Practical-1

Basics of R-Software

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

1) Solve the following:

i) $4+6+8\div2-5$

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

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

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

v) round off $46\div7+9\times8$

Sol^n

```
① >4+6+8/2 -5
[1] 9

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

③ >5^3 + 7*5*8 + 46/5
[1] 4.4.2

④ >sqrt(4^2+5*3 + 7/6)
[1] 5.671567

⑤ round(46/7 + 9*8)
[1] 79
```

Solve:

```
① c(2,3,5,7) * 2
② c(2,3,5,7) * (2,3)
③ c(2,3,5,7) *( 2,3,6,2)
④ c(1,6,2,3) * c(-2,-3,-4, -1)
5) c (2,3,5, 7)^2
⑥ c( 4,6,8,9,4,5)^c(1, 2,3)
7) c(6, 2, 7,5)/ c(4,5)
```


sol

```
①>c(2, 3,5,7) * 2
[1] 4 6 10 14

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

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

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

⑤>c( 2, 3, 5, 7)^2
[1] 4 9 25 49

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

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

Q. III)

ac

1) $x = 20$
$y = 30$
$z = 2$

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

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

problem 4 :-
create matrix

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

sol^n
```
> 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
```

Q) problem 5.

find $x+y$ and $2x+3y$ where

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

Sol^n:

```
> 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,-6,7,9,7))

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

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

Q) problem :- crete table.

Q) marks of statistic of CS students
59, 20, 35, 24, 46, 56, 55, 45, 27,
22, 47, 58, 54, 40, 50, 32, 36, 29, 35,
39

Sol^n

```
> x = c(59,20,35,24,46,56,55,45,27,22,
        47,58,54,40,50,32,36,29,35,
        39)
> length(x)
[1] 20
> breaks = seq(20,60,5)
> a = cut(x, breaks, right = FALSE)
> b = table(a)
> c = transform(b)
> c
```

output:

```
        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
```

AM 5/12

Sol^n

```
>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,-6,7,9,7))
```

```
> x + y
```

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

```
>2 * 2 + 3 * y
```

	[,1]	[,2]	[,3]
[1,]	38	-19	33
[2,]	50	-12	41
[3,]	63	-28	21

Q) problem :- crete table

Q) marks of statistic of CS students
59, 20, 35, 24, 46, 56, 55, 45, 27,
27, 47, 58, 54, 40, 50, 32, 36, 29, 35,
39

Sol^n

```
>x = c(59,20,35,24,46,56,55,45,27,27,
      47,58,54,40,50,32,36,29,35,
      39)
>length(x)
[1] 20
>breaks = seq(20,60,5)
>a = cut(x, breaks, right=FALSE)
>b = table(a)
>c = transform(b)
>c
```

output:

	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

Practical - 2

Probability Distribution

Check whether the following are PMf or not

1)

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

2)

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

3)

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

Sol^n:

1) $\because P(2) = -0.5$, can not be a probability mass function.

$\because P(x) \geq 0$ for all x.

2) $\therefore$ It can be a probability mass function. as is PMf the sumation of $P(x)$ should be equal to 1.

```
> prob = c(0.2, 0.2, 0.3, 0.2, 0.2)
> sum (prob)
[1] 1.1
```

Scanned with CamScanner

③ ∴ $\Sigma P(x)=1$, It is a pmf
>prob=c(0.2,0.2,0.35,0.15,0.1)
>Sum(prob)
[1] 1

Hence, it is a probability mass function.

Q.II Find cdf.

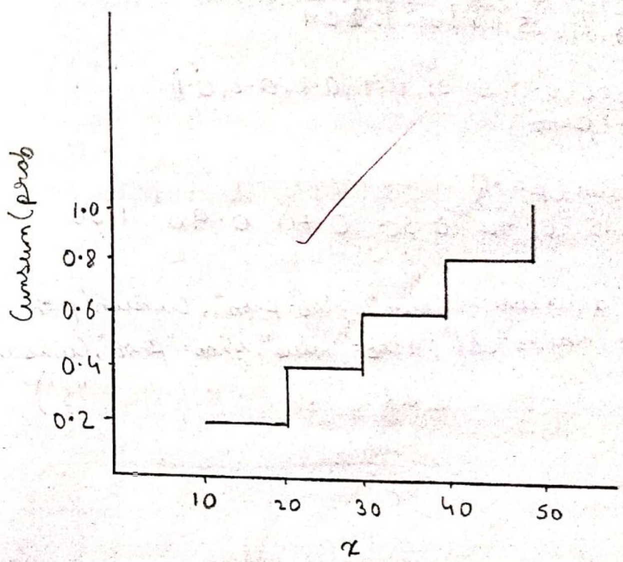
① Find the cdf for the following p.m.f and sketch the graph

