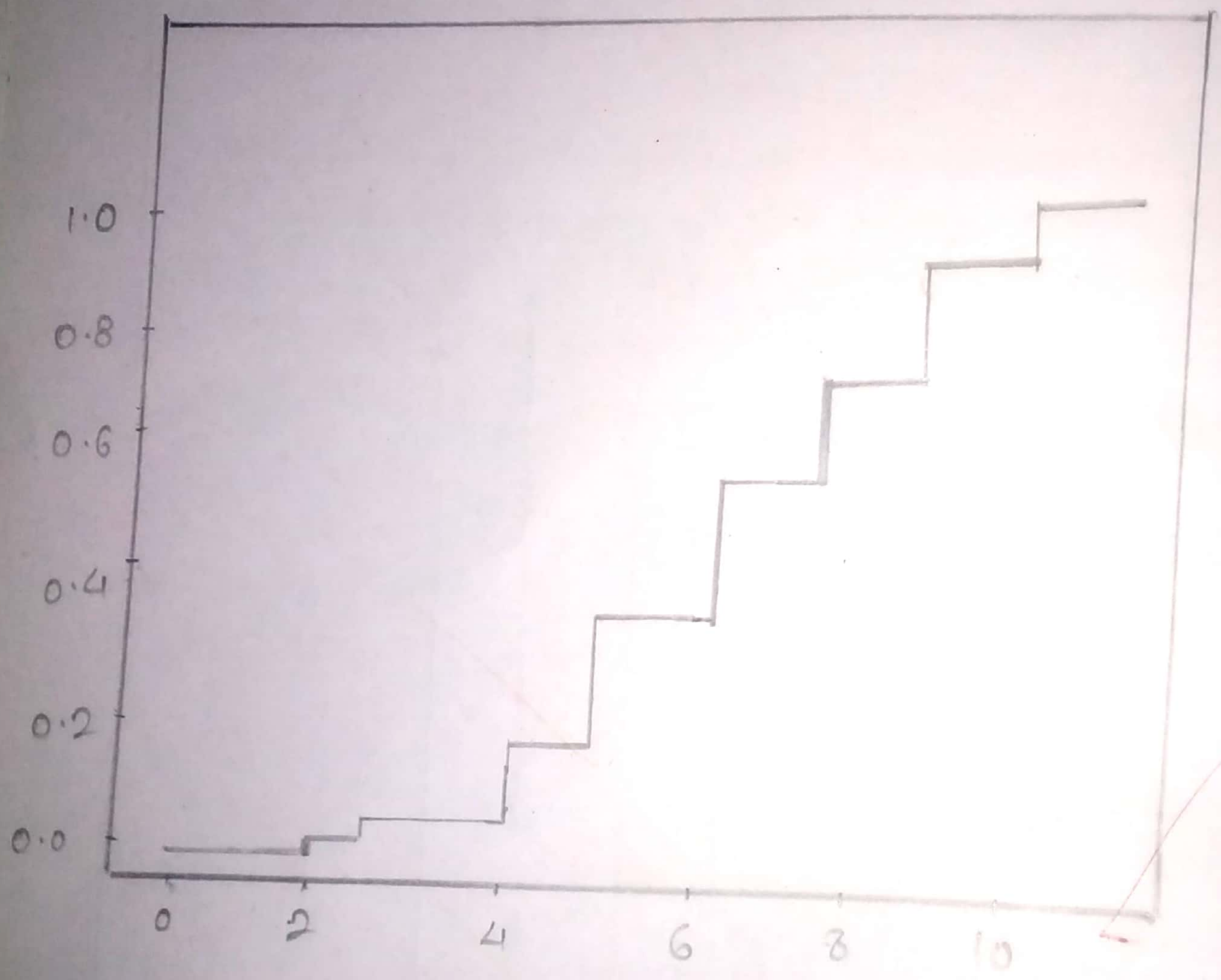
x.	values	probability
1	0	0.0001048576
2	1	0.0015728640
3	2	0.0106168320
4	3	0.0424673280
5	4	0.1114767320
6	5	0.2006581248
7	6	0.2508226568
8	7	0.2149908480
9	8	0.1214908840
10	9	0.1204323520
11	10	0.006047 [illegible]

Scanned by CamScanner

> plot(x, bp, "h")



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



Practical no.3

1) check the followings are p.m.f (probability mass function) 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

solution:.

(i) $0 \le P(x) \le 1$

(ii) $\sum p(x) = 1$

(i) since, it not satisfy the first condition.

(ii) soln:.

since all values of $P(x)$ are more than zero and less then one, 1st condition is satisfied.

Also,

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

$\therefore$ 2nd condition is also satisfied.

→ 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) is between 0 to 1 therefore condition 1 is satisfied.

