★★ INDEX ★★

No.	Title	Page No.	Date	Staff Member's Signature
1]	Basic of R software		26/11/19	
2]	Probability distribution		3/12/19	
3]	Binomial distribution		10/12/19	
4]	Normal Distribution		17/12/19	
5)	Normal and t-test		7/1/19	
6)	large sample test		14/1/19	
7]	Small sample test		21/1/20	
8]	large and small test		04/2/20	
9]	Non - Parametric tests		14/2/20	

PRACTICAL - I

Aim : Basic of R software

1) R is a software for stastical analysis and data computing.

2) It is an effective data handling software and Outcome storage is possible.

3) It is capable of graphicle display.

4) It is a free software.

Q.1. Solve the followings.

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

```
> 4+6+8/2-5
[1] 9
```

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

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

3] $\sqrt{4^2+5\times3+7\div6}$

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

4] round off
$46 \div 7 + 9 \times 8$

```
> round (46/7 +9* 8)
[1] 79
```

Q.2

```
> c(2,3,5,7) * 2
[1] 4 6 10 14

> c (2,3,5,7) * c(2,3,6,2)
[1] 4 9 30 14

> c(6,2,7,5)/c(4,5)
[1] 1.50  0.40  1.75  1.00

> c(4,6,8,9,4,5) ^ c(1,2,3)
[1] 4 36 512 9 16 25
```

Q.3.

```
> x=30  > y=30   > z=2
> x^2 + y^3 + z
[1] 27402
```

```
sqrt(x^2 +y)
[1] 20.73644
> x^2 + y^2
[1] 1300
```


Q.4

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

Q.5 find x+y and 2x+3y where $x=\begin{bmatrix} 4 & -2 & 6 \\ 7 & 0 & 7 \\ 9 & -5 & 3 \end{bmatrix}$

$y=\begin{bmatrix} 10 & -5 & 7 \\ 12 & -4 & 9 \\ 15 & -6 & 5 \end{bmatrix}$

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

```
> x+y   [,1]  [,2]  [,3]
[1,]     14   -19    33
[2,]     19   -12    41
[3,]     24   -28    21

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

Q6. marks of statics of cs Batch A

```
x=c(59,20,35,24,46,56,55,45,27,22,47,58
    54,40,50,32,36,29,35,39)
> x=c(data)
> breaks=seq(20,60,5)
> a=alt(x,break,right=FALSE)
> b=table(a)
> c=transform(b)
> c
```

	a	Feq
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

PRACTICAL = 2

Aim: Proabability Distribution.

Q.1 Check whether the following are p.m.d or not.

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

If the given ~~data~~ is p.m.d then $\Sigma P(x) = 1$

$\therefore P(0)+P(1)+P(2)+P(3)+P(4)+P(5) = P(x)$

$= 0.1 + 0.2 + (-0.5) + 0.4 + 0.3 + 0.5$

$= 1.0$

$\because P(2) = -0.5$ it can be a probability mass function

$\because P(x) \geq 0 \quad \forall, x$

2]

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

The condition for p.m.f is

1) $p(x) \geq 0 \quad \forall x$ satisfy
2) $\sum p(x) = 1$

$\sum p(x) = p(10) + p(20) + p(30) + p(40) + p(50)$

$= 0.2 + 0.2 + 0.35 + 0.15 + 0.1$

$= 1$

The given data is p.m.f

code:

```
> Prob = c( 0.2, 0.2, 0.35, 0.15, 0.1)
> sum ( prob)
[1] 1
```

2. find the 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

$F(x) = 0 \quad x < 10$

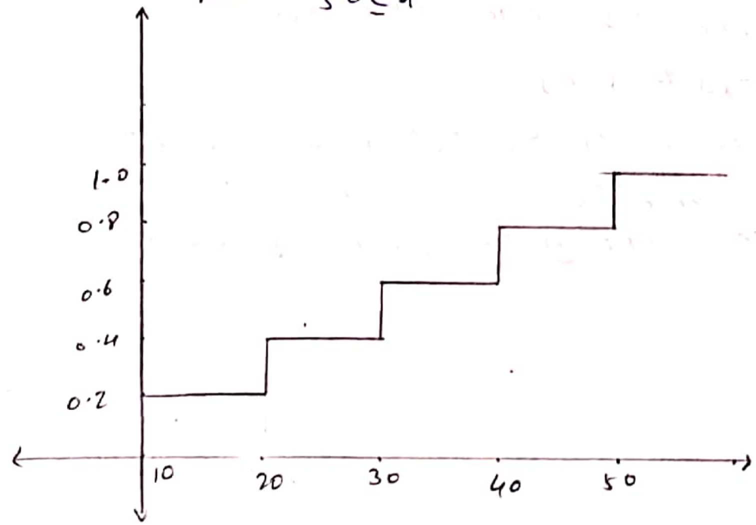
$0.2 \quad 10 \leq x < 20$

$0.4 \quad 20 \leq x < 30$

$0.75 \quad 30 \leq x < 40$

$0.90 \quad 40 \leq x < 50$

$1.0 \quad 50 \leq x$




