Practical - 1

Aim: Basics 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 graphical display.

4. It is a free software.

Solve the followings

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

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

2. $2^2 + |-3| + \sqrt{46}$

```
> 2^2 + abs(-3) + sqrt(46)
[1] 13.7082
```

3. $5^3 + 7 \times 5 \times 8 + 46/5$

```
> 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] 5.671567
```

Scanned with CamScanner

6) round off

46 ÷ 7 + 9 × 8

> round (46 / 7+9 * 8)

[1] 79

Q2.

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

[1] 4 9 25 49

> c (6,2,7,5) / c (4,5)

[1] 1.50 0.40 1.75 1.00

> c (2,3,5,7) * c (2,3)

[1] 4 9 10 21

> c (1,6,2,3) * c (-2,-3,-4,-1)

[1] -2 -15 -8 -3

> c (4,6,8,9,4,5) ^ c (1,2,3)

[1] 4 36 512 9 16 125

Scanned with CamScanner

Q3

```
> x <- 20        > y <- 30        > z <- 2
> x^3 + y^3 + z
[1] 27402

> log2(x^2 + y)
[1] 2.073644

> x^2 + y^2
[1] 1300
```

Q4

```
> 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.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, 6,7,3))
> x     [,1]  [,2]  [,3]
[1,]     4    -2     6
[2,]     7     0     7
[3,]     9    -5     3
```

Scanned with CamScanner

```
> 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    -7    13
[2,]   19    -4    16
[3,]   24   -11     8

> 2 * x + 3 * y
     [,1]  [,2]  [,3]
[1,]   38   -19    33
[2,]   50   -12    44
[3,]   63   -28    21
```

Q.6. Marks of statistics of CS Batch D.

```
x = c (59,20,35,24,46, 56,55,45, 27,22, 41,
       58,54,40,60,32,36,29,35,39)
> x = c (data)
> breaks = seq (20,60,5)
```

