Basics of R Software

1. R is a Software for Statistical analysis and data Computing
2. It is an effective data handling Software and outcome Storage is Possible.
3. It is Capable of graphical display.
4. Its a free Software

Problem 1 :- Solve the following

(1) $4+6+8\div2-5$

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

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

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

(3) $5^3+7\times5\times8+46/5$

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

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

```
> Sqrt(4^2+5*3+7/6)
[1] 5671567
```

(5) Round off

$46-7+9\times8$

Problem ②

(i) c(2,3,5,7) * 2
(ii) c(2, 3,5,7) * (2,3)
(iii) c(2,3,5,7) * c(2,3,6,2)
(iv) c(1,6,2,3) * c(-2,-3,-4,-1)
(v) c(2,3,5,7)^2
(vi) c(4,6,8,9,4,5)^ c(1,2,3)
(vii) c(6,2,7,5) / c(4,5)

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

(ii) c(2,3,5,7) * c(2,3)
[1] 4 9 10 21

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

(iv) c(1,6,2,3) * c(-2,-3,-4,-1)
[1] -2 -18 -8 -3

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

(vi) c(4,6,8,9,4,5)^ c(1,2,3)
[1] 4 36 512 9 16 125

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

Problem 3:

Q./ $x=20$
$y=30$
$z=2$

find i) x^2+y^3+z
ii) $\sqrt{x^2+y}$
iii) x^2+y^2

Ans

```
>x=20
>y=30
>z=2
>x^2 + y^3+z
[1] 27402
> sqrt(x^2+y)
[1] 20.73644
> x^2+ y^2
[1] 1300
```

Problem 4:
Create matrix

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


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

Q5) Find x + Y and 2x+3Y Such that

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

```
> x = matrix (nrow = 3, ncol=3 data = c(4,7,9,-2,0,-5,6,7,3))
> Y = matrix (nrow = 3, ncol =3, data = c(10,12,15,-5,-4,-6,7,9,5))

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

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

Q6) Marks of Statistics of Computer Science Students

59, 20, 35, 24, 46, 56, 55, 45, 27, 22, 47, 58, 54,
40, 50, 32, 36, 29, 35, 39

```
> x = c(data)
> length (x)
[1] 20
> breaks = Seq (20, 60, 5)
> a = cut (x, breaks, right = FALSE)
> b = table (a)
> c = transform (b)
> c
```

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

M

Practical-2

(Probablity Distribution)

Check whether the following are pmf or not

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

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

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

Since $P(2)$ -0.5, it cant be a pmf
Since pmf $P(x) \geq 0 \; \forall x$

It cannot be a pmf as in pmf
$\Sigma P(x) \neq 1$
Prob = c (0.2, 0.2, 0.3, 0.2, 0.2)
> Sum (Prob)
[1] 1.1

Ans Since $\Sigma p(x) = 1$, ∴ is a Pmf

> Prob = c(0.2, 0.2, 0.35, 0.15, 0.1)
> Sum (Prob)
> [1] 1

Q2)

i) Find c.d.f for the following P.m.f and Sketch graph.

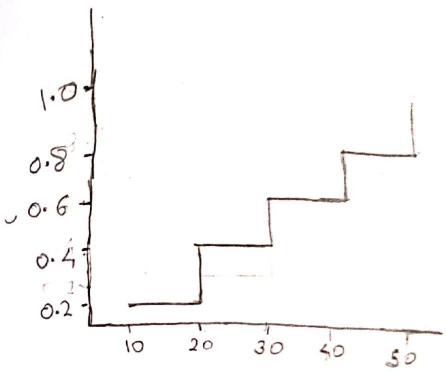
x	10	20	30	40	50
$P(x)$	0.2	0.2	0.35	0.15	0.1

Ans

>Prob = c(0.2, 0.2, 0.35, 0.15, 0.1)
>Sum (Prob)
[1] 0.20 0.40 0.75 0.90 1.00

$f(x) = 0$	$x < 10$
$= 0.2$	$10 \le x < 20$
$= 0.4$	$20 \le x < 30$
$= 0.75$	$30 \le x < 40$
$= 0.95$	$40 \le x < 50$
$= 1.0$	$x \ge 50$

>x = c (10, 20, 30, 40, 50)
>Plot (x, CumSum (Prob), "S").



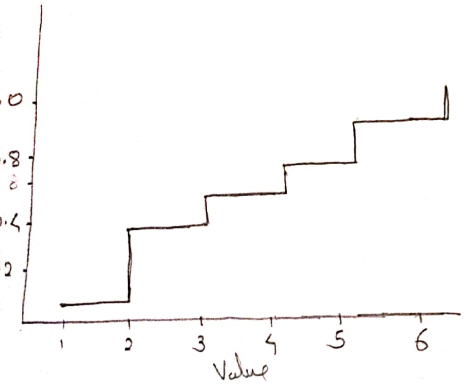
x	1	2	3	4	5	6
$P(x)$	0.15	0.25	0.1	0.2	0.2	0.1

Prob = c (0.15, 0.25, 0.1, 0.2, 0.2, 0.1)
Sum (Prob)
[1] 1
CumSum (Prob)
[1] 0.15 0.40 0.50 0.20 0.90 1.00

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

Scanned with CamScanner

