★★ INDEX ★★

No.	Title	Page No.	Date	Staff Member's Signature
	2nd Sem.			
~~1.~~	~~Basic~~			
~~2.~~	~~Problem of Binomial~~			
~~3.~~	~~Binomial distribution and probability distribution~~			
	IInd Sem.			
1.	Basic of R software	33		AM
2.	Problem on cdf and pdf	36		
3.	Binomial distribution and probability distribution	39	11/12/19	
4.	Binomial distribution	41.	18/12/19	AM
5.	Normal Distribution	43	8/1/20	AM
6.	Z - distribution	46.	22/1/20	AM
7	large sample Test	48	5/2/20	AM
8.	Small sample test	51	12/2/20	AM

★★ INDEX ★★

No.	Title	Page No.	Date	Staff Member's Signature
9.	chisquare and Anova	55	19/2/20	AM 11.3
10	non-parametric test	59	26/2/20	

PRATICAL-1.

Basics of R software.

- R is a software for data analysis and statistical computing.
- This software is used for effective data handling and output storage is possible.
- Is capable for graphical display.
- It is free software.

Q.1 $2^2 + \sqrt{25} + 35$

Q.2 $2 \times 5 \times 3 + 62 \div 5 + \sqrt{49}$

Q.3 $\sqrt{76 + 4 \times 2 + 9 \div 5}$

Q.4 $42 + |-10| + 7^2 + 3 \times 9$

solution:

1.
```
x = c (2^2 + sqrt (25)+35)
x
[1] 44
```

2.
```
y=c(2*5 *3 + 62/5 + sqrt(49))
y
[1] 49.4
```

3.
```
z= c(sqrt (76+4 *2+ 2+9/5))
z
[1] 9.262829
```

4.
```
a= c(42 + abs(-10)+ 7^2 +3*9)
a
[1] 128
```

Q.5 $x=20$

$y=30$

find $x+y$, x^2+y^2, $\sqrt{y^3-x^3}$, $|x-y|$

```
c=x+y
c
[1] 50

d = x^2 +y^2
d
[1] 1300

sqrt (y^3 -x^3)
[1] 137.8405

abs (x-y)
[1] 10

c (2,3,4,5)^2
[1]  4 9  16  25

c (4,5, 6,8) *3
[1]  12   15   18  24

c(2,3,5,7) * c (-2,-3,-5,-4)
[1]  -4   -9   -25   -28

c (2,3,5,7) * c (8,9)
[1] 16  27  40  63
```



```
c(2,3,5,7) * c (1,2,3)
warning message
In c (2,3,5,7)* (1,2,3): longer object length is
not a multiple of shorter object length.
```

Q. Find the sum product of maximum, minimum of a values 5, 8, 6, 7, 9, 10, 15, 5

```
x= c (5,8,6,7,9,10,15,5)
>> length (x)
[1] 8
>> sum(x)
[1] 65
>> min(x)
[1] 5
>> prod (x)
[1] 11340000
>> max (x)
[1] 15

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

Q. $x = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$

$y = \begin{bmatrix} 2 & 4 & 10 \\ -2 & 8 & -11 \\ 10 & 6 & 12 \end{bmatrix}$

$x+y$, $x*y$, $2x+3y$

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

> y <- matrix (nrow=3, ncol=3, data=c(2,-2,10,4,8,6,10,...
x+y   [,1] [,2] [,3]
[1,]   4    8    20
[2,]  -4   16   -22
[3,]  20   12    21

> x * y
      [,1] [,2] [,3]
[1,]   4    16   100
[2,]   4    64   121
[3,]  100   36   144

2x+3y [,1] [,2] [,3]
[1,]   10   20   50
[2,]  -10   40  -55
[3,]   50   30   60
```


Q. x = c(2, 4, 6, 1, 3, 5, 7, 18, 16, 14, 17, 19, 3, 3, 2, 5, 0, 15, 9, 14, 18, 10, 12)

```
> length (x)
[1] 23
> a = table (x)
> transform (a)
   x   freq
   0    1
   1    1
   2    2
   3    3
   4    1
   5    2
   6    1
   7    1
   9    1
  10    1
  12    1
  14    2
  15    1
  16    1
  17    1
  18    2
  19    1
> breaks = seq (0, 20, 5)
> b  cut (x, breaks, right= FALSE)
> c= table (b)
> transform (c)   →
         b      freq
  1   [0,5)      8
  2   [5,10)     5
  3   [10,15)    4
  4   [15,20)    6
```

PRACTICAL-2

Problem on Pdf and Cdf

Can the following be pdf ?

(i) $f(x) = \begin{cases} 2-x \;; & 1 \le x < 2 \\ 0 & o.w \end{cases}$

(ii) $f(x) = \begin{cases} 3x^2 \;; & 0 < x < 1 \\ 0 & o.w \end{cases}$

(iii) $f(x) = \begin{cases} \frac{3x}{2}\left(1-\frac{x}{2}\right) ; & 0 \le x \le 2 \\ 0 & o.w \end{cases}$

Solution:

i) $\int f(x)\, dx = 1$

$\Rightarrow \int_1^2 (2-x)\, dx$

$\Rightarrow \int_1^2 2\,dx - \int_1^2 x\,dx$

$\Rightarrow 2x \Big|_1^2 - \frac{x^2}{2}\Big|_1^2$

$\Rightarrow (4-2) - (2-0.5)$

$\neq 1$

$\therefore$ It is not a pdf

ii) $\Rightarrow \int_0^1 (3x^2)\, dx$

$\Rightarrow \cancel{\int_0^1 (3)dx - \int_0^1 (x^2) dx}$ $\qquad \Rightarrow 3\int_0^1 x^2\, dx$

