5/11/19

PRACTICAL NO:- 1

BASIC OF R SOFTWARE

- R is a software for data analysis and statistical computing.
- It is a software by which effective data handling and outcome storage is possible.
- It is capable of graphical display.
- It is a free software.

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

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

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

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

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

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

```
> x = 10
> y = 15
> z = 5
> x + y + z
[1] 30
```

> a = x*y*z

> a

[1] 750

> sqrt (a)

[1] 27.38613

> round (sqrt (a))

[1] 27

- A vector in R Software is denoted by the Syntax c( )

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

> x = c(2, 3, 5, 7)

> x^2

[1] 4 9 25 49

4) > c (2, 3, 5, 7, 9, 11) ^ c(2, 3)

[1] 4 27 25 343 81 1331

5) > c(1, 2, 3, 4, 5, 6) ^ c (2, 3, 4)

[1] 1 8 81 16 125 1296

> c(2, 4, 6, 8) * 3

[1] 6 12 18 24

> c(2, 4, 6, 8) * c(-2, -3, -5, -7)

[1] -4 -12 -30 -56

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

> c(1, 3, 5, 7) + 10
[1] 11  13  15  17

> c(1, 3, 5, 7) + c(-2, -3, -1, 0)
[1] -1  0  4  7
```

Sum, Product

1) find the sum, product, square root of sum of for the following values.

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

```
> x = c(given data)
> y = sum(x)
> y
[1] 125
> z = prod(x)
> z
[1] 8.559323e+12
> sqrt(y)
[1] 11.18034
> sqrt(z)
[1] 2925632
```

2) find the sum, product, maximum & minimum values of

```
> x = c(2, 8, 9, 11, 10, 7, 6)^2
> sum(x)
[1] 455
> prod(x)
[1] 442597478400
> max(x)
[1] 121
> min(x)
[1] 4
```

Matrix

1)
```
> 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
```

2) $\begin{bmatrix} 2 & 8 & 5 & 1 \\ 6 & 9 & 0 & 4 \\ 7 & 4 & 2 & 5 \end{bmatrix}$

```
> x <- matrix(nrow = 3, ncol = 4, data = c(2,6,7,8,9,4,5,
                                            0, 2, 1, 4, 5))
```

```
>x
      [,1]  [,2]  [,3]  [,4]
[1,]   2     8     5     1
[2,]   6     9     0     4
[3,]   7     4     2     5

3) >x = matrix(nrow = 3, ncol = 3, data = c(4,5,6,7,8,9,4,0,2))
>y = matrix(nrow = 3, ncol = 3, data = c(6,9,5,11,12,8,9,7,4))
>x + y
      [,1]  [,2]  [,3]
[1,]   10    18    13
[2,]   9     20    7
[3,]   11    17    6
>x*2
      [,1]  [,2]  [,3]
[1,]   8     14    8
[2,]   10    16    0
[3,]   12    18    4
>y*3
      [,1]  [,2]  [,3]
[1,]   18    33    27
[2,]   12    36    21
[3,]   15    24    12
```

7.12.19

2.12.19

PRACTICAL No:- 2

BINOMIAL DISTRIBUTION

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

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

dbinom(x, n, p)	P(x)
n * p	E(x)
n * p * q	V(x)
pbinom(x, n, p)	$P(X \le x)$ Probability of atmost x
1 - pbinom(x, n, p)	$P(X > x)$ Probability of atleast

1) Toss a coin 10 times with probability (head = 0.6). Let x be the no. of heads. find the probability of
i) 7 heads ii) 4 heads iii) atmost 4 heads
iv) atleast 6 heads v) no head vi) all heads
also find E(x) and V(x)

```
> n = 10
> p = 0.6
> q = 0.4
> dbinom(7, n, p)
[1] 0.2149908
> dbinom(4, n, p)
[1] 0.1114767
```