```
x : c(1, 2, 3, 4, 5, 6)
Splot(x, main= "Prob dist", xlab= "Value", Y/ab "
      umSum (Prob), "s")
```



Check whether the following p.d.f or not

$f(x) = 3 - 2x \quad : \quad 0 < x < 1$

$f(x) - 3x^2 \quad\quad 0 < x < 1$

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

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

$\int_3^1 (3dx - \int_0^1 2x\,dx$

$: [3x - x^2]_0^1 \quad : 2$

$= 2$

ii) $\int_3 3x^2$

$\int_3^1 x^2$

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

$\frac{3}{3}\,[x^3]_0^1$

$= [1 - 0]$

$= 1$

AM

Practical - 3

Topic:- Binomial distribution

i) $P(X = x) = dbinom(x, n, p)$
ii) $P(X \leq x) = pbinom(x, n, p)$
iii) $P(X > x) = 1 - pbinom(x, n, p)$
iv) $P_1 = P(X \leq x)$ qbinom (P_1, n, p)

Q1) Find the Probability of exactly 10 Success in 100 trials with $P = 0.1$

Q2) Suppose there are 12 MCQ each question have 5 option out of which 1 is Correct Find the Probability of having
i) exactly 4 Correct answer
ii) atmost 4 Correct answer
iii) more than 5 Correct answer

Q3) Find the Complete distribution when $n = 5$ and $P = 0.1$

Q4) $n = 12$, $P = 0.25$ Find the following Probability
i) $P(x = 5)$
ii) $P(x \leq 5)$
iii) $P(x > 7)$
iv) $P(5 < x < 7)$

Ans
Q1) dbinom (10,100,0.1)
[1] 0.1318653

2) dbinom (4,12,0.2)
[1] 0.1328756
> pbinom (4,12,0.2)
[1] 0.9274445
> 1 - pbinom (4,12,0.2)
[1] 0.0725555

3) >dbinom (0:5,5,0.1)
[1] 0.59049 0.32805 0.07290 0.00810
0.00045 0.00001

4) >dbinom (5,12,0.25)
[1] 0.1032414
> pbinom (5,12,0.25)
