B.Sc.I.T and B.Sc.C.S. Department

Subject: R- Language (R-studio)
Subject Code: USCSP106

Semester-I

Class: F.Y.B.Sc.C.S.

Name of Professor: Asst. Prof. Kshitij Lalan Jha

INDEX

Sr. NO	Practical	Page No.	Date	Remark
1	FUNCTION IN R-LANGUAGE		7-08-2019	
2	A: FREQUENCY DISTRIBUTION B: GRAPHICAL REPRESENTATION		21-08-2019	
3	MEASURE OF CENTRAL TENDENCY AND DISPERSION		28-08-2019	
4	MEASURE OF CENTRAL TENDENCY AND DISPERSION		4-09-2019	
5	CORRELATION AND REGRESSION ANALYSIS		18-09-2019	
6	PROBABILITY		25-09-2019	

PRACTICAL 01 FUNCTION IN R-LANGUAGE

```
> rm()                    #Remove an object
> ls()                    #List of all existing objects
> x=c(2,3,4,1,8)                  # to set a vector named x containing 5 nos. 2,3,4,1,8.
> sort(x)                 # sort a vectors in ascending and descending order
[1] 1 2 3 4 8
> sort(x,decreasing=T)
[1] 8 4 3 2 1
> x=c(2,3,4,1,8)
> rank(x)                 #return the ranks of the values in a vector
[1] 2 3 4 1 5
> print(x)        #Enables to print an object using a different format than simply typing its
name
[1] 2 3 4 1 8
> cat("number of observation=",10)                    #Enables to print
number of observation= 10>
> length(x)                       #Returns the number of values in an R object
[1] 5
> mean(x)                         #Return the arithmetic average of a vector
[1] 3.6
> median(x)                       #Return the median of a vector
[1] 3
> max(x)                          #Return the maximum value of a vector
[1] 8
> min(x)                          #Return the minimum value of a vector
[1] 1
> range(x)                        #Returns the range of a vector
[1] 1 8
> var(x)                                  #Returns the variance of a vector
[1] 7.3
> sd(x)                           #Returns the standard deviation of a vector
[1] 2.701851
> summary(x)            gives min, Q1, median, mean, Q3 and descriptive statistics of an object
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0     2.0     3.0     3.6     4.0     8.0
> sqrt(x)                                 #Returns the square root of a values of a vector
[1] 1.414214 1.732051 2.000000 1.000000 2.828427
> log(x)                                          #Returns the natural logarithm of a values  of a
vector
[1] 0.6931472 1.0986123 1.3862944 0.0000000 2.0794415
> log10(x)                                #Returns the base 10 logarithm of a values vector
[1] 0.3010300 0.4771213 0.6020600 0.0000000 0.9030900
> exp(x)                                          #Returns the exponential of a values of a vector
[1]    7.389056   20.085537   54.598150    2.718282 2980.957987
> x=c(-2,-3,-4,-5,6,9)
> abs(x)
[1] 2 3 4 5 6 9
> choose(10,4)
[1] 210
```

```
> x=c(2,3,6,8,10)

> cumsum(x)                          #Returns the cumulative sum
[1]  2  5 11 19 29
> x=c(12,14,15,19,22,24)
> y=c(5,16,24,30,38,45)
> cor(x,y)                           #Return the correlation coefficient between two vectors
[1] 0.9746549
> x=c(1,2,4,8,6,3,4,1,4,2,1,3,2,1,2)
> table(x)                           #generate frequency table
x
1 2 3 4 6 8
4 4 2 3 1 1
> t=table(x)
> transform(t)                       #transform () function converts table column wise
  x Freq
1 1    4
2 2    4
3 3    2
4 4    3
5 6    1
6 8    1
```

PRACTICAL 02

A: Frequency Distribution

Q.1) In a survey of 40 families in a village, the number of children per family was recorded and the following data obtained.
1,0,3,2,1,5,2,2,1,0,3,4,2,1,6,3,2,1,5,3,3,2,4,2,2,3,0,2,1,4,5,3,3,4,4,1,2,4,5,5.
Represent the data in the form of discrete frequency distribution.
Solution: R code is

```
x=c(1,0,3,2,1,5,2,2,1,0,3,4,2,1,6,3,2,1,5,3,3,2,4,2,2,3,0,2,1,4,5,3,3,4,4,1,2,4,5,5)
> table(x)
X                                   #Output
 0  1  2  3  4  5  6
 3  7 10  8  6  5  1
> transform(table(x))
  x Freq                            #Output
1 0    3
2 1    7
3 2   10
4 3    8
5 4    6
6 5    5
7 6    1
```

Q.2) The weight in Kg of 50 college students are given below.