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

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

$F(x) = 0$	$x < 10$
0.2	$10 \leq x < 20$
0.4	$20 \leq x < 30$
0.75	$30 \leq x < 40$
0.90	$40 \leq x < 50$
1.0	$x \geq 50$

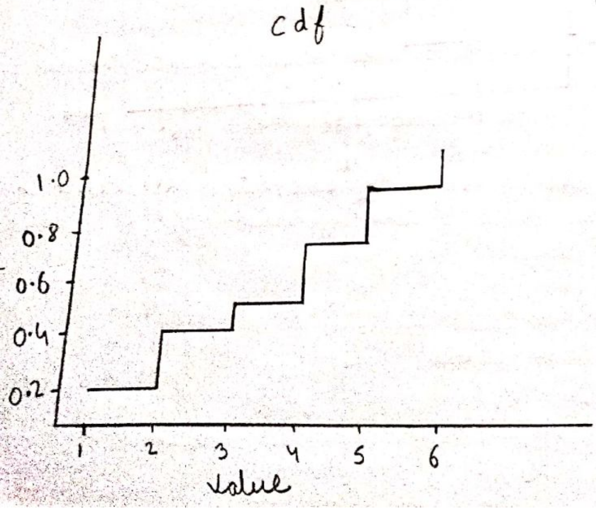
>x=c(10,20,30,40,50)
>plot(x, Cumsum(prob),"S")



② find the cdf for the following pmf and sketch the graph.

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  0.70  0.90  1.00
> x = c(1, 2, 3, 4, 5, 6)
> plot (x, xlab = "Value", ylab="prob", cumsum(prob), "s")
> plot(x, main="cdf", xlab="Value", ylab="prob", cumsum(prob), "s")
```



Q III) Check whether the following is pdf or not

① $f(x) = 3-2x \quad ; 0 \le x \le 1$

② $f(x) = 3x^2 \quad ; 0 < x < 1$

Sol^n:

① $f\,x = \int_0^1 f(x)\,dx$

$= \int_0^1 3-2x$

$= \int_0^1 3\,dx - \int_0^1 2x\,dx$

$= 3\int_0^1 dx - 2\int_0^1 x\,dx$

$\cancel{= 3(1-0)}$

$= [3x]_0^1 - 2\left[\frac{x^2}{2}\right]_0^1$

$= 3(1-0) - (1-0)^2$

$= 3(1) - 1$

$= 3-1$

$= 2$

$\therefore$ It is not a p.d.f.

(2) $f x = \int_0^1 f(x)\,dx$

$= \int_0^1 3x^2\,dx$

$= 3\int_0^1 x^2\,dx$

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

$= \left(\frac{1}{3} - 0\right)3$

$= 1$

$= 1$

$\therefore$ It is a p.d.f.

Practical 3
Binomial Distribution
~~Probability Distribution~~

Commands

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

Q1) Find the probability of exactly 10 success in 100 trials with $P = 0.1$

Q.2) Suppose there are 12 mcq. Each question has 5 option out of which 1 is correct. find the probability of getting
i) Exactly 4 correct answer.
ii) Atmost four correct answer.
iii) more than 5 correct answer

Q.3) Find the complete distribution, when $n = 5$ and $p = 0.1$

Q.4) $n = 12$, $p = 0.25$ find the following probabilitys
i) $P(X = 5)$
ii) $P(X \le 5)$
iii) $P(X > 7)$
iv) $P(5 < X < 7)$

Solⁿ:

1) >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(5,12,0.2)
[1] 0.01940528

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] 0.00278151
>dbinom(6,12, 0.25)
[1] 0.04014945

Q.5) The probability of a salesman making a sold to a customer is 0.15. find the probability of (1) no sells out of 10 customer.
(2) more than 3 sells out of 20 customer.

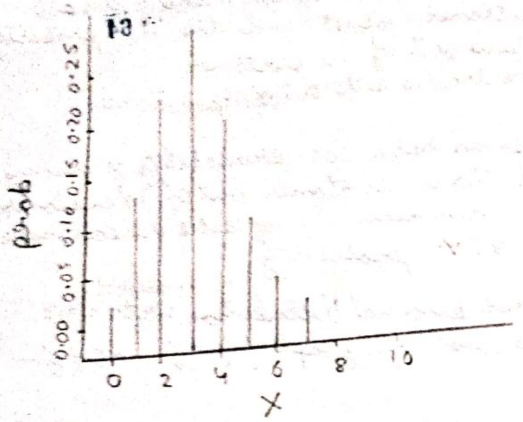
Q.6) A salesman has a 20% probability of making a sell to a customer out of 30 customer what minimum no. of sells he can make with 88% probability.

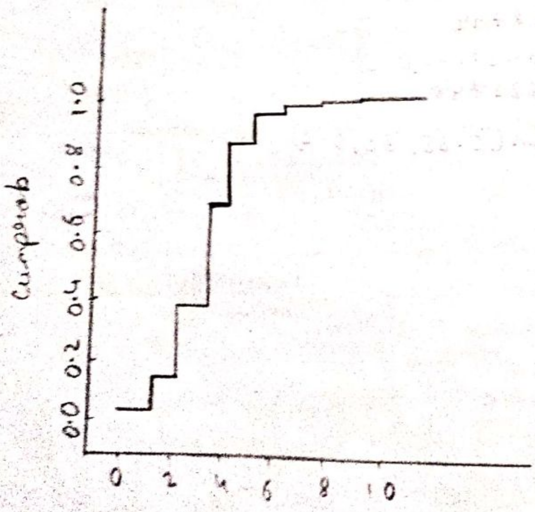
Q.7) x follows binomial distribution with n=10, p=0.3, plot the graph of p.m.f and c.d.f.

Solⁿ:

5) >dbinom(0,10,0.15)
[1] 0.1968744
> 1-pbinom(3,20,0.15)
[1] 0.3522748

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





7)
```
> n = 10; p=0.3
> x = 0:n
> prob= dbinom (x,n,p)
> cumprob = pbinom (x,n,p)
> d= data.frame ("xvalues"= x ,"probability"= prob)
> print(d)
```

	xvalues	probability
1	0	~~0.0282 0~~ 0.0282
2	1	~~0.1210~~ 0.1210
3	2	~~20~~ 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

Topic - Normal Distribution

1) P(X = x) = dnorm (x, μ, σ)
2) P(X ≤ x) = pnorm (x, μ, σ)
3) P (X > x) = 1 - pnorm (x, μ, σ)
4) To generate random numbers from a normal distribution (n random numbers) the R code is rnorm (n, μ, σ)

1) A random variable x follows normal distribution with mean = μ = 12 SD = σ = 3
find :- i) P(X ≤ 15)
ii) P(10 ≤ X ≤ 13)
iii) P(X > 14)
iv) Generate 5 observations (random numbers).

Solⁿ.