[1] 0.9455978
> 1 - pbinom (7,12,0.25)
[1] 0.0027815,
> dbinom (6,12,0.25)
[1] 0.04014945

Q5) The Probablity of a Salesman making a Seel the a Customer of 0.5

1) No Sells out of 10 customer
2) More than 3 Sales out of 20 customer

Q6) A Salesman has a 20% Probablity of making a Sell to a Customer out of 30 customers what minimum number of Sells with 88% Probablity.

Q7) x follows binomial distribution with x P=0.3 Plot the graph of P.m.f and c.d.f.

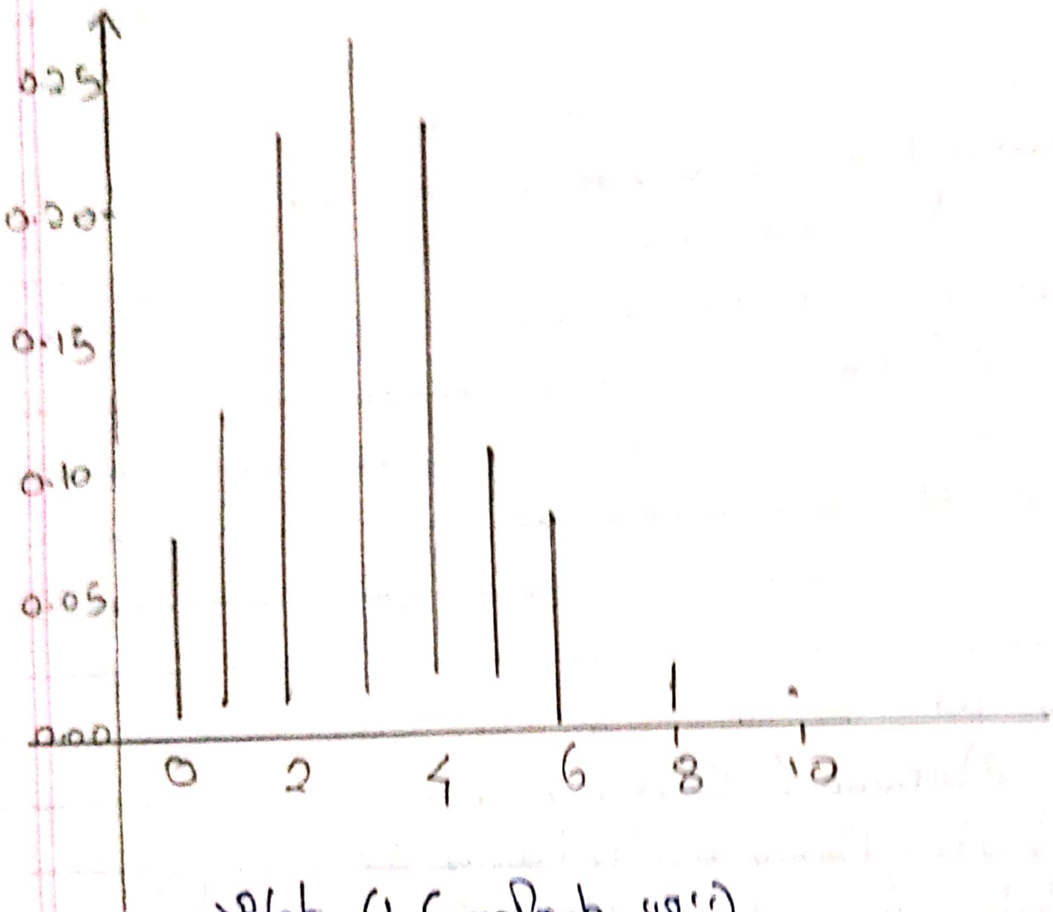
Ans

Q5
```
>dbinom (0,10,0.5)
[1] 0.0009765625
>1-Pbinom (3,20,0.5)
[1] 0.9987116
```

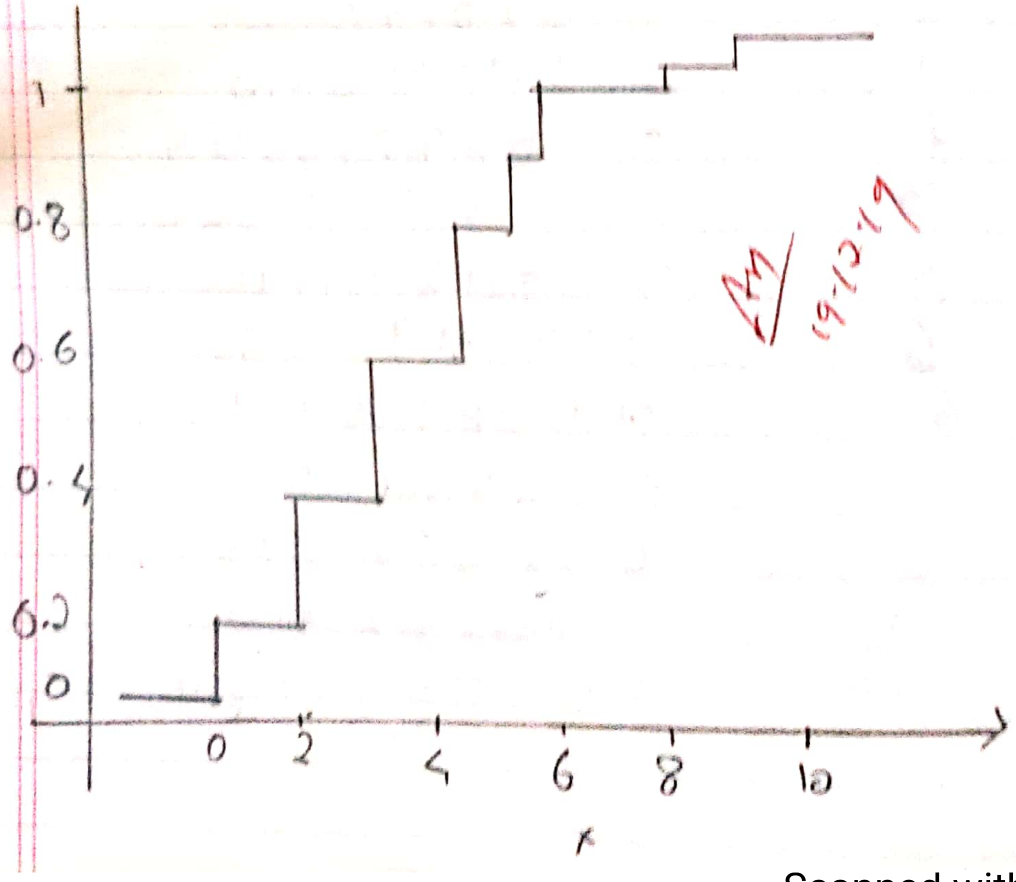
6)
```
>qbinom(0.88, 30,0.2)
[1] 9
```

7)
```
>x = 0:10
>Prob = dbinom (x,10,0.3)
> CumProb =Pbinom (x,10,0.3)
> d = data.frame (''x values'' : x, ''Probability'': Prob)
>Print (d)
```

	x Values	Probablity
1	0	0.028247
2	1	0.1210608210
3	2	0.2334744405
4	3	0.2668274320
5	4	0.2001209490
6	5	0.1029193452
7	6	0.0367569090
8	7	0.0090016920
9	8	0.0014467005
10	9	0.0001377810
11	10	0.0000059049



>plot (1, CumProb, "S")



19-12-19

Practical. 4

TOPIC : Normal distribution

$P(X = X)$: dnorm $(x, \mu, 6)$
$P(X \leq X)$ = Pnorm (X, μ, σ)
$P(X > x)$ = 1 - Pnorm $(x, \mu, 6)$

To generate random numbers from a normal distribution (n random no.)
the R Code is rnorm (n, t, σ)

A random variable x follows normal distribution with mean = μ = 12 and SD = σ = 3 find

i) $P(X \leq 15)$ ii) $P(10 \leq X \leq 13)$
iii) $P(X > 14)$ iv) generate 5 observations (random nos.)