```
> pbinom(4, n, p)
[1] 0.1662386

> 1 - pbinom(6, n, p)
[1] 0.3822806

> dbinom(0, n, p)
[1] 0.0001048576

> dbinom(10, n, p)
[1] 0.006046618

> n * p
[1] 6

> n * p * q
[1] 2.4
```

2) Suppose there are 12 MCQ in an english question paper each Question has 5 answers and only one of them is ~~cor~~ correct. find the probability of

i) 4 correct ans

ii) atmost 4 correct ans

iii) atleast 3 correct ans

3) find the complete binomial distribution where ~~a=5~~ $n=5$ $p=$ ~~0.5~~ 0.1

4) find the probability of exactly 10 successes out of 100 trials with $p=0.4$

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

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

7) A salesman has a 20% probability of making a sale to a customer on a typical day he will meet 30 customer what min no. of sales he will make with 88% probability

8) for $n=10$ and $p=0.6$ find the binomid probabilities and plot the graphs of p.m.f and c.d.f

* Note:

1) $P(X=x)$: dbinom(x, n, p)
2) $P(X \leq x)$ = pbinom(x, n, p) almost
3) $P(X > x)$ = 1 - pbinom(x, n, p) atleast
4) P_1 qbinom(P_1, n, q)

Solution

2) i) >dbinom (4, 12, 1/5)

[1] 0.1328756

ii) > pbinom (4, 12, 1/5)

[1] 0.92744845

iii) P(x ≥ 3) = P (x > 2)

>1 - pbinom (2, 12, 1/5)

[1] 0.4416543

3) > 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] 0.00045

>dbinom (5, 5, 0.1)

[1] 1e-05

4) > dbinom(10, 100, 0.1)

[1] 0.1318653

5) n = 12 , p = 0.25

i) P(x ≤ 5)

> pbinom (5, 12, 0.25)

[1] 0.9455978

ii) P(x > 7)

> 1 - pbinom(7, 12, 0.25)

[1] 0.00278151

iii) P(5 < x < 7)

> dbinom (6, 12, 0.25)

[1] 0.04014945

6) P(x ≥ 7) = P(x > 6)

> 1 - pbinom(6, 10, 0.9)

[1] 0.9872048

7) p1 = 0.88, n = 30, p = 0.2

> qbinom (0.88, 30, 0.2)

[1] 9

8) > n = 10

> p = 0.6

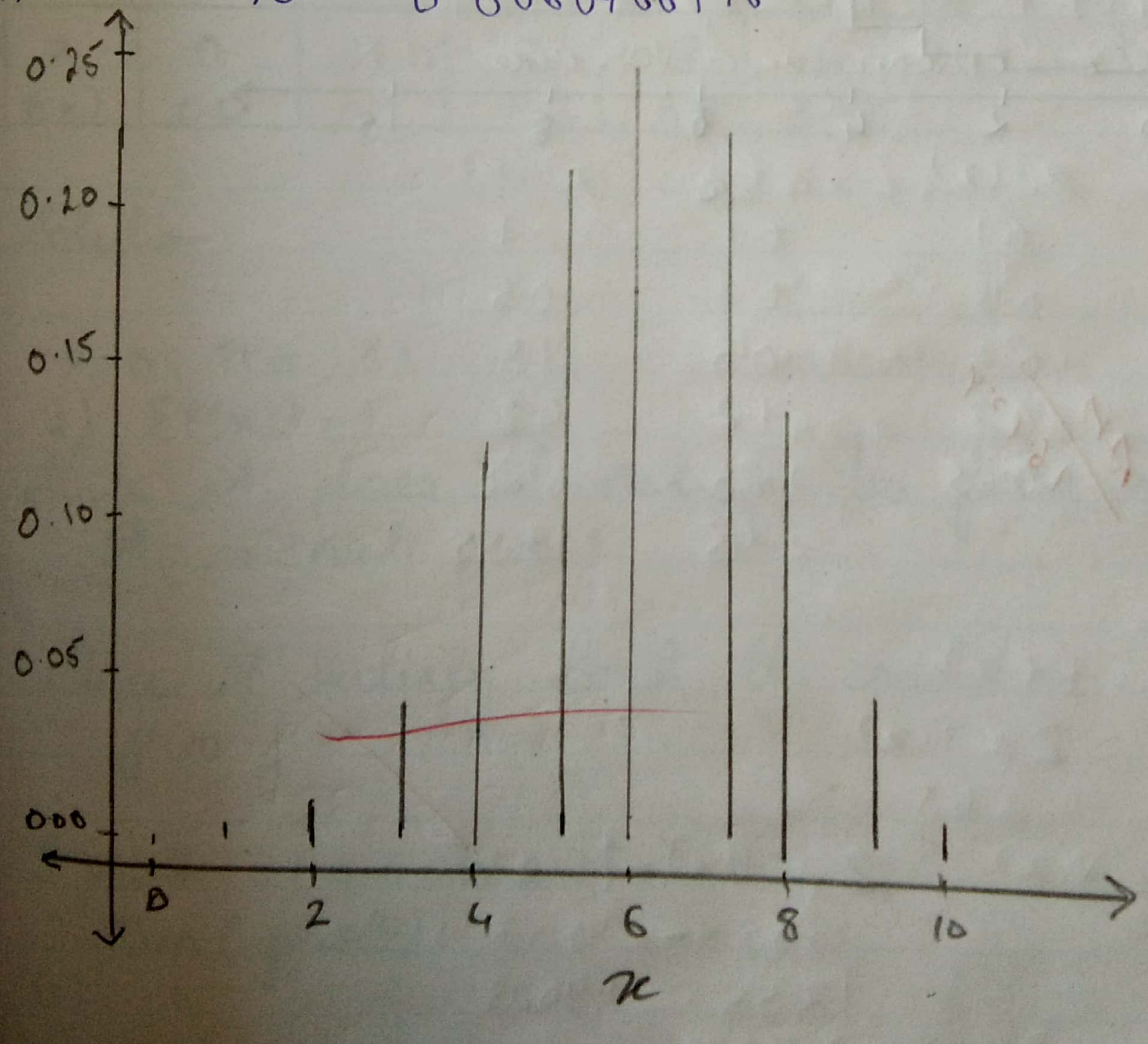
> x = 0:n

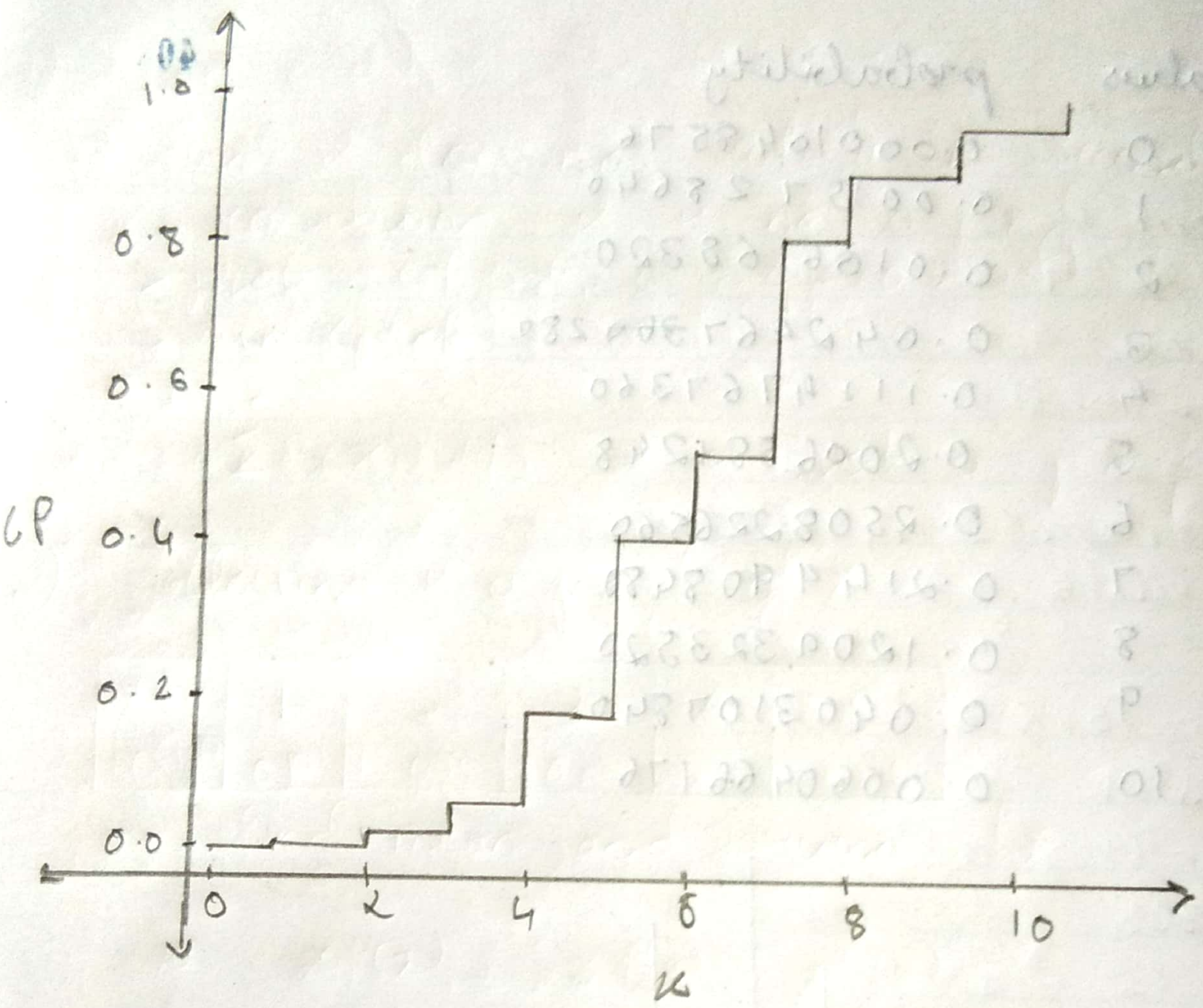
> bp = dbinom (x, n, p)

> 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	0.042467~~700~~280
5	4	0.1114767360
6	5	0.2006581248
