PRACTICAL NO. 1

Aim:- Basics of R software.

(1) R is a software for statistical analysis and data computing.
(2) It is an effective data handling software and outcome storage is possible.
(3) It is capable of graphical display.
(4) 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 $5^3+7\times5\times8+46\div5$

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

④ $\sqrt{4^2+5\times3+7\div6}$

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

⑤ round off 46 ÷ 7 + 9 * 8

> round (46/7 + 9 * 8)

[1] 79

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

[1] 4 6 10 14

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

[1] 4 9 10 21

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

[1] 4 36 512 9 16 125

Q.3) > 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

Q.4] >x <- matrix [nrow=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
```

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]     [,1] [,2] [,3]
  [1,]    4   -2    6
  [2,]    7    0    7
  [3,]    9   -5    3
>y <- matrix [nrow=3, ncol=3, data=c(10,12,15,-5,-4,-6,7,9,5)]
> y
[1]     [,1] [,2] [,3]
  [1,]   10   -5    7
  [2,]   12   -4    9
  [3,]   15   -6    5
>x+y
[1]     [,1] [,2] [,3]
  [1,]   14   -7   13
  [2,]   19   -4   16
  [3,]   24   -11   8
```

```
>2*x+3*y
[1]     [,1] [,2] [,3]
[1,]    38   -19   35
[2,]    50   -12   41
[3,]    63   -28   21
```

Q.6] Marks of statistics of CS Batch B

```
x=c(58,20,35,24,46,56,55,45,27,22,47,58,54,40,50,32,36,
    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 NO. 2

TOPIC : Probability Distribution.

Q.1] Check whether the following are pmf or not.

(1)

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. then $\Sigma P(x) = 1$

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

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

$= 1.0$

Hence the given data is p.m.f.

(2)

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

The condition for p.m.f is $\Sigma P(x) = 1$.

So,

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

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

$= 1.1$

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

(3)

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

The condition for p.m.f. is

① $P(x) \geq 0$ $\forall x$ satisfy

② $\Sigma P(x) = 1$

$\Sigma P(x) = 0.2 + 0.2 + 0.35 + 0.15 + 0.1$

$= 1$

∴ The given data is p.m.f.

Code:-

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

Q.2 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

$F(x) = 0 \quad x < 10$

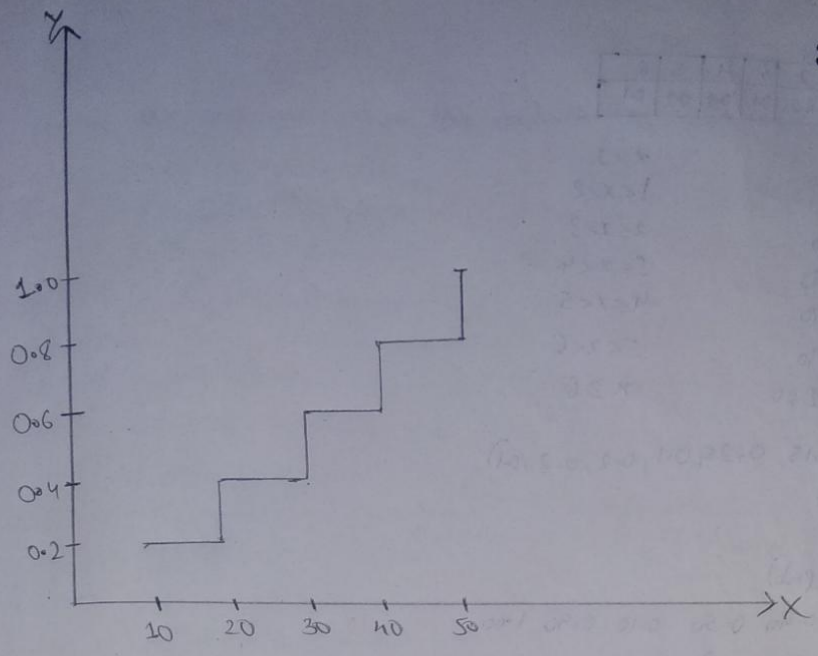
$= 0.2 \quad 10 \leq x < 20$

$= 0.4 \quad 20 \leq x < 30$

$= 0.75 \quad 30 \leq x < 40$

$= 0.90 \quad 40 \leq x < 50$

$= 1.0 \quad x \geq 50$



graph.

Q(2)

38

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

$F(x) = 0$ $x < 1$

$= 0.15$ $1 \le x < 2$

$= 0.40$ $2 \le x < 3$

$= 0.50$ $3 \le x < 4$

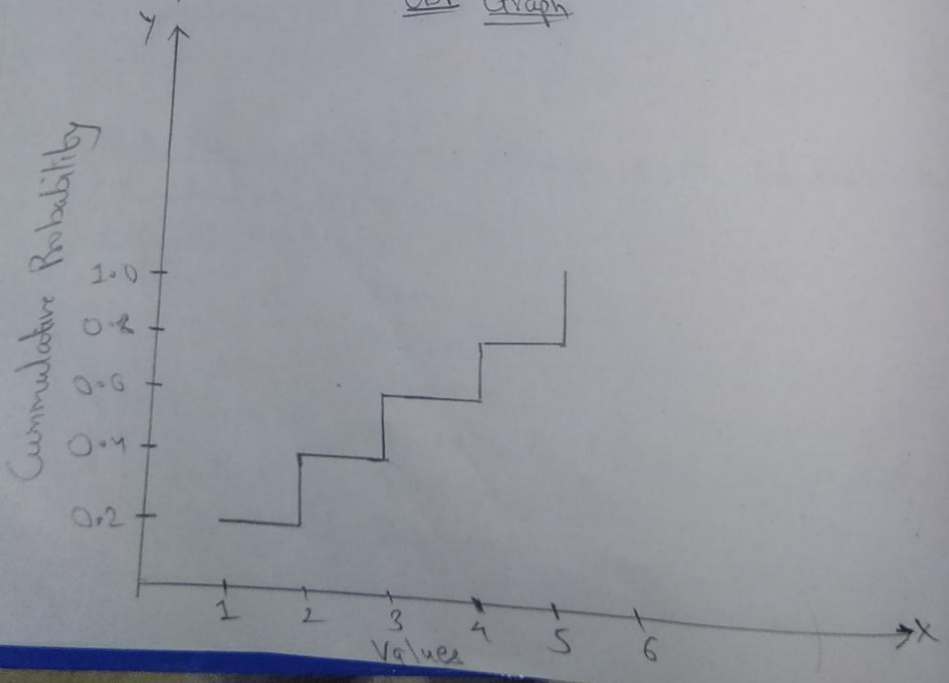
$= 0.70$ $4 \le x < 5$

$= 0.90$ $5 \le x < 6$

$= 1.00$ $x \ge 6$