```
>P1 = Pnorm(15,12,3)
[1] 0.8413447
>Cat ("P(X<=15) =", P1)
[1] P(X< =15) = 0.8413447
>P2 = Pnorm (13,12,3
>Cat ("P(10<= X <=13) =", P2)
[1] P(10<=X<=13) = 0.3780661
>P3 = 1 - Pnorm (14,12,3)
>Cat ("P(x>14) =", P3)
[1] P(x>14) = 0.2524925
>rnorm (5,12,3)
[1] 12.639812      13.041270
    6.865470
```

2) X follows normal distribution with $\mu = 10, \sigma = 2$
Find
i> $P(X \le 7)$ ii) $P(5 < X < 12)$
iii> $P(X > 12)$ iv> generate 10 obs.
v> Find K such that Probability $P(X < K) = 0.5$

```
>P1 = Pnorm (7,10,2)
>P1 [1] 0.668072
>P2 = Pnorm (5,10,2) - Pnorm (12,10,2)
> P2 [1] 0.8361351
> P3 = 1- Pnorm (12,10,2)
>P3 = [1] 0.1586553
> P4 = rnorm (10,10,2)
> P4
[1] 10.266560    10.596867
    9.045499     9.978724
    11.929820    5.833428
    11.123124
>P5 = qnorm (0.4,10,2)
>P5
[1] 9.493308
```

3) generate 5 random numbers with mean = 15 & $\sigma = 4$
Find Sample mean, median, S.D, and Print it

4) X follows normal
$X \sim N(30, 100)$
Find i> $P(X \le 40)$ ii> $P(X > 35)$
iii> $P(25 < X < 35)$ iv> Find K such that $P(X < K) = 0.6$

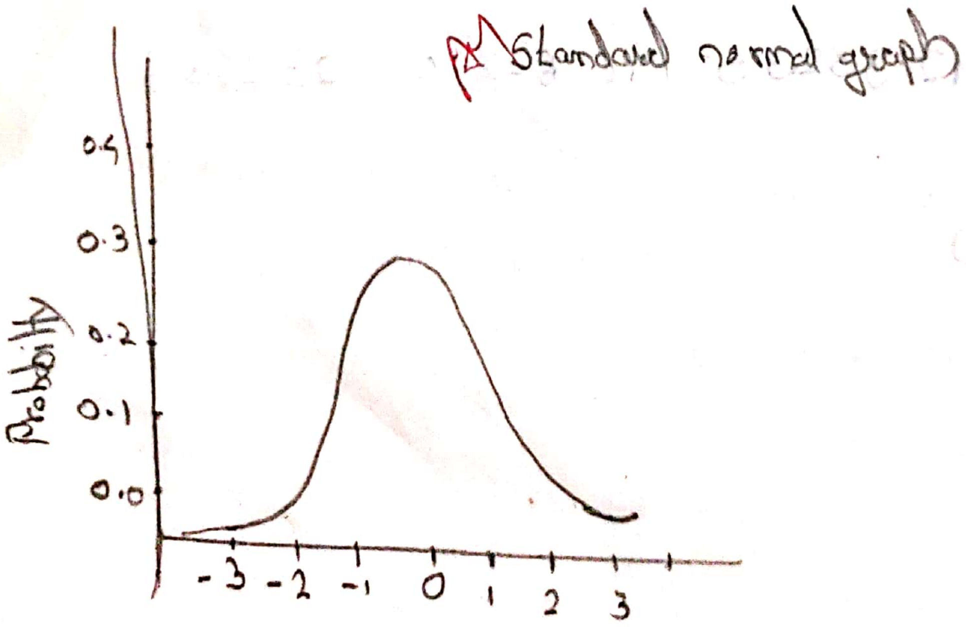
```
>P1 = Pnorm (40,30,10)
>P1
[1] 0.8413447
>P2 = 1 - Pnorm (35,30,10)
>P2
[1] 0.3085375
>P3 = Pnorm (25,30,10) - Pnorm (35,30,10)
>P3
[1] -0.382949
>P4 = qnorm (0.6,30,10)
>P4
[1] 32.53347
```

3)
```
>x = rnorm (5,15,4)
>x
[1] 12.73337    12.30235    16.62240    23.24314
    13.08828
>am = mean (x)
>am
[1] 15.33276
```

```
>me=median(x)
>me
[1] 14.86556
>n=5
>V=(n-1)*Var(x)/n
>V
[1] 2.574805
>Sd=Sqrt(V)
>Sd
[1] 1.60462
>Cat("Sample mean", am)
 Sample mean 15.33276
>Cat("Sample Sd", Sd)
 Sample Sd 1.60462
```

5) Plot the normal graph (Standard)