$\Rightarrow \cancel{3x\Big|_0^1 - \frac{x^3}{3}\Big|_0^1}$ $\qquad \Rightarrow 3\cdot\frac{x^3}{3}\Big|_0^1$

$\Rightarrow x^3\Big|_0^1$

$\Rightarrow \cancel{(3 - 0.33)}$ $\qquad \Rightarrow 1$

$\Rightarrow$

It satisfy Pdf

iii) $\frac{3/x}{2} \int_0^2 \frac{3x}{2}\left(\frac{1-x}{2}\right)$

$\frac{3}{2}\int_0^2 x\left(1-\cancel{\frac{x}{2}}\right)$ $\qquad \Rightarrow \int_0^2 \left(\frac{3x}{2} - \frac{3x^2}{4}\right) dx$

$\cancel{\frac{3}{2}\int_0^2 x \int_0^2 \left(1-x/2\right)}$ $\qquad \Rightarrow \int_0^2 \frac{3x}{2}\, dx - \int_0^2 \frac{3x^2}{4}\, dx$

$\cancel{\frac{3}{2}}\int_0^2$ $\qquad \Rightarrow \cancel{\int_0^2 3x}\; \frac{3}{2}\int_0^2 x\, dx \; \frac{3}{4}\int x^2 dx$

$\Rightarrow \frac{3}{2}\Big[\frac{x^2}{2}\Big]_0^2 - \frac{3}{4}\Big[\frac{x^3}{3}\Big]_0^2$

$\Rightarrow \frac{3}{2}\Big[\left(\frac{2^2}{2}\right) - 0\Big] - \frac{3}{4}\Big[\frac{2^3}{3}\Big]$

$\Rightarrow \frac{3}{2} \times \frac{4}{2} - \frac{3}{4} \times \frac{8}{3}$

Pdf $\Rightarrow 3 - 2 = 1$

2) Can the following be pmf?

i)

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

→ Since one probability is negative, hence it is not a pmf

ii)

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

→ since $p(x) \geq 0 \ \forall x$, it is a pmf.

```
> x=c(0,1,2,3,4,5)
> prob= c(0.1,0.3,0.2,0.2,
  0.1,0.1)
> sum(prob)
[1] 1
```

iii)

x	10	20	30	40	50
$p(x)$	0.2	0.3	0.3	0.2	0.2

→ It is not a pmf

```
> x=c(10,20,30,40,50)
> prob=c(0.2,0.3,0.3,0.2,0.2)
> sum(prob)
[1] 12
```

Q3) Find $p(x \leq 2)$, $P(2 \leq x \leq 4)$, $P(\text{atleast } 4)$, $P(3<x<6)$

x	0	1	2	3	4	5	6
$p(x)$	0.1	0.1	0.2	0.2	0.1	0.2	0.1

1) $P(x \leq 2)$

$P(x=2)+P(1)+P(0)$

$0.2+0.1+0.1 = 0.4$

2) $P(2 \leq x \leq 4)$

$\cancel{p(4)} \cdot p(2)+p(3)$

$= 0.2+0.2 = 0.4$

3) $P(\text{atleast } 4)$

$p(4)+p(5)+p(6) = 0.1+0.2+0.1 = 0.4$

4) $P(3<x<6)$

$P(4)+P(5)$

$0.1+0.2 = 0.3$

Q.4 Find cdf.

i)

x	0	1	2	3	4	5	6
$p(x)$	0.1	0.1	0.2	0.2	0.1	0.2	0.1

ii)

x	10	12	14	16	18
$p(x)$	0.2	0.35	0.15	0.2	0.1

solution.

i)

```
prob = c(0.1,0.1,0.2,0.2,0.1,0.2,0.1)
cumsum(prob)
[1] 0.1 0.2 0.4 0.6 0.7 0.9 1.0
```

Ans:

$F(x) = 0$ if $x<0$

$= 0.1$ if $0 \leq x < 1$

$= 0.2$ if $1 \leq x < 2$

$= 0.4$ if $2 \leq x < 3$

$= 0.6$ if $3 \leq x < 4$

$= 0.7$ if $4 \leq x < 5$

$= 0.9$ if $5 \leq x < 6$

37

$= 1 \quad \text{if } x \geq 6$

ii)
```
> prob1 = c(10, 12, 14, 16, 18)
> cumsum(prob)
[1]  10  22  36  52  70
```