7	6	0.2508226560
8	7	0.2149908480
9	8	0.1209323520
10	9	0.0403107840
11	10	0.0060466176





AH 16-1-24

7/12/19

PRACTICAL NO.: 3

TOPIC : PROBABILITY DISTRIBUTION

1) Check the folling 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) 1) $0 \le P(x_i) \le 1$

2) $\sum P(x_i) = 1$

Since it does not satisfy the first condition it is not p.m.f.

ii) Since it satisfy both the condition it is a p.m.f

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

$\because \sum P(x_i) = 1$ it is a p.m.f

iii) Since it does not satisfy the second condition it is not a p.m.f

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

$\because \sum P(x_i) \neq 1$ it is not a p.m.f.

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	13.55

Mean $= E(x)$

$= \sum x\,P(x)$

$= 3.45$

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

$= 13.55 - 3.45 \times 3.45$

$= 13.55 - 11.9025$

$= 1.6475$

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

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

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

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

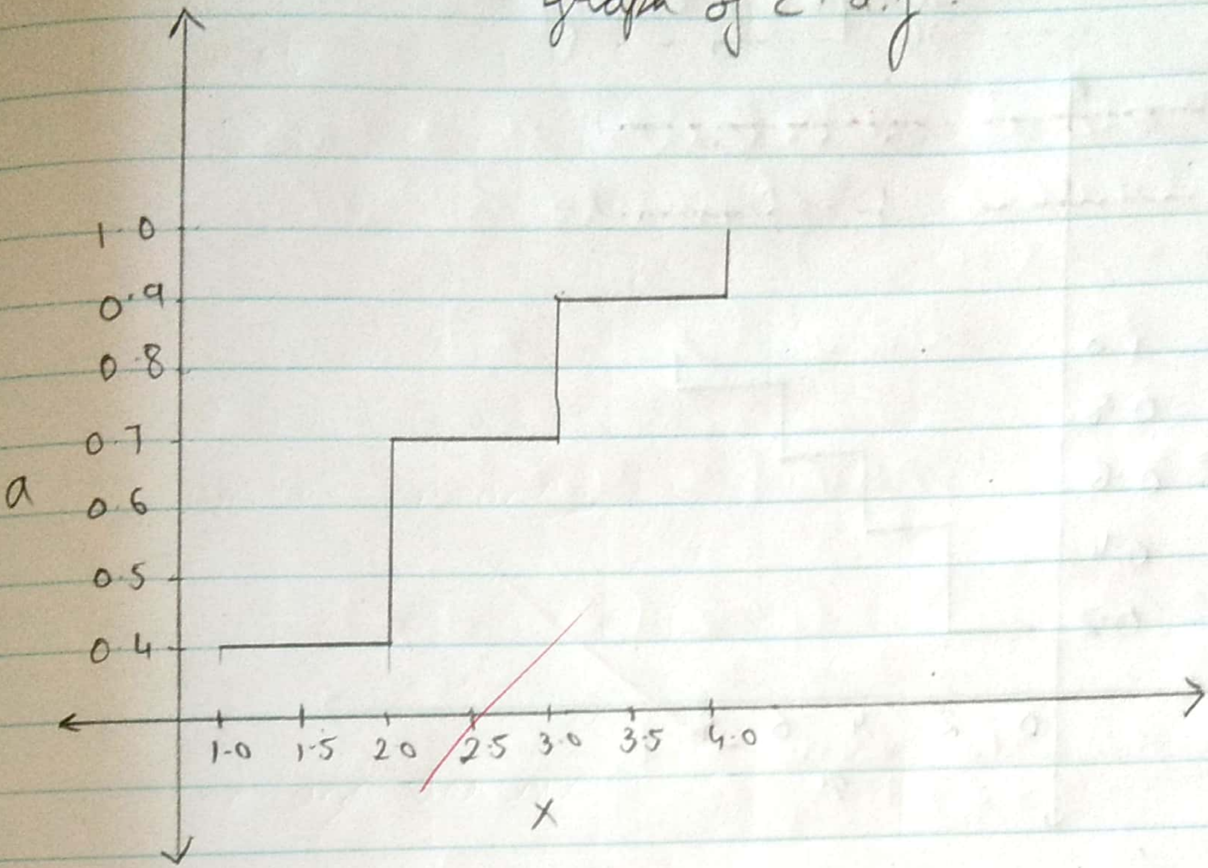
x	1	2	3	4
$P(x)$	0.4	0.3	0.2	0.1

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

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ 0.4 & 1 \le x < 2 \\ 0.7 & 2 \le x < 3 \\ 0.9 & 3 \le x < 4 \\ 1.0 & x \ge 4 \end{cases}$$

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

Graph of c.d.f.

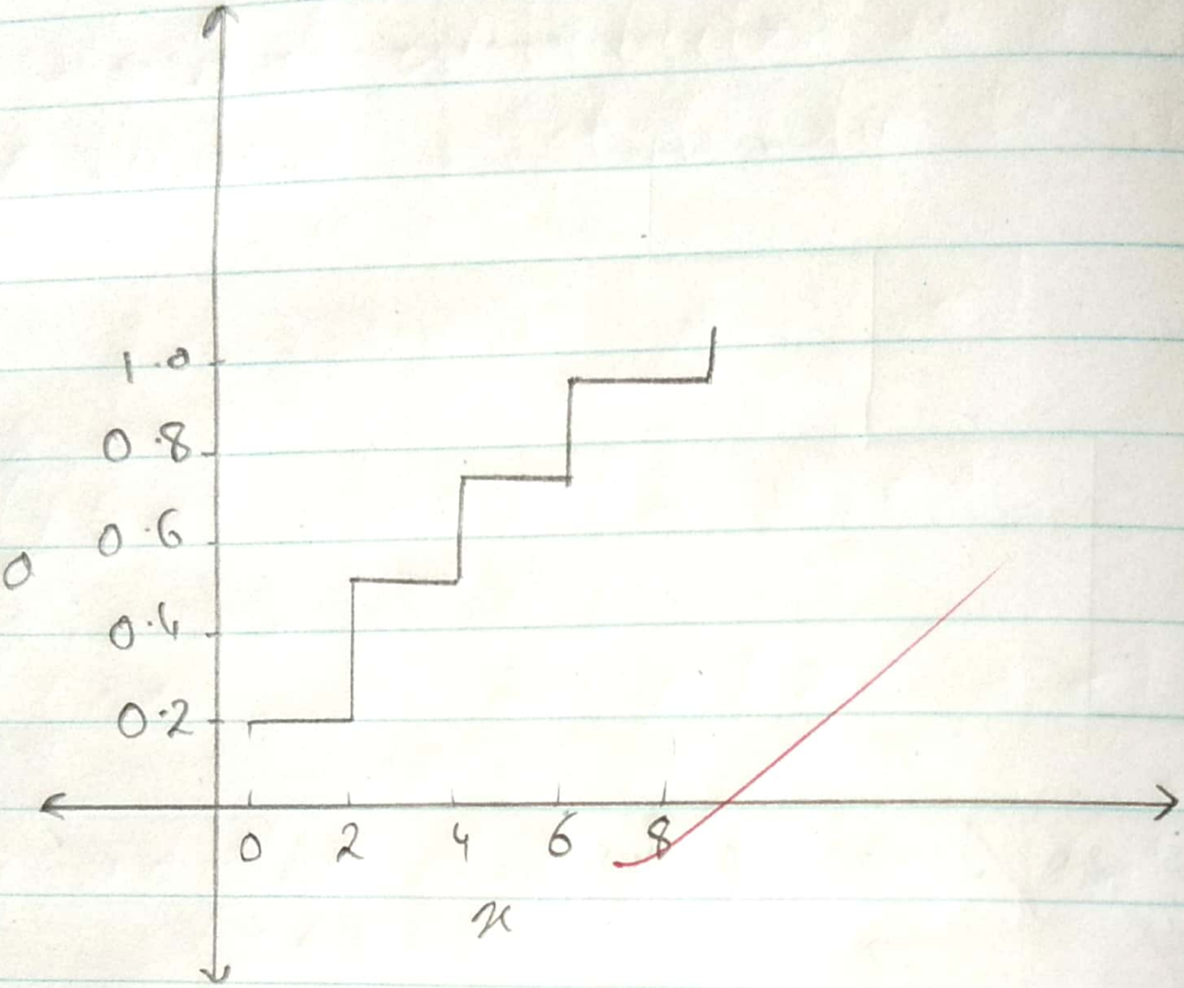


ii)

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

Graph of c.d.f



A.V.
9/12/19

PRACTICAL No:- 4

1) X follows binomial distribution with parameters $n = 8$, $p = 0.6$ find

i) $P(X = 7)$

ii) $P(X \le 3)$

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

Solution: $n = 8$ $p = 0.6$

$q = 1 - p = 1 - 0.6 = 0.4$

i) $P(X = 7)$

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

$\therefore P(X = 7) = {}^8C_7 \, (0.6)^7 \, (0.4)^1$