```
>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
```



```
>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)>rnorm(5, 12, 3)
[1] 7.429476  7.382370  12.491955  10.711917  8.086499
```

2) X follows Normal Distribution with μ = 10 σ = 2, find :- i) P(X ≤ 7) ii) P(5 < X < 12) iii) P(X>12) iv) Generate 10 observations v) find k such that probability P(X<k) = 0.4

```
> p1 = pnorm ( 7,10, 2)
>p1
[1] 0.668072
> p2 = pnorm(12,10,2) - pnorm(5,10, 3)
> p2
[1] 0.7935544
> p2 = pnorm( 12, 10, 2) - pnorm (5, 10, 2)
>p2
[1] 0.8351351
>p3 = 1 - pnorm(12,10,2)
>p3
[1] 0.1586553
>rnorm(10,10,2)
[1] 8.482086 6.833916  11.300854  10.062350  7.727590  11.865895  8.394346
 11.912815  10.609953  10.686826
```

~~3) Generate 5 random numbers from a~~

```
> qnorm(0.4, 10, 2)
[1] 9.493306
```

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

4) X follows $X \sim N(30, 100)$ $\mu = 30, \sigma = 10$
find i) $P(X \leq 40)$
ii) $P(X > 35)$
iii) $P(25 < X < 35)$
iv) find k such that $P(X < k) = 0.6$

4.Sol^n)