Ans:- $F(x) = 10 \quad \text{if } x < 10$

$= 0.2 \quad \text{if } 10 \leq x < 12$

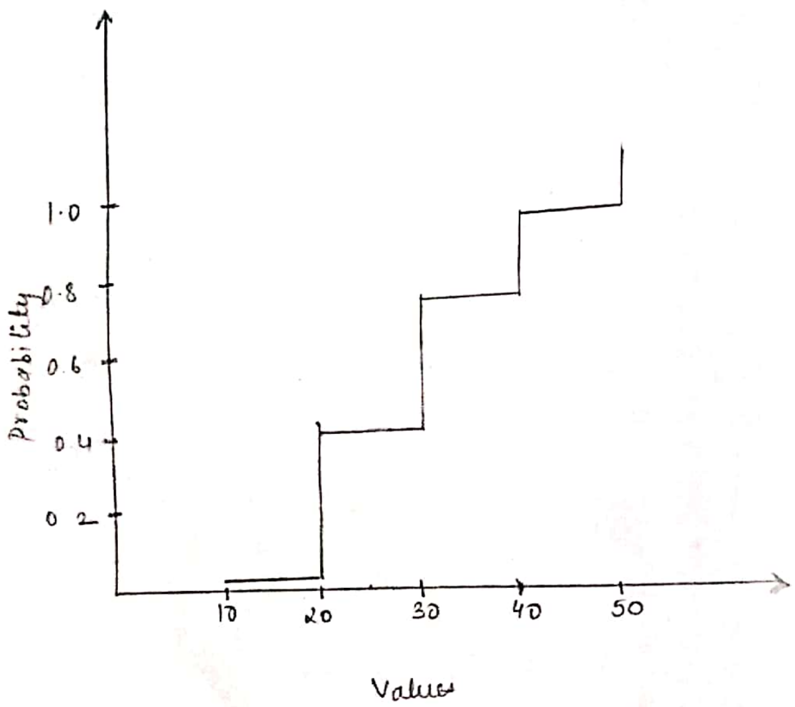
$= 0.35 \quad \text{if } 12 < x < 14$

$= 0.15 \quad \text{if } 14 < x < 16$

$= 0.2 \quad \text{if } 16 < x < 18$

$= 0.1 \quad \text{if } x \geq 18$

graph of cdf



PRACTICAL- 3

Probability Distribution and binomial distribution

1) Find the cdf of the following pmf and draw the graph

x	10	20	30	40	50
p(x)	0.15	0.25	0.3	0.2	0.1

```
> X = c(10,20,30,40,50)
> x
[1] 10  20  30  40  50
> prod = c(0.15,0.25,0.3,0.2,0.1)
> prod
[1] 0.15, 0.25  0.3  0.2  0.1
> cumsum(prod)
[1] 0.15  0.40  0.70  0.90  1.00
```

$F(x) = 0$ if $x < 10$

$= 0.15$ $10 \le x < 20$

$= 0.40$ $20 \le x < 30$

$= 0.70$ $30 \le x < 40$

$= 0.90$ $40 \le x < 50$

$= 1.0$ $x \geq 50$

```
> plot(x, cumsum prod, xlab= "values", ylab=" probability",
  main=" graph of cdf", "s")
```

Binomial distribution

1) Suppose there are 12 MCQ in a test. each question has 5 options and only one of them is correct. Find the probability of having
i. five correct answer
ii. Atmost 4 correct answer

Soln:-
It is given that, $n = 12$, $p = 1/5$, $q = 4/5$
X = Total no. of correct answers
$X \sim B(n, p)$

```
> n=12; p= 1/5 ; q= 4/5; x=5
> dbinom (5, 12, 1/5)
[1] 0.05315022
> pbinom (4,12, 1/5)
[1] 0.9274445
```

2) there are ten members in a Comunity the probability of any member attending a meeting is 0.9. Find the probability
1) Seven members attended
2) Atleast five member attended
3) Atmost six member attended

```
> n= 10; p= 0.9, q=0.1
> dbinom (7, 10, 0.9)
[1] 0. 05739563
```



```
> 1- pbinom(4,10,0.9)
[1] 0.99~~8365~~98531

> pbinom (6,10,0.9)
[1] 0.0127952
```

$p(x \geq 5)$
$= 1 - p(x < 5)$
1- pbinom (5,

3) Find the c.d.f and draw the graph

x	0	1	2	3	4	5	6
$p(x)$	0.1	0.1	0.2	0.2	0.1	0.2	0.1

```
> x= c(0,1,2,3,4,5,6)
> x
[1] 0 1 2 3 4 6
> prob=c(0.1,0.1,0.2,0.2,0.1,0.2,0.1)
> prob
[1] = 0.1 0.1 0.2 0.2 0.1 0.2 0.1
> cumsum (prob)
[1] 0.1 0.2 0.4 0.6 0.7 0.9 1 0
> plot (x, cumsum (prob), xlab=" values", ylab="probability"
  main = " graph of cdf " , " s ")
```

