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

Q1 Solve the following

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

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

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

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

2) round off

$46 \div 7 + 9 \times 8$

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

Q2)

```
>c(2,3,5,7)*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 -18 -8 -3
>c (4,6,8,9,4,5)^c(1,2,3)
[1] 4 36 512 9 16
25
```

Q3

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

Q4

```
>x <- matrix (n row = 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
```

Q5) Find $x+y$ and $2x+3y$ where $x = \begin{bmatrix} 4 & -2 \\ 7 & 0 \\ 9 & -5 \end{bmatrix}$

$y = \begin{bmatrix} 1 & -5 & 7 \\ 12 & -4 & 9 \\ 15 & -6 & 5 \end{bmatrix}$

```
> x <- matrix (nrow = 3, ncol= 3, data
  = c(4,7,9,-2, 0, -5, 6, 7, 3))
> x
       [,1]   [,2]  [,3]
[1,]     4     -2
[2,]     7      0     6
[3,]     9     -5     7
                      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     -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 CS Batch B

```
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 = 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

Practical : 02

Topic : Probability distribution

1) 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

If the given data is pmf then $\sum P(x) = 1$

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

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

$= 1.0$

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

$\therefore$ probability mass function

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

2)

x	$P(x)$
1	0.2
2	0.32
3	0.3
4	0.2
5	0.2

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

so,

$$\Sigma P(x) = P(1) + P(2) + P(3) + P(4) + P(5)$$
$$= 0.2 + 0.2 + 0.3 + 0.2 + 0.2$$
$$= 1.1$$

$\therefore$ The given data is not a pmf because $P(x) \neq 1$

3)

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

The condition for pmf is

1) $P(x) \geq 0 \quad \forall x$ satisfy
2) $\Sigma P(x) = 1$

$$\Sigma P(x) = P(10) + P(20) + P(30) + P(40) + P(50)$$
$$= 0.2 + 0.2 + 0.35 + 0.15 + 0.1$$
$$= 1$$

$\therefore$ The given data is pmf

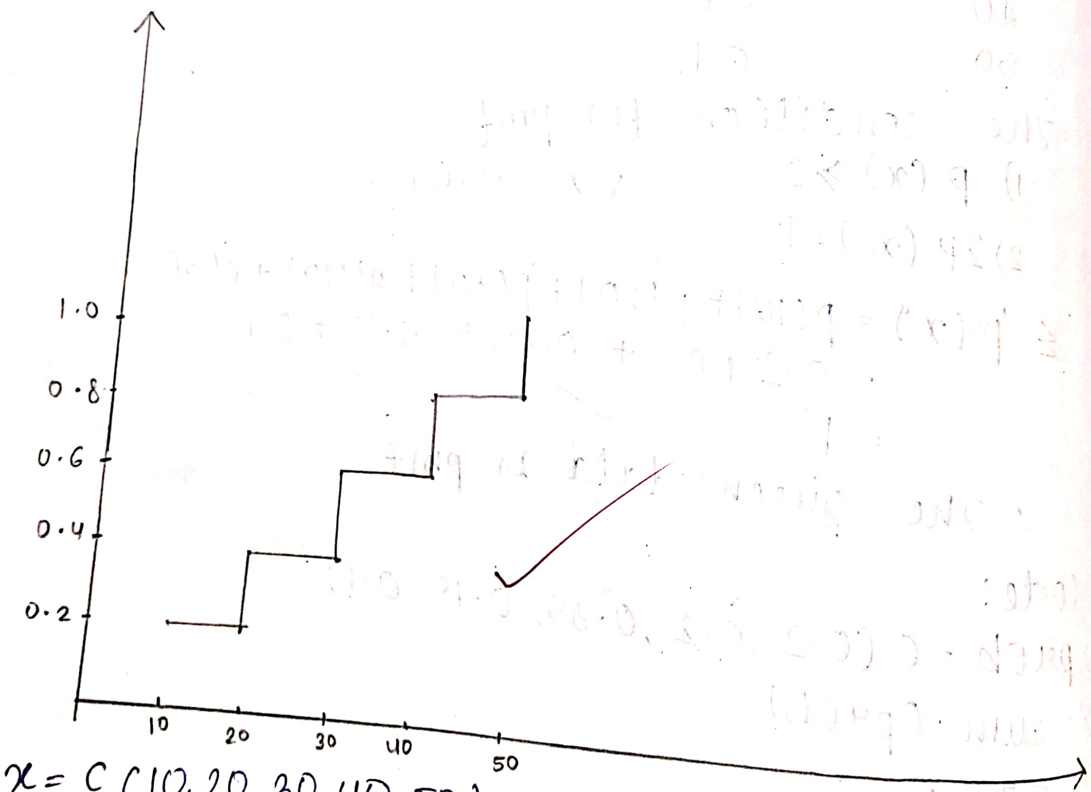
Code:

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

Q2) i) Find the cdf for ~~pmf~~ pmf. Sketch the graph.

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

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




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

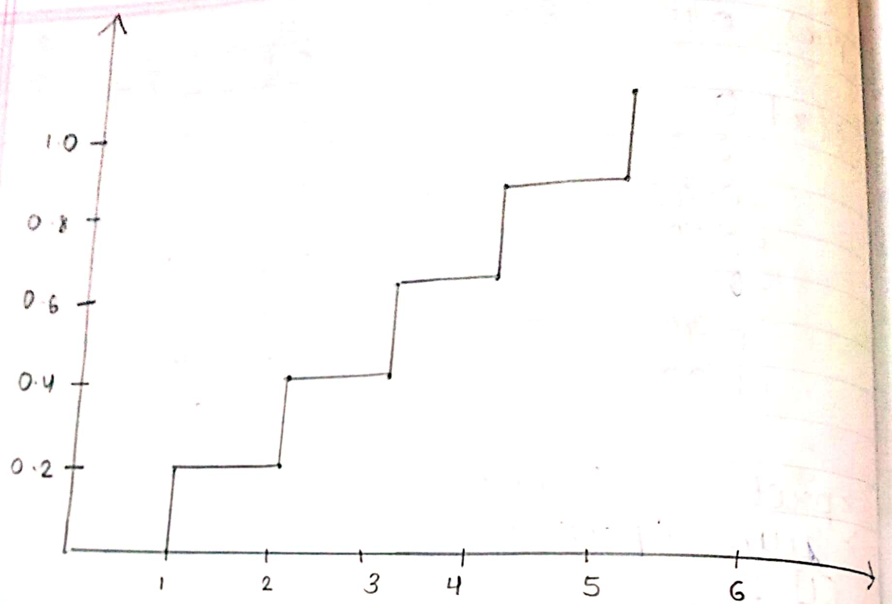
Scanned with CamScanner

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

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

```
>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 , 0.50 , 0.70 , 0.90 , 1.00
> x = c (1, 2, 3, 4, 5, 6)
> plot (x, cumsum (prob) , "S" xlab
    = "value",
  y lab = "cumulative probability"
  main = "CDF graph", "col = "brown")
```

Scanned with CamScanner



Q3 Check that whether the following is pdf or not

I $f(x) = 3-2x \; ; \; 0 \leq x \leq 1$

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

1 $f(x) = 3-2x$

$= \int_0^1 f(x)$

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

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

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

$\therefore$ The $\int_0^1 f(x) = 1$ $\quad\therefore$ It is not a pdf 40

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

$\int_0^1 f(x)$

$= \int_0^1 3x^2$

$= 3\int_0^1 x^2$

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

$= x^3$

$= 1$

The $\int_0^1 f(x) = 1$ $\quad\therefore$ It is a pdf.

