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

```
> 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] 6.67156

5] round off
$46 \div 7 + 9 \times 8$
> round(46/7+9*8)
[1] 79

Q2 > c(2, 3, 5, 7)*2
[1] 4 6 10 14

> c (2, 3, 57) + c(2,3)
[1] 4 9 10 21

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

Q3 > x=20 > y=30 > z=2
> x^2 + y^3 + z
[1] sqrt(x^2+y)
20.73644

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

Q4
> x <- matrix(nrow=4, ncol=2, data=c(1,3,3,4,5,6,
7,8))

[1] x	[,1]	[,2]
[1] [1,]	2	5
[2,]	2	6
[3,]	3	7
[4,]	4	8

Q5 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))
```

Q

Practical 2

Topic:
Probability Distribution

Q1 Check whether the following 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

Since, $p(2) = -0.5$ it cannot be a p.m.f as in p.m.f $p(x) \geq 0 \ \forall x$

2)

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

The cond^n for p.m.f 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 cond^n for pmf is
1) $P(x) \geq 0 \quad \forall x$ satisfy
2) $\Sigma P(x) = 1$

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

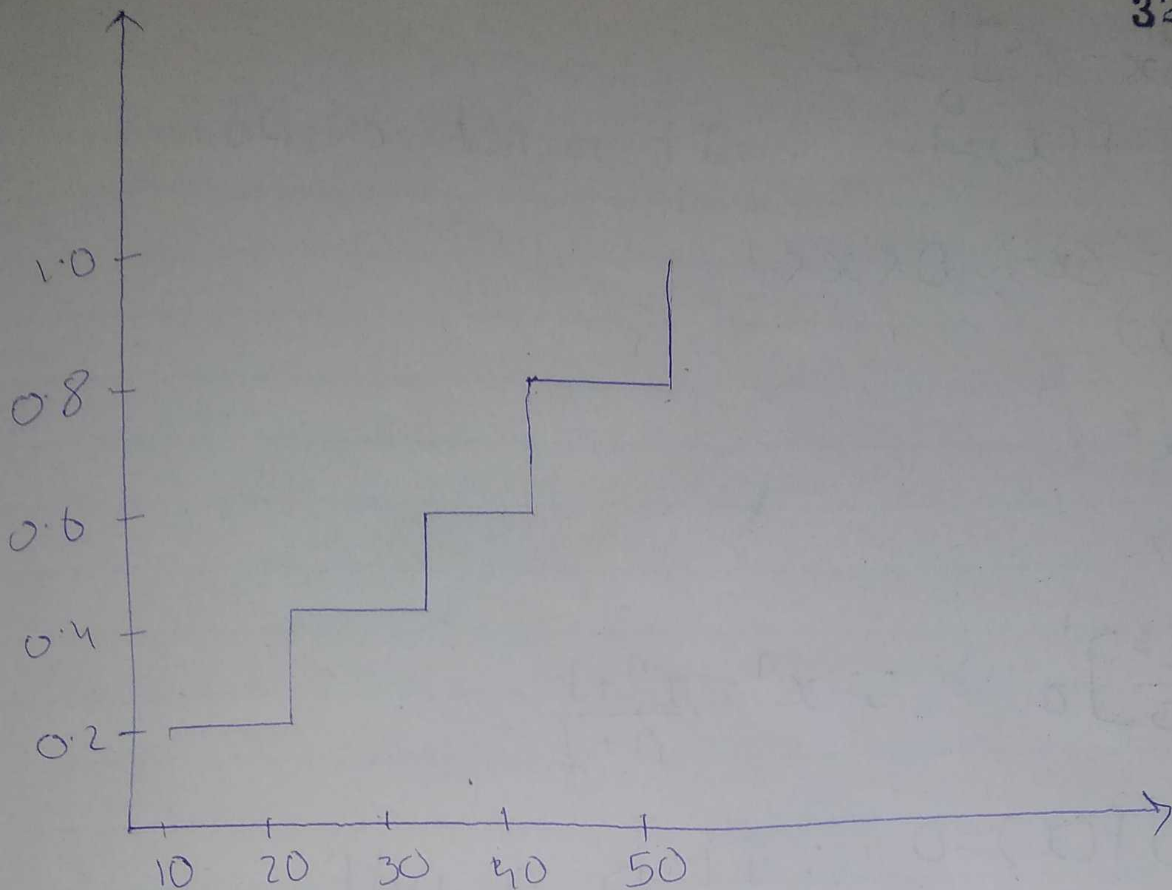
The given data is pmf.

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

Q2 Find the cdf for the following pmf and sketch the graph

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

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



Q3 Check that whether the following is p.d.f or not

i) $f(x) = 3 \cdot 2x \;;\; 0 \leq x \leq 1$

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

i) $f(x) = 3 \cdot 2x$

$= \int_0^1 f(x)$

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

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

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

$\because$ The $f(x) = 1$ $\therefore$ It is not a pdf

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

$\int_0^1 f(x)$

$= \int_0^1 3x^2$

$= 3\int_0^1 x^2$

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

The $\int_0^1 f(x) = 0$ $\therefore$ It is a pdf.