$F(x) = 0$ if $x < 0$

$= 0.1$ $\quad 0 \le x < 1$

$= 0.2$ $\quad 1 \le x < 2$

$= 0.4$ $\quad 2 \le x < 3$

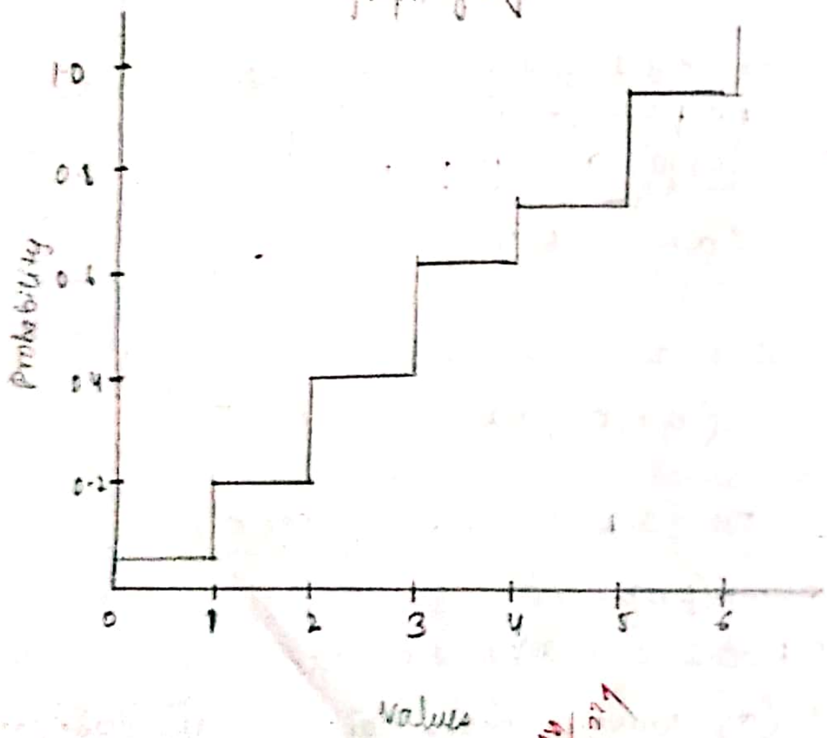
$= 0.6$ $\quad 3 \le x < 4$

$= 0.7$ $\quad 4 \le x < 5$

$= 0.9$ $\quad 5 \le x < 6$

$= 1.0$ $\quad x \ge 6$

graph of cdf



PRATICAL-4

AIM: Binomial distribution

1) Find the complete binomial distribution when $n = 5$ ~~and~~ when $p = 0.1$

2) Find the probability of exactly 10 success in 100 trial with $p = 0.1$

3) X follow binomial distribution with $n = 12$ $p = 0.25$ find i) $p(x = 5)$ ii) $p(x \le 5)$ iii) $p(x > 7)$ iv) $p(5 < x < 7)$

4) The probability of a salesman makes a sell to a customer is 0.15 Find the probability i) no sale for 10 customer ii) more than 3 sale in 20 customers

5) A student writes 5 MCQ each question has 4 options, out of which 1 is correct Calculate the probability for atleast 3 correct answer

Note: $p(x \le x) =$ pbinom(x , n, p)
$p(x > x) =$ 1 - pbinom(x , n, p)
$p(x = x) =$ dbinom(x , n, p)
To find the value of x for which the probability is given as p1
> qbinom (p, n, p)

Solution:

1)
n= 5 ; p= 0.1 ;
> dbinom (0:5, 5 , 0.1)
[1] 0.59049 0.32805 0.07290 0.00810 0.00045 1e

2)
> n= 100;
> x= 10;
> p= 0.1;
> dbinom(10, 100, 0.1)
[1] 0.1318653

3) i) > n = 12 ; p= 0.25 ; x=5;
> dbinom (5, 12, 0.25)
[1] 0.1032414

ii) > n=12 ; p=0.25 ; x=5
> pbinom (5, 12, 0.25)
[1] 0.9455978

iii) $p(x>7) = 1 - p(x \le 7)$
> 1- pbinom (7, 12 , 0.25)
[1] 0.00278151

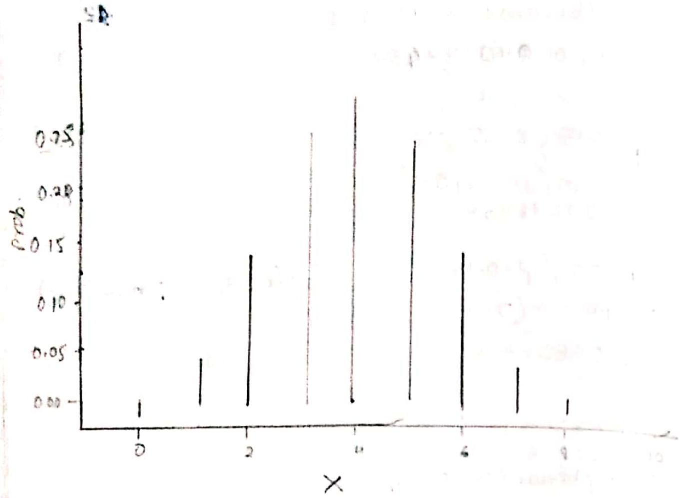
iv) > n = 12; p = 0.15 ; x = 6
> dbinom (6, 12, 0.15
[1] 0.0040149 45

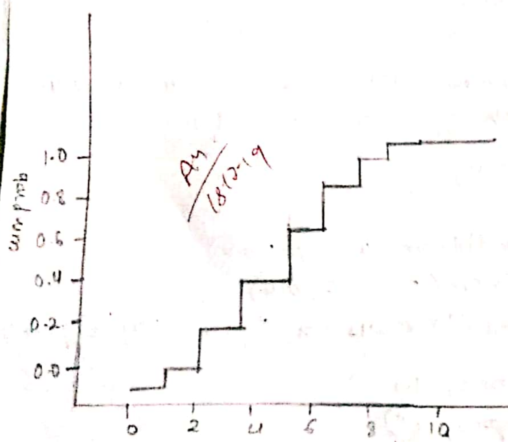
4) i) > p = 0.15 ; n = 10 , x=0;
> dbinom (0, 10, 0.15)
[1] 0.1298337

ii) > n = 20 , p=0.15 ; [$p(x>3) = 1 -$ pbinom $(x \le 3)$
> 1- pbinom(3, 20, 0.15)
[1] 0.3522748

5) $p(x \ge 3) = 1 -$ pbinom$(x \le 2)$
x=2 ; n=5, p=1/4
> 1- pbinom (2 , 5, 1/4)
[1] 0.1035156