Scanned with CamScanner

```
> 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

Scanned with CamScanner

Practical - 2

Aim: Probability Distribution

1) Check whether the followings are P.m.f 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.f the

$\Sigma P(x) = 1$

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

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

$= 1.0$

$\because P(2) = -0.5$ it cant be a

$\therefore$ Probability mass function.

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

Scanned with CamScanner

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

The condition for pmf is $\Sigma P(x) = 1$

So.

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

$= 0.2 + 0.2 + 0.3 + 0.2 + 0.2$

$= 1.1$

∴ The given data is not a P.m.f because the $P(x) \neq 1$

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) $\Sigma P(x) = 1$

$\Sigma P(x) = P(10) + P(20) + P(30) + P(40) + P(50)$

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

$= 1$

∴ The given data is P.m.f.

Scanned with CamScanner

SNO

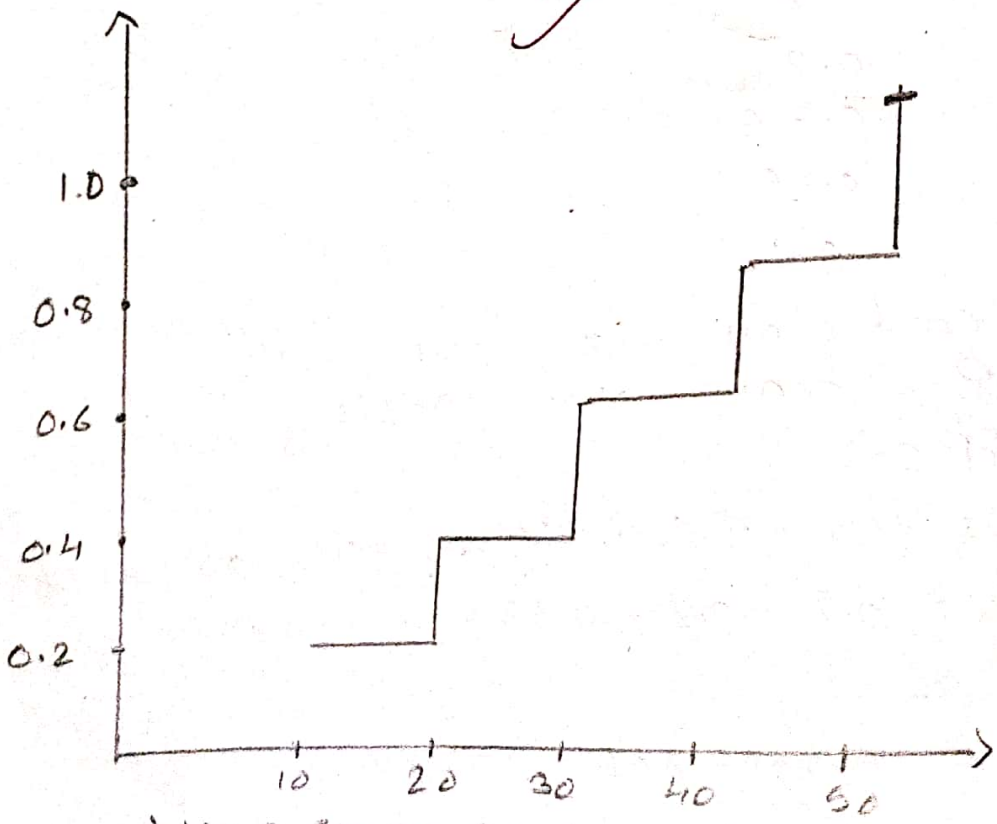
Code:

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

Q2. 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

$$F(x) = \begin{cases} 0 & x < 10 \\ 0.2 & 10 \le x < 20 \\ 0.4 & 20 \le x \; 30 \\ 0.75 & 30 \le x \; 40 \\ 0.90 & 40 \le x \; 50 \\ 1.0 & x \geq 50 \end{cases}$$




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

Scanned with CamScanner

Find

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

$$F(x) = \begin{cases} 0 & x < 1 \\ = 0.15 & 1 \le x < 2 \\ = 0.40 & 2 \le x < 3 \\ = 0.50 & 3 \le x < 4 \\ = 0.70 & 4 \le x < 5 \\ = 0.90 & 5 \le x < 6 \\ = 1.00 & x \ge 6 \end{cases}$$

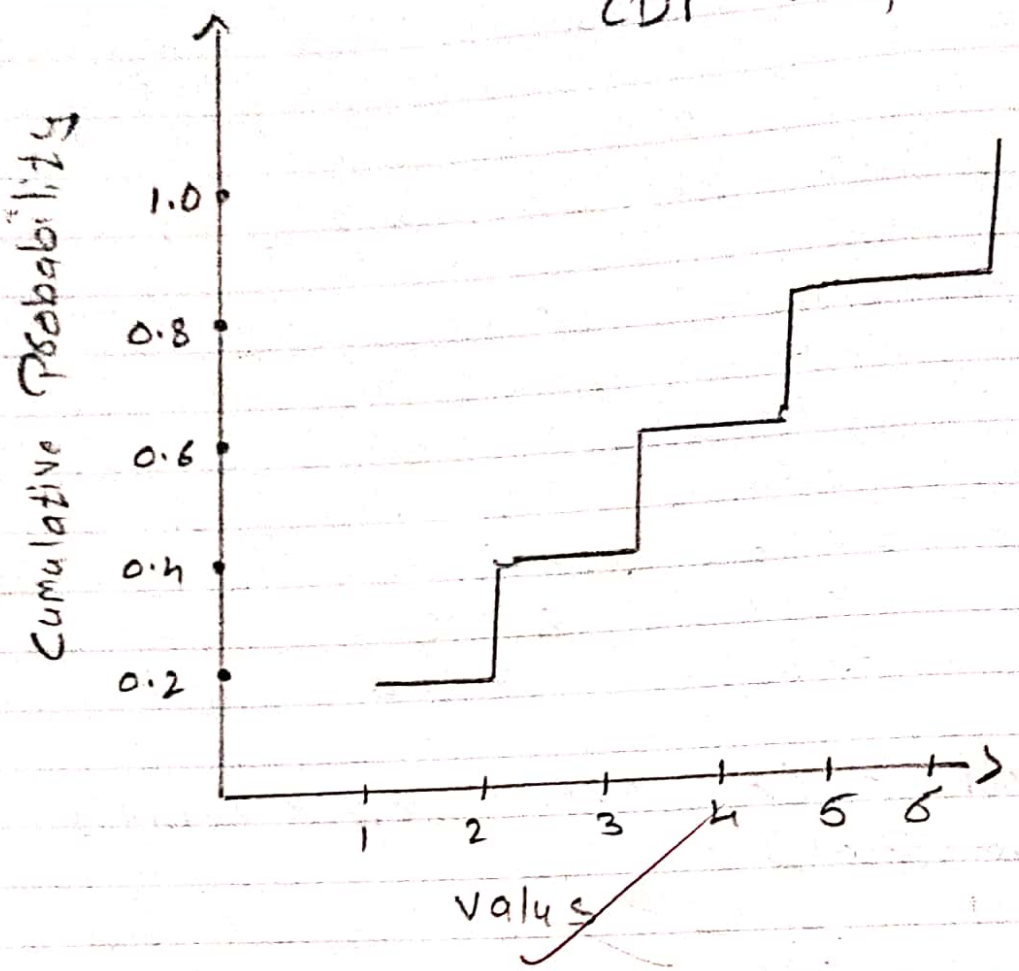
```
> Prob = c (0.15, 0.25, 0.1, 0.2, 0.2, 0.1)
> sum (Prob)
[1] 1
> cumsum (Prob)
[1] 0.15 , 0.40 , 0.50 , 0.70 , 0.90 , 1.00
> x = c (1, 2, 3, 4, 5, 6)
> plot ( x, cumsum (Prob), "s", xlab = "value",
    ylab = "Cumulative Probability",
  main = " CDF Graph,          = " brown")
```

cumu freq

for line

Scanned with CamScanner

CDF graph



3. Check that whether the following is P.d.f or not.

i) $f(x) = 3-2x \quad ; \quad 0 \le x \le 1$

ii) $f(x) = 3x^2 \quad ; \quad 0 < x < 1$

i) $f(x) = 3-2x$

$= \int_0^1 f(x)$

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