Scanned with CamScanner

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

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

ii) pbinom (4,12, 0.2)
[1] 0.92744445

iii) 1-pbinom (5,12,0.2)
[1] 0.01940528

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

4) i) 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

Practical :03

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 100 trials with $P = 0.1$

2 Suppose there are 12 MCQ, each question has 5 option out of which 1 is correct. Find the probability of having i) exactly 4 correct answers.
ii) atmost 4 correct answers.
iii) 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 probabilities

1 $P(x = 5)$
2 $P(x > 7)$
3 $P(x \leq 5)$
4 $P(5 < x < 7)$

Scanned with CamScanner

5) The probability of a salesman making a sale to a customer 0.15
Find the proability of 10 customer.
1 NO sales out of
2 More 3 sales out of 20 customer

6) A salesman has 20% probability of making a sale to customer. out of 30 customers what minimum number of sales he can make with 88% probability.

7) X follows binomial distribution. with $n = 10$, $p = 0.3$, Plot the graph of pmf and cdf.

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

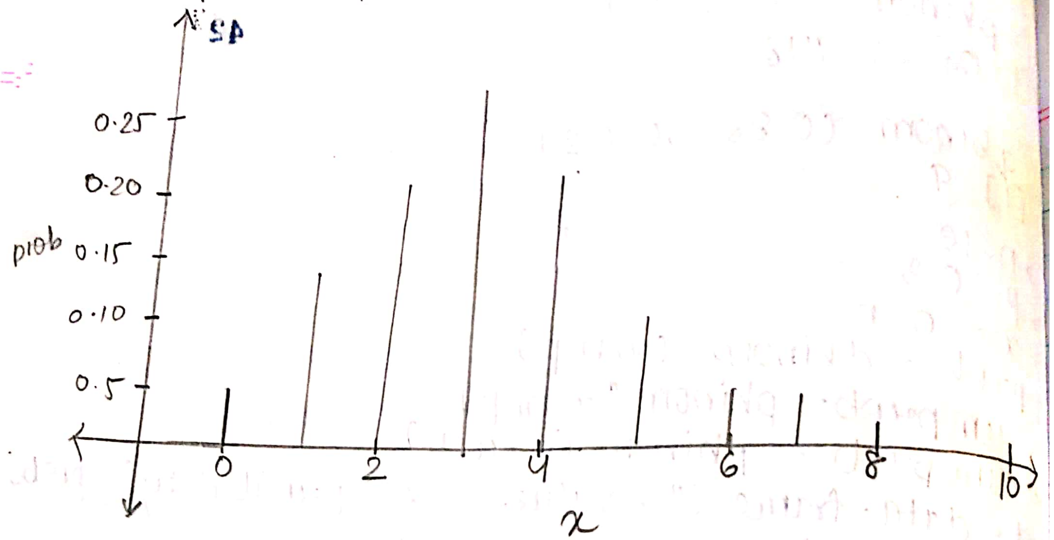
6)
```
qbinom (0.88, 30, 0.2)
[1] 9
```

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

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

Scanned with CamScanner

> plot (x, prob, "h")



> plot (x, cumprob, "s")

Practical: 04

Topic: Normal distribution

p (X = x) = d norm (x, μ, σ)
p (X ≤ x) = pnorm (x, μ, σ)
p (X > x) = 1 - pnorm (x, μ, σ)

To generate random no. from a normal distribution (in Random nos) The r code is r norm (n, μ, σ)

(1) A random variable X follows normal distribution with mean = μ = 12 and s.d = σ = 3. find.

i p (X ≤ 15)
ii p (10 < X ≤ 13)
iii p (X > 14)
iv generate 5 observation (Random nos).