38

Practical -3

TOPIC : Binomial distribution

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

1\. Find the probability of exactly 10 success in hundred trials with $p = 0.1$.

2\) Suppose there are 12 mcq, Each question has 5 option out of which 1 is correct find the probability of having exactly 4 correct answers
ii) atmost 4 correct answer.
iii) More than 5 correct answer

3\) Find the complete distribution when $n = 5$ & $p = 0.1$.

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

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

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

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

Answers:

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] 0.4274445.
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 - 0.0045

5 - 0.00001

4) 1) ~~bb~~ 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.

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)

cum prob = pbinom (x, n, p)

```
d = data.frame("xvalues"= x, "probibility"=
                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	0.0014
10	9	0.0001
11	10	0.0000.

Practical - 4

Aim : Normal Distribution.

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

Q.1] A random variable x follows normal distribution with Mean $= \mu = 12$ & S.D $= \sigma = 3$. find i. $P(x \le 15)$ ii $P(10 \le x \le 13)$ iii. $P(x > 14)$ iv. generate 5 observation (random numbers)

CODE.

```
> P1 = pnorm (15, 12, 3)
> P1
[1] 0.8413447
> cat ("P(X <= 15) = ", P1)
P(X<=15) = 0.8413447
> P2 = pnorm (13, 12, 3) - pnorm (10, 12, 3)
> P2
[1] 0.3780661
> cat ("P(10<=x<=13) = ", P2)
P (10<=X<=13) = 0.3780661
> P3 = 1 - pnorm (14, 12, 3)
> P3
[1] 0.2524925
```

```
>cat("P(x>14)="P3)
P(x>14) = 0.2524925
>P4 = rnorm(5,12,3)
>P4
[1] 15.254723   16.548505   11.280515   6.419944
                                         12.272400
```

2] X follows normal distribution with $\mu = 10, \sigma = 2$
find i P ($x \leq 7$) ii) P($5 < x < 12$) iii) P ($x > 12$)
iv) generate 10 observation v. find k such that
P [$x < k$] = 0.4

CODE:

```
> a1 = Pnorm (7,10,2)
> a1
[1] 0.668072
> a2 = Pnorm (5,10,2) - Pnorm (12,10,2)
> a2
[1] 0.835135.
> a3 = 1 - Pnorm (12,10,2)
> a3
[1] 0.1586553
> a4 = rnorm (10,10,2)
> a4
[1] 11.608931    9.120417    12.637741   8.073354
                 9.193226    9.366824    11.707106
    8.721380
                 10.715000
    9.537584
> a5 = qnorm (0.4,10,2)
> a5
[1] 9.493306.
```

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

CODE:

```
> rnorm (5, 15, 4)
[1] 10.7649   7.793249   9.953444  13.345904
    17.509668
> am <- mean (x)
> 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] 3.33163
> cat ("SD is = ", SD)
SD is = 3.331613.
```

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

i) $P(X \leq 40)$

ii) $P(X > 35)$

iii) $P(25 < X < 35)$

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

```
> f1 = pnorm (40, 30, 10)
> f1
[1] 0.8413447
> f2=1 - pnorm (35, 30, 10)
> f2
[1] 0.3085375
> f3 = pnorm (25, 30, 10) - pnorm (35, 30, 10)
> f3
[1] 0.3085375
> f4 = qnorm (0.6, 30, 10)
> f4
[1] 32.53347
```

Q.5) Plot the standard normal graph.

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

Practical 5.

Topic :- Normal & t-test.

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

Test the hypothesis.

Random sample of size 400 is drawn & it is calculated. the sample mean is 14 & S.D is 13 test the hypothesis at 5 % level of significance.

0.05 > accept the value

0.05 < less than Reject.

```
>mo=15
>mx=14
>n=400
>sd=3
>zcal = (mx-mo)/(sd/(sqrt(n)))
>zcal
[1] -6.666667
>cat ("Calculated value of z is =", zcal)
calculated value of z is = -6.666667
> p value = 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 Ho = μ = 10 against H_1 : μ≠10
A random sample size of 400 is drawn. with sample mean = 10.2 & SD = 2.25
Test the hypothesis at

```
> mo = 10
> n = 400.
> mn = 10.2
> sd = 2.25
> zcal = (mn - mo) / (sd / (sqrt(n)))
> zcal
[1] 1.77778
> Pvalue = 2 * (1 - Pnorm (abs (zcal)))
> Pvalue
[1] 0.07544036.
```

∵ The Pvalue is greater than 0.05
∴ the value is accepted.

3) Test the hypothesis Ho: Proportional of smokers in college is 0.2 A sample is collected & calculated. the sample proportion 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)

4) Last year farmer's lost 20 % of their crops. A random sample of 60 fields are collected & it found that a fields crops are insect polluted. Test the hypothesis of 1 % level of significance.