But,

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

It is more than 1

It does not satisfy condition 2

It is not a ~~prob~~ p.m.f.

```
> prob= c(0.4,0.2,0.3,0.2,0.1)
> sum(prob)
[1] 1.2
```

Mean $= E(x) = \sum x\,p(x) = 3.45$

Var $= V(x) = \sum x^2 p(x) - [E(x)]^2$

$= 13.55 - (3.45)^2$

$= 1.6475$

---

code:

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

Q.2 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 and variance of x

x	$p(x)$	$x\,p(x)$	$x^2\,p(x)$
1	0.1	0.1	0.1
2	0.15	0.3	0.6
3	0.2	0.6	1.8
4	0.3	1.2	4.8
5	0.25	1.25	6.25
		3.45	13.555

Q.3 find mean and variance of x

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

4. 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

18

```
> 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
```

Q4 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
```

$p(x) = 0 \quad x \leq 1$

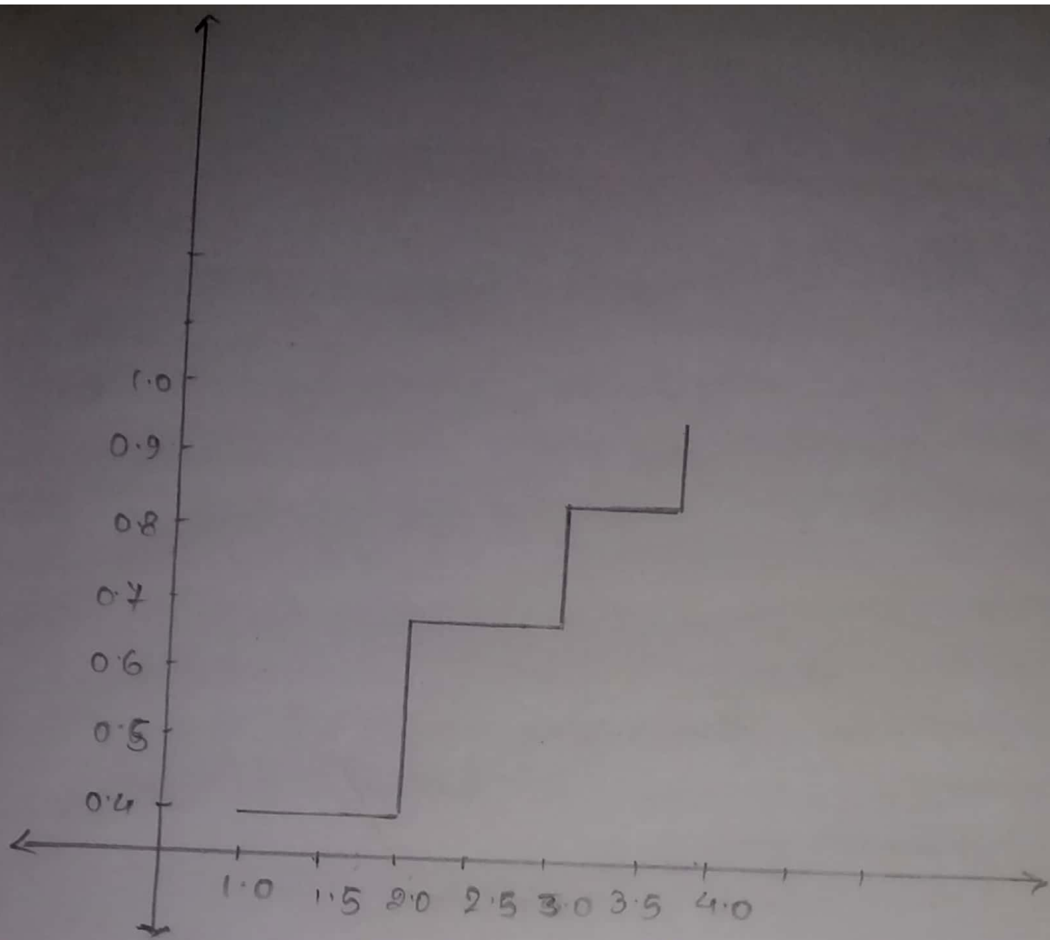
$= 0.4 \quad 1 \leq x < 2$

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

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

$= 1.0 \quad x \geq 4$

plot(x, a, "s")



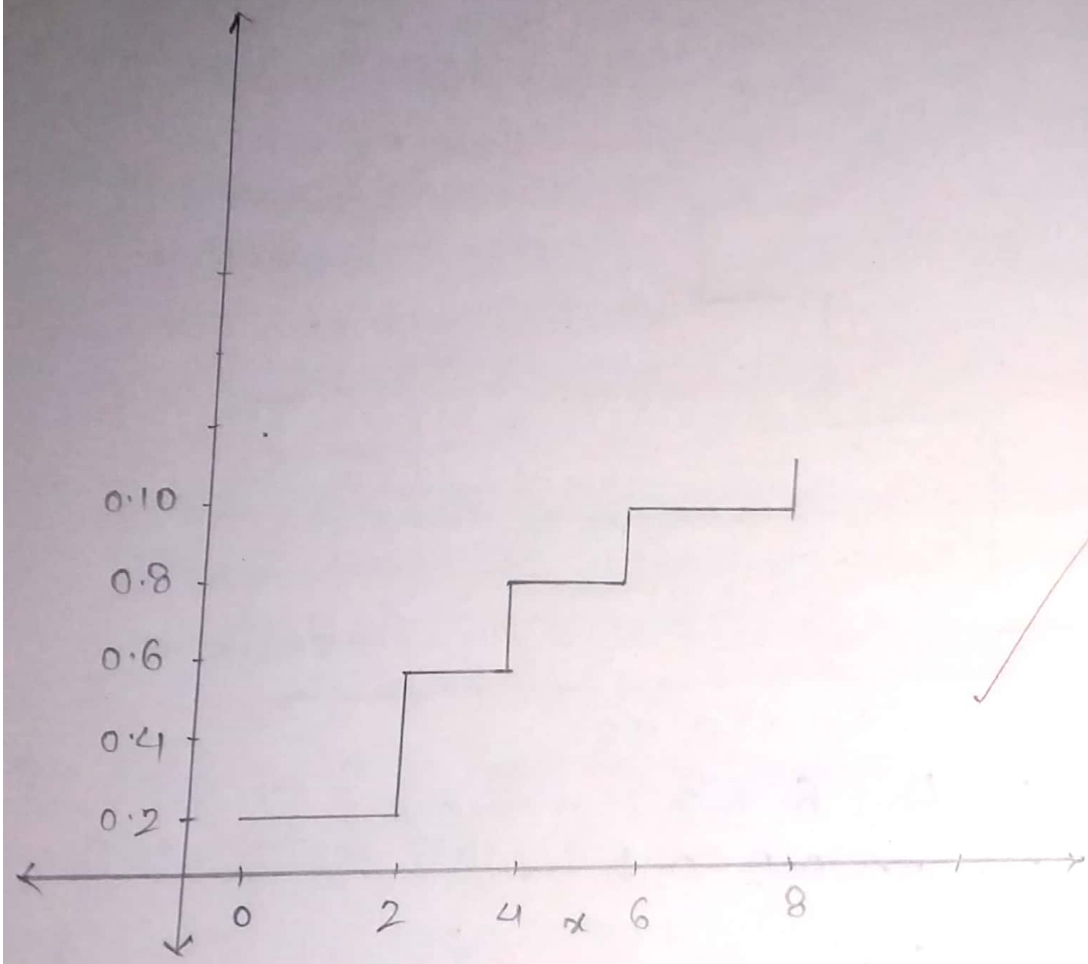
Q.4

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
```