$= 8 \times 0.0279936 \times 0.4$

$= 0.08957952$

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

$= {}^8C_0 \, (0.6)^0 \, (0.4)^8 + {}^8C_1 \, (0.6)^1 \, (0.4)^7 + {}^8C_2 \, (0.6)^2 \, (0.4)^6 + {}^8C_3 \, (0.6)^3 \, (0.4)^5$

$= 0.00065536 + 8 \, 0.00786432 +$

~~0.0619152 + 0.08257536~~

$+ 0.04128768 + 0.12386304$

AM 17.12.19

$= 0.1736704$

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

$= 0.04128768 + 0.12386304$

$= 0.36515072$

PRACTICAL NO:- 5

NORMAL DISTRIBUTION

Normal distribution is a example of continuous probability distribution.

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

1) $P(X=x)$ = dnorm (x, μ, σ)

2) $P(X \leq x)$ = pnorm (x, μ, σ)

3) $P(X > x)$ = 1 - pnorm (x, μ, σ)

4) To find the value of k so that $P(X \leq k) = p$
   qnorm (p, μ, σ)

5) To generate random sample of size n.
   rnorm (n, μ, σ)

Q1) A random variable x follows ND with $\mu = 10$, $\sigma = 2$

find i) $P(X \leq 7)$

ii) $P(X > 12)$

iii) $P(5 \leq X \leq 12)$

iv) $P(X < k) = 0.4$

Q2) $X \sim N(100, 36)$

i) $P(X \leq 110)$

ii) $P(X > 105)$

iii) $P(X \leq 92)$

iv) $P(95 \leq X \leq 110)$

v) $P(X < k) = 0.9$

Scanned with CamScanner

3) $X \sim N(10, 3)$
generate 10 random sample and find the sample mean, median, and variance and standard deviation.

4) Plot the Standard Normal Curve

5) $X \sim N(50, 100)$
find i) $(P \leq 60)$
ii) $P(X > 65)$
iii) $P(45 \leq X \leq 60)$

Solution

1) i)
```
>p1 = pnorm(7, 10, 2)
> cat ("P(x<=7) = ", p1)
P(x<=7) = 0.0668072
```

ii)
```
> p2 = 1- pnorm (12, 10, 2)
> cat ("P(x>12) = ", p2)
P(x>12) = 0.1586553
```

iii)
```
>p3 = pnorm(12, 10, 2) - pnorm(5, 10, 2)
>cat ("P(5<= x <= 12) = ", p3)
P(5<=x<=12) = 0.8351351
```

iv)
```
>k = qnorm(0.4, 10, 2)
>cat ("P(x<k) = 0.4, k is", k)
P(x<k) = 0.4 , k is 9.493306
```

2) i)
```
> p1 = pnorm (110, 100, 6)
>cat ("P(x<=110) = ", p1)
P(x<=110) = 0.9522096
```

ii)
```
> p2 = 1- pnorm (105, 100, 6)
>cat ("P(x>105) = ", p2)
P(x>105) = 0.2023284
```