42,62,46,54,41,37,54,44,32,45,47,50,58,49,51,42,46,37,42,39,54,39,51,58,47,64,43,48,49,48,49, 61,41,40,58,49,59,57,57,34,56,38,45,52,46,40,63,41,51,41

Prepare Frequency table. Also obtain relative frequencies.

Solution: R Code is

```
w=c(42,62,46,54,41,37,54,44,32,45,47,50,58,49,51,42,46,37,42,39,54,39,51,58,47,64,43,48,49,4
8,49,61,41,40,58,49,59,57,57,34,56,38,45,52,46,40,63,41,51,41)
> range(w)
[1] 32 64                                   #Output
> breaks=seq(30,65,5)
> score=cut(w,breaks,right=FALSE)
> freq=table(score)
> transform(table(score))
    score Freq                              #Output
1 [30,35)    2
2 [35,40)    5
3 [40,45)   11
4 [45,50)   13
5 [50,55)    8
6 [55,60)    7
7 [60,65)    4
To add Realtive frequency column
```

```
> relafreq=freq/length(w)
> cbind(freq,relafreq)
      freq relafreq                          #Output
[30,35)    2     0.04
[35,40)    5     0.10
[40,45)   11     0.22
[45,50)   13     0.26
[50,55)    8     0.16
[55,60)    7     0.14
[60,65)    4     0.08
```

to add percentage frequecy and cumulative frequecny column

```
> percenfreq=100*relafreq
> cumfreq=cumsum(freq)
> cbind(freq,relafreq,percenfreq,cumfreq)
      freq relafreq percenfreq cumfreq                    #Output
[30,35)    2     0.04         4       2
[35,40)    5     0.10        10       7
[40,45)   11     0.22        22      18
[45,50)   13     0.26        26      31
[50,55)    8     0.16        16      39
[55,60)    7     0.14        14      46
[60,65)    4     0.08         8      50
```

B: GRAPHICAL REPRESENTATION

Q.1) Draw

1. Histogram
2. Frequency curve
3. Frequency polygon
4. Ogives(Less than and more than type) for the following data.

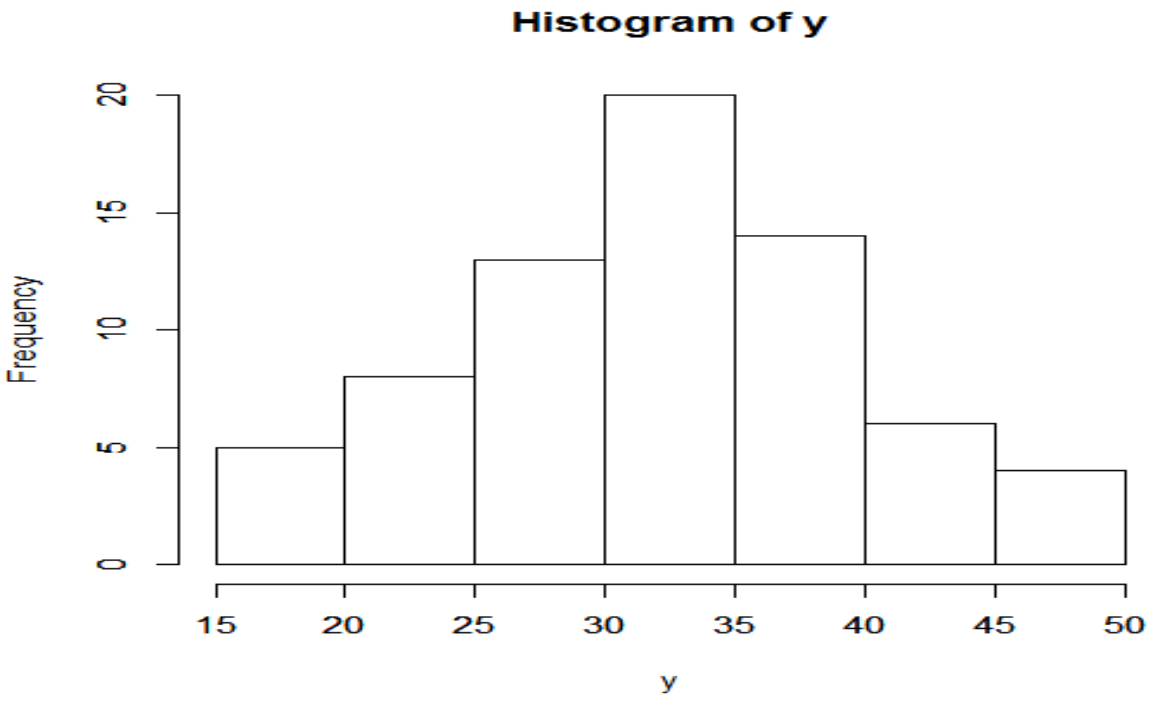
Life of tyres (in '000kms)	15-20	20-25	25-30	30-35	35-40	40-45	45-50
No. of tyres	5	8	13	20	14	6	4