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

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



23.12.19

Practice :-

Q] X follows binomial distribution with $p=0.6$ $q=0.4$
find

a] $P(X=7)$ b] $P(X\leq 3)$

$\rightarrow P(X=7) = ({}^8C_7)\ (0.6)^7\ (0.4)^{8-7}$
$= {}^8C_7 \times 0.2799 \times 0.4$
$= 0.8957$

$P(X\leq 3) = P(0) + P(1) + P(2) + P(3)$
$= {}^8C_0 (0.6)^0 (0.4)^8 + {}^8C_1 (0.6)^1 (0.4)^7$
$+ {}^8C_2 (0.6)^2 (0.4)^6 + {}^8C_3 (0.6)^3 (0.4)$

$= 1 \times 1 \times 0.00065536 \times 8 \times 0.6 \times 0.001638$
$+ 428 \times 0.36 \times 0.0004096 + 56 \times$
0.216×0.101024
$= 0.1736704$

$P[X=2 \text{ or } 3] = P(2) + P(3)$
$= {}^8C_2 (0.6)^2 (0.4)^6 + {}^8C_3 (0.6)^3 (0.4)$
$= 28 \times 0.36 \times 0.004096 + 56 \times$
0.216×0.01024
$= 0.04128768 + 0.12386304$
$= 0.16515012$