```
> p1 = pnorm(40,30,10)
> p1
[1] 0.8413447
> p2 = 1-pnorm (35, 30,10)
> p2
[1] 0.3085375
> Cat("p(X<=40)=", p1)
p(X<=40)= 0.8413447> Cat ("p(x>35)=", p2)
p(x>35) = 0.3085375> p3 = pnorm( 35,30,10)-pnorm(25,30,10)
> p3
[1] 0.3829249
> qnorm(0.6, 30,10)
[1] 32.53347
```


3.Sol^n)

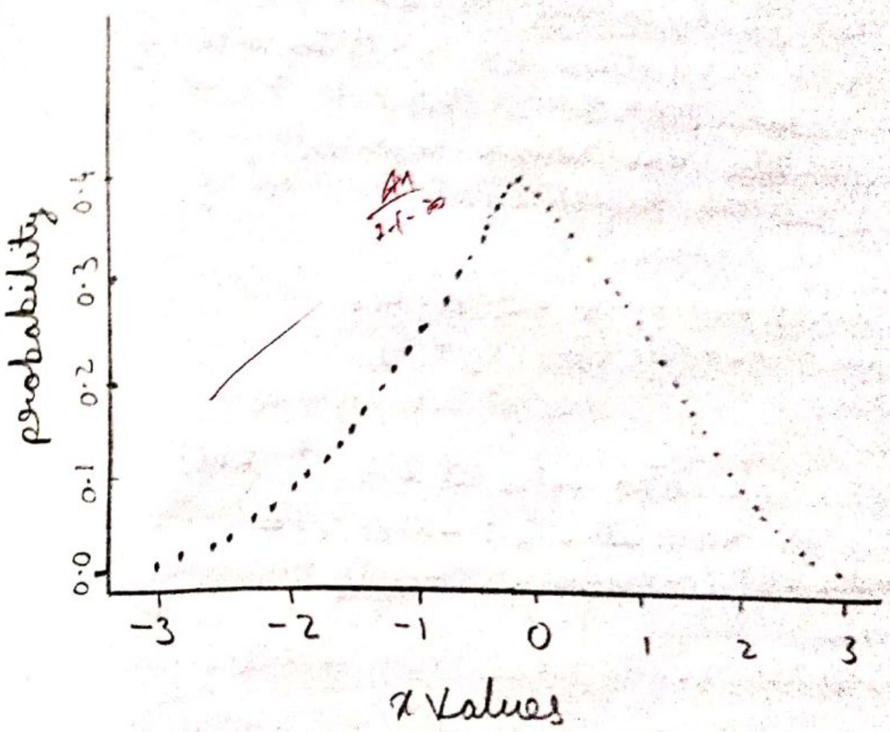
```
> x = rnorm(5,15,4)
> am = mean(x)
> me = median(x)
> n=5
> Variance = (n-1)* var(x)/n
> sd = sqrt (variance)
> x
[1] 15.775925  12.619834  14.568656  7.745604  8.488287
> am
[1] 11.83966
> me
[1] 12.61983
> Variance
[1] 10.30865
> sd
[1] 3.210709
> cat("Sample mean is= ", am)
Sample mean is= 11.83966 > cat("Sample median is", me)
Sample median is 12.61983> Cat("Sample sd is", sd)
Sample sd is 3.210709
```

5) plot the standard normal graph

Sol^n.

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


Standard Normal Graph



Practical V

Normal and t-Test

1) Test the hypothesis $H_0: \mu = 15$, $H_1: \mu \neq 15$ where H_0 - Null Hypothesis and H_1 = Alternative Hypothesis. A 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

```
> m0 = 15; mx = 14; sd = 3; n = 400
> zcal = (mx - m0) / (sd / (sqrt(n)))
> zcal
[1] -6.666667
> cat ("Calculated value at z is: ", zcal)
Calculated value at z is : -6.666667
> pvalue = 2 * (1 - pnorm( abs (zcal)))
> pvalue
[1] 2.616796e-11
```

∴ pvalue is less than 0.05, we reject $H_0: \mu = 15$.

2) Test the Hypothesis $H_0: \mu = 10$ against $H_1: \mu \neq 10$. If random sample of size 400 is drawn with sample mean 10.2 and s d 2.25, test the Hypothesis at 5% level of significance.

⇒
```
> m0 = 10 ; mx = 10.2 ; sd = 2.25 ; n = 400
> zcal = (mx - m0) / (sd / (sqrt (n)))
> zcal
[1] 1.777778
> cat ("Calculated value at z is : ", zcal)
Calculated value at z is 1.777778
> pvalue = 2 * (1 - pnorm (abs (zcal)))
> pvalue
[1] 0.07544036
```

∴ p value is more than 0.05, we accept $H_0: \mu = 10$

3) Test the Hypothesis [H_0]. proportion of smokers in our college is 0.2. A sample is calculated and it is collected and calculated as 0.125. Test the Hypothesis at 5% level of significance

⇒
```
> P = 0.2 ; p = 0.125 ; n = 400 ; Q = 1 - P
> zcal = (p - P) / (sqrt (P * Q / n))
> zcal
[1] -3.75
> cat ("Calculated value at z is ", zcal)
Calculated value at z is -3.75
> pvalue = 2 * (1 - pnorm (abs(zcal)))
> pvalue
[1] 0.0001768346
```

∴ pvalue is less than 0.05, we reject $H_0: \mu = 0.2$

4) Last year, farmers lost 20% of their crops. A random sample of 60 fields are collected and it is found that 9 fields are insect polluted. Test the Hypothesis at 1% level of significance

```
>p = 20/100 ; P = 9/60 ; n = 60 ; Q = 1-P
>zcal = 2 * (1 - pnorm(abs(zcal)))
[1] -0.9682458
>cat ("Calculated Value at 2 is", zcal)
Calculated Value at 2 is -0.9682458
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.3329219
```

∴ pvalue is more than 0.05, we accept H_0.

5] Test the Hypothesis $H_0 : u = 12.5$ from the following sample at 5% level of significance.