Solution:

1) Histogram

```
x=seq(17.5,48.5,5)
> f=c(5,8,13,20,14,6,4)
> w=5
> lb=x-w/2; ub=x+w/2
> brks=c(lb[1],ub)
> y=rep(x,f)
> hist(y)
```

Output:

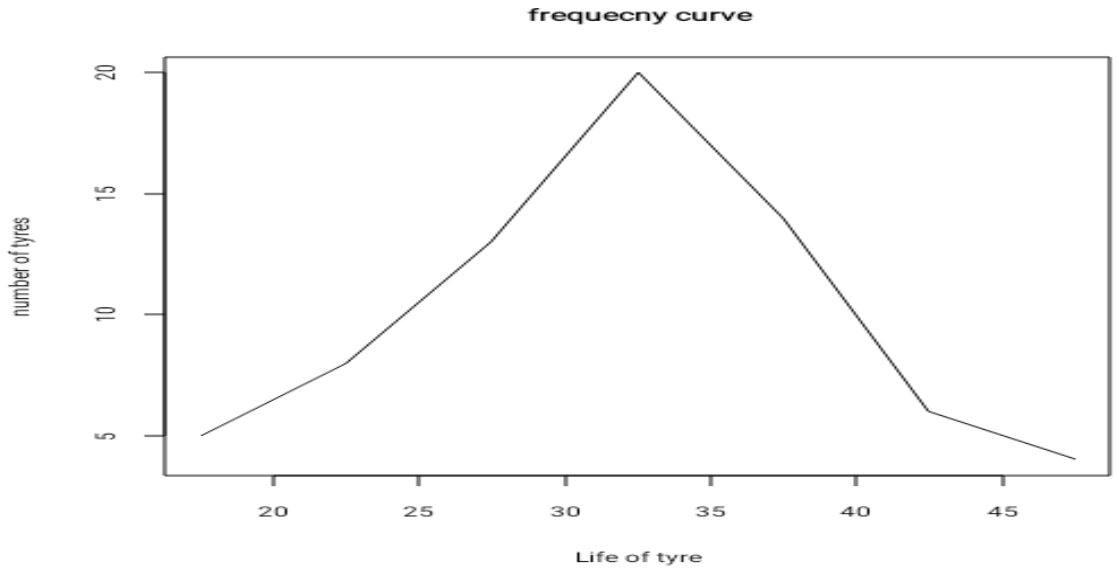


2) Frequency curve

Solution: R Code is

```
x=seq(17.5,48.5,5)
> f=c(5,8,13,20,14,6,4)
> plot(x,f,"l",xlab="Life of tyre",ylab="number of tyres",main="frequecny curve")
```