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

$= [3x - x^2]_0^1 = 2$

$\therefore$ The $\int_0^1 f(x) = 1$

$\therefore$ It is not a Pdf

ii) $f(x) = 3x^2 \; ; \quad 0 < x < 1$

$\int_0^1 f(x)$

$= \int_0^1 3x^2$

$= 3\int_0^1 x^2$

$= \left[3 \cdot \frac{x^3}{3}\right]_0^1 \qquad \because x^n = \frac{x^{n+1}}{n+1}$

$= x3$

$= 1$.

The $\int_0^1 f(x) = 1$

$\therefore$ It is a Pdf.

Scanned with CamScanner

Practical 3

Topic : Binomial distribution

$P(X = x)$ = dbinom (x, n, P)
$P(X \leq x)$ = pbinom (x, n, P)
$P(X > x)$ = $1 -$ pbinom (x, n, P)
If x is unknown
$P_1 = P(X \leq x)$ = 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) $n = 12$, $P = 0.25$, find the following probability.
   i) $P(X = 5)$
   ii) $P(X \leq 5)$
   iii) $P(X > 7)$
   iv) $P(5 < X < 7)$

Scanned with CamScanner

1) 1) > dbinom(10, 100, 0.1)

> [1] 0.1318683

2) 1) dbinom(4, 12, 0.2)

[1] 0.1328796

2) pbinom(4, 12, 0.2)

[1] 0.9272445

3) 1-pbinom(5, 12, 0.2)

[1] 0.01940528

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

0 - 0.69049
1 - 0.32805
2 - 0.07290
3 - 0.00810
4 - 0.00045
5 - 0.00001

4) 1) dbinom(5, 12, 0.25)

[1] 0.1032414

2) pbinom(5, 12, 0.25)

[1] 0.9455978

3) 1-pbinom(7, 12, 0.25)

[1] 0.00278151

4) dbinom(6, 12, 0.25)

[1] 0.04014945

Scanned with CamScanner

5) The probability of a salesman making a sale to customer 0.15. Find the probability of 10 customer.
i) No sales out of 10
ii) More than 3 sales out of 20 customer.

6) A sales man has 20% probability of making a sale to customer. Out of 30 customer's what minimum number of sales he can make with 88% of probability.

7) X follows binomial distribution with $n=10$, $p=0.3$. Plot the graph of p.m.f. and c.d.f.

x= c
sum(x)

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 ("x values" = x, "Probility" = Prob)
> Print (d)

	Xvalues	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	0.0015
10	9	0.0001
11	10	0.0000

Scanned with CamScanner

Practical 4

Aim: Normal Distribution

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

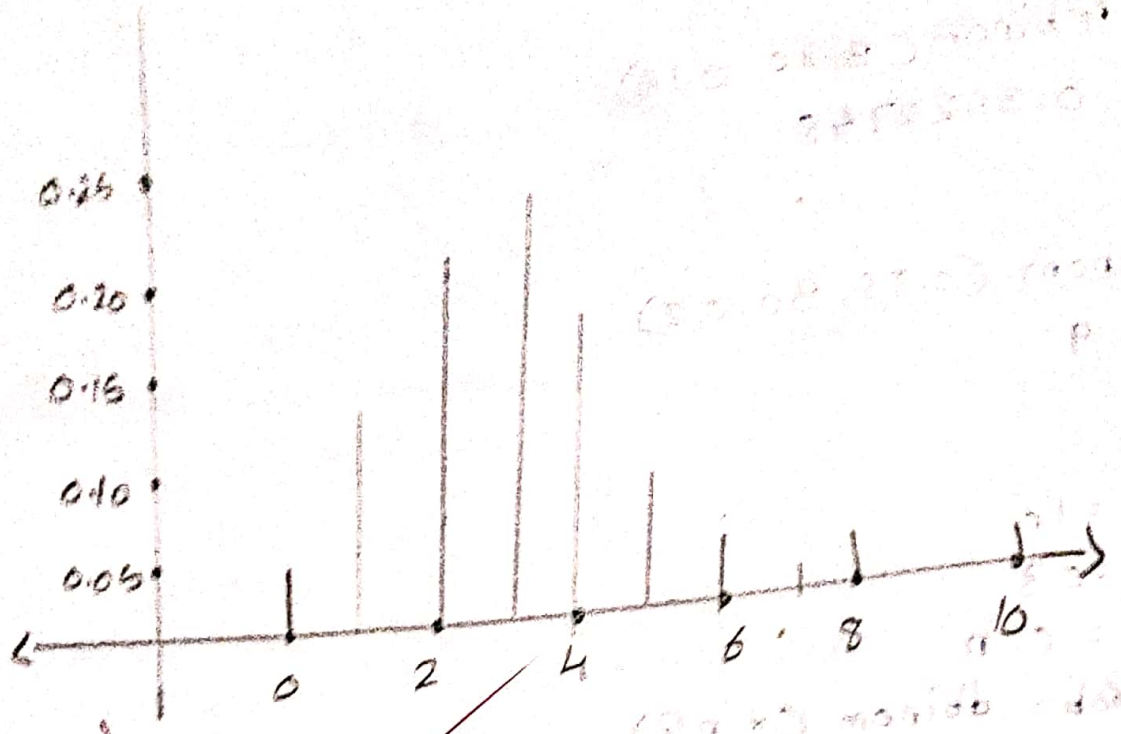
- A random variable x follows normal distribution with mean = $\mu = 12$ and S.D = $\sigma = 3$. Find
  i) $P(x \le 15)$ ii) $P(10 \le x \le 13)$ iii) $P(x > 14)$
  iv) generate 5 observations (random ~~variat~~ numbers)

Code:

```
> P1 = Pnorm ( 15, 12, 3)
> P1
[1]     0.8413447
> cat ( " P (x <= 15) = " , P1)
P (x <= 15) = ~~Cat~~ 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
```

Scanned with CamScanner

> plot (x, Prob, t)



> plot (x, prob, "s")

```
> cat ("P(x>14)", P3)
P(x>14): 0.2624485
> P4 = rnorm (5, 12, 3)
> P4
[1] 15.254723  16.648505  11.280515  6.419944  12.27246
```

Q2 X follows normal distribution with $\mu=10$, $\sigma=2$
find i) $P(x<7)$ ii) $P(5<x<12)$
iii) $P(x>12)$ iv) Generate 10 observations
v) Find k such that $P(x<k)=0.4$

```
> a1 = pnorm (7, 10, 2)
> a1
[1] 0.0668072
> a2 = pnorm (6, 10, 2) - pnorm (12, 10, 2)
> a2
[1] -0.8351351
> a3 = 1-pnorm (12, 10, 2)
> a3
[1] 0.1586553
> a4 = rnorm (10, 10, 2)
> a4
[1] 11.608931  9.920417  12.63774  8.078354  8.721385
    9.193726  9.366824  11.70716  9.537584  10.71500

> a5 = qnorm (0.4, 10, 2)
> a5
[1] 9.493306
```

Scanned with CamScanner

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

```
> rnorm(5, 15, 4)
[1] 10.76499   7.793249   9.953444   13.345904
    17.509668
> am = mean(rnorm(
> am
[1] 11.87345
> cat("sample mean is = ", am)
sample mean is = 11.87345
> me = median(x)
> me
[1] 10.76499
> cat("Median is = ", me)
Median is = 10.76499
> n = 5
> v = (n-1) * var(x)/n
> v
[1] 11.09965
> sd = sqrt(v)
> sd
[1] 3.331613
> cat("S.D is = ", sd)
S.D is = 3.331613
```

Scanned with CamScanner

Q5. Plot the standard normal graph.

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

Scanned with CamScanner

Practical - 5

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.

⇒
```
> 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 $H_0 = \mu = 15$

2) Test the hypothesis $H_0 : \mu = 10$ against $H_1 : \mu \neq 10$. A random sample size of 400 is drawn with sample mean = 10.2 and S.D = 2.25. Test the hypothesis.