Ans:- i) p1 = pnorm (15, 12, 3)

```
> p1
[1] 0.8413447
> cat ("p (X <= 15) =", p1)
> p (X <= 15) = 0.8413447
```

ii) p2 = pnorm (13, 12, 3) - pnorm (10, 12, 3)

```
> p2
[1] 0.3780661
> p (10 <= X <= 13) = 0.3780661
```

iii) p3 = 1- pnorm (14,12,3)
> p3
[1] ~~cat~~ 0.25 24925
> cat (" P(X>14) =", p3)
p(x>14) = 0.25 24925

> rnorm (5,12,3)
[1] 14.98663 12.51616 13.14904
14.98075 19.73518

② X follows normal distribution with
μ = 10 σ = 2 Find = ?

1 p (x ≤ 7)
2 p (5 < x < 12)
3 p (x > 12)
4 generate 10 random observation
5 Find k such that probability
p (X < K) = 0.4

Ans- i) p1 = pnorm (7,10,2)
p1
[1] 0.0668072
> cat ("p (X<=7) =", p1)
p (X <= 7) = 0.0668072

ii) p2 = pnorm (12,10,2) - pnorm (5,10,2)
p2
[1] 0.8351351
> cat ("p (15 < X <12) =", p2)
p (15 < X <12) = 0.8351351

Scanned with CamScanner

```
iii) p3 = 1 - pnorm (12,10,2)
p3
[1]  0.1586553
> cat ("p(p(X >12)= ", p3)
p(p(X >12) = 0.1586553
> rnorm (10,10,2)
[1]  11.986324   12.155504   7.130492
     10.519359   5.637988    10.959963
     7.988735    9.336612    8.115370
     13.052133
> qnorm  (0.4, 10, 2)
   [1]  9.493306.
```

3) generate 5 random no. from a normal distribution with mean = 15 and sd = 4 find sample mean, median, sd and (print (cat)).

4). X ~ N (30,100) (~~6² - 6~~ = 10) find only. p_1, p_2 (~~no~~ cat command).

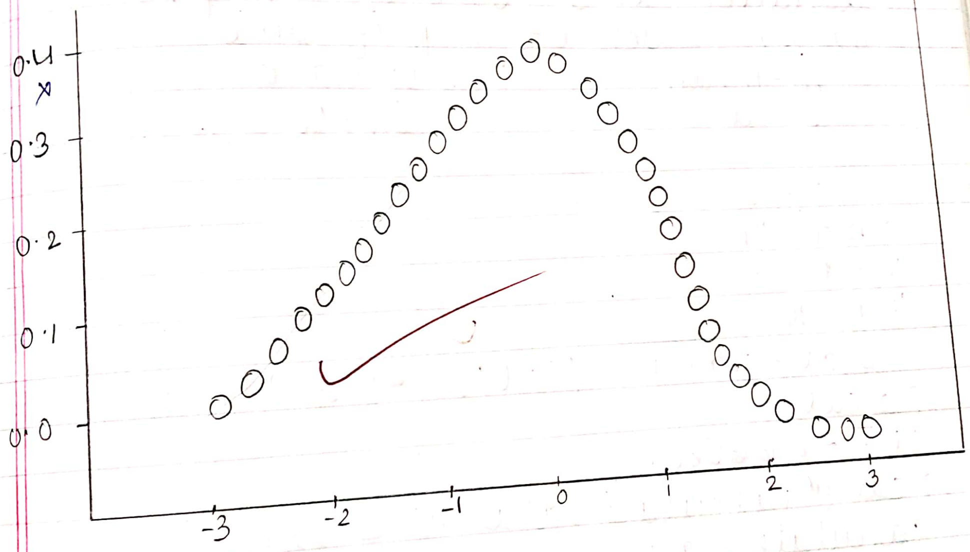
```
ans->  >p1 = pnorm (40,30,10)
       > p1
       [1]  0.8413447
       > p2 = 1 - pnorm (35, 30, 10)
       > p2
       [1]   0.3085375
> p3 = pnorm (35,30,10) - pnorm (25,30,10)
> p3
  [1]   0.3829249
> qnorm (0.6, 30,10)
  [1]  32.53347
```