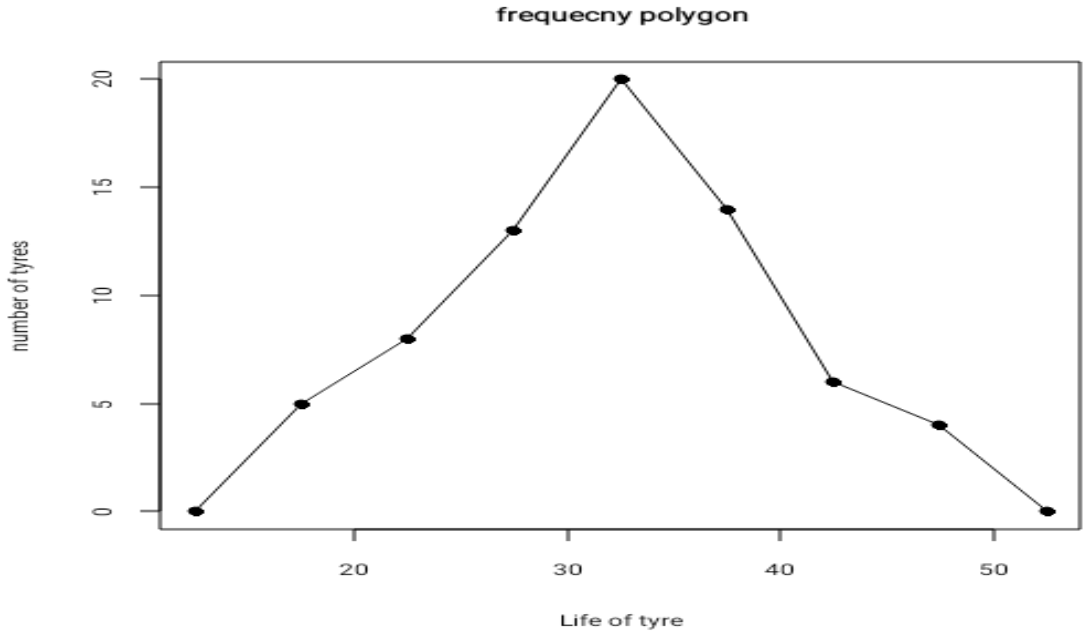
Output:



3) Frequency Polygon

```
x=seq(17.5,48.5,5)
> f=c(5,8,13,20,14,6,4)
> x1=c(12.5,x,52.5)
> f1=c(0,f,0)
> plot(x1,f1,"l",xlab="Life of tyre",ylab="number of tyres",main="frequecny polygon")
> points(x1,f1)
```

Output:



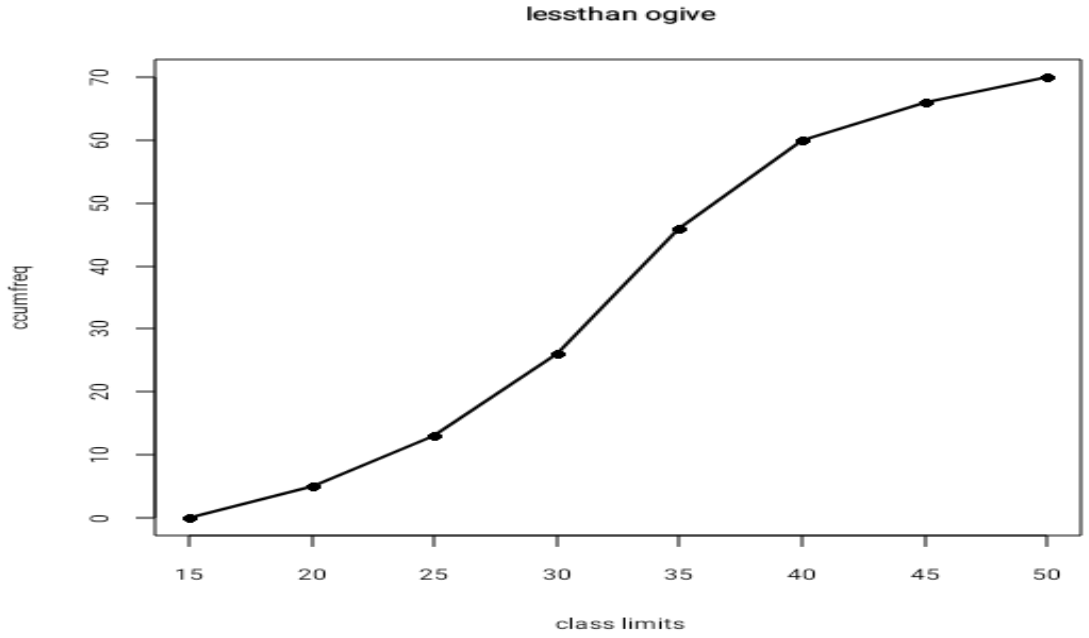
4)

a. Ogive (Less Than type)