```
iii) > p3 = pnorm (92, 100, 6)
     > cat("P(X ≤ 92) = ", p3)
     P(X<=92) = 0.09121122

iv) > p4 = pnorm(110,100,6) - pnorm(95, 100, 6)
    > cat(" P(95 <= X <=110) = " p4)
    P(95<=X<=110) = 0.7498813

v) > k = qnorm (0.9, 100, 6)
   > cat (" P(X<k) = 0.9 , k is ", k)
   P(X<k) = 0.9 , k is 107.6893

3) > x = rnorm (10, 10, 3)
   > x
   [1] 9.556779  9.088768  11.539242  10.523844  12.951400
   [8] 13.538153  8.935227  7.941349  13.964728  11.447649
   > am = mean (x)
   > am
   [1] 10.94877
   > me = median (x)
   > me
   [1] 10.98575
   > n=10
   > var = (n-1) * var(x) / n
   > var
   [1] 3.0908883
   > sd = sqrt (var)
   > sd
   [1] 1.977089
```

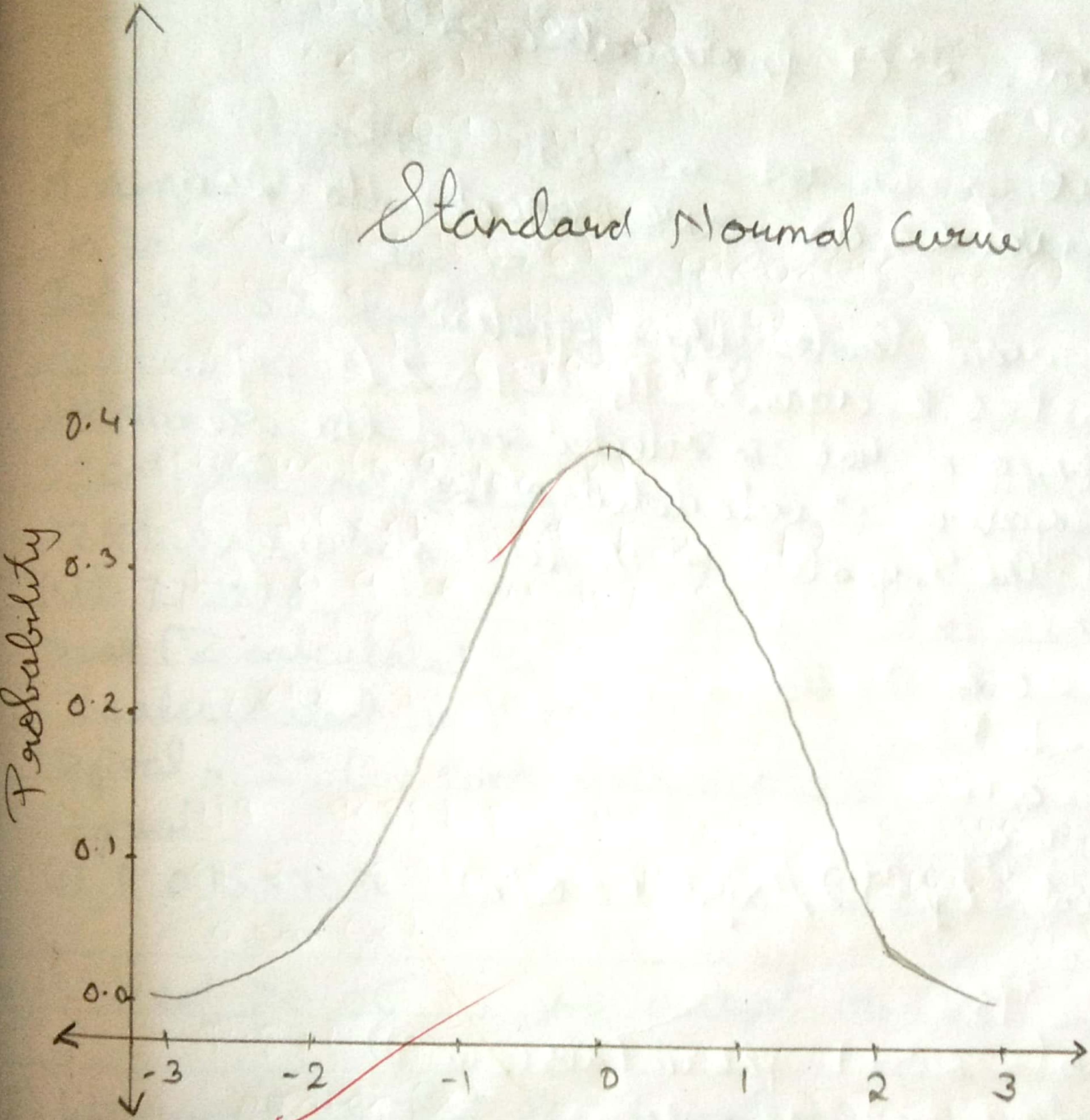
4) >x = seq(-3, 3, by=0.1)
> y = dnorm (x)
> plot (x, y, xlab = "x values", ylab = "Probability"
main = "Standard normal curve")

5) i) > p1 = pnorm(60, 50, 10)
> cat ("P(X<=60) = ", p1)
P(X<=60)= 0.8413447

ii) > p2 = 1 - pnorm (65, 50, 10)
> cat ("P(X >65) = ", p2)
P(X >65) = 0.0668072

iii) > p3 = pnorm(60, 50, 10) - pnorm(45, 50, 10)
> cat("P(45 ≤ X ≤ 60) = ", p3)
P(45 ≤ X ≤ 60) = 0.5328072

Standard Normal Curve



A.14
6-1-20

PRACTICAL No: 6

Z and t distribution

1) Test the hypothesis $H_0: \mu = 20$ against $H_1: \mu \neq 20$. A sample of size 400 is selected and the sample mean is 20.2 and the SD is 2.25. Test at 5% ~~leg~~ level of significance.