```
>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", col="brown")
```

CDF Graph



Q.3] Check ~~that~~ whether the following is p.d.f. or not.

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

$f(x) = 3-2x$

$= \int_0^1 f(x)$

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

$= \int_0^1 3\,dx$

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

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

$= 2$

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

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

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

$\int_0^1 f(x)$

$= \int_0^1 3x^2$

$= 3\int_0^1 x^2$

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

$= x^3$

$= 1$

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

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

(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] 0.0940528

(3) > dbinom(0:5, 5, 0.1)
0 - 0.59049
1 - 0.32805
2 - 0.0729
3 - 0.00810
4 - 0.00045
5 - 0.00001

(4)
(i) > dbinom(5, 12, 0.25)
[1] 0.1032414
(ii) > pbinom(5, 12, 0.25)
[1] 0.9455978
(iii) > 1 - pbinom(7, 12, 0.25)
[1] 0.0029815
(iv) > dbinom(6, 12, 0.25)
[1] 0.0401494

PRACTICAL NO. 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

  $p1 = P(X \leq x)$ = 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 one is correct. Find the probability of having (i) exactly 4 correct answers.
(ii) Almost 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 $P(X = 5)$ ③ $P(X > 7)$
② $P(X \leq 5)$ ④ $P(5 < X < 7)$

(1) P.8

(5) The probability of a salesman making a sale to customer is 0.15. Find the probability of
① No sales out of 10 customer
② More than 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% of probability.

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

(5) >dbinom (0, 10, 0.15)
[1] 0.1968744
>1-pbinom (3, 20, 045)
[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, "probability" = prob)
>print (d)

	x values	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 No. 4

TOPIC: Normal Distribution.

- $P(X=x)$ = dnorm (x, μ, σ)
- $P(X<x)$ = pnorm (x, μ, σ)
- $P(X>x) = 1 - \text{pnorm}(x, \mu, \sigma)$
- To generate random numbers from a normal distribution (in 9 random numbers) the R code is rnorm (n, μ, σ).

Q.1 A random variable X follows normal distribution with mean = $\mu = 12$ and S.D = $\sigma = 3$. Find
(i) $P(x \leq 15)$ (ii) $P(10 \leq x \leq 13)$. (iii) $P(x > 14)$
(iv) Generate 5 observations (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
> 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  12.272460
```

Q.3] X follows normal distribution with $\mu=10$, $\sigma=2$.
Find ① $P(X\leq 7)$ ② $P(5<x<12)$. ③ $P(X>12)$
④ Generate 10 observations. ⑤ 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] -0.8351351
>a3 =1 - pnorm(12,10,2)
>a3
[1] 0.1586553
>a4=rnorm(10,10,2)
>a4
[1] 11.608931  9.920417  12.837741  8.073354
    8.721380   9.193726  9.386824   11.707106
    9.537584   10.715006
>a5=qnorm(0.4,10,2)
>a5
[1] 9.493306
```

Q.3] Generate 5 random numbers from a normal distribution. $\mu = 15$, $\sigma = 4$. find sample mean, median, standard deviation and print it.

CODE:-

```
>norm(5,15,4)
[1] 10.7649  7.793249  9.953444  13.345904
    17.509668
>an=mean(x)
>an
[1] 11.87345
>cat("Sample mean is=", an)
 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("Standard deviation is=", SD)
Standard deviation is= 3.33163
```

Q.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 $P(X < k) = 0.6$

CODE:-

```
> 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] -0.3829249
> f4 = qnorm(0.6,30,10)
> f4
[1] 32.53347
```

PRACTICAL NO. 5

Topic : 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 SD. is 3. Test the hypothesis at 5% level of significance
> 0.05 accept the value
< 0.05 reject the value.

CODE :-

```
>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
>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 SD = 2.25. Test the hypothesis at 5% los.

CODE:-

```
>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 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 and calculated the sample proportional as 0.125. Test the hypothesis at 5% los [sample size is 900].

CODE:-

```
>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] 0.0001768345 (Reject)
```

4] Last year farmer's lost 20% of their crops. A random sample of 60 fields are collected and it is found that a field crops are insect polluted. Test the hypothesis at 1% los.

CODE:-

```
>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 $H_0: \mu = 12.5$ from the following sample at 5% los.

CODE:-

```
>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
>mo=12.5
>t= (mx-mo)/(sd*sqrt(n))
>t
[1] -8.944909
>pvalue = 2*(1- pnorm(abs(t)))
>pvalues
[1] 0
```

∴ The value is less than 0.05, the value is accepted.