Scanned with CamScanner

```
>mo = 10
>n = 400
>mx = 10.2
>sd = 2.25
>zcal = (mx - mo) / (sd / sqrt(n)))
>zcal
[1] 1.77778
>pvalue = 2 * (1 - pnorm (abs (zcal)))
>pvalue
[1] 0.07544036
```

∵ The value p value is greater than 0.05

∴ The value is accepted.

) Test the hypothesis Ho : 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 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] 0.0001768346 (Reject)
```

Scanned with CamScanner

4) Last year farmer's lost 20% of their crops. A random sample of 60 fields are counted. And it is found that 9 filled crops are insect polluted. Test the hypothesis at 1% level of significance.

=> >P = 0.2
>p = 9/60
>n = 60
>zcal = (p-P) / (sqrt(P * Q/n))
>z.cal
[1] - 0.9682458
>P value = 2 * (1-Pnorm(abs(zcal)))
>pvalue
[1] 0.3329216
∵ The value is 0.1 so value is accepted.

5) Test the hypothesis Ho: μ = 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
>mx = mean(x)
>mx
[1] 12.107

Scanned with CamScanner

```
> variance = (n-1) * var(x)/n
> variance
[1]  0.019521
> sd = sqrt(variance)
> sd
[1]  0.1397176
> mo = 12.5
> t = (mx - mo) / (sd/sqrt(n))
> t
[1]  -8.894909
> pvalue = 2*(1-pnorm(abs(t)))
> pvalue
[1]  0
```

∵ The value is less than 0.05 the value is accepted.

Practical VI

Large sample test

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

=> >mo = 250; mx = 275; sd = 30; n = 100
>Zcal = (mx - mo) / (sd/sqrt(n))
calculate value is 8.2333
>pval = 2 * (1-pnorm (abs(zcal)))
>pval
[1] 0
> cat ("since, pvalue is than 0.05, Hence we reject Ho)

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

=> >P = 0.8; Q = 1-P; p = 0.01/1000; n = 1000
>zcal = (p.p) / (sqrt (P * Q/n))
>cat ("calculated value is -3.492847
>pval = 2 * (1-pnorm (abs(zcal)))
>pval
[1] 7.72268e-05

Scanned with CamScanner

3) 2 random sample spaces of size 1000 and 2000 are drawn from 2 population with the same S.D = 2.5. The sample means are 67.5 and 68, respectively. Test the Hypothesis at 5% level of significance.

=> > n1 = 1000; n2 = 2000; mx1 = 675; mx2; 68;
sd1 = 2.5; sd2 = 2.5

> zcal = (mx1 - mx2) / sqrt ((sd1^2)/n1) + (sd2^2)/n2))
> zcal ("calculated value is", zcal)
> calculated value is -6.163978
> pval = 2 * (1 - Pnorm (abs (zcal)))
> pval

[1] 2.417564e-07

> cat ("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 Ho at 5% level of significance.

4) A study of noise level of 2 hospitals is given below. Test the claim that the 2 hospitals have 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

Scanned with CamScanner

SCO

⇒ >n = 84; n2 = 34, mx1 = 61.2; mx2 = 59.4; sd1 = 7; sd2 = 15

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

>cat ("calculated value is", zcal)

calculated value is 1.162928

>Pval = 2 * (1 - Pnorm (abs (zcal)))

>Pval

[1] 0.2460211

>cat("since, Pvalue is greater than 0.05, Hence we accept Ho at 1% level of significance)

since, Pvalue is greater than 0.05, Hence we accept Ho at 1% level of significance.

5) In a sample of 500 students in a college, 400 use blue ink in another College from a sample of 900, 450 use blue ink. Test the hypothesis that the Proportion of students using blue ink using 2 colleges are equal or not at 1% level of significance

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

>n1 = 600; n2 = 900; p1 = 400/600; p2 = 450/900

>p = (n1 * p1 + n2 * p2) / (n1 + n2)

>q = 1 - p

>zcal = (p1 - p2) / sqrt (p * q * (1/n1 + 1/n2))

>cat ("calculate value is", zcal)

Calculated value is 6.381324.

>p

[1] 0.5666667

Scanned with CamScanner