Ans→ > x = rnorm (5, 15, 4)

```
> x
[1] 16.43602   12.58878   17.37383
    15.04210   16.91096
> am = mean (x)
> am
[1] 15.67034
> med = median (x)
> me
[1] 16.43602
> variance = (n-1) * var (x) / n
> variance
[1] 2.9836
> sd = sqrt (variance)
> sd
[1] 1.72731
> cat (" sample mean is = ", am)
→ sample mean is = 15.67034
> cat ("sample median is = ", me)
→ sample median is = 16.43602
> cat ("sample sd is = ", sd)
Sample sd is = 1.72731
```

5 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")
```



STANDARD NORMAL GRAPH

Practical - 05

Aim - Normal and T-test

Q1 Test the hypothesis $H_0 : \mu = 15$, $H_1 : \mu \neq 15$. Random sample of size 400 is drawn and it is calculated the sample mean is 14. And the standard deviation is 3. Test the hypothesis at 5% level of significance.

m - mean   0 - population
s - sd   n - sample size

Ans-

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

Since p value is less than 0.05 we reject $H_0 . \mu = 15$.

Q2) Test the hypothesis $H_0: \mu = 10$, $\mu \neq 10$. Random sample of size 400 is drawn with sample mean 10.2 and standard deviation 2.25. Test the hypothesis at 5% level of significance.

46

```
> m0 = 10
> mx = 10.2
> sd = 2.25
> n = 400
> zcal = (mx-m0)/(sd/(sqrt(n)))
> zcal
[1] 1.777778
> cat(" calculate value at z is=", zcal
calculate value at z is = 1777778
> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 0.07544036
```

Q3) Test the hypothesis Ho proportion of smokers in a college is 0.2 a sample is collected and sample proportion is calculated as 0.125. Test the hypothesis at 5% level of significance (sample size = 400)

P - population

p - sample

```
> P = 0.2
> p = 0.125
> n = 400
> Q = 1-P
```

Scanned with CamScanner

```
> zcal = (p-P)/(sqrt (P * Q/n))
> z cal
[1]  -3.75
```

Q4) Last year farmers lost 20% of their crops a random sample of 60 fields are collected. It will found that a field crops are insect populated. Test the hypothesis at 1% level of significance

P - 0.2 (capital)

p - 9/60

→

```
> P = 0.2
> p = 9/60
> n = 66
> z cal = (p-P) / (sqrt (P * Q/n))
> z cal
[1]  -0.9682458
> pvalue = 2 * (1 - pnorm(abs (zcal)))
> p value
[1] 0.3329216
```

∴ The value is 0.1 so value is accepted.

Scanned with CamScanner

Q5) 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,
         11.94, 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] 0.019521
> sd = sqrt (variance)
> sd
[1] 0.1397176
> m0 = 12.5
> t = (mx - m0) /(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 - 06

Aim: Large sample Test

Q1 Let the population mean (the amount spent by customer in a restaurant is 250. A sample of 100 customers selected. Sample mean is calculated as 275 and sd as 30. Test the hypothesis that population mean is 250 or not at 5% level of significance.

→ $H_0 : \mu = 275$
against $H_1 : \mu \neq 275$

```
> m0 = 250
> mx = 275
> n = 100
> sd = 30
> zcal = (mx-m0)/(sd/sqrt(n)))
> zcal
> [1] 8.333333
> pvalue = 2 * (1 - pnorm(abs(zcal)))
> pvalue
  [1] 0

> cat(" Z cal:", zcal)
> Zcal: 8.333333
> cat("pvalue:", pvalue)
```

> pvalue : 0

∴ p value = 0 < 0.05 we reject Ho at 5% level of significance.

Q2) In a random sample of 1000 students it is found that 750 use blue pen. Test the hypothesis that population proposition is 0.8 at 1% level of significance

→ Ho : P = 0.8 against H1 : P ≠ 0.8