```
> x = c (10,20,30, 40,50)
> plot( x, cumsum (prob), "s")
```

Q.3. find

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

$F(x) = 0 \quad x < 1$

$= 0.15 \quad 1 \leq x < 2$

$= 0.40 \quad 2 \leq x < 3$

$= 0.50 \quad 3 \leq x < 4$

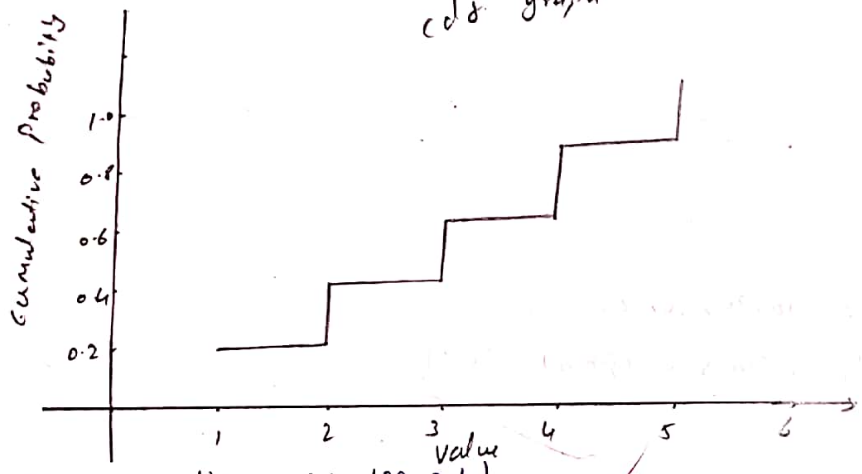
$= 0.70 \quad 4 \leq x < 5$

$= 0.90 \quad 5 \leq x < 6$

$= 1.00 \quad x \geq 6$