```
>pval = 2 * (1-pnorm(abs(zcal)))
>pval
[1] 1.75322 e-10
>cat ("since Pvalue is less than 0.05, Hence we reject the Ho at 1% level of significance.")
```

since, Pvalue is less than 0.05, Hence we reject Ho at 1% level of significance

6) For sample size, n1 = 200, n2 = 200; P1 = 44/200, P2 = 30/200 Test at 5% level of significance

=>
```
>n1 = 200; n2 = 200; P1 = 44/200, P2 = 30/200
>P = (n1 * P1 + n2 * P2) / (n1+n2)
>P
[1] 0.185
>q
[1] 0.815
>zcal = (P1-P2) / sqrt(P * q * (1/n1 + 1/n2))
> cat ("calculated value", zcal)
calculate value is 1.802741
>pval = 2 * (1-pnorm(abs(zcal)))
>pval
[1] 0.0714288
>cat ("since Pvalue is less than 0.05, Hence we reject Ho at 1% level of significance.")
```

Practical VII

Small sample Test.

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

=> Ho : $\mu = 66$

> x = c (66, 63, 66, 67, 68, 69, 70, 77, 72)

> t. test (x)

One sample t-test

data : x

t = 72.125, df = 8, P-value = 1.521e-12

alternative hypothesis : true mean is not equal to 0.

95 percent confidence interval :

65.82588 70.17412

Sample estimates :

mean of x :

68

∴ The value is less than 0.05 we reject the hypothesis at 5% level of signiface.

Q2. Two groups of student scored the following marks. Test the hypothesis that there is no significance 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

Scanned with CamScanner

-> Ho: There is no difference between 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 hypthesis: true difference in means
is not equal to 0.

95 percent confidence interval:
0.1628200          6.0371795

Sample estimates:
mean of x   mean of y
   20.1        17.5

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

Scanned with CamScanner

(Paired T. Test)

3) The sales data of 86 shops before and after a special campaign are given below.

Before: 53, 28, 31, 48, 60, 42

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

Test the hypothesis that the campaign is affective or not.

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

=> > x = c (Before)

> y = c (After)

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

Paired t-test

data: x and y

t = -2.7315, df = 5; P value = 0.9906

alternative hypothesis:

True difference in means is greater than 0.

95 percent confidence interval: -6.033547, inf

sample estimates.

Mean of difference

-3.5

∵ Pvalue is greater than 0.05, we accept the hypothesis at 5% Level of significance

Scanned with CamScanner

Q.4, Following are the weights before and after the diet Program. Is the diet Program effective:
Before: 120, 125, 115, 130, 123, 119
After: 100, 114, 95, 90, 119, 99.

Sol^n: Ho: There is bio significance difference
> x = c(Before)
> x = c (After)
> t.test (x, y) Paired = T, alternative = "Less")
Paired t-test
data: x and y
t = 4.3458, df = 5, P value = 0.993
alternative hypothesis: true difference in means is less than 0.
95 percent confidence interval:
-Inf 29.0295
Sample estimates:
mean of the differences
19.8333

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

Scanned with CamScanner

Practical No. 8

Topic: Large and Small Test

FYCS Practical 8
Large and Small Sample tests

Q.1 The arithmetic mean of a sample of 100 items from a large population is 52. If the standard is 7, test the hypothesis that the population mean is 55 against the alternative it is more than 55 at 5% LOS.

Q.2 In a big city 350 out of 700 males are found to be smokers. Ddoes the information supports that exactly half of the males in the city are smokers ? Test at 1% LOS.

Q.3 Thousand article from a factory:A are found to have 2% defectives ,1500 atricles from a 2nd factory:B are found to have 1% defective. Test at 5% LOS that the two factory are similar are not.

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

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

Q.6. Two random samples were drawn from 2 normal populations and there values are A-66,67,75,76,82,84,88,90,92 B-64,66,74,78,82,85,87,92,93,95,97. Test whether the population have the same variance at 5% LOS.

7. A company producing light bulbs finds that mean life span of the population of bulbs is 1200 hours with s.d. 125. A sample of 100 bulbs have mean 1150 hours. Test whether the difference between population and sample mean is significantly different?

8. From each of two consignments of apples, a sample of size 200 is drawn and number of bad apples are counted. Test whether the proportion of rotton apples in two assignments are significantly different at 1 % LOS.?

	Sample size	No. of bad apples
Consignment 1	200	44
Consignment 2	300	56

Scanned with CamScanner

6) 2 medicines are applied to two groups of patient respectively.

grp 1: 10, 12, 13, 11, 14
grp 2: 8, 9, 12, 14, 15, 10, 9
Is there any significance difference between 2 medicines.

He: There is no significance difference

```
> x = c (grp 1)
> y = c(grp2)
> t. test (x, y)
  data: x and y
  t = 0.80384        df = 9.7594        Pvalue = 0.4405
alternative hypothesis: true difference in means is not equal to 0.
95 percent confidence interval:
        -0.9698563        4.2981886
  Sample estimates:
    Mean of x        Mean of y
    12.0000          10.33333
```

∵ Pvalue is greater than 0.05 we accept the hypothesis at 6% level of significance

->) $H_0: \mu = 65$, $H_1: \mu \neq 65$

```
>n=100
>mx=62
>mo=65
>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 H_1 at 5% level of significance.