```
> P = 0.8
> Q = 1-P
> p = 750/1000
> n = 1000
> zcal = (p-P)/(sqrt(P*q/n)
> zcal
[1] -3.952847
> pvalue = 2* (1- pnorm(abs(zcal)))
> pvalue
  [1] 7.72268e-05
```

we reject !

∴ The value is less 0.01

Q3) Two random sample of size 1000 and 2000 are drawn from two population with the same sd = 2.5. The sample means are 67.5 and 68 respectively. Test the hypothesis $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$ at 5% level of significance.

→ $H_0: \mu_1 = \mu_2$ as $H_1: \mu_1 \neq \mu_2$

```
> 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]  -5.163978
> zcal("zcal:",zcal)
> zcal: -5.163978
> pvalue = 2 * (1-pnorm(abs(zcal)))
> pvalue
[1]  2.417564e-07
> cat ("pvalue:", pvalue)
> pvalue :2.417564e-07
```

∴ Rejected

The study of noise level in two hospitals is given below. Test the claim that the two hospital have same level of noise at one (1%) level of significance.

	hospital A	hospital B
size	84	34
mean	61.2	59.4
sd	7.9	7.5

```
>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/n)
>zcal
[1] 1.162528
>cat("zcal :", zcal)
  zcal : 1.162528
>pvalue = 2 * (1-pnorm (abs(zcal)))
> pvalue
  [1] 1.753222 e-10
```

∴ value is less than 0.01 the value is rejected.

Q5 In a sample of 600 student in a college 400 use blue ink in another college from a sample of 900 students 450 use blue ink. Test the hypothesis that the proportion of students using blue ink in two colleges are equal or not. At 1% level of significance.

```
> n1 = 600
> n2 = 900
> P1 = 400/600
> P2 = 450/900
> Zcal = (p1-p2)/sqrt (p * q * (1/n1 + 1/n2))
> Zcal
[1]  6.381534
> pvalue = 2 * (1-pnorm (abs(zcal)))
> pvalue

[1] 1.753222e-10
```

∴ value is less than 0.01 the value is rejected.

Practical - 07

Topic: small sample test

1) 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.

→ $H_0 : \mu = 66$

> 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 = 1.558 e^{-13}

alternative hypothesis

True 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 that at 5% level of significance

3

2) Two groups of students 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

Ho : There is not difference b/w the 2 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    pvalue = 0.03798
alternative hypothesis :
    True difference in means is not equal
to 0. 95% confindence 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 Ho")}
  else {cat ("reject Ho")}
  reject Ho.
```

3 The sales data of 6 shops before and 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 effective or not.

H0: There is no significance difference of sales before and after campaign

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

paired t-test
data: x and y
t = -2.7815, df = 5, pvalue = 0.9806
alternative hypothesis:
True difference in means is greater than 95 percent confidence interval:
-6.035547 inf
sample estimates
mean of the difference
-3.5

∵ pvalue is greater than 0.05, we accept the hypothesis at 5% level of significance

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, 115, 99

→ H0 : There is no significance difference

\> x = c (before)

\> y = c (after)

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

paired t.test

data : x and y

t = 4.3458, df = 5, pvalue = 0.9963

alternative hypothesis : true difference in means is less than 0.

95 % confindence interval :

inf 29.0295

Sample estimate :

mean of the differences

19.83333

∵ pvalue is greater than 0.05 we accept the hypothesis at 5% level of significance.

5) 2 medicines are applied to two groups of patient respectively.

grp1 : 10, 12, 13, 11, 14

grp2 : 8, 9, 12, 14, 15, 10, 9

Is there any significance difference between 2 medicines.

H0 : There is no significance difference

> x = c (grp1)

> y = c (grp2)

> 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 :

mean of x mean of y

12.0000 10.33333

∵ pvalue is greater than 0.05 we accept the hypothesis at 5% level of significance

Practical - 08

Topic: large and small test.