```
> prob = c( 0.15, 0.25, 0.1, 0.2, 0.2, 0.1)
> sum(prob)
[1] 2
> cumsum(prob)
[1] 0.15 , 0.40   0.5   0.10   0.90   1.00
> x = c (1,2,3,4,5,6)
> plot (x, cumsum (prob) ,"s", xlab= "value"
     ylab= "Cumulative probability "
     main = "CDF graph", col= "brown")
```

cdf graph



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

2] dbinom (4, 12, 0.2)
[1] 0.1328756

3] dbinom (0:5, 5, 0.1)

0 - 0.59049
1 - 0.32805
2 - 0.07290
3 - 0.00810
4 - 0.00045
5 - 0.00001

PRACTICAL = 3

Aim: Binomial distribution

$P(X=x)$ = dbinom (x, n, p)
$P(X \le x)$ = pbinom (x, n, p)
$P(X > x)$ = 1 - pbinom (x, n, p)
If x is unknown
$p_1 = P(X \le x)$ = qbinom (p_1, n, p)

1] find the probability of exactly 10 success in hundred trials with p = 0.1

2] Suppose there are 12 mcq, Each question has 5 options out of which 1 is correct. find the probability of having exactly 4 correct answers.
i) Atmost 4 correct answers
ii) More than 5 correct answers.

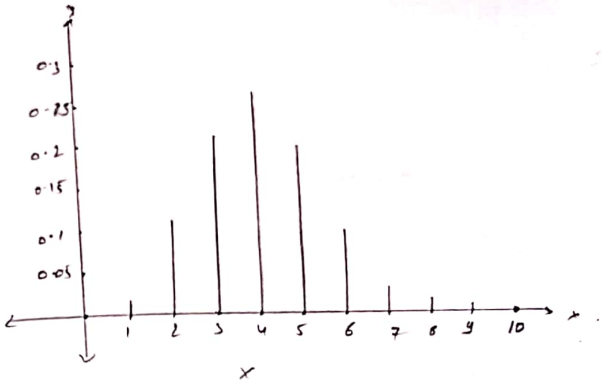
3] find the complete distribution when n = 5 and p = 0.1

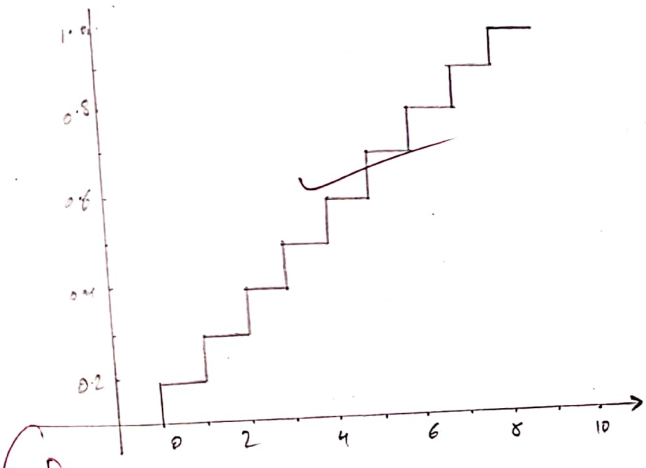
4] The probability of a Salesman making a sale to customer 0.15 find the probability
1) No sales out of 10 customer.
2) More than 3 sales out of 20 customer.

3] > dbinom (0, 10, 0.15)
[1] 0.1968744
> 1 - pbinom (3, 20, 0.15)
[1] 0.3522798

5] qbinom (0.88, 20, 0.2)
[1] 9
> n = 10
> p = .3
> x = 0:n
> prob = dbinom (x, n, p)
> cumprob = pbinom (x, n, p)
> d = data frame ("x value", x", "probability = prob)
> print (d)

	x value	probability
1]	0	0.0282
2]	1	0.1210
3]	2	0.2334
4]	3	0.2668
5]	4	0.2001
6]	5	0.1029
7]	6	0.0367
8]	7	0.0090
9]	8	0.0014
10]	9	0.0001
11]	10	0.0000





Practical = 4

Aim : Normal Distribution

i) $P(x = x)$ = dnorm (x, μ, σ)
ii) $P(X \le x)$ = pnorm (x, μ, σ)
iii) $P(X > x)$ = 1 - pnorm (x, μ, σ)
iv) To generate random number from a normal distribution (n random numbers) the R code is rnorm (n, μ, σ)

Q.1. A random variable X follows normal distribution with mean = μ = 12 and S.D = σ = 3 find

i) $P(X \le 15)$
ii) $P(10 \le X \le 13)$
iii) $P(X > 14)$
iv) generate 5 observation (random variable)

Code :

```
> P1 = pnorm ( 15, 12, 3 )
> P1
[1]   0.8413447
> cat ( " P(X <=15) = ", P1 )
P( X <= 15 ) = 0.8413447
> P2 = pnorm (13, 12, 3) - pnorm (10, 12, 3)
> P2
[1] 0.3780661
```

8

```
>cat ("P(10<=x<=13) = ", P2)
 P(10<=x<=13) = 0.3780661
>P3 = 1 - pnorm(14,12,3)
>P3
[1] 0.2524925
>cat ("P(x>14) = ", P3)
 P(x>14) = 0.2524925
>P4 = rnorm (5,12,3)
>P4
[1] 15.25413   16.548505   12.272460
    11.280515  6.419844
```

Q.2. Generate 5 random number from a normal distribution $\mu = 15$, $\sigma = 4$ find sample mean, median, S.D and print it.

Code :

```
> rnorm (15,15,4)
[1] 10.7649   7.793249   9.963444   13.3459   17.504
> am = mean (x)
> am
[1] 11.87345
> cat (" Sample mean is = ", am)
 Sample mean = 11.87345
> mc = median
> mc
[1] 10.76499
```