Q6) X follows binomial distribution with n=10, p=0.4, plot the graph of pmf and cdf
> n=10 ; p=0.4;
> x = 0:10
> prob= ~~dbni~~ dbinom (0:10, 10, 0.4)
> cumprob= pbinom(0:10 , 10 , 0.4)
> d= data.frame ("x values" = x, "probability" = prob)
> print (d)
> plot (x, prob, "h")
> plot (x, cumprob, "s")





PRATICAL - 5

AIM: NORMAL DISTRIBUTION

1. $P[X=x] = dnorm(x, \mu, \sigma)$
2. $P(X \le x) = pnorm(x, \mu, \sigma)$
3. $P(X > x) = 1 - pnorm(x, \mu, \sigma)$
4. $P(x_1 < x < x_2) = pnorm(x, \mu, \sigma)$
5. To find the value of k so that $P(X \le k) = p_1$; $qnorm(p_1, \mu, \sigma)$
6. To generate n random numbers $rnorm(n, \mu, \sigma)$

Q) 1] $X \sim N(\mu = 50, \sigma^2 = 100)$

find i) $P(X \le 40)$

$P(X > 55)$

$P(42 \le X \le 60)$

$P(X \le k) = 0.7$; $K = ?$

2] $X \sim N(\mu = 100, \sigma^2 = 36)$

i) $P(X \le 110)$

ii) $P(X \le 95)$

iii) $P(X > 115)$

iv) $P(95 \le X \le 105)$

v) $P(X \le K) = 0.4$; $K = ?$

3. Generate 10 random numbers from a normal distribution with ~~mean equals to~~ m= 60, SD = 5. Also calculate sample mean, median, variance and S.D.

4. Draw the graph of standard normal distribution.

Solutions.

Q1)

i) a = pnorm (40, 50,10)

> cat("p(x<=40) is = ", a)

p(x<=40) is = 0.1586553

ii) b= 1- pnorm (55,50,10)

> cat("p (x>55) is =", b)

~~p (x>55) is = ", b)~~ p(x>55) is = 0.3085375

iii) c= pnorm(60,50,10) - pnorm (2,50,10)

> cat(" p(42<x< 60) is = ", c)

p (42 <x<60) is = 0.841344

iv) d= qnorm (0.7, 50,10)

>cat(" p(x<=k)= 0.7, k is = ", d)

p(x<= k)=0.7, k is = 55. 24401

Q2)

i) >e = pnorm(110, 100, 6)

> cat (" p (x<=100) is =", e)

p (x <= 110)is = 0.9522096

ii)> f= pnorm(95,100,6)

> cat (" p (x<= 95) is=", f)

p (x< =95 is = 0.2023284

iii) > g=1-pnorm (115,100,6)
> cat ("p(x>115) is=", g)
p(x>115) is = 0.006209665

iv)> h= pnorm (95,100,6) - pnorm(105,100,6)
> cat (" p (95<=x<=105) is = ", h)
p (95<=x<=105) is =-0.5953432

v) > i= qnorm (0.4, 100, 6)
> cat (" p(x<=k)=0.4, k is =", i)
p (x<=k)=0.4, k is = 9.847992

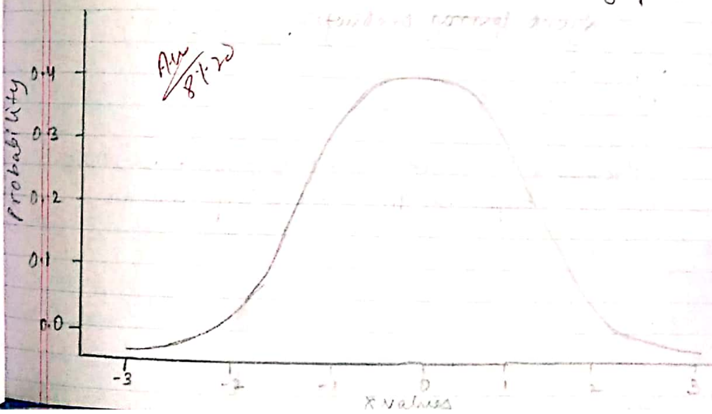
Q3
n= 10
m= 60
SD=5
> x= rnorm (10,60,5)
> x
[1] 60.746774 57.11559 63.74548 52.63748 55.41122 63.57193 57.71972 66.89140 59.01766 61.89085

[1] 58.21355 61.95681 52.56121 55.72794 57.96677
54.41589 50.60946 57.92217 55.58918 57.85833

> n=10
> ~~r~~= 60
>sd=5

> am= mean(x)
> am
[1] 56.28213
> m= median(x)
> m
> 56.79313
> var= (n-1)* var (x)/n
> var
[1] 9.423902
> sd= sqrt(var)
> sd
[1] 3.069837

4) > x= seq(-3, 3, by= 0.1)
> y= dnorm (x)
> plot (x, y, xlab=" x values", ylab=" probability', main=" standard Normal graph ')



A.W. 8/1/20

71

PRACTICAL - 6

Topic - Z- Distribution.