```
>x = c(12.25, 11.97, 12.15, 12.08, 12.31,
       12.28, 11.94, 11.89, 12.16, 12.04)
>n = length (x)
>n
[1] 10
>mo = 12.5
>mo
[1] 12.5
```



```
>mx = mean (x)
>mx
[1] 12.107
>var = (n-1)* var(x)/n
>var
[1] 0.019521
>sd = sqrt (var)
>sd
[1] 0.1397176
>zcal = (mx-mo)/(sd/(sqrt(n)))
>zcal
[1] -8.894909
>cat ("Calculated Value is", zcal)
Calculated value is -8.894909
>pvalue = 2*(1-pnorm (abs (zcal)))
>pvalue
[1] 0
```

∴ pvalue is less than 0.05. we reject H_0.

1)
```
>mo = 250; mx = 275 ; sd = 30; n = 100
>zcal = (mx-mo)/(sd/sqrt(n))
>cat ("Calculated Value is", zcal)
Calculated value is 8.333333
>pval = 2* (1-pnorm (abs (zcal)))
>pval
[1] 0
>cat ("Since, p value is less than 0.05,
     Hence, we reject, H0)
```

Since, P value is less than 0.05, Hence, we reject H_0.

07

2) >p=0.8; q=1-P; p=80/100; n=100
>zcal= (p-P)/(sqrt(P*q/n))
>cat ("Calculated value is", zcal)
Calculated value is -3.452847
>pval = 2* (1-pnorm(abs(zcal)))
> pval
[1] 7.722689-05
>cat(" Since, Pvalue is less than 0.05, Hence, we reject H_0 at 5% level ")
Since, Pvalue is less than 0.05, Hence, we reject H_0 at 5% level of significance.

3) >n_1 = 1000; n_2 = 2000; mx_1=67.5; mx_2=68; sd1 = 2.5; sd2 = 2.5;
>zcal = (mx_1 - mx_2)/sqrt ((sd1^2/n_1) + (sd2^2/n_2))
>Cat (" Calculated value is ",zcal)
>Calculated value is -5.163978.
>pval = 2* (1-pnorm (abs(zcal)))
>pval.
[1] 2.4175649-0.7
>Cal (" Since, Pvalue is less than 0.05. Hence, we reject Ho at 5% level of significance .
> Since, Pvalue is less than 0.05. Hence, we reject H_0 at 5% level of significance

Practical - 6

Large Sample Test

Q.1) Let the population mean (the amt spent per customer in a restaurant) is 250. A sample of 100 customer selected. The sample mean is calculated as 275 & S.D is 30. Test the hypothesis at the population mean is 250 or not at 5% level of significance.

Q.2) In a random sample of 1000 student, it is found that 750 use blue pen. Test the hypothesis that the population proportion is 0.8 at 1% level of significance.

→1) H_0: μ = 275 against H_1: $\mu \neq 275$.
> m_0 = 250; mx = 275; n=100; sd = 30
> zcal = (mx - m_0)/(sd/(sqrt(n)))
> zcal
[1] 8.333333
> cat ("Calculated value at 2 is = ", zcal)
Calculated value at 2 is = 8.333333
> pvalue = 2* (1-pnorm(abs(zcal)))
> pvalue
[1] 0
∵ pvalue is less than 0.05, we reject H_0 at 5% level of significance.

Scanned with CamScanner

15.

2) H_0 : P = 0.8 against H_1 : P ≠ 0.8

```
P = 0.8
Q = 1-P
Q
[1] 0.2
> p = 750/1000 = 0.75
> n = 1000
> zcal = (p-P)/(sqrt(P*Q/n))
> zcal
[1] -3.952847
> cat ("Calculated value at 2 is = ", zcal)
calculated value at 2 is -3.952847.
> pvalue = 2* (1-pnorm(abs(zcal)))
> pvalue
[1] 7.72268e-05
```

∴ pvalue is less than 0.05, ∴ we reject H_0 at 1% level of significance.

Q 3) Two random sample of size 1000 & 2000 are drawn from 2 population with same S.D i.e 2.5. The sample means are 67.5 & 68 respectively. Test the hypothesis $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$ at 5% level of significance

Q.4) A study of noise level in 2 hospital given below. Test the claim that the 2 hospital has same level of noise at 1% level of significance

	Hospital A	Hospital B
Size	84	34
mean	61.2	59.4
S.D	7.9	7.5

Q.5) In a sample of 600 student in a clg, 400 use blue ink another clg from a sample of 900 student 450 use blue ink. Test the hypothesis that the proportion of student using blue ink in 2 clgs are equal or not at 1% level of significance

→ 3) $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$
$n_1 = 1000$; $n_2 = 2000$, $mx_1 = 67.5$, $mx_2 = 68$; $sd_1 = sd_2 = 2.5$

```
>zcal=(mx1-mx2)/sqrt((sd1^2/n1)+(sd2^2/n2))
>zcal
[1] -5.163978
>cat("zcalculated is = ",zcal)
calculated is = -5.163978
>pvalue=2*(1-pnorm(abs(zcal)))
>pvalue
[1] 2.417564e-07
```

∴ pvalue is less than 0.05, we reject H_0 at 5% level of significance.

→ 4) $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$.