```
> cat ("Sample median = ", mc)
> n=5
> v=(n-1)* var(x)/n
> v
[1] 11.09965
> S.D= sqrt(v)
  S.D
[1] 3.33163
> cat ("SD= ", SD)
> SD = 3.33163
```

Q.4. $X \sim N(30,100)$, $\sigma = 10$

i) $P(X \le 40)$; ii) $P(X > 35)$ iii) $P(25 < x < 35)$

iv) find k such that $P(X < k) = 0.6$

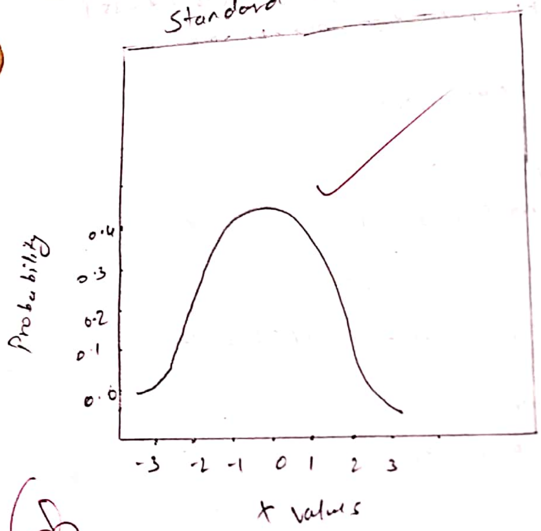
code :

```
> d1 = pnorm (40, 30, 10)
> d1
[1] 0.8413444
> d2 = 1 - pnorm (35, 30, 10)
> d2
[1] 0.30853
> d3 = pnorm (25,30,10) - pnorm (35, 30, 10)
> d3
[1]   0.3829244
> d4 = qnorm (0.6,30,10)
> d4
[1] 32.53317
```

Q.5 plot standard normal graph

```
> x = seq (-3, 3, by = 0.1)
> y = dnorm (x)
> plot (x, y, xlab = "x value", ylab = "probability",
     main = "Standard normal graph")
```

Standard normal graph



Practical = 5

Aim = Normal and t-test

1] $H_0 : \mu = 15$ $H_1 : \mu \neq 15$

Test the hypothesis

Random sample of size 400 is drawn and it is calculated the sample mean is 14 and S.D is 3 test the hypothesis at 5% level of significance at 5%

```
# 0.05 > accept the value
# 0.05 < less the value
> m0 = 15
> mx = 14
> n = 400
> sd = 3
> zcal = (mx - m0) / [sd / sqrt(n)]
> zcal
[1] -6.667
> cat ("calculated value of z is = ", zcal)
  Calculated values Z is = -6.666
> Pvalue = 2 * (1 - Pnorm (abs (zcal)))
> pvalue
[1] 2.616796
```

The value is less than 0.05 we will reject the value of $H_0 = \mu = 15$

11.

Q.2 Test the hypothesis Ho: proportion of smokers in college is 0.2. A Sample is collected and calculated the sample as 0.125 test the hypothesis of 5% level of significance (sample size is 400)

```
> P=0.2
> p=0.125
> n=400
> Q=1-P
> zcal = (p-P) (sqrt (P* Q(n)))
> cat ("calculated value of z is:", zcal )
[1] calculated value of z is -3.75
> pvalue = 2* [1- pnorm (abs (zcal))]
pvalue
[1] 0.000176834 (Reject)
```

Q.3) last year farmer's lost 20% of their crops. A random sample of 60 fields are collected and it test hypothesis of 1% level of significance.

```
>P=0.2
> p=9/60
```



```
> n=60
> zcal =( p-P)/ [sqrt(P* Q /n )]
> zcal
[1] -0.96824
> pvalue= 2 * (1- pnorm (abs (zcal) )]
> pvalue
[1] 0.332916
```

Q.4 Test the hypothesis Ho: µ=12.5 from the followed Sample at 5% level of significance

```
> x= c (12.25, 11.97, 12.25, 12.08, 12.31, 12.28, 11.94
        11.84, 12.16, 12.04)
> n= length (x)
> n
[1] 10
> mx = mean (x)
>mx
[1] 12.107
> variance= (n-1)* var(x)/n
> variance
[1] 0.019521
> sd = sqrt (variance)
> sd
[1] 0.1397176
```