Practical no. 4

Practice:-

Q] X follows binomial distribution with $P=0.6$ $q=0.4$ find

a] $P(X=7)$ b] $P(X \leq 3)$

$\rightarrow P(X=7) = ({}^{8}C_{7})\ (0.6)^{7}\ (0.4)^{8-7}$

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

$= 0.8957$

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

$= {}^{8}C_{0}(0.6)^{0}(0.4)^{8} + {}^{8}C_{1}(0.6)^{1}(0.4)^{7}$

$+ {}^{8}C_{2}(0.6)^{2}(0.4)^{6} + {}^{8}C_{3}(0.6)^{3}(0.4)$

$= 1 \times 1 \times 0.00065536 \times 8 \times 0.6 \times 0.001638$

$+ 428 \times 0.36 \times 0.0004096 + 56 \times$

0.216×0.101024

$= 0.1736704$

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

$= {}^{8}C_{2}(0.6)^{2}(0.4)^{6} + {}^{8}C_{3}(0.6)^{3}$
(0.4)

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

0.216×0.01024

$= 0.04128768 + 0.12386304$

$= 0.16515012$

Practical No 5

Normal distribution:-

1] Normal distribution is an example of continuous probability distribution

$X \sim N(\mu, \sigma^2)$

a] $P(X=x)$ dnorm(x, μ, σ)

b] $P(X<x)$ pnorm (x, μ, σ)

c] $P(X>x)$ = 1-pnorm (x, μ, σ)

d] To find the value of k so that the command is qnorm (p, μ, σ)

k $P(X \geq k) = p$

qnorm(p, μ, σ)

e] To generate a random sample of size n, rnorm (n, μ, σ)

Q.1] A random variable $X \sim N(10, 2)$ find

1] $P(X \leq 7)$ 2] $P(X > 12)$ 3] $P(5 \leq x \leq 12)$

4) $P(x < K) = 0.4$ with $\mu = 10$, $\sigma = 2$

Q.2] $X \sim N(100, 36)$ σ = sqrt(36)

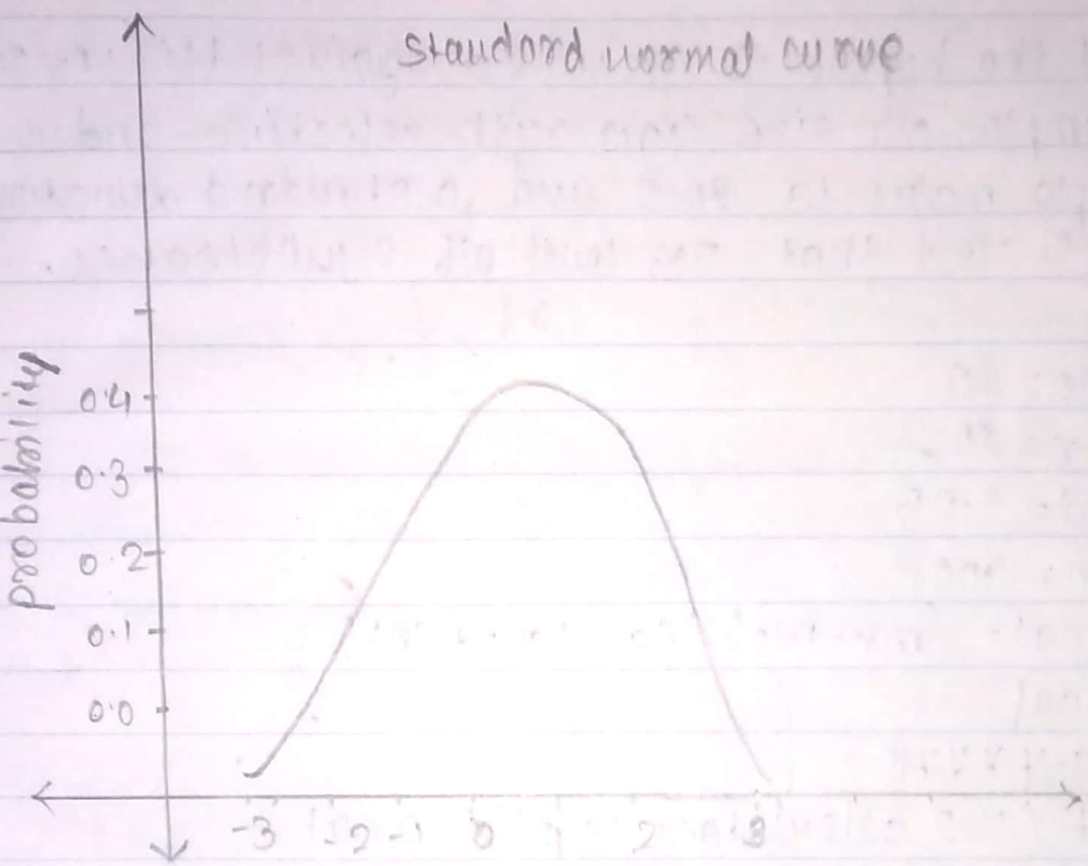
1] $P(X \leq 110)$ 2] $P(X > 105)$ 3] $P(X \leq 92)$

4] $P(95 \leq X \leq 110)$ 5] $P(X \leq K) = 0.9$

Q.3] Generate 10 random sample and find the sample mean, median and variance, standard deviation.