Sol^n 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_1 at 1% level of significance.

Scanned with CamScanner

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

```
>n1 = 1000
>n2 = 1600
>p1 = 2/1000
>p2 = 1/1600
>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
```

∵ Pvalue is greater than 0.05 we accept H_0 and 5% level of significance

4) $H_0 : \mu = 100$ against $H_1 : \mu \neq 100$

```
>var = 64
>n = 400
>m0 = 100
>mx = 99.0
>sd = sqrt (var)
>sd
[1] 8
>zcal = (mx-m0) / (sd / (sqrt (n)))
>zcal
```

Scanned with CamScanner

SCD

```
>zcal = (mx - mo) / (sd (sqrt (n)))
>zcal
[1] - 2.5
>pvalue = 2 * (1-Pnorm (abs (zcal)))
>pvalue
[1] 0.01241933
```

Since pvalue is less than 0.05 we reject H_0 at 5% level of significance

->5) $H_0 : \mu = 66$ against $H_1 : \mu \neq 66$

```
>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 = 5.622e-8
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 P value is less than 0.05 we reject H_0 at 1% level of significance.

Scanned with CamScanner

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, 74, 78, 82, 85, 87, 92, 93, 95, 97)
>var.test(x,y)
```

F test to compare two variances

data: x and y

F=0.788803, num df = 7 denom. df = 10, P value=0.7137

alternative hypothesis : true ratio of variances is not equal to 1.

95 Percent confidence interval:

0.199509 3.751881

sample estimates:

ratio of variances

0.7880295

∵ Pvalue is greater than 0.06 we accept H_0 at 6% level of significance.

Sol^n ⑦ $H_0: \mu = 1150$ against $H_1: \mu \neq 1150$

```
>n=100
>mx = 1150
>mo = 1200
>sd = 125
>zcal = (mx- mo) / (sd /sqrt(n))
>zcal
[1] -4
>Pvalue = 2*(1-Pnorm(abs(zcal)))
```

Scanned with CamScanner

Practical - IX

Chi-square Test And Anova

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

		Condition of Home	
		Clean	Dirty
Condition of child	Clean	70	50
	Family 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

```
> pv = chisq.test(y)
> pv
```

Pearson's chi-squared test

data: y

X-squared = 25.646, df = 2, pvalue = 2.698e-06

pvalue is less than 0.05

∴ We reject H_0

Scanned with CamScanner

>Pvalue

[1] 6.334248p-06

∵ Pvalue is less than 0.05 we reject H_0

Sol^n (8): $H_0: P_1 = P_2$ 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] 0.9128709
>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.

8

2) Test the hypothesis that vaccines and diseases are independent or not.

		Vaccines	
		Affer	Not Aff.
Diseases	Aff	70	46
	Not Aff	35	37

=> H_0: vaccines and diseases are independent.

```
> 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
> chisq.test(y)
        Pearson's chi-squared test with Yates' continuity correction
data:  y
X-squared = 2.0275, df = 1     Pvalue = 0.1545
```

∵ Pvalue is greater than 0.05, ∴ we reject H_0.

3) Perform ANOVA (analysis of variance) for following data.

Type	Observation
A	50, 62
B	53, 55, 53
C	60, 58, 57, 56
D	52, 54, 54, 55

Scanned with CamScanner

=> Ho: The means are equal for A, B, C, D.

```
> x1 = c(50, 62)
> x2 = c(53, 55, 53)
> x3 = c(60, 55, 57, 56)
> x4 = c(62, 54, 54, 56)
> 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
```

∵ pvalue is less than 0.06 ∴ we reject H_0

```
> anova = aov(values ~ ind, data = d)
> summary(anova)
```

	Df	Sum sq	Mean sq	F value	Pr(>F)
ind	3	71.06	23.688	11.73	0.00153 *
Residuals	9	18.17	2.019		

---

signif. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

4) The following data gives the life of tyres of brands

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

Scanned with CamScanner

080

⇒ H_0: The average life A, B, C and D are equal

```
> x1 = c(20, 22, 18, 17, 15, 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"
> one.way.test(values ~ ind, data = d, var.equal = T)

        One-way analysis of means

data:  values and ind
F = 6.945, num df = 3, denom df = 20, p-value = 0.00234
> 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    0.00235 **
Residual    20  89.06   4.453
---
> x = read.csv("c:/Users/Administrator/Desktop/Marks.csv")
> x
   stats  Maths
1     40     60
2     45     45
3     42     47
4     15     20
5     37     25
6     36     27
7     49     57
8     69     69
9     20     26
10    27     27
```