- Test for hypothesis

Q1 $H_0: \mu=10$ against $H_1: \mu \neq 10$. A sample of size 400 is selected which gives the mean = 10.2 and SD = 2.25. Test the hypothesis at 5 pcent level of significance

```
> mo = 10;          (mean population)
> mx = 10.2;        (mean sample)
> n = 400;
> sd = 2.25;
> zcal = (mx-mo)/(sd/(sqrt(n)))
> cat ("Zcal is =", zcal)
[1] 1.777778
> pvalue = 2* (1-pnorm(abs(zcal)))
> pvalue
[1] 0.07544036
```

∴ ~~Hypothesis~~ If result tested hypothesis is > 0.05 assumed hypothesis $H_0: \mu=10$ is expected as verified.

Q2 Test the hypothesis $H_0: \mu = 75$ against $H_1: \mu \neq 75$.

A sample of size a 100 is selected by & the sample mean is 80 with sd = 3. Test the hypothesis at 5% level of significance

```
> mo = 75;
> mx = 80
> n = 100
> sd = 3
> zcal = (mx - mo)/(sd/(sqrt(n)))
> cat("Zcal is =", zcal)
  zcal is ~~1.77777~~ 16.66667
> pvalue = 2* (1- pnorm(abs(zcal)))
> pvalue
  [1] 0
```

Q.3 Test the hypothesis $H_0: \mu = 25$ against $H_1: \mu \neq 25$ At 5% level of significance. The following of sample of 30 is selected

20, 24, 27, 35, 30, 46, 26, 29, 10, 20, 30, 37, 35, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 39, 27, 15, 19, 22, 20, 18

```
> x = c( 20, 24, 27, 35, 30, 46, 26, 29, 10, 20, 30, 37, 35, 21, 22,
  23, 24, 25, 26, 27, 28, 29, 30, 39, 27, 15, 19, 22, 20, 18)
> mx = mean(x)
> mx
[1] 26.06667
```

```
> n= length (x)
> n
[1] 30
> variance= (n-1)* var (x)/n
> variance
[1] 52.94556
> sd = sqrt (variance)
> sd
[1] 7.27905
> zcal = (mx-mo)/(sd/(sqrt(n)))
> zcal
[1] 0.802
> pvalue= 2* (1-pnorm(abs(zcal)))
> pvalue
[1] 0.422
```

Q.4 Experience has shown that 20% students in a college smoke. A sample of 400 students reveal that out of 400 only 50 smoke test the hypothesis that the experience gives the correct proportion or not.

```
> P= 0.2
> Q= 1-P
> p= 50/400
> n=400
> zcal = (p-P)/(sqrt(P*Q/n)))
> zcal
[1] -3.75
> pvalue= 2* (1-pnorm (abs(zcal)))
```


Q.5 Test the hypothesis $H_0 : P = 0.5$ against $H_1 : P \neq 0.5$. A sample of 200 is selected and sample proportion is calculated p= 0.56 test the hypothesis at the 1% level of significance.

```
> P= 0.5
> p= 0.56
> Q = 1-P
> n= 200
> zcal = (p-P)/(sqrt(P*Q/n))}
zcal
[1] 1.697056
> pvalue= 2* (1-pnorm (abs (zcal)))
> pvalue
[1] 0.08968602
```

∴ The value of p hypothesis is greater than 0.01 the hypothesis is accepted.

Am
22.01.20

PRATICAL- 7

Topic large sample test

Q.1 A study of noise level in 2 hospitals is calculated below the test hypothesis that the noise level into hospitals are same or not.

	HOS A	HOS B
no of sample obs	84	34
mean	61	59
S.D.	7	8

H_0 : The noise levels are same

$n_1 = 84$
$n_2 = 34$
mx=61
my=59
sdx=7
sdy=8
z= (mx-my)/sqrt((sdx^2/n1)+(sdy^2/n2))
cat ("z calculated is : ", z)
pvalue= 2 * (1- pnorm (abs (z)))
Since value is < 0.0051 we reject H_0 at 5.% level of significance.
z= 1.273
pvalue = 0.1291361

48

2) 2 Random samples of size 1000 and 2000 are ~~mean than~~ drawn from 2 population which mean of 67.5 and 68 respectively and the same S.D of 2.5 test the hypothesis that the mean of 2 population are equal.

	A	B
no. of obs	1000	2000
mean	67.5	68
SD	2.5	2.5

```
> n1=1000
> n2 = 2000
> mx= 67.5
> my= 68
> sdx= 2.5
> sdy= 2.5
> z (mx-my)/ sqrt (sdx ^ 2/n1)+(sdy^2/n2))
= -5.163
pvalue = 2* (1-pnorm (abs (z)))
= 2.41 * 5.64e-07.
```

3) In FYBSC 20% of a random sample of 400 students had defective eyesight, the SYBSC class had 15.5% of 500 sample had the same ~~effect~~ defect; is the difference of proportion is same?

H_0 the proportion of the population are equal

```
> n1 = 400
> n2 = 500
> p1 = 0.2
> p = 0.
> p = (n1 * p1 + n2 * p2) / (n1+n2)
  = 0.175.
> q = 1-p
  = 0.825
> z = (p1-p2) / sqrt ( p* q * (1/n1 + 1/n2))
> z = 1.76547.
  cat ( " z calculated is ", z)
> pvalue = 2* (1- pnorm-(abs (z)))
  = 0.07748
```


a) In a class out of the sample of 60 mean height is 63.5 inch, with SD = 2.5. In an M com, out of 50 student, mean height is 69.5 inch with 2.5. Test hypothesis the mean of MA class are same.