```
>x=Seq(-3,3,by=0.1)
>Y=dnorm(x)
>Plot(x,Y,xlab="x Value", Ylab="Probablity",
    main="Standard normal graph")
```



PRACTICAL - 5

TOPIC : Normal and t-test

i) $H_0 : \mu = 15 \quad H_1 : \mu \neq 15$

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

\# 0.05 > accept the Value

\# 0.05 < less than Reject

```
>mo=15
>mx=14
>Sd=3
>n=400
>zCal=(mx-mo)/(Sd/(Sqrt(n)))
>zCal
[1] -6.666667
>Cat("Calculated Value of Z is :", zcal)
Calculated Value of Z is : -6.666667
>Pvalue=2*(1-Pnorm(abs(zCal)))
>Pvalue
[1] 2.616796e-11
```

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

8

2) Test the hypothesis $H_0: \mu = 10$ against $H_1: \mu \neq 10$
A random Sample Size of 500 is drawn with
Sample mean = 10.2 and S.D = 2.25
Test the hypothesis at

```
>mo=10
>n=500
>mx = 10.2
>sd =2.25
> z Cal = (mx - mo) /(sd/(sqrt(n)))
> z Cal
[1] 1.77778
>Pvalue = 2*(1-Pnorm(abs(zCal)))
>Pvalue
[1] 0.07555036
```

∵ The Value pvalue is greater than 0.05
∴ The Value is accepted

3) Test the hypothesis H_0: Proportional of Smokers in College is 0.2
A Sample is Collected and Calculated the Sample Proportional as 0.125. Test the hypothesis at 5% level of Significance
(Sample Size is 500)

```
>P=0.2
>P=0.125
>n=500
>Q=1-P
```



```
>zCal= (P-P)/(sqrt ...)
>Cat ("Calculated Value of z is =", zCal)
[1] Calculated Value of z is = -3.75
>PValue = 2 *(1-Pnorm (abs (zCal)))
>P Value
[1] 0.0001768346 (Reject)
```

4) Last Year farmer's lost 20% of their Crops. A random Sample of 60 fields are cauted And it is found that a field Crops are insect Polluted.
Test the hypothesis at 1% level of Significan

```
>P=0.2
>P=9/60
>n = 60
> z Cal = (P-P) /(Sqrt (P*Q/n))
> z Cal
[1] -0.9682458
> P Value = 2* (1-Pnorm (abs (zCal)))
>P value
[1] 0.3329216
```

∵ The Value is 0.1 So Value is accepted

6) Test the hypothesis $H_0: \mu = 12.5$ from the following Sample at 5% level of Significance

```
> X= c(12.25,11.97, 12.15, 12.08, 12.31, 12.28, 11.95, 11.89, 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] 0.01952)
> Sd = Sqrt (Variance)
>Sd
[1] 0.1397176
>mo =12.5
>t = (mx-o)/(Sd/Sqrt(n))
>t
[1] -8.895909
>PValue = 2*(1-Pnorm(abs(t)))
>P Value
[1] 0
∴ The Value is less than 0.05 the Value is accepted

Practical - 6

Topic :- Large Sample test

Let the Population mean (amount Spent per Customer in a restaurant) is 250 A Sample of 100 Customers Selected the Sample mean is Calculated as 275 and Standard d is 30. Test the hypothesis that Population is 250 and not at level of 5% of Significance

In a random Sample of 1000 Student it is found that 7.50 Used blue pen test hypotesis that the population Proposition is 0.8 at level of Significance

>mo = 250 ;
>mx = 275 against $H_1 : \mu \neq 275$
> Sd = 30
> n=100
> Z Cal = (mx - mo)/(Sd/(Sqrt(n)))
>Z Cal
[1] 8.333333
>Pvalue = 2 * (1-Pnorm(abs(zcal)))
>Pvalue
[1] 0
Pvalue is less than 0.05 We reject. Ho at 5% level of Significance.

3.

2) $H_0: P = 0.8$ against $H_1: P > 0.8$

P=0.8

q=1-P

p=750/1000

n=100

```
>Zcal =(p-P)/(sqrt(P*q/n))
[1] -3.952847
>Pvalue=2*(1-Pnorm(abs(zcal)))
>Pvalue
[1] 7.72268-05
```

3) Two random Sample of Size 1000 and 2000 are drawn from two population with Same Standard deviation 2.5. The Sample means are 67.5 and 68 respectively Test the hypothesis at 5% level of Significance

4) A Study of Noise level in a 2 hospital is given below test the claim that the two hospitals are Same level of Noise at 1% level Significance

	Hos A	Hos B
Size	84	34
Mean	61.2	59.4
S.D	7.9	7.5

5) In a Sample of 600 Student in a Clg 400 used blue ink in another clg from a Sample of 900 Student 450 used blue ink. Test the hypothesis that the Proportion of Student using blue ink in 2 college are equal or not at 1% level of Significance

67

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

$n_1 = 1000$

$n_2 = 2000$

mx1=67.5

mx2=68

sd1 = 2.5

sd2 = 2.5

```
>zcal=(mx1-mx2)/sqrt(sd1^2/n1)+(sd2^2/n2))
>zcal
[1] -5.163978
>PValue=2*(1-Pnorm(abs(zcal)))
>PValue
[1] 2.417564e-07
```

P Value is <0.05

We reject Ho at 5%.