Q.1] >m=10
>s=2
>p1=pnorm(7,m,s)
>cat c("p(x<=7) is=", p1)
p(x<=7) is =0.06680727>
>p2=1-pnorm(12,m,s)
>cat("p(x>12) is =", p2)
p(x>12) is =0.1586553>
>p3=pnorm(12,m,s)-pnorm(5,m,s)
>cat("p(5<=x<=12) is =", p3)
p(5<=x<=12) is =0.8351351
>k=qnorm(0.4,m,s)
>cat("p(x<k)=0.4, k is=", k)
p(x<k)=0.4, k=9.493306>

Q.2) >m=100
>s=sqrt(36)
>p1=pnorm(110,m,s)
>cat("p(x<=110) is =", p1)
~~>cat~~
p(x<=110) is = 0.9522096>
>p2=1-pnorm(105,m,s)
>cat("p(x>105) is", p2)
p(x>105) is=0.2023284>
>p3=pnorm(92,m,s)
>cat("p(x<=92) is", p3)
p(x<=92) is=" 0.09121122>
p4=pnorm(110,m,s)-pnorm(95,m,s)
>cat c("p(95<x<=110) is =", p4)
p(95<=x<=110) is=0.7498813



Q.5] $X \sim N(50, 100)$

find 1] $P(X \le 60)$ 2] $P(X > 65)$

3] $P(45 \le X \le 60)$

```
> u = 50
> s = sqrt(100)
> s
[1] 10
> p1 = pnorm(60, u, s)
> p1
[1] 0.8413447
> p2 = 1 - pnorm(65, u, 5)
> p2
[1] 0.668072
> p3 = pnorm(60, u, 5) - pnorm(45, u, 5)
> p3
[1] 0.5328072
```

Practical no 6

Q.1) Test the hypothesis 0: u = 20 against Hi: u≠2
A sample of size 400 with selective and sample mean is 20.2 and a standard deviatio 2.25 test that 5% level of significance.

→ > mo = 20
> mx = 20.2
> sd = 2.25
> n = 400
> zcal = (mx - mo)/(sd/sqrt(n))
> zcal
[1] 1.77778
> cat ("2 calculator is 2", 2cal)
> calculated is = 1.777798
> pvalue
[1] 0.07544056
Since, p value is more than 0.05 we accept
40: u = 20

Q.2] we want to test the hypothesis 40: u = 250 against 40: u ≠ 260 a sample of size 100 has a mean of 275 and sd as 30 test the hypothesis at 5% level of significance

→ > mb = 250
> mx = 275
> sd = 30
> n = 100
> zcal (mx - mb) / (sd - sqrt(n))

```
>zcal
[1] 8.333333
>cat("z calculated is = ", zcal)
>cat z calculated is = 8.333333
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0
```

since p value is less than 0.05 we reject H0

Q.3 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 0.625. 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)))
>pvalue
[1] 0.000176834
```

Since pvalue is less than 0.05 we reject H0

```
>zcal
[1] 8.333333
>cat("z calculated is =", zcal)
>cat z calculated is= 8.333333
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0
```

Since p value is less than 0.05 we reject H_0

Q.3 we want to test the hypothesis H0: P:2 againt H1: P≠0.2 & p= population proportion A sample of 400 is selected and the sample proportion is calculated 0.625 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/n)e
>zcal
[1]-3.75
>pvalue=2*(1-pnorm(abs(zcal))
>pvalue
[1]0.001768346
```

Since pvalue is less than 0.05 we reject H0

4> In a big city of 385 men ~~of~~ out of 600 men where found to be self-employed. Does this information support the conclusion that exactly that half of the men are self-employed

> >P=0.5
> >n=600
> >p=325/600
> >Q=1-P
> >zcal=(p-P)/(sqrt(P*Q/n))
> >zcal
> [1] 2.041241
> >pvalue= 2*(1-pnorm(abs(zcal)))
> >pvalue
> [1] 0.04121683

Since, pvalue is less than 0.505 we reject H0.