```
n1=84;n2=34;mx1=61.2;mx2=59.4;sd1=7.5;
sd2=7.5
>zcal=(mx1-mx2)/sqrt((sd1^2/n1)+(sd2^2/n2)
>zcal
[1] 1.162528
>pvalue= 2*(1-pnorm(abs(zcal)))
>pvalue
[1]0.2450211
```

∴ pvalue is ~~greater~~ greater than 0.05, ∴ we accept H_0 at 1% level of significance.

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

$n_1 = 600$, $n_2 = 900$
$p_1 = 400/600 = 0.666667$
$p_2 = 450/900 = 0.5$
$p = (n_1 * p_1 + n_2 * p_2)/(n_1 + n_2) = 0.5666667$
$q = 1 - p = 0.4333333$

```
>zcal=(p1-p2)/sqrt(p*q*(1/n1+1/n2))
>zcal
[1] 6.381534
>cat("calculated is= ",zcal)
calculated is 6.381534
>pvalue =2*(1-pnorm(abs(zcal)))
>pvalue
[1] 1.753222e-10
```

∴ pvalue is less than 0.05, we reject H_0 at 1% level of significance.

6) First, $n_1 = n_2 = 200$, $p_1 = 44/200$, $p_2 = 30/200$
Test hypothesis at 5% level of significance

```
→ >n1=n2=200; p1=44/200; p2=30/200
>p=(n1*p1+n2*p2)/(n1+n2)
>p
[1] 0.185
>q=1-p
>q
```

```
[1] 0.815
> zcal = (p1-p2)/sqrt(p*q*(1/n1+1/n2))
> zcal
[1] 1.802741
> cat("Calculated is =", zcal)
   Calculated is =1.802741
> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 0.07142888.
```

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

Practical 7

Small sample test

Q. 1) 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.

→ $H_0 : \mu = 66$

```
> 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, pvalue = 1.558e-13
alternative hypothesis: true mean is not equal
to 0, 95 percent confidence interval: 65.65871
70.14829
sample estimates:
  mean of x
    67.9
```

∴ pvalue is less than 0.05. Hence, we reject the H_0 at 5% level of significance

```
> if(pvalue > 0.05)
   { cat ("Accept H0")} else(cat("Reject H0")}
```

Q.2) Two groups of students 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 B : 16, 20, 14, 21, 20, 18, 13, 15, 17, 21 -

→ H_0 : There is no difference between two groups

x = c(18, 22, 21, 17, ..., 22, 21)
x = c(16, 20, ..., 17, 21)
t.test(x, y)
Welch Two sample t-test
data : x and y
t = 2.2573, df = 16.376, pvalue = 0.03798.
alternative hypothesis : true difference in means is not equal to 0.
0.1628205 5.0371795.
sample estimates :
mean of x mean of y
20.1 17.5
> if (pvalue > 0.05) {cat("Accept H_0")}
else {cat("Reject H_0")}.
Reject H_0.

Q 3) 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.
H_0 : 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 & Y
t = -2.7815, df = 5, pvalue = 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 H_0")} else
{cat("Reject H_0")}
Accept H_0.

35

Q.4) Two medicines are applied to 2 group of patient
Group1 : 10, 12, 13, 11, 14
Group 2 : 8, 9, 12, 14, 15, 10, 9
is there any significance difference between two groups

Q.5) The following weights before and after a diet program, is diet program effective
Before : 120, 125, 115, 130, 123, 119.
After : 100, 114, 95, 90, 115, 99.

Q.4) H_0 = There is no significance difference between medicines

```
>a =c(10,12,13,11,14)
>b =c(8,9,12,14,15,10,9)
>t.test(a,b)
        Welch Two samplet-test.
data: a and b.
t= 0.65591, df= 9.567, pvalue = 0.5273.
alternative hypothesis: true difference is mean is not
equal to 0 95 percent Confidence interval :
 -1.934382   3.534382
 sample estimates:
 mean of x        mean of y
  11.8              11.0
>pvalue =0.5273
>if (Pvalue>0.05) {cat("Accept H0")} else {cat("Reject H0")}
Accept H0.
```


5) H_0 = There is no difference between before and After