Scanned with CamScanner

```
>am1 = mean (x $ stats)
>am1
[1] 37
> am2 = mean (x $ Math)
>am2
[1] 39.4
> med1 = median (x $ stats)
>med 1
[1] 38.5
> med2 = median (x $ Maths)
>mod 2
[1]
>n1 = length (x $ Maths)
>n1
[1] 10
>n2= length (x $ stats)
>n2
[1] 10
>sd1= sqrt ((n1-1) * var (x $ stats) /n1)
>sd1
[1] 12.64911
>sd2= sqrt ((n2-1) * var (x $ Maths) /n2)
>sd2
[1] 15.20000
> cor (x $ stats, x $ Maths)
[1] 0.830618
```

Scanned with CamScanner

Practical 10

Non-Parametric Test

1) Following are the amounts of sulphur dioxide emitted by the industry in 20 days. Apply sign test to test the hypothesis that the population meditation is 21.5 at 5% LOS

17, 15, 20, 29, 19, 18, 22, 25, 27, 9, 24, 20, 17, 6, 24, 8, 15, 23, 24, 26

=> Ho : Population median is 21.5

```
x = c(17, 15, 20 .....  , 23, 24, 26)
> mc = 21.5
> sp = length(x[x > mc])
> sn = length(x[x < mc])
> n = sp + sn
> pv = pbinom(sp, n, 0.5)
> pv
[1] 0.4119015
```

∵ we accept Ho. ∴ P value is grater than 0.05.

* If H1 : Po me ≠ or me <= then pv = pbinom(sp, n, 0.5) else pv = pbinom(sn, n, 0.5)

2) Following are the data of 6 observation Apply sign test to test the hypothesis that the population median 625 against alternative greater than 625

Scanned with CamScanner

612, 619, 631, 628, 643, 640, 665, 649, 670, 663

Ho: Population median is 629

```
> x = c(612, 619, 631, ..... 670, 663)
> me = 629
> sp = length(x[x > me])
> sn = length(x[x < mo])
> n = sp + sn
> pv = pbinom(sn, n, 0.5)
> pv
[1] 0.0546875
```

∴ Pvalue is greater than 0.05 we accept Ho.

3) Following are the data given. Test the hypothesis that population median is 60 against alternative greater than 60 at 5% LOS using Wilcoxon signed Rank Test.

63, 65, 60, 89, 61, 71, 58, 91, 69, 62, 63, 39, 72, 69, 48, 66, 72, 63, 87, 69

Ho: Population median is 60
Hi: Population median is greater than 60.

```
> x = c(63, 65, ..... 87, 69)
> wilcox.test(x, alter = "greater", me = 60)
  Wilcoxon rank sum test with continuity correction
data: x and me
W = 155, Pvalue = 0.2031
alternative hypothesis: true location shift is greater than 0
Warning message.
```

Scanned with CamScanner

In wilcox.test default (x, alter = "greater", mu):
cannot compute exact Pvalue with ties
∵ Pvalue is greater than 0.05 we accept H_0

4) Following are the data given. Test the hypothesis that population median is 12 against alternative less than 12 using Wilcox Test
15, 17, 24, 25, 26, 21, 32, 28, 12, 25, 24, 26

⇒ H_0: population median is 12
H_1: population median is less than 12

```
x = c (15,17,24, ..... : 24,26)
>me = 12
>wilcox.test (x, alter: "less", mu)
    Wilcoxon rank sum test with continuity correction
data: x and me
W = 11.5 , Pvalue = 0.9463
alternative hypothesis : true location shift
is less than 0
Warning message:
In wilcox.test.default (x, alter, "less", mu):
    Cannot compute exact P-value with ties
```

∵ Pvalue is greater than 0.05 we accept H_0.

5) The weighs of students b/4 stopped smoking using wilcox Test, Test that there is no sign change

W_0	W_1
65	72
75	74
75	72
62	66
72	73

H_0: Before and After, There is no significant change
H_1: There is " siginificant change.

```
x = c (65, 75, 75, 62, 72)
y = c (72, 74, 72, 66, 73)
d = x - y
mu = 0
>wilcox.test (d, alter = "two.sided", mu)
        W = 2                    p value = 1
alternative hypothesis: true location shift is
                        not equal to 0.
```

∵ p value is greater than 0.05 we accept H_0.

Scanned with CamScanner