H_0: The mean of MA & M.com ~~are~~ class are same

```
> n1 = 60
> n2 = 50
> mx = 63.5
> my = 69.5
> sdx = 2.5
> sdy = 2.5
> z = (mx-my) / sqrt ((sdx^2/n1)+(sdx^2/n2))
> cat (" z calculated : ", z)
  [1] z calculated z = -12.53359
> pvalue = 2* (1-pnorm (abs (z)))
> pvalue
  [1] 0.
```

∴ pvalue < 0.05, we ~~reject~~ reject the H_0 at 5% level of significance.

Q5 From each of box of apples, a sample size of 200 is collected. It is found that there are 44 bad apples in the first sample & 30 bad apples in second sample. Test the hypothesis that 2 boxes are equivalent in terms of no bad apples.

H_0: The two boxes are equivalent in terms of no. of bad apples

```
> n1 = 200
> n2 = 200
> p1 = 44/200
> p2 = 30/200
> p = (n1* p1 + n2* p2)/(n1+n2)
[1] 0.185
> q = 1-p
> q
[1] 0.815
> z = (p1-p2) / sqrt (p* q *(1/n1 + 1/n2))
> cat ('zcalculate = 1.802741
> pvalue = 2* (1- pnorm (abs(z)))
> pvalue
[1] 0.07142888
```

∴ pvalue > 0.05, we accept the H_0 at 5% level of significance.

Scanned with CamScanner

PRATICAL-8

Small sample test.

Q. The flowers stems are selected and the heights are found to be 63, 63, 68, 69, 71, 71, 72 cms. Test the hypothesis that the mean height is 66 cms or not at 1% l.o.s.

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

One sample t-test

data: x
t= 47.94, df=6, pvalue= 5.522e-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.

∴ Since ~~null hypothesis~~ is pvalue is less than 0.01 we reject H_0 ~~Ho~~ at 1%.

Q. Two random sample were drawn from two different population

Sample 1: 8, 10, 12, 11, 16, 15, 18, 7
Sample 2: 20, 15, 18, 9, 8, 10, 11, 12
Test the hypothesis that there is no difference between the two population mean at 5% los.

H_0 - There is difference in the population means.

Welch Two sample t-test

data = x and y
t = -0.36277, df = 13.837; p-value = 0.7225
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-5.192719 3.692719
sample estimates:
mean of x means of y
12.125 12.875
∴ Since pvalue is more than ~~5%~~ 0.05 we accept H_0 at 5%.

Q. Following are the weights of 10 people before and after a diet program. 52

Test the hypothesis that a diet program is effective or not.
Before after (kg)
Before: (100, 125, 95, 96, 98, ~~88~~ 112, 115, 104, 109, 110)

After: (95, 80, 95, 98, 90, 100, 110, 85, 100, 101)

soln: H_0: the diet program is not effective.

> x = c(100, 125, 95, 96, 98, 112, 115, 104, 109, 110)
> y = c(95, 80, 95, 98, 90, 100, 110, 85, 100, 101)
> t.test (x, y, paired = T, alternative = "less")

paired t-test

data: x and y
t = 2.6089, df = 9 ; p-value = 0.9858
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
-Inf 18.729088
sample estimates:
mean of the differences
11
∴ Since pvalue is ~~more than~~ less than 0.05 we accept H_0 at that diet program is not effective.

Q: The marks before and after a training program are given below.

before: (20, 25, 32, 28, 27, 36, 35, 25)

After: (30, 35, 42, 37, 37, 40, 40, 23)

Test the hypothesis that training program is effective or not.

Soln: H_0: Training program is not effective.

```
> x = c(20, 25, 32, 28, 27, 36, 35, 25)
> y = c(30, 35, 42, ~~33~~ 37, 37, 40, 40, 23)
> t.test(x, y, paired = T, alternative = "greater")

          paired t-test

data : x and y
t = -4.5252 , df = 7 , p-value = 0.9986
alternative hypothesis : true difference in means is
                         greater than 0
95 percent confidence interval:
 -9.930698      Inf
sample estimates:
mean of the differences
        -7
```

=> pvalue is greater than 0.05, Hence H_0 is accepted
∴ Training program is not effective.

Q. 2 random sample were drawn from a normal population and the 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 same variance at 5% level of significance.

H_0: ~~Two populati~~ The variances of two population are equal.

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

     F test to compare two variances

data : A and B
F = 0.76421 , num df = 8 , ~~denote~~ denom df = 10 ,
    p-value = 0.7192
alternative hypothesis : true ratio of variances
is not equal to 1
95 % confidence interval ~~tend~~:
0.1982442     3.2823785
sample estimates:
ratio of variance
~~0.7624099~~ 0.7642099.
```

∴ Since pvalue > 0.05 we accept H_0 at 5% level of significance

Q. The arithmatic mean of sample of 100 observations is 52. If the standard Deviation is 7 test the hypothesis that population is 55 or not at 5% los.

```
> n= 100
> mx= 52
> mo= 55
> sd = 7
> zcal = (mx-mo)/(sd/sqrt(n))
> zcal
[1] -4.285
> pvalue = 2* (1-pnorm(abs(zcal)))
> pvalue
[1] 1.82153e-05
```

Since pvalue < 0.05, Hence H_0 we reject H_0 at 5% level of significance.

PRATICAL-9.

chi-square test. and ANOVA

Q. Use the following data to test whether the cleanliness of the home depends upon the child conditions or not.

		Condition of home	
		clean	Dirty
cond of child	clean	70	50
	Fairly clean	80	20
	Dirty	35	45

H_0 = condition of home and 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
> pvalue = chisq.test(y)
> pvalue
          Pearson's chi-squared ~~value~~ Test
data: y
X-squared = 25.646, df = 2, p-value = 2.698e-06
```

∴ ~~As~~ Since pvalue is less than 0.05 ∴ ~~reject~~, we reject H_0 at 5% LOS.