```
>> mo=20; mx=20.2; sd=2.25; n=400
>zcal=(mx-mo)/(sd/sqrt(n))
>zcal
[1]1.777778
>cat("Z calculated is",zcal)
Zcalculated is 1.777778
>pval=2*(1-pnorm(abs(zcal)))
>pval
[1] 0.07544036
```

$\because$ pval > 0.05, we accept $H_0 = 20$

2) We want to test the hypotesis $H_0: \mu = 250$ against $H_1: \mu \neq 250$. A sample of size 100 has a mean of 275 and sd is 30. Test the hypotesis at 5% level of significance

```
>mo=250; mx=275; sd=30; n=100
>zcal=(mx-mo)/(sd/sqrt(n))
>cat("Z calculated is",zcal)
Z calculated is 8.333333
```

29

```
>pval= 2*(1-pnorm(abs(zcal)))
>pval
[1] 0
```

Ans ∵ pval < 0.05, we reject $H_0 = 250$

3) We want to test the Hypotesis
$H_0: P = 0.2$ against $H_1: P \neq 0.2$ (P = population proportion)
A sample of 400 is selected and the sample proportion is calculated 0.125.
Test the hypotesis 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)))
>pval
[1]0.0001768346
```

Ans ∵ Pval < 0.01, we reject P = 0.2

4) In a big city 325 men out of 600 men were found to be self-employed. Thus, this ~~inn~~ information support the conclusion that half of the men in the city are self-employed?

$P = 0.5$ $n = 600$ $p = 325/600$

```
>P=0.5
>Q=1-P
>p=325/600
>n=600
>zcal=(p-P)/(sqrt(P*Q/n))
>zcal
[1] 2.041241
>pval=2*(1-pnorm(abs(zcal)))
>pval
[1] 0.04122683
```

Ans ∵ Pval < 0.05, We reject P = 0.5

5) Test the hypotesis $H_0: \mu = 50$ againt $H_1: \mu \neq 50$
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.

```
>mo = 50
>x = c( given data ... )
>n = length(x)
>mx = mean(x)
>variance = (n-1) * var(x)/n
>variance
[1] 30.95556
>sd = sqrt(variance)
>sd
[1] 5.563772
>mo =
>zcal = (mx - mo)/(sd/ sqrt(n))
>zcal
[1] -0.6562965
>pval = 2*(1 - pnorm(abs(zcal)))
>pval
[1] 0.5116334
```

Ans : Pval > 0.05 , we accept $H_0 = 50$

20.01.20

PRACTICAL NO:- 7

LARGE SAMPLE TEST

1) Two & random sample of size 1000, 2000 are drawn from 2 population with the SD 2 & 3 respectively. Test the hypotesis that the 2 population means are equal or not at 5% level of significance. The sample means are 67 and 68 respectively.

$H_0\ \mu_1 = \mu_2$ 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)
Z calculated is = -10.84652
>pval= 2* (1 - pnorm(abs(zcal)))
>pval
[1]  0
```

$\therefore$ Pval < 0.05, we reject $H_0\ \mu_1 = \mu_2$

Scanned with CamScanner

2) A study of noise level in two hospitals is done following data is calculated. first sample size 84, mean 61.2, SD 7.9.
Second sample size 34, mean 59.4, SD 7.8
Test H_0 $\mu_1 = \mu_2$ against H_1 $\mu_1 \neq \mu_2$ at 1% level of significance.

H_0: $\mu_1 = \mu_2$ 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)
z calculated is = 1.131117
>pval=2*(1-pnorm(abs(zcal)))
>pval
[1] 0.258006
```

∴ Pval > 0.01 we accept H_0 $\mu_1 = \mu_2$

3) from each of 2 population of oranges the following sample are collected. Test whether the proportion bad oranges are equal or not
first sample is size = 250
Second sample size = 200
No. of bad oranges in 1st sample is 44 & in 2nd sample is 30

H_0 $P_1 = P_2$ against H_1 $P_1 \neq P_2$

```
>n1 = 250
>n2 = 200
>p1 = 44/250
>p2 = 30/200
>p = (n1*p1 + n2*p2)/(n1+n2)
>p
[1] 0.1644444
>q = 1-p
>zcal = (p1-p2)/sqrt(p*q*(1/n1+1/n2))
>zcal
[1] 0.7393581
>pval = 2*(1-pnorm(abs(zcal)))
>pval
[1] 0.4596896
```

∵ Pval > 0.05, we accept H_0 $P_1 = P_2$

4) Random sample of 400 males and 600 females were asked whether they want the atm nearby. 200 male & 390 females were in the favour of the proposal. Test the Hypothesis that the proportion of males and females favouring the proposal are equal or not at 5% level of significance.

5) following are the 2 independent sample of 2 Population test equality of 2 Population means at 5% level of significance

Sample 1 - 74, 77, 74, 73, 79, 76, 82, 72, 75, 78, 77, 78, 76, 76.