Q1) $H_0 : \mu = 55$, $H_1 : \mu \neq 55$

```
> 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 H_0 at 5% level of significance.

Q2) $H_0 : P = 0.5$ against $H_0 : P \neq 0.5$

```
> P = 0.5
> q = 1-P
> n = 100
> 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 1% level of significance

Q3 Ho : $P_1 = P_2$ against Ho : $P_1 \neq P_2$

```
>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
```

∵ pvalue is greater than 0.05 we accept Ho and 5% level of significance

Q4 Ho : $\mu = 100$ ~~against~~ $H_1 : \mu \neq 100$

```
>var = 64
> n = 400
> mo = 100
> mx = 900
>sd = sqrt (var)
> sd
```

88

```
>zcal = (m1-m0)/(sd/(sqrt(n)))
> zcal
 [1] 2.5
> pvalue = 2*(1-pnorm(abs(zcal)))
>  pvalue
[1]  0.01241933
```

∵ pvalue is less than 0.05 we reject Ho at 5% level of significance

Q5 Ho : $\mu = 66$ against $H_1 : \mu \neq 66$

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

One sample test

data: x
t = 47.94
df = 6
p-value = 5.522 e-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
∵ pvalue is le...

Ho at 1% level of significance.

Q6) $H_0 : \sigma_1 = \sigma_2$ 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 and y

f = 0.788803

num df = 7,

dnum df = 10

pvalue = 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

∵ pvalue is greater than 0.05 we accept Ho at 5% level of significance.

Q7) $H_0 : \mu = 1150$ against $H: \mu \neq 1150$

```
> n = 100
> mx = 1150
> mo = 1200
> sd = 125
> zcal = (mx-mo) /(sd/(sqrt(n)))
> zcal
[1] -4
> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue
[1] 6.334248 e-05
```

∴ pvalue is less than 0.05 we reject

Q8) $H_0 : P_1 = P_2$ 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
```

Scanned with CamScanner

```
> pvalue = 2*(1-pnorm(abs(zcal)))
> pvalue
[1]  0.3613104
```

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

Practical - 09

Topic - Non parametric testing of hypothesis using R-cout

1) The following data represent earnings (in dollars) for a random sample of five common stocks listed on the new york stock exchange test whether median earnings is 4 dollars.

data : 1.68, 3.35, 2.50, 6.23, 3.24

```
> x <- c(1.68, 3.35, 2.50, 6.23, 3.24);
> n <-length(x);
> n
> [1]5
> x >4;
[1] FALSE FALSE FALSE TRUE FALSE
> s <-sum(x >4); s;
[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.value = 0.9688

alternative hypothesis: true probability of success is greater than 0.5

95 % confidence interval:

0.01020622 1.00000000

sample estimates:

probability of success 0.2

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	9	7	11	12
score after :	13	16	15	13	9	10	13	10

CODE :

```
> b <- c (10,15 , 16, 12 , 12, 9 , 7 , 11, 12);
> a <- c (13, 16, 15, 13, 8 , 10, 13, 10);
> D <- b - a ;
> wilcox.test (D, alternative = "greater");
```

Wilcoxon signed rank list with continuity correction data : D
V = 10.5 P.value = 0.8722
alternative hypothesis : true location is greater than 0.
warning message :
In wilcox.test.default (D, alternative = "greater");
cannot compute exact p-value with ties.
∴ p-value is greater than 0.05 we accept it.

3) The diameter of a ball bearing was measured by 6 inspectors each using two different kinds of calipers. the results were, test whether average ball bearing for.

Inspector :	1	2	3	4	5	6
Caliper 1 :	0.265	0.268	0.266	0.267	0.269	0.264
Caliper 2 :	0.263	0.262	0.270	0.261	0.271	0.260

Caliper 1 and Caliper 2 are same.

CODE :

```
>x <- c(0.265, 0.268, 0.266, 0.267 ,0.269 ,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 = 0.197
```

alternative hypothesis : true location shift is greater than 0.