5 Test the hypothesis of H0 = 50 we reject H0.
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.

> mo=50
> >x=c(50,49,52 _ _ _ 48,49)
> >n=length(x) >n [1] 30
> >mx=mean(x) > mx[1] 49.3333
> >variance=(n-1)*var(x)/n[1] 30.45556

> sd = sqrt(variance > sd [1] 5.563772
> zcal = (mx - mo) / sd / sqrt(n)
> zcal
[1] -0.6562964
> pvalue < 2 * (1-pnorm(abs(zcal)))
> pvalue
[1] 0.5116234
Since, p value is more than 0.05 we accept $H_0 : U = 50$

Practical no. 7

large sample test

1] two random sample of size 1000 and 2000 area drawn from two population with a standard deviation 2 and 3 repectively test the hypothesis that 2 population means are equal or not at 5% level of significant sample means of are 67 and 68 % respectively

→ > n1 = 1000
> n1
[1] 1000
> n2 = 2000
> n2
[1] 2000
> mx1 = 67
> mx1
> 67
> mx2 = 68
> mx2
[1] 68
> sd1 = 2
> sd1
[1] 2
> sd2 = 3
> sd2
[1] 3

38

> zcal = [mx1 - mx2] / (sqrt(sd1^2/n1)) | (sd2^2/n2)

> zcal

[1] - 15.80684

>cat ('z calculated is= ', zcal)

> z calculated is -15.80689

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

> pvalue

[1] = 0

< 0.05 we reject H0: $u_1 \neq u_2$

Q.2 A study on noise level to hospitals [1] cat close following data is calculated sample size and 4 sample mean 61% sd1 = 7.9 sample size = 34 sample man 59.4 sd2 = 7.8 H0 = $u_1 = u_2$ against $H_1 = u_1 \neq u_2$ At 1% level of significance

> n1 = 84

> n2

[1] 84

> n2 = 34

> n2

[1] 34

mx1 = 61.2

> mx1

[1] 61.2

> mx2 = 59.6

> mx2

[1] 59.4

sd1 = 7.8

sd1

[1] 7.8

sd2 = 7.8
[1] 7.8
> zcal = (mx1 - mx2) / sqrt(sd1^2(n1) + (sd^2/n2))
> zcal
[1] 1.31117
> cat ("z calculated is", zcal)
> calculated is 1.31117
> pvalue 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 0.258066
since pvalue is greater than 0.01 we accept

Q.3] from each of two population of oranges the following sample are collected whether the proportion of bad oranges are equal or not
1) sample size = 250
2) sample size = 200
3) No. of bad orange in 1st sample = 44
4) No. of bad orange in 2nd sample = 30

$H_0: P_1 - P_2$
$H_1: P_1 = P_2$

> n1 : 256
> n1
[1] 256
> n2 : 200
> n2
[1] 200

Ho ?

> P1 . 44 /250
> P1
[1] 0.176
> P2 = 30/ 200
> P2
[1] 0.15
> P(n1 * p1 + n2 * P2)/ n1 + n2
> P
[1] 0.164444
> Q = 1-p
> Q
[1] 8.35556
> zcal = (P1-P2)/sqrt (p * p + (1/n1 + n/n2))
> zcal
[1] 0.739358

Since p value is greater we accept
Ho : P1 = P2

Accepted

a) >n1 >400
>n2=600
>P1=200/600
>P2=390/600
> p(n*P1+n2*P2)/(n1+n2)
> q=1-p
>zcal=(p1*P2)/sqrt(p*Q*(1/n1+1/n2))
>zcal
[1]4.24751
>pvalue = 2*(1-pnormabs(zcal)))
>pvalue
[1] 2.303972e-06
since pvalue accepted H0: P1=P2

(5) >x1=c(74,72,74,73,79,76,82,72,85,78,
77,78,76,76)
>x1
[1]74,77,74,73,79,76,82,72,75,78,
77,76,76
>n1=length(14)
>n1
[1]14
mx1=mean(x1)
>mx1
[1]76.21429
>variance=(ne-1)* var(x1)/n1
>variance
[1] 6.311224

>sd1 =sqrt(variance)
>sd1
[1] 2.512215
[1] 2.512215
>x= c(72,16,74,70,70,78,70,72,75,79,79,74,
75,78,72,74,86)
>x1
[1] 22,76,74,70,70 78 70 72 75 79 79 74 75
78 72 74 80
> n2=length(x2)
>n2
[1] 1
>mx2=mean(x2)
>mx2
[1] 74.58824
>variance=(n2-n2)*var(x2)n^2
>variance
[1] 10.47751
>sd2=sqrt(variance)
>sd2
[1] 3.6236898
>t.test(x1,x2)
p-value 0.1387
>0.03 is accepted Ho:P1=P2

pm

Practical :- 8

Topic :- small sample test

1) A random sample of 15 observation is given by
80, 100, 110, 105, 122, 170, 120, 110, 101, 88, 83, 95,
89, 107, 125

Due this data support the assumption i.e. popular
mean = 100

→ $H_0 : \mu = 100$

x = c(80, 100, 110, 105, 122, 70, 120, 110, 101, 88, 83,
89, 107, 125)

t.test(x)

one sample t.test

data: x

t = 24.029 df = 14 p.value = 8.819e-13

ternative hypothesis: true mean is not equal to as
confidence interval:

91.37475   107.28892

ample estimates:

mean of x : 100.3333

Since, p value is less than 0.05 we reject

2] > 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)

data: g1 and g2

t = 2.2573 df = 16.376 p.value = 0.03798

Alternative hypothesis: true difference in mean not equal to 0

ample estimates:

mean of x mean of y

20.1 = 17.5

~~The arithmetic mean of~~ sample

Two type of medicines are used on two or seven patience for reducing their weight. the decreasing in the weight after using the medicines given below:-

med A = 10, 12, 13, 11, 14
med B = 8, 9, 12, 14, 15, 10, 9
is there a significance distance in the efficiency of the medicines.
> x = c (10, 12, 15, 11, 14)
> y = c(5, 9, 12, 14, 15, 18, 9)
Z = 0.80384 , df = 9.7594 pvalue = 0.4406
alternative hypothesis; true difference in mean is not equal to 95 percent confidence interval:
= 1.78117 3.78117
sample estimates
mean of x mean of y
12 11
since, pvalue is more than 0.05 we accept
ho. w1 = 10 at 20.5

Q.4> The weight reducing diet program is conducted and the observation is noted for 10 participant 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,

Ho, there is no significant different in weight H, the diet program reduce weight

> x=c(120,125, - - - - 118)

> y = c(111, 114, - - - - 112)

> t.test(x,y, paired=T, alternative="less")

data: x and y

t=17, df=9 pvalue=1

alternative hypothesis: true difference in in means is less than 0 as percent confidence interval:

Inf = 9.416556

sample estimates

mean of the difference 8.5

Q.5) Sample(A)=66, 67, 75, 76, 82, 84, 85, 90, 92

Sample (B)=64, 66, 74, 78, 82, 85, 87, 92, 93, 95, 97

Test the population mean are equal as no.

> x = c (66, 67, - - - 90, 92)

> y = c(64, 66, - - - 95, 97)

> t.test(x, y)

data = x and y

t = -0.63938 , df = 17.974 pvalue = 0.5304

42

Alternative hypothesis: True difference in means is mean equal to 0 as percentage

confidence interval:

-12.853992 6.853992

sample estimate:

mean of x mean of y

80 83

since, pvalue is greater than 0.05 we accept x and 41% level of significance.

AM

Scanned by CamScanner

Practical 9

Aim:- large and small sample space

1] The arithmetic mean of a sample of a item from against an alternative it is more than 55 at 5% LOS (level of significance)

2] In a big city 350 out of 750 males are found to be smokers, thus this information support that exactly half of the males in the city are smokers? Test at 1% level of significance.

3] Thousand article from a factory A are found to have 2% defectives 500 articles from a second factory B are found to have 1% defective. Test at 5% LOS that the two factory are similar or not.

4] A sample of size 400 was drawn and the sample mean is 99. Test at 5% LOS that the sample comes from a population with mean 100 and variance 64?

5] The flowers steams are selected and the weights are found to be (cm) 63, 63, 68, 69, 71, 71, 72. Test the hypothetic that the mean wight is 66 or not 1% LOS.

6] Two random sample where drawn from to normal population and their value 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 a same variance at 5% LOS

Solution:

$H_0 . \mu_1 = \mu_2$ against .

```
> n=100
> mx=52
> mo=55
> sd=7
> zcal= (mx-mo)/(sd/sqrt(n))
> zcal
[1] -4.285714
> pvalue = 2*(1-pnorm (abs (zcal)))
> pvalue
[1] 1.821530e-05
```

pvalue is < 0.05 we rejected

2) $H_0 . \mu_1 = \mu_2$ against $H_0 : \mu_1 \neq \mu_2$

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

pvalue > 0.05 we accept.

3) $H_0 \cdot \mu_1 = \mu_2$
```
>n1=1000
>n2=1500
>p1=20/100
>p2=(n1*p1+p2*p2)/(1/n1+1/n2)
>p
[1] 0.014
>zcal=(p1-p2)/sqrt(p*q*(1/n1+1/n2)
>zcal
[1] 2.082731
>pvalue=2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.03727577
```
since, pvalue < 0.05
we reject pvalue.

4) $H_0 \; \mu_1 = \mu_2$
```
>n=400
>mx=99
>mo=100
>variana=64
>sd=sqrt(variance)
>sd
[1] 8
>zcal=(mx-mo)/(sd/sqrt(n))
>zcal
[1] -2.5
```

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

> pvalue

[1] 0.01241933

Since, pvalue < 0.05

we reject.

5) $H_0 \cdot \mu_1 = \mu_2$

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

> t.test(x)

One sample t.test

data = x

t = 47.54, df = 6, p.value = 5.522e-09

alternative hypothesis :- True mean is not equal to 0.