```
>A=c(120,125,115,130,123,119)
>B =c(100,114,95,90,115,99)
>t.test(A,B, paired =T, alternative = "less")
            Paired t-test
data : A and B.
t = 4.3458 , df = 5, pvalue = 0.9963
alternative hypothesis: true difference in mean is
less than 0. 95 percent confidence interval:
       -Inf  29.0295.
Sample estimates:
mean of the differences
      19.83333
>p value = 0.9963
>if (pvalue>0.05) {cat ("Accept H0")} else
            {cat ("Reject H0")}
 Accept H0.
```

Scanned with CamScanner

Practical - 8

Topic - Large and small sample tests

1) $H_0 : \mu = 55$ against $H_1 : \mu \neq 55$

```
>n=100 , mx=52, mo = 55, sd = 7
>zcal = (mx - mo)/(sd / sqrt(n))
>zcal
[1] - 4.285714
>cat ("Calculated value at 2 is = ", zcal).
calculated value at 2 is = -4.285714.
>pvalue = 2* (1- pnorm(abs(zcal)))
>pvalue
[1] 1.82153e -05
```

∴ pvalue is less than 0.05, we reject H_0 at 5% level of significance.

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

```
>P = 0.5
>Q = 1-P
>Q
[1] 0.5
> p = 350/700
> p
[1] 0.5
> n = 700
> zcal = (p-P)/sqrt(P*Q/n)
>zcal
[1] 0
>cat ("Calculated value at 2 is = ", zcal)
Calculated value at 2 is = 0 >
>pvalue = 2* (1-pnorm (abs(zcal)))
>pvalue
[1] 1
```

∴ pvalue is greater than 0.5, we accept H_0 at 1% level of significance.

4) $n = 400$, $H_0 = \mu = 99$ against $H_1 : \mu \neq 99$

```
mo = 100
mx = 99
Var = 64
>sd = sqrt(64)
>sd
[1] 8
zcal = (mx-mo)/sd/sqrt(n)
>zcal
[1] -2.5
>cat ("Calculated value at 2 is = ", zcal)
Calculated value at 2 is = 2.5 >
>pvalue = 2* (1-pnorm(abs (zcal)))
>pvalue
[1] 0.01241933
```

∴ p value is less than 0.5, we reject H_0 at 5% level of significance.

5) H_0: ~~85~~ 66

```
> x = c((63,63,68,69,71,71,72)
> n = length(x)
> n
[1] 7
> t.test(x)
```

One sample t - test

data : x

t = 47.94, df = 6, pvalue = 5.5220-09

alternative hypothesis : true mean is not equal to 0.

95 percent confidence interval :

64.66479   71.62092

sample estimates :

mean of x

68.14286

∵ if (pvalue > 0.05){cat("Accept H0")} else
{cat("Reject H0")}

Reject H_0

7) $H_0 : \mu = 1150$ against $H_1 : \mu \neq 1150$.

n = 100

```
> mx = 1150
> m0 = 1200
> sd = 125
> zcal = (mx - m0)/(sd/sqrt(n))
> zcal
[1] -4
> cat("calculated value at z is:" ; zcal).
```

calculated value at z is = -4.

```
> pvalue = 2* (1 - pnorm (abs (zcal)))
> pvalue
[1] 6.334248e-05.
```


∵ pvalue is less than 0.5, we reject H_0.

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

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

F test to compare two variance

data : x and y

F = 0.78803, num df = 7, denom df = 10, pvalue = 0.7737

alternative hypothesis : true ratio of variances if not equal to 1. 95% confidence interval :

0.199509   3.751881

Sample estimates :

ratio of variance

0.7880255

∵ p value is greater than 0.05, Hence we accept H_0.

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

```
> n1 = 200, n2 = 300
> p1 = 44/200, p2 = 50/300
> 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] 0.9128709
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.3613104
```

∴ pvalue is greater than 0.05. Hence, we accept H_0

3) $H_0: p_1 = p_2$ against $H_1: p_1 \neq p_2$

```
>n1 =1000 ; n2 =1500
>p1 = 2/1000 , p2 =1/150
>p= (n1*p1 +n2*p2)/(n1 +n2)
>p
[1] 0.0012
>q =1-p
>q
[1] 0.9988
>zcal = (p1 -p2)/sqrt(p*q*(1/n1+1/n2))
>zcal
[1] 0.94337522
>pvalue =2*(1-pnorm (abs (zcal)))
>pvalue
[1] 0.345489
```

∴ pvalue is greater than 0.05,
Hence we accept H_0.

Practical 9

Chi-square tests and ANOVA

1) Use the following data to test whether the condition of home & the condition of the child are independent or not.

		Condition of home	
		clean	Dirty
C of child	clean	70	50
	fairly c	80	20
	Dirty	35	45

→ H_0: Condition of home & 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
> pv = chisq.test(y)
> pv
        Pearson's chi-squared test
data: y
X-squared = 25.646, df = 2 p value = 2.698e-06
```

18

∴ pvalue is less than 0.05, we reject H_0 at 5% level of significance.