∴ p value is greater than 0.05 we accept it.

11) An officer has 3 electric typewriters A, B and C. In a study of machine usage, firm has kept records of machine usage rate of seven weeks. Machine A was out of repair for two weeks. It is of interest to find out which machine has better usage rate. Analyse the following data on usage rates and determine if there is a significant difference in avg rate.

A	B	C
12.3	15.7	32.4
15.4	10.8	41.2
10.3	45.0	35.1
08.0	12.3	25.0
14.6	8.2	8.2
–	20.1	18.4
–	26.3	32.5

```
>x<-c(12.3,15.4,10.3,8.0,14.6)
>n1<-length(x);
>n1
 [1] 5
>y<-c(15.7,10.8,45.0,12.3,8.2,20.1,26.3);
>n2<-length(y);
>n2
[1] 7
```

Q1

```
>z <-c (32.4, 41.2, 35.1, 25.0, 8.2, 18.4, 32.
>n3 <-length (z);
> n3
[1] 7
> x <-c(x,y,z);
> g <-c(rep(1,n1), rep(2,n2), rep (3,n3));
> kruskal.test (x, g).
```

Kruskal - wallis rank sum test

data : x and g

Kruskal-Wallis chi-squared = 5.217, df = 2

p value = 0.0736

∴ p-value is greater than 0.05 we accept it.

Practical-10

Aim : Chi square test and ANOVA (Analysis of Variance).

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

cond child	cond home clean	dirty
clean	70	50
fairly clean	80	20
dirty	35	45

Ho: 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
```

88

Pearson's chi-squared test

data : y
x-squared = 25.646
df = 2
p-value = 2.698e-06
They are dependent
∵ p value is less than 0.05 we reject the hypothesis at 5% level of significance.

Q2 Test the hypothesis that vacation and disease are independent or not.

vaccine

disease	affected	not affected
affect	70	46
non-affected	35	37

Ho : disease and vaccine 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

>pv = chisq.test (y)
>pv

Pearsons chi-squared test with yates continuity correction data : y

x-square = 2.0275
df = 1
p-value = 0.1545

∴ p-value is more than 0.05 we accept the hypothesis at 5% level of significance

// They are INDEPENDENT.

Q3) Perform a ANOVA for the following data :

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

H0: the means are equal for A, B, C, D.

Q4

```
> 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
```

one way-analysis of means.

data: values and ind

F = 11.735

df = 3

denom df = 9,

p value = 0.00183

p value is less than 0.05 we reject the hypothesis

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

	Df	sum	mean sq	F value	pr(>F)
ind	3	71.06	23.688	11.73	0.00
Residuals	9	18.17	2.019		

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

Scanned with CamScanner

Q4 Following data gives a life of tires of 4 brands.

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

H0 : the means 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, 14, 22, 17, 20)
> x4 = c(15, 14, 16, 18, 14, 16)
> d = stack(list(b1 = x1, b2 = x2, b3 = x3, b4 = x4))
> name(d)
[1] "values"  "ind"
> oneway.test(values ~ ind, data = d,
      var.equal = T)
```

One way analysis of means.

data: value and ind.

F = 6.8445

num df = 3

denom df = 20

p-value = 0.002349

∵ p-value is less than 0.05 we reject the hypothesis

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

            df     sumsq    meansq    Fvalue   Pr(
ind         3      9444     30.479    6.845    6.0
residuals   20     89.06    4.453

signif code: 0 '***' 0.001 '**' 0.01 '*'
             0.05 '.' 0 ' ' 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        57
8    59        58
9    20        25
10   27        27
```

d

```
> am = mean (x $ stats)
> am
[1] 37
> am1 = mean (x $ maths)
> am1
[1] 39.4
> m1 = median (x $ stats)
> m1
[1] 38.5
> m2 = median (x $ maths)
> m2
[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
```

Scanned with CamScanner