Sample 2 - 72, 76, 74, 70, 70, 78, 70, 72, 75, 79, 79, 74, 75, 78, 72, 74, 80

<u>Solution</u> Ho: Ho: P1 = P2 against ~~P~~ H1 P1 ≠ P2

4)
```
> n1 = 400
> n2 = 600
> p1 = 200/400
> p2 = 390/600
> p = (n1 * p1 + n2 * p2) / (n1 + n2)
> q = 1 - p
> zcal = (p1 - p2) / sqrt(p * q * (1/n1 + 1/n2))
> cat ("Z calculated is = ", zcal)
```

* Z calculated is = -4.724751
>pval = 2 * (1 - pnorm(abs(zcal)))
>pval
[1] 2.303972e-06

∴ pval < 0.05 we reject H_0 $P_1 = P_2$

5) H_0 $\mu_1 = \mu_2$ against H_1 $\mu_1 \neq \mu_2$
>x1 = c(Given data . . .)
>n1 = length(x1)
>mx1 = mean(x1)
>variance = (n1-1) * var(x1)/n1
>sd1 = sqrt(variance)
>x2 = c(given data . . .)
>n2 = length(x2)
>mx2 = mean(x2)
>variance = (n2-1) * var(x2)/n2)
>sd2 = sqrt(variance)
>t.test(x1, x2)

Welch Two Sample t-test

data: x1 and x2
t = 1.5227, df = 28.941, p-value = 0.1387

∴ p-value > 0.05 we accept H_0 $\mu_1 = \mu_2$

AM
27.01.20

PRACTICAL No: 8

SMALL SAMPLE TEST

1) The random sample of 15 observation 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.

$H_0: \mu = 100$

```
> x = c(80, 100, 110, . . . . . , 125)
> t.test(x)
```

One Sample t-test

data = x
t = 24.029 , df = 14 , p-value = 8.819e-13
alternative hypothesis: true mean is not equal to 0
95 % confidence interval:
91.37775 109.28892
sample estimates:
mean of x
100.3333

$\because$ P-value < 0.05 we will reject $H_0: \mu$

2) 2 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 hypotesis that there is no significant difference between the score at 1 % level of significance.

$H_0 : \mu_1 = \mu_2$ against $H_1 = \mu_1 \neq \mu_2$

```
> 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)
        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 ~~to~~ 95% confidence interval:
 0.1628205         5.0371795
sample estimates:
mean of x   mean of y
    20.1         17.5
```

$\because$ p-value $>$ 0.01 we accept $H_0 : \mu_1 = \mu_2$

3) 2 types of medicines are used ~~to~~ on 5 & 7 patients for reducing their weight. The decrease in the weight after using the medicine are given below.

Medicine A: 10, 12, 13, 11, 14

Medicine B: 8, 9, 12, 14, 15, 10, 9

Is there a significant difference in the efficiency of the medicine.

$H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$

~~Welch~~

```
>m1 = c(10, 12, 13, 11, 14
>m2 = c(8, 9, 12, 14, 15, 10, 9)
>t.test(m1, m2)
        Welch Two Sample t-test
data: m1 and m2
t = 0.80384    , df = 9.7594, p-value = 0.4406
alternative hypothesis: true difference in means
is not equal to 0
95 % confidence interval:
 -1.781171        3.781171
sample estimates:
mean of x   mean of y
  12           11
```

∵ p-value > 0.05 we accept $H_0: \mu_1 = \mu_2$

4) The weight reducing diet program is conducted and observation are 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.

H_0: There is no significant difference in weight against

H_1: The diet program reduce the weight.

>x = c(Before Data)

>y = c(after Data)

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

paired t-test

data : x and y

t = 17 , df = 9, p-value = 1

alternative hypothesis: true difference in means is less than

95 % confidence interval:

-Inf 9.41655b

sample estimates:

mean of differences

8.5

$\because$ P-value > 0.05 we accept H_0.

36

5) Sample A = 66, 67, 75, 76, 82, 84, 88, 90, 92

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

Test the population means are equal or not.

$H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$

6) The following are the marks before & after a training program test the program is effective or not.

Before = 71, 72, 74, 69, 70, 74, 76, 70, 73, 75

After : 74, 77, 74, 73, 79, 76, 82, 72, 75, 78

Solution

5) $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$

```
> a = c(66, 67, ... ..., 92)
> b = c(64, 66, ........ ..., 97)
> t.test(a,b)

        Welch Two Sample t-test
data: a and b
t = -0.63968 , df = 17.974 , p-value = 0.5304
alternative hypothesis: true difference in means is not
   equal to 0
95 % confidence interval:
   -12.853992   6.853992
```

Sample estimates:
mean of x   mean of y
80   83

$\because$ p-value > 0.05 we accept $H_0: \mu_1 = \mu_2$

6) H_0: There is no significant difference in marks against
H_1: The program increases the marks.

```
> x = c(71, 72, . . . . . , 75)
> y = c(74, 77, . . . . . , 78)
> t.test(x, y, paired = T, alternative = "greater")
        Paired t-test
data: x and y
t = -4.4691, df = 9, p-value = 0.9992
alternative hypothesis: true difference in means is
   greater than 0
95 % confidence interval:
  -5.076639   Inf
Sample estimates:
mean of the difference
          -3.6
```

$\because$ P-value > 0.05 we accept H_0

AN
07-2-20

PRACTICAL No 9

1) The AM of sample of 100 items from a large population is 58. If a SD is 7. Test the hypothesis that the population mean is 55 against the alternative that that mean is more than 55 at 5% LOS

2) In big city 350 out of 700 males are found to be smokers thus this information supports that exactly half of the males in the city are smokers? Test at 1% LOS

3) 1000 articles from a factory A are found to have 2% defectives. 15000 articles from a second factory B are found to have 1% defective. Test that 5% level of significance that the two factory similar or not?

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

5) The flower stems are selected and the heights are found to be (cm) 63, 63, 68, 69, 71, 71, 72. Test the hypothesis that the mean height is 66 or not at 1% LOS

6) Two random sample were drawn from two normal population and there 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 has the same variance at 5% LOS.