```
> t = (mx - m0) / (sd/sqrt (n))
> t
[1] - 8.53503
> pvalue = 2 * (1 - pnorm (abs (t)))
> pvalue
> [1] 0
```

∴ This value is less than 0.05
the value is accepted.

Practical = 6

Aim : Large Sample test

1] Let the population mean (the amount spent per customber in a restorant is 250. A sample of 100 customber selected the Sample mean is customber of ~~100~~ 275 custo. and sp 30 test the hypothesis that the population mean is 250 or not on 5% level of significants.

2] In a random sample of 100 students it is found that 250 use value pen b/w test the hypothesis that the population proportion is 0.8 at 1% level of significance.

Solu:-

1]
```
> m0 = 250
> mx = 275
> sd = 30
> n = 100
> zcal = (mx - m0) / [sd / (sqrt (n))]
> cat (" calculated value is z is = ", zcal)
[1] calculated value is z is = 8.333
> pvalue = 2 * (1-pnorm (abs (zcal)))
> pvalue
[1] 0
```
∴ The value is less that 0.05 we will return $H_0 = \mu = 250$

```
2] > p = 0.8
> Q = 1-P
> p = 750/1000
> n = 1000
> zcal = (p-P)/(sqrt(P* Q/n))
> cat ("calculated value is Z is = ";zcal)
[1] calculated value of Z = -3.95528
> pvalue = 2 * (1-pnorm (abs (zcal)))
> pvalue
[1] 7.77628
```

∴ The value is less 0.01 use reject!

3] A study sample at size 1000 & 2000 are drawn from two population with same sd 2.5 the sample mean are 67.5 and 68 Test hypothesis $\mu_1 = \mu_2$ at 5% of significance

4] A study of noise level in 2 hospital given below test the claim that 2 hospital have same level of noise at 1% level of significance.

Hos A	Hos B
84	84
61.2	59.4
7.9	7.5


```
3] > n1 = 1000
> n2 = 2000
> mx1 = 67.5
> mx2 = 68
> sd1 = 2.5
> sd2 = 2.5
> zcal = (mx1-mx2)/sqrt (sd1^2/n1) + (sd2^2 /n2))
> zcal
[1] -5.16397
> pvalue = 2 * (1-pnorm (abs( zcal )))
> pvalue
[1] 2.417564 e-07
```

n Rejected.

```
4] > n1 = 84
> n2 = 34
> mx1 = 61.2
> mx2 = 59.4
> sd1 = 7.9
> sd2 = 7.5
> zcal = (mx1 - mx2) / sqrt( (sd1^2/n1)+ (sd2^2))
> zcal
[1] 1.16252
> pvalue = 2* [1- pnorm (abs (zcal)))
> pvalue
[1] 0.2450211
```

∴ The value is greater than 0.01 we accept the value.

Practical = 7

Topic : Small Sample test

Q1. The marks of 10 students are given by 63, 63, 66, 67, 68, 69, 70, 71, 72 test the hypothesis that the sample comes from the population with average 66.

Sol: $H_0 : \mu = 66$

```
> x = c(66, 63, 66, 67, 68, 69, 70, 70, 71)
> t.test (x)
```

one sample t-test

data : x

t = 68.319 , df = 9 , pvalue = $1.558e^{-13}$

alternative hypothesis

three mean is not equal to 0

95 percent confidence interval

65.65171 70.14829

Sample estimates

mean of x

67.9

∴ The pvalue is less than 0.05 we reject the hypothesis at 5% level of significance

2 Two groups of students scored the following marks test the hypothesis that there is no significant difference between the 2 groups.

GR1 = 18, 22, 21, 17, 20, 17, 23, 20, 22, 21
GR2 = 16, 20, 14, 21, 20, 18, 13, 15, 17, 21

Ho : There is not difference between the 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
alternative hypothesis :
True difference is 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


```
> p-value = 0.03798
> if (pvalue > 0.05) { cat ("accept ") }
else { cat ("reject ") }
reject
```

* PAIRED T-Test

Q.3. The sales data of 6 shops before & after a special campaign are given below.

Before : 53, 28, 31, 48, 50, 42
After : 58, 29, 30, 55, 56, 45

Test the hypothesis that the campaign is affective or not.

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

```
> 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 & y
t = -2.7815 df = 5 pvalue = 0.9806
alternative hypothesis :
True difference in mean is greater than 0

iii)

95 percent confidence interval

-6.035547 Inf

Sample estimate

mean of the difference

-3.5

∴ Pvalue is greater than 0.05, we accept the hypothesis at 5% level of significance.

4) 2 medicin are applied to two groups of patient respectively

group 1 : 10, 12, 13, 11, 14

group 2 : 8, 9, 12, 14, 15, 10, 9

Is There is any significance difference between 2 groups medicines.