4) $H_0: \mu_1 = \mu_2$ against;

$n_1 = 84$

$n_2 = 34$

mx1 = 61.2

mx2 = 59.4

sd1 = 7.9

sd2 = 7.5

```
>zcal (mx1-mx2)/sqrt((sd1^2/n1)+(sd2^2/n2))
>zcal.
```

Scanned with CamScanner

[1] 1.162528.

>Pvalue = 2 * (1- Pnorm (abs(z(al)))

> Pvalue

[1] 0.2405211

∵ Pvalue is greater than 0.05 We accept H_0 at 1% level of Significance

5) $H_0 : P_1 = P_2$ against $H_1 : \mu_2 \neq \mu_1$

$n_1 = 600$

$n_2 = 900$

$P1 = 400/600$

$P^2 = 450/900$

$P = (n_1 * P_1 + n_2 * P_2)/(n_1+n_2)$

$q = 1-P$

>Z(al) = (P'-P^2)/Sqrt (P * q(1/n1+1/n2))

>Z(al)

[1] 6.381634

> PValue = 2 *(1-Pnorm (abs (Zcal)))

> Pvalue

[1] 1.75322e-10...

Pvalue is <0.05 We reject H_0 at 1% level of Significance

For Sample Size

$n_1 = 200$

$n_2 = 200$

$P_1 = 44/200$

$P_2 = 30/200$

$P = (n_1 * P_1 + n_2 * P_2)/(n_1+n_2)$

>Z(al) = (P1-P2) /Sqrt (P * q * (1/n1+1/n2)

> Z(al)

[1] 1.802741

> P

[1] 0.185.

> P Value = 2 (1- Pnorm (abs (z(al)))

> PValue

[1] 0.07142888

P Value is >0.05 We accept.

H_0 at 5% level of Significance.

3

Practical - 7

Topic: Small Sample Test

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

⇒ Ho: μ = 66 > x = c(63, 63, ..., 71, 72)
>n = length (x)
>n
[1] 10
> t.test (x),
One Sample t-test
data: x
t = 68.319, df = 9, P.Value = 1.658e-13 ~~alter~~
alternative hypothesis: true mean is not equal to 0
95 Percent confidence interval: 65.65177 70.14829
Sample estimates:
mean of x
67.9

∵ Pvalue is less than 0.05. Hence, we reject the Ho at 5% level of Significance.
>if (Pvalue > 0.05) {cat("Accept H.")}else {Cat("Reject H.")}
Reject H.

69

Q2) Two groups of Student Score the following marks. Test the hypothesis that there is no level of Significance difference between 2 groups.

Group A : 18, 22, 21, 17, 20, 17, 23, 20, 22, 21
Group A : 16, 20, 14, 21, 20, 18, 13, 15, 17, 21

⇒ Ho: There is no difference between two groups
x = c(18, 22, 21, 17, 20, 17, 23, 20, 22, 21)
y = c(16, 20, 14, 21, 20, 18, 13, 15, 17, 21)
t.test (x, y)

Welch Two Sample t-test
data: x and y
t = 2.2573, df = 16.376, P-Value = 0.03798
alternative hypothesis: true difference in means is not equal to 0
0.1628205 6.0371795
Sample estimates:
mean of x mean of y
20.1 17.5

>if (Pvalue > 0.05) {Cat ("Accept H.")}else{Cat ("Reject H.")}
Reject Ho.

Scanned with CamScanner

Q3) Sales data of 6 shop. before and after are Special Campaign given below:-

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

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

Test the hypothesis the Campaign is effective or not.

H0: There is no Significance difference of Sales before and after the 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 and Y
t=-2.7815, df=5  P-Value=0.9806
alternative hypothesis: true difference in means is greater than 0.
-6.035547  Inf.
Sample estimates:
mean of the differences
        -3.5
>Pvalue=0.9806
>if (Pvalue>0.05) {Cat("Accept H0")} else {Cat ("Reject H0")}
Accept H0.
```

Two medicines are applied to 2 group of Patient

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

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

is there any Significance difference between two groups.