Q. Table below shows the relation ~~b/w~~ performance ~~b/w~~ b/w Mathematics and computer

		Maths		
		HG	MG	LG
	HG	56	71	12
comp	MG	47	163	38
	LG	14	42	85

H_0 = performance in math and computer.

```
> x = c(56, 47, 14, 71, 163, 42, 12, 38, 85)
> m=3
> n=3
> y= matrix (x, nrow=m, ncol=n)
> y
      [,1]  [,2]  [,3]
[1,]   56    71    12
[2,]   47   163    38
[3,]   14    42    85
> pvalue= chisq.test (y)
> pvalue
        pearson's chisquared test
data: y
X-squared = 145.78, df=4, pvalue < 2.2e-16
```

∴ Since pvalue less than 0.05 we reject H_0 at 5% los

Q. Perform Anova for the following data.

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

~~means A, B, C, D~~

H_0 = The means of variety A, B, C, D 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 = 12.779, num df = 3, denom df = 10, p-value = 0.0009

> anova = aov (values ~ ind, data = d)
> summary (anova)
           Df   Sum Sq   Mean Sq   F value   Pr(>F)
ind         3   92.96    30.988    12.78     0.000933 ***
Residuals  10   24.25    2.425
```

Hence pvalue is less than 0.05, we reject at 5% level.

Q. Peform Anova

Types	observations
A	6, 7, 8
B	4, 6, 5
C	8, 6, 10
D	6, 9, 9

```
> x1 = c(6,7,8)
> x2 = c( 4,6,5)
> x3 = c(8,6,10)
> x4 = c(6,9,9)
> 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= 2.6667 , num df = 3 , denom df =8, pvalue=0.1189
> anova = aov (values ~ ind , data = d)
> summary (anova)
            Df    sumsq   Mean Sq   F value  Pr(>F)
ind         3     18      6.00      2.667    0.119
Residuals   8     18      2.25
```


Q. Type the data in Excel and save as in desktop filename . CSV (MS-DOS)

Then open R-software and type x= read.csv(" ")

Minimize R-software then right click the on the file saved on desktop, click on properties, copy the location and paste it on

X= read.csv(" (location) ')

Replace \ this with / then after location add ~~marks.csv~~ filename.CSV)

Enter x

The data written in Excel sheet will be seen on R-software.

PRATICAL-10.

Topic: Non parametric test.

Q. Following are the amongs of sulphur oxide emitted by factory. 17, 15, 20, 29, 19, 22, 25, 27, 9, 24, 20, 17, 6, 24, 14, 15, 23, 24, 26 apply sign to test the types that population medium is 21.5 against the alternative less than 21.5

H_0: population median = 21.5.
H_1: less than 21.5

x: c (17, 15, 20, 29, 19, 22, 25, 27, 9, 24, 20, 17, 6, 24, 14, 15, 23, 24, 26).

```
> m= 21.5
> sp= length(x[x>m])
> sp
[1] 9
> sn = length (x[x<m])
> sn
[1] 20
> pv = pbinom (sp, n, 0.5)
> pv
[1] 0.4119015
```

Since pvalue is greater than 0.05 we accept.

Note :- If the alternative is greater than median
pv= pbinom (sn, n, 0.5)

Q.2 For the observation (12, 19, 31, 28, 43, 40, 55, 49, 70, 63) apply sign test to test population median is 25 against the alternative. It is more than 25.

```
> x= c (12, 19, 31, 28, 43, 40, 55, 49, 70, 63)
> m=25
> sp= length ( x[x>m])
  sn= length ([x<m])
  n= sp+sn
  pv= pbinom (sp, n, 0.5)
  pv= 0.1+1
> sp
  [1] 7
> sn = length ([x<m])
> sn
  [1] 2
> n= sp +sn
> pv= pbinom (sn,n,0.5)
> pv
  [1] 0.0546875.
```

since pvalue > 0.005, we accept at 5% LOS.

(Q3) For the following data (60, 65, 63, 89, 61, 71, 58, 51, 48, 66), Test the hypothesis using willcokson Test. for testing the alternative it is greater than 60.

> H_0: median is 60

> H_1: It is greater than 60

```
> x= c( 60, 65, 63, 80, 89 61, 71, 58, 48, 66)
> m0=60
  wilcox test (x.alter = 'greater', mu= 60)
  wilcox signed rank test with continuity correction
data: x
  x= 29 , p-value = 0.2386
alternative hypothesis true location is greater than 60
```

since pvalue > 0.05, we accept at 5% LOS.

Q.4 Test the hypothesis, median is 12 against the alternative that is less than 12 using wilcoxons value = 12, 13, 10, 30, 15, 5, 1, 7, 6, 11, 9, 10

H_0 : median is 12

H_1 : It is less than 12

> x = c(12, 13, 10, 30, 15, 5, 1, 7, 6, 11, 9, 10)

> mu = 12

wilcox Test (x, alter = "less", mu = 12)

wilcoxon signed rank test with continuity correction

data : x

x = 25   pvalue = 0.2521

alternative hypothesis true location less than 12

Since, pvalue > 0.05 we accept the H_0 at 5% los