H0: There is no significance difference

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

data : x and y

t = 0.80384 df = 9.7594 Pvalue = 0.4406

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

95 percent confidence interval

-0.9698553 4.2981886

Sample estimates :

means of x and y

12.0000 10.3333

∴ Pvalue is greater than 0.05 we accept the hypothesis at 5% level of significance.

Practical = 8

Topic : large and small test

Q.1. $H_0: \mu = 55$, $H_1 = \mu \neq 55$

```
> 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]    1.82153e-05
```

As pvalue is less than 0.05 we reject μ_1 at 5% value of significance.

Q.2. $H_0: p = 0.5$ against $H_0 : p \neq 0.5$

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


As pvalue is greater than 0.05 we accept H_0 at 1% level of significance.

Q.3] $H_0 : \mu = 66$ against $H_1 : \mu \neq 66$

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

[1] One sample test

data = x

t= 47.94   df = 6   p-value = 5.522e-09

alternative hypothesis : True mean is not equal to 0

95 percent confidence interval

64.6647   71.6292

Sample estimates :

mean of x

68.14286

Since p value is less than 0.05 we reject H_0 at 1% level of significance.

Q.4]

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

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

F test to compare two variance.

data : x and y

F = 0.7888 , num df = 7 , denom. df = 10

p value = 0.7737

alternative hypothesis : True ratio of variance is not equal to 1

95 percent confidence interval

0.199509 3.75188

Sample estimates :

ratio of variance

0.7880255

∵ p-value is greater than 0.05 we accept H_0 at 5% level of significance.

Q.5] $H_0 : P_1 = P_2$ against $H_1 : P_1 \neq P_2$

```
> n1 = 200
> n2 = 300
> p1 = 44/200
> p2 = 56/300
> p ( n1 * p1 + n2 * p2) / (n1+n2)
> p
[1]   0.2
> q = 1-p
[1]   0.8
> zcal = ( p1 - p2) / sqrt ( p * q * (1/n1 + 1/n2))
> zcal
[1]   0.91287
> pvalue = 2 * (1- pnorm (abs (zcal)))
> pvalue
[1]   0.3613104
```

∵ Pvalue is greater than 0.05 we accept H_0 at 1% level of significance.

Practical = 9

Topic: Non-Parametric test

Sign test:

Q.1 The following data present earnings (in dollar) for a random sample of five common stocks listed on the New york Exchange. Test whether median earnings 4 dollars.
(1.68, 2.35, 2.30, 6.23, 3.24)

Sol:

```
> x <- c (1.68, 3.35, 2.50, 6.23, 3.24)
> n <- length (x); po <-0.5;
[1] FALSE FALSE FALSE TRUE TRUE

> s<- sum (x>4); s;
∴ x>4
[1] 1
> binom.test(s,n,p=0.5, alternative = "greater");
```

Exact binomial test

data : s and n
number of successes = 1, number of trials=5
p-values = 0.9688

alternative hypothesis : true probability of successes is than 0.5, 95 % confidence interval

0.01020622 1.0000

Sample estimates
probability of success

wilcoxon test:

Q.2. The scores of 8 students in reading before and after lesson are as follows: Test whether there is effect of reading:

Student no:	1	2	3	4	5	6	7	8
Score before:	10	15	16	12	09	07	11	12
Score after :	13	16	15	13	09	10	13	10

```
> b <- c(10,15,16,12,09,07,11,12);
> a<- c(13,16,15,13,05,10,13,10);
> D<- b-a;
> wilcox.test (D, alternative = "greater")
```

data: 0
V= 10.5, P-values = 0.8722
alternative hypothesis : True location is greater than 0

warning message : In wilcoxon.test default 10 alternative = "greater");
Cannot computer exact p-value with ties.

Mann-whitney - wilcoxon's test

Q.3) The diameter of a ball bearings was measured by 6 inspector each using two different kinds of calipers. The result were arrange ball bearing for Test whether

Inspector	1	2	3	4	5	6
caliper 1	0.265	0.268	0.266	0.267	0.069	0.264
caliper 2	0.263	0.262	0.270	0.261	0.271	0.260

```
> # Given
> x <- c (0.265, 0.268, 0.266, 0.267, 0.069, 0.264)
> y <- c (0.263, 0.262, 0.270, 0.261, 0.271, 0.260)
> wilcox. test ( x, y , alternative = "greater")
```

wilcoxon rank sum test

data : x and y

w : 24 , p.value = 0.197

alternative hypothesis : True location shift is greater than 0.