```
>a = c(10,12,13,11,14)
>b = c(8,9,12,14,15,10,9)
>t.test(a,b)
        Welch Two Sample t.test
data. a and b
t = 0.65591    df=9.567,  P.Value=0.5273
alternative hypothesis: true difference in mean is not equal to 0, 95 percent Confidence interval:
: -1.954382    3.534382
Sample estimates:
   mean of x mean of y
      11.8     11.0
>Pvalue = 0.5273
>if (P Value >0.05) {Cat("Accept H0")} else
{cat("Reject H0")}
Accept H0.
```

Q5) H_0: There is no Significance difference between before and after

```
>A = c(120,125,115,130,123,119)
>B = c(100,113,95,90,115,99)
>t.test (A,B, Pared = T, alternative = "less")
```

Pared t.test

data : A and B.

t = 4.3451 , df = 5, P-Value, 0.9963

alternative hypothesis : True difference in mean is less than 0 95. Percent Confidence interval:

Inf 29.0293

Sample estimates :

mean of the differences

19.83333

```
>P Value : 0.9963
>if (PValue > 0.05) {Cat ("Accept H.")else
{Cat ("Reject H.")}
```

Accept H_0.

AN

Practical - 8

Topic: Large and Small Sample tests

$H_0 : \mu = 55$, $H_1 : \mu \neq 55$

```
>n=10
>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-0.5
```

As Pvalue is less than 0.05 we reject H at 5% level of Significance

(2) $H_0 : P = 0.5$ against $H_1 : P \neq 0.5$

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

> Pvalue = 2 * (1-Pnorm (abs (zcal)))
> P Value
[1] ,
As Value is greater than 0.05 We accept Ho at 1% level of Significance

3 Ho : P1 = P2 against H1 : P1 ≠ P2
>n1 = 1000
>n2 = 1500
>P1 = 2/1000
> P2 = 1/1500
> P = (n1 * P1 + n2 * P2) / (n1 + n2)
> P
[1] 0.0012
>q = 1-P
[1] 0.9988
>zcal = (P1-P2) / Sqrt (P*q* (1/n1 + 1/n2))
>zcal
[1] 0.9433752
>Pvalue = 2 * (1-Pnorm (abs (zcal)))
>Pvalue
[1] 0.345489

∴ P Value is greater than 0.05 we accept Ho and 5% level of Significance

④ Ho
>Var = 64
> n = 400
> m0 = 100
> mx = 99
> sd = Sqrt (Var)
> sd
[1] 8
>zcal = (mx - m0) / (sd / (Sqrt (n)))
>zcal
~~>zcal (m.~~
[1] -2.5
> Pvalue = 2 * (1-Pnorm (abs (zcal)))
> Pvalue
[1] 0.01241933

Since PValue is less than 0.05 we reject Ho at 5% level of Significance

⑤ Ho: μ = 66 against H1 : μ ≠ 66
> x = c (63, 63, 68, 69, 71, 71, 72)
> t.test (x)
One Sample t-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 :
~~0~~ 64.66479 71.62092
Sample estimates :
mean of x
68.14286

Sol^n (6) $H_0 : \sigma_1 = \sigma_2$ 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, 85, 87, 92, 93, 95, 97)
var.test (x, y)
        F test to Compare two Variances
data : x & y
F = 0.788803, num df = 7, denom.df = 10, P.Value = 0.7737
alternative hypothesis: true ratio of Variances
 is not equal to 1
 95 Percent Confidence interval:
    0.199509 3.751881
  Sample estimates :
  ratio of Variances
        0.7880255
```

P-Value is greater than 0.05 We accept H_0 at 5% level of Significance

Sol^n (7) $H_0 : \mu = 1150$ against $H : \mu \neq 1150$

```
>n = 100
>mx = 1150
>mo = 1200
>Sd = 125
>Zcal = (mx-mo)/(Sd/(sqrt(n)))
```



```
> Z Cal
[1] -4
> P Value
[1] 6.334248 P-0.5
```

∴ P Value is less than 0.05 we reject H_0 at

(8) $H_0 : P_1 = P_2$ against $H_1 : P_1 \neq P_2$

```
> n = 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))
[1] 0.9128709
> Pvalue = 2*(1-Pnorm(abs(zcal)))
> P Value
[1] 0.3613104
```

My 17.1.20

P Value is greater than 0.05 so we accept H_0 at 1% level of Significance

Practical - 9

Aim :- chi Square tests & ANOVA (Analysis of Variance)

Q1) Use the following data to test whether the Condition of the home and child are independent or not.

	Condition of Home	
	clean	Dirty
Clean	70	50
Fairly Clean	80	20
Dirty	35	45

```
> 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

∴ P Value is less than 0.05

∴ We reject H_0

75

Q2) Test the hypotesis that vaccines and deaseases are independent or not

	Vaccies	
	Affected	Not Affected
Affected	70	46
Not Affected	35	37

```
> X = c (70,35,46,37)
> m = 2
> n = 2
> Y = matrix (x, nrow = m, ncol = n)
> Y
     [,1] [,2]
[1,]   70   46
[2,]   35   37
> PV = chisq.test (y)
> PV
```

Pearson's chi squared test with Yates' Continuity Correction

data : y

X - Squared = 2.0275, df = 1, P. Value = 0.1545

Q3) Perform ANOVA for following data

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

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

```
> x1 = c(50,52)
> x2 = c(53,55,53)
> x3 = c(60,58,57,56)
> x4 = c(52,54,54,55)
> d = Stack(list(b1=x1, b2=x2, b3=x3, b4=x4))
> names(d)
[1] "Values" "ind"
> One way.test(Values ~ ind, data=d, Var.equal=T)
  One-way analysis of means
data: Values and ind
  f: 11.735, num df: 3, denom df = 9 PValue = 0.00183
> anova = aov(Values ~ ind data=d)
> Summary (anova)
```

	df	Sum Sq	mean Sq	f Values	Pr(>f)
ind	3	71.06	23.688	11.73	0.0083 **
Residual	9	18.17	2.019		

Signif Codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

The following data gives the life of the tyre of 4 brands

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

H_0 = The average life A, B, C and D are equal.