> P = 0.2
> p = 9/60
> n = 60
> zcal = (p - P) / (sqrt (P * Q / n))
> zcal
[1] -0.9682458
> Pvalue = 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 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] 12.167.
> 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 - 6

AIM:- Large Sample test

1) Let the population mean (the amount spent per customer in a restaurant) is 250. A sample of 100 customers selected the sample mean is calculated on 275 and ~~so~~ SD 30. Test the hypothesis that the population mean is 250 or not on 5% level of significance.

2- In a random sample of 1000 students it is found that 750 use blue pen test the hypo-the hypothesis that the population proportion in 0.8 at 1% level of significance.

1 Solution:

```
>m0 = 250
>mx = 275
>sd = 30
>n = 100
>zcal = (mx-m0)/(sd)/(sqrt(n)))
>cat ("calculated value of z is = " zcal)
[1] Calculated value of z is = " zcal) 8.3333
>pvalue = 2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0
```

∴ The value is less than 0.05 we will reject the value of $H_0 = \mu = 250$.

2) Sol^n

```
> p=0.8
> Q=1-p
> p=750/1000
> n=1000
> zcal (p-p) /(sqrt(p*Q /n))
> cat("Calculated value of z is:", zcal)
[1] Calculated value of z is:- -3.952847
> pvalue =2 * (1-pnorm(abs(zcal)))
> pvalue
[1] 7.72268e-05
```

3) To random sample of size 1000 & 2000 are draw from two population with same sd 2.5 the sample means are 67.5 and 68 test the hypothesis $\mu_1=\mu_2$ at 5% < of significance

Sol^n

```
> 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
```

```
>pvalue=2*(1-pnorm(abs(zcal)))
>pvalue
[1] 2.417564e-07 ∴ (Rejected)
```

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

HOS.A	HOS.B
84	34
61.2	59.4
7.9	7.5

Sol^n

```
>n1=84
>n2=34
>mx1=61.2
>mx2=59.4
>sd1=7.9
>sd2=7.5
>zcal=(mx1-mx2)/sqrt((sd1^2/n1)+(sd2^2))
>zcal
[1] 1.162528
>pvalue=2*(1-pnorm(abs(zcal)))
>pvalue
[1] 0.245021
```

∴ The value is greater than 0.01 we accept the value

Practical-7

TOPIC :- Small Sample Test.

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 = 1.558e-13
alternative hypothesis
True mean ~~is not~~ equal to 0 95%
confidence interval 65.65171 to 14829
Sample estimates
mean of x
67.4

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

2) Two groups of students scored the following masks test the hypothesis that there is no significant diffe. b/w the 2 grps

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 no diff 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)
```

Watch two sample t-test

data: x & y

t = 2.2573 df = 16.376 p-value = 0.0378
alternative hypothesis.
True different in means is not equal two 95 percent confidence interval 0.1628205 5.0371795
sample estimates:
mean of x mean of y
20.1 17.5

```
> pvalues = 0.03798
> if pvalue > 0.05)
```

Practical-8

Topic :- Large & small test

1) Ho : $\mu = 55$, $H_1 = \mu \neq 55$

```
> n = 100
> mx = 52
> mo = 5
> sd = 7
> zcal = (mx - mo) / (sd / (sqrt (n)))
[1] -4.285714
> pvalue = 2 * (1 - pnorm (abs (zcal)))
> pvalue
[1] 1.82153e-05
```

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

2) Ho = P = 0.5 against Ho : P ≠ 0.5

```
> P = 0.5
> q = 1-p
> n = 760
> zcal = (p-p) / (sqt (p * q /n))
> zcal
[1] 0
> pvalue = 2 * (1 - pnorm (abs (zcal)))
> pvalue
[1] 1
```

74

As value is greater than 0.05 we accept Ho at 1% level of significance

```
> 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
```

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

4) Ho: u = 100 against H1: u ≠ 100

```
> var = 64
> n = 400
> mo = 100
> mx = 99
> sd = sqrt(var)
> sd
```

[1] 8
> zacl = (mx-mo) /(sd) /(sqrt(n)))
> zcal
> zcal = (mx-mo) / (sd)/ sqrt (n)))
> zcal

[1] 2.5

Practical-9

Aim:- ~~Sq~~ Chi-square test & ANOVA

Q1) Using the following data test whether the $cond^n$ of home & the $cond^n$ of child are idependen or not

$Cond^n$ Child	$Cond^n$ Home	
Clean	Clean	Dirty
family	70	50
Clean	80	20
Dirty	35	45

Ho: $cond^n$ of home & chid are independent

```
> x=c(70, 80, 35, 50, 45)
> m=3
> n=2
> y=matrix(x, nrow=n
> y
> pvalue= chi sqrt
> pvalue
  persons chi-square test data: y
```

4

x. squared = 25.040

df = 2

p-value = 2.048

They are dependent.