95% confidence interval

sample estimates

mean of x

68.14286

pvalue is < 0.05 we reject.

(6) $H_0 \cdot \mu_1 = \mu_2$ against $H_0 \cdot P_1 = P_2$

> x = c(66, 67, - - - 90, 92)

> y = c(64, 66, - - - 85, 97)

> f.var.test(x, y)

> f

F test to compare two variance

data: x and y

F = 0.70686, num df = 8, denom df = 10, p-value.
Alternative hypothesis: true ratio of variance is not equal to 1.
95 % confidence interval.
0.1833662 3.0360393
Sample estimates
ratio of variance
0.7068567
$H_0 . \mu_1 = \mu_0$ against $H_0 . \mu_1 = \mu_2$
pvalue > 0.05 we accept $H_0 . \mu_1 = \mu_2$

Practical 10

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

		C.H Clean	Dirty
C.C.	clean	70	50
	fairly clean	80	20
	Dirty	35	45

solution:- H_0 : CC and CH

```
> 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
```

Since, pvalue is less than 0.05, we get reject.
Hence, child cleanliness and home cleanliness are dependent.

Q2 use the following data to find a vaccination and a particular disease are independence or not.

		Aff	not aff
vac	Given	20	30
	not given	25	35

Soln:- H0: vac & Dis

```
> 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
```

pearson's chi-squared test with yates' continuity correction

data=y

X-squared=0, df=1, pvalue=1

since, pvalue is more than 0.05, we accept Hence, the vaccine and disease independence of each other.

Q:3 perform ANOVA for the following data

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

Soln: H0: 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 and ind
    F = 3.326, numdf = 3, denomdf = 9, pvalue = 0.0
> anova = aov(values~ind, data = d)
> anova
  call:
   aov(formula = values ~ ind, data = d)
```

Terms:

Sum of squares ind Residuals
Deg. of freedom 79.26923 71.50000
3 9

Residual standard error: 1.420776
Estimated effects may be unbalanced.

Q4) The following data gives life of tired of four branded.

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 hypothetic that the average life of four branded is same.

H0: the average of four branded 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)
> d = stack(list(b1=x1, b2=x2, b3=x3, b4=x4))
> names(d)
[1] "values" "ind"
> 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 : 2.110236
Estimated effects may be unbalanced.

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)

Q:- one thousand student of a college graded according to the IQ and economic conditional of their home. check ~~that~~ is there any association between IQ and ~~the~~ economic condition of the home.

		I.Q	
E.C		High	low
	Hight	460	140
	pred	330	200
	low	240	160

soln:. H0: IQ & E.C.

```
> 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 pvalue=2.364e-07

Since, pvalue is less than 0.5, it rejected.

Ans

q. no

Ho: population median is 625.

```
>x=c(612,619,631,628,643.----)
> 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
```

Practical no. 11

Non-parametric test

Following are the amounts of sulphur oxide emitted 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, 23, 24, 26.

Ho: population median is 21.5.

```
> x=c(17,15,20,29,19,18,22,25,____)
> median=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
```

(2) Following are the 10 distribution,
612, 619, 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 and 1% LOS.

Note:- If alternative is greater
pv=pbinom(sn)

Q.3 Following are the distributions.
36, 32, 21, 30, 24, 25, 20, 22, 20, 18

Using sign test, test the hypothesis that the population median is 25 against the alternative is less than 25 at 1% LOS.

```
> 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
```

[Ho: population median is 25]

Q4 Following are the distributions
63, 65, 80, 89, 61, 71, 58, 51, 69, 62, 63, 39, 72, 65
Using wilcoxon signed rank test, test hypothesis that the population median is 60. Against the alternative it is greater 60 at 5% LOS.

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) 50

wilcoxon signed rank test with continuity correction
data: x

V = 68, p-value = 0.06186

alternative hypothesis: true location is greater than 60

~~warning messages~~

Q.5 Following are the

20, 25, 27, 30, 18, test the hypothesis.

→ x = c (20, 25, 27, 30, 18) H0: population median is 2

> wilcox.test (x, alt = "two sided", mu=25)
wilcoxon. test. rank. test with continuity correcti
data: x

V = 3.5 , p-value = 0.7127

Alternative hypothesis: true location not equal to 25.

Q.6 Following are the observation.
15, 17, 24, 25, 20, 21, 32, 28, 12, 25, 24, 26, use wilcoxon Rank test.

> x = c(15, 17, 24, ------)
> wilcox.test(x, alt = "less", mu = 20)
> ~~x = c(15, 17~~, 24, 25,
data = x
V = 48.9, p.value = 0.9257
Alternative hypothesis true location is less than 20.

4-3-20