Solution

1) $H_0: \mu_1 = 55$ against $H_1: \mu > 55$

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

$\because$ p-value < 0.05 we reject $H_0: \mu = 55$

2) $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 = \mu_2$

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

$\because$ Pvalue $= 1$ we accept $H_0: \mu_1 = \mu_2$

Scanned with CamScanner

Q3) $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$

```
>n1 = 1000
>n2 = 1500
>p1 = 20/1000
>p2 = 10/1500
>p = (n1*p1+n2*p2)/(n1+n2)
>p
[1] 0.014
>q = 1-p
>zcal = (p1 - p2) /sqrt (p* q*(1/n1 + 1/n2)
>zcal
[1] 2.084842
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.0376364
```

∵ pvalue < 0.05 we reject $H_0: \mu_1 = \mu_2$

4)

```
>m0 = 100 , mx = 99 , n = 400, variance = 64
>sd = sqrt (variance)
>sd
[1] 8
>zcal = (mx - m0)/(sd/ sqrt(n))
>zcal
[1] -2.5
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.01241
```

∴ pvalue < 0.05 we reject H_0.

5)

```
>x=c(63 . . . . , 72)
>t.test(x)
```

One Sample t-test

data: x

t = 47.94 , df = 6 , p-value = 5.522e-04

alternative Hypothesis: true mean is not equal to 0

95 percent confidence interval:

64.66479 71.62092

sample estimates:

mean of x

68.44286

∵ pvalue < 0.05 we reject H_0

6)

```
>x=c(66, . . . ,92)
>y=c(64, . . . ,97)
>f=var.test(x,y)
>f
```

f test to compare two variances

data: x and y

f = 0.70686 , num df = 8 , denom df = 10, pvalue = 0.6359

alternative Hypothesis: true ratio of variance is not equal to 1

95 percent confidence interval:

0.1833662 3.0360393

Sample estimates:

ratio of variances

0.7061567

∵ p-value > 0.05 we accept H_0

PRACTICAL No:- 10

ANOVA and CHI-SQUARE TEST

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

Cleanliness of child \ Cleanliness of Home	clean	Dirty
Clean	70	50
Fairly clean	80	20
Dirty	35	45

Sol^n: H_0: Cleanliness of child and Cleanliness of home.

```
> x = c(70, 80, 35, 50, 20, 45)
> m = 3
> n = 2
> y = matrix(x, nrow=m, ncol=n)
> pv = chisq test (y)
> pv
        Pearson's Chi-squared test
data: y
X-squared = 25.646 , df = 2 , p-value = 2.698e-06
```

∵ P-value < 0.05 we reject H_0: Cleanliness of child and cleanliness of home are independent.

) Use the foll data to find a vaccination & a disease are independent or not.

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

Sol: Ho: Vaccination and disease are independent

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

Pearson's Chi-squared test with Yates continuity correction

data: y
X-squared = 0 , df = 1 , p-value = 1

∴ P-value > 0.05 we accept Ho: Vaccine and disease are independent.

3) Perform ANOVA for the foll data.

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

Sol^n Ho: 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 mean
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 may be unbalanced.

∵ P-value < 0.05 we reject H_0: The means of the varieties are equal.

4) The foll data gives the ~~ft~~ life of tyres of 4 brands

Type	Observations
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 life for 4 brands is same.

Sol^n^ H_0: The average ~~f~~ life of 4 type of tyres is same

```
>x1 = c( Data of A ...)
>x2 = c( Data of B ...)
>x3 = c( Data of C ...)
>x4 = c( Data of D ...)
>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 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: 2.110236
Estimated effects may be unbalanced.
```

∴ P-value < 0.05 we reject Ho: The average life of 4-type of tyres is same.

5) 1000 students of a college are graded according ~~of~~ to the IQ & economic condition of their home check that is there any association between IQ & economic condition of the home

		IQ High	IQ low
Economic Condition	High	460	140
	Med	330	200
	low	240	160

Soln Ho: IQ and Economical condition ~~is~~ are independent.

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

Pearson's Chi-squared test

data: y

X-squared = 39.726, df = 2, p-value = 2.364e-09

∵ p-value < 0.05 we reject H0: IQ and Economic condition are independent.

AM
17.2.22

Practical No. 11
Non-Parametric Test

1) 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

Sol: Ho: population median is 21.5

```
> x = c(data ...)
> 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
```

∴ p-value > 0.05 we accept Ho: population median is 21.5

2) Following are the 10 observation 612, 619, 631, 628, 643, 640, 685, 649, 670, 663.

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

Note: If the alternative greater than

pv = pbinom(sn, n, 0.5)

Sol:- Ho: population median is 625 against
H1: population median is greater than 625

```
>x = c(data.....)
> 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
```

∴ p value > 0.01 we accept Ho: population median is 625.

3) 10 observations are 36, 32, 21, 30, 24, 25, 20, 22, 20, 18. Using sign test. Test the hypothesis that population median is 25 against the alternative it is less than 25 at 5% LOS.

Sol:- Ho: population median is 25 against
H1: population median is less than 25.

```
>x = c(data...)
>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

∴ p-value ≥ 0.05 we accept Ho: population median is 25.

4) The foll are some measurments. 63, 65, 60, 89, 61, 71, 58, 51, 69, 62, 63, 39, 72, 65. Using Wilcoxon sign rank test. Test the hypothesis that the population median is 60 against the alternative it is greater than 60 at 5% LOS

Sol: Ho: population median is 60 against

H_1: population median > 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

∴ p-value > 0.05 we accept Ho: population median is 60.

5) obs - 15, 17, 24, 25, 20, 21, 32, 28, 12, 25, 24, 25
Use WSRT to test the hypothesis that Population median is 20 against alternative it is less than 20.

Sol^n H_0: population median is 20 against
H_1: population median < 20

```
>x = c(data......)
>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.

$\because$ p-value > 0.05 we accept H_0: population median is 20.

6) obs: 20, 25, 27, 30, 18. Test the hypothesis that population median is 25 against alternative it is not 25.

Sol^n H_0: population median is 25 against
H_1: population median is not 25

> x = c (data.. ..)
> 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 25

∵ p-value > 0.05 we accept H0 ∴ population median is 25.

9.3.20