Solution: R code is (Less than Ogives)

```
>x=seq(17.5,48,5); w=5
> f=c(5,8,13,20,14,6,4)
> lb=x-w/2;ub=x+w/2;k=length(x)
> lb1=c(lb,50);ub1=c(15,ub)
> f1=c(0,f)
> lcf=cumsum(f1)
> plot(ub1,lcf,"l",xlim=c(15,50),xlab="class limits",ylab="ccumfreq",main="lessthan
ogive",lwd=2)
> points(ub1,lcf,pch=16)
```

Output:



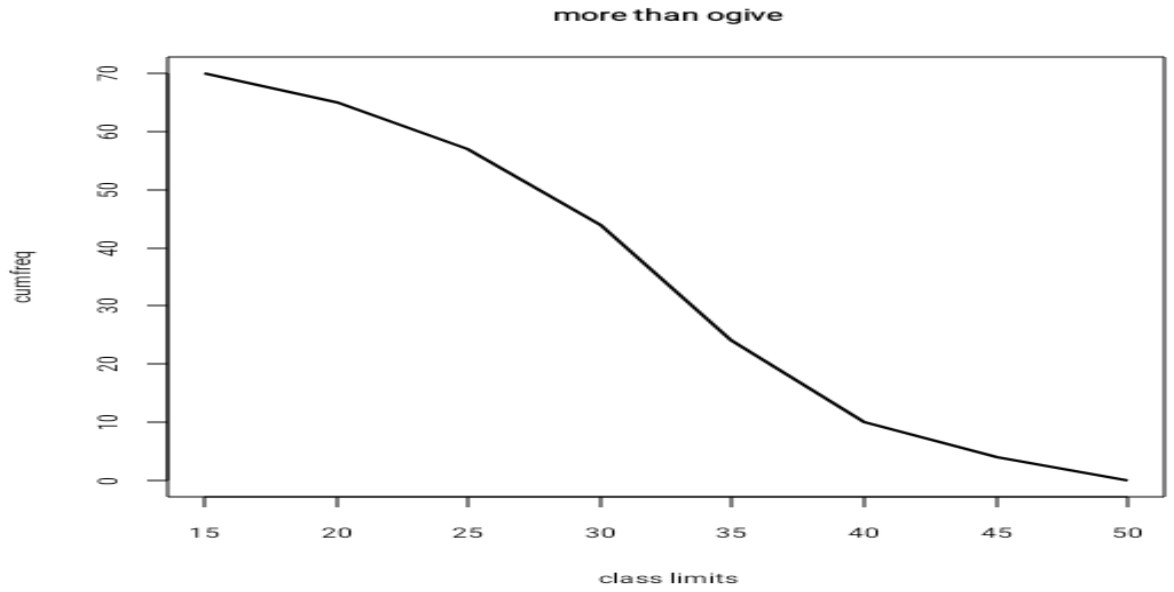
b. Ogives (more Than type)

Solution: R code is (More than Ogive)

```
>x=seq(17.5,48,5); w=5
> f=c(5,8,13,20,14,6,4)
> lb=x-w/2;ub=x+w/2;k=length(x)
> f2=c(f,0);k1=k+1
> mcf=1:k1
```

```
> for(i in 1:k1){mcf[i]=sum(f2[k1:i])}
> lb1=c(lb,50);ub1=c(15,ub)
> plot(lb1,mcf,"l",xlim=c(15,50),xlab="class limits",ylab="cumfreq",main="more than
ogive",lwd=2)
```

Output:



Practical 03: Measure of central tendency and dispersion

Q.1) Monthly sales (in '000Rs.) of 20 small shops are given below:
120, 115, 130, 140, 180, 210, 180, 120, 130, 150, 