2) Test the hypothesis that the vaccination & the diseases are independent or not

		Vaccine	
		Aff.	Not Aff
Disease	Aff	70	46
	NA	35	37

→ H_0: Vaccination and Diseases are independent or not.

```
> 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.

∴ pvalue is greater than 0.05, we accept H_0 at 5% L.O.S

(3) Compare ... following data

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

→ H_0: The means 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	Fvalue	Pr (>F)	
ind	3	71.06	23.688	11.73	0.00183	**
Residuals	9	18.17	2.019			

∵ Pvalue is less than 0.05, ∴ we reject H_0 at 5% L.O.S

(4) Following data gives the lifes of the types of 4 brands

Type	Life
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

→ H_0 : Average life of A, B, C, D are equal

```
> x1 = c(20, 23, 18, 17, 18, 22, 24)
> x2 = c(19, 15, 17, 20, 16, 17)
> x3 = c(21, 19, 22, 17, 20)
> x4 = c(15, 14, 16, 18, 14, 16)
> d = stack(list(b1 = x1, b2 = x2, b3 = x3, b4 = x4))
> names(d)
[1] "values" "ind"
> oneway.test(values ~ ind, data = d, var.equal = T)
```

oneway analysis of means

data : values and ind

F = 6.8445, num df = 3, denom df = 20, p value = 0.002349

∵ p value is less than 0.05, we reject H_0 at 5% L.O.S

(5)
```
> x = read.csv("C:/Users/admin/Desktop/marks.csv")
> x
```

	stats	math
1	40	60
2	45	48
3	42	47
4	15	20
5	37	25
6	36	27
7	49	57
8	59	58
9	20	25
10	27	27

```
> am = mean(x$stats)
> am
[1] 37
> median = median(x$stats)
> median
[1] 38.5
> n = length(x$stats)
> n
[1] 10
> sd = sqrt((n-1)*var(x$stats)/n)
> sd
[1] 12.64911
> am1 = mean(x$maths)
> am1
[1] 39.4
```

```
>median = median(x$maths)
>median
[1] 37
>n = length(x$maths)
>n
[1] 10
>sd = sqrt((n-1) * var(x$maths)/n)
>sd
[1] 12.64911
```

Practical No. 10

Non - Parametric test

(1) Following are the amount of sulphur oxide emitted by some industry in 20 days. Apply sign test to test the hypothesis that the population median is 21.5 at 5% L.O.S

17, 15, 20, 29, 19, 18, 22, 25, 27, 9, 24, 20, 17, 6, 24, 14, 15, 23, 24, 26.

→ H_0: Population median is 21.5

```
> x = c (17,15 . . . . 24,26)
> 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.
```

∵ pvalue is greater than 0.05, we accept H_0 at 5% L.O.S.

(2) Following is the data of 10 observation. Apply sign test, Test the hypothesis that the population median is 625 against the alternative which is more than 625.

612, 619, 631, 628, 643, 640, 655, 649, 670, 663

→ H_0: population median is 625

```
> x = c(612, 619 . . . 663)
> 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
```

∵ pvalue is greater than 0.05, ∴ we accept H_0.

(3) Following are a sample. Test the hypothesis, that the population median is 60 against the alternative which is more than 60 at 5% L.O.S 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.

→ H_0: population median = 60

H_1: population median > 60

```
> x = c(63, 65, . . . , 69)
> wilcox.test(x, alter = "greater", mu = 60)
```

Wilcoxon signed rank test with continuity correction

data: x

V = 128, pvalue = 0.03338

alternative hypothesis: true location is greater than 60.

∵ pvalue is less than 0.05, we reject H_0 at 5% L.O.S

(4) Using wilcoxon test, Test the hypothesis, population median is 12 or less than 12

15, 17, 24, 25, 20, 21, 32, 28, 12, 25, 24, 26

→ H_0: population median is 12.

```
> x = c(15, 17, . . . , 26)
> wilcox.test(x, alter = "less", mu = 12)
```

Wilcoxon signed Rank test with Continuity correction data: x

V = 66, pvalue = 0.9986.

alternative hypothesis: true location is less than 12.

∵ pvalue is greater than 0.05, ∴ we accept H_0 at 5% L.O.S.

38

(5) The weights of students before and after they stop smoking are given below. Using WSRT test, there is no significance change.

weight before	weight after
65	72
75	74
75	72
62	66
72	73

→ H_0 : Before & after, there is no change

H_1 : There is change.

```
> x = c(65, ..., 72)
> y = c(72, ..., 73)
> d = x - y
> wilcox.test(d, alter = "two.sided", mu = 0)
Wilcoxon signed rank test with continuity
   correction.
data : d
V = 4.5, p value = 0.4982
alternative hypothesis: true location is not equal to 0
```

∵ pvalue is greater than 0.05, ∴ we accept H_0 at 5% L.O.S.