```
> x1 = c(20,23,18,17,18,20,24)
  x2 = c(19,15,17,20,16,17)
  x3 = c(21,19,22,17,20)
  x4 = c(D)
> d = Stack(list(b1=x1, b2=x2, b3=x3, b4=x4))
> names(d)
[1] Values" "ind"
> one way.test(Values ~ ind, data = d, Var.equal = T)
    one-way analysis of means
data: Values & ind
  F: 6.8445, num df: 3  denom df: 20
  P-Value = 6.002349
```

∵ P-Value is less than 0.05 we reject the hypothesis

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

	Df	Sum Sq	mean Sq	F Value	Pr(>F)
ind	3	91.44	30.479	6.845	6.00235 **
Residuals	20	89.06	4.953		

Signif. Codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

>x = read.csv("C:/users/admin/Desktop/marks.csv")

>x

	Stats	Maths
1	40	60
2	45	48
3	42	47
4	15	20
5	37	25
6	36	27
7	49	67
8	59	58
9	20	25
10	27	27

```
>am = mean(x$Stats)
>am
[1] 37
>am1 = mean(x$Maths)
>am1
[1] 39.5
>m1 = median(x$Stats)
>m1
[1] 38.5
>m2 = median(x$Maths)
[1] 37
>n = length(x$Stats)
>n
[1] 10
```



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

Practical-10

Aim:- Non Parametric test

Q1) Following are the amounts of Sulphur oxide emmited by industrys in 20 days. Apply Sign test, to test the hypothesis that the population median is 21.5 at 5% LOS

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, ... 26)
> me=21.5
> sp= length(x[X>me])
> sp
[1] 9
> Sn= length(x[X<me])
> Sn
[1] 11
> n = SP + Sn
> n
[1] 20
> Pv = Pbinom(SP, n, 0.5)
> Pv
[1] 0.4119015
```

∴ PV is less than 0.05 we accept Ho at 5% level of Significance

$H_1: m_e \neq$
or $m_e <$
then, PV = Pbinom (Sp, n, 0.5)

If $H_1: me >$ then,
Pv = Pbinom (Sn, n, 0.5

Q2) Following is the data of 10 observation, apply Sign test. The Population median 625. The alternative H_1 is more than 625

612, 619, 631, 623, 643, 640, 655, 649, 620, 663

Ho Population median is 625
H_1: me > 625

X = c(612, 619, 631, 628, 643, 640, 653, 649, 670, 663)
>me = 625
>SP = length (x[x>me])
>Sn = length (x[x<me])
[1] n = SP+n
>n
[1] 10

>PV = Pbinom (Sn, n, 0.5)
>P
[1] 0.0546875

∴ PV is greater than 0.05 we accept Ho at

Q3) Following are the data of Samples. Test the hypothesis that the Population median is 60 against the alternative i.e. more than 60 at 5% level of Significance using wilcoxon Signed Rank test

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

→ Ho: me 60 against H_1: me > 60
> x = c(63, 65, 60, 89 .... 69)
> wilcox.test (x, alter = "greater", mu = 60)
Wilcoxon Signed rank test with Continuity Correction
data = x

V = 145, P-Value = 0.02298
alternative hypothesis: true location is greater than 60
P Value is less than 0.05 We reject Ho at 5% level of Significance

Note:- If ① H_1: me < then alter = "less"
② H_1 = me ≠ then alter = "two sided"

Q4) Using WSR test the population median is 12 or less than 12
13, 12, 24, 25, 20, 21, 32, 23, 12, 25, 24, 26

Soln :- H_0 : me = 12 against H_1 : me < 12.

```
> x = c (15, 17, 24, 25, 20, 21, 32, 28, 12, 25, 24, 26)
> wilcox.test (x, alter = "less", mu = 12)
```

data : x

V = 66, P.Value = 0.9986

alternative hypothesis : true location is less than 12

∵ Pvalue is greater than 0.05 we accept H_0

Q5) The weight of Student before and after they stopped Smoking are given below using WSR test that there is no Significant Change.

Before	After
65	72
75	74
75	72
62	66
72	73

Soln :- H_0 : Before and after there is no change against H_1 : There is change

```
> x = c(65, 75, 75, 62, 72)
> y = c(72, 74, 72, 66, 73)
> d = x - y
> wilcox.test (d, alter = "two sided", mu = 0)
```


data : d

V = 4.5, P.Value = 0.4982

alternative hypothesis : true location is not equal to 0

∴ Pvalue is greater than 0.05 therefore we accept H_0.

An
4/3/21

☆☆ INDEX ☆☆

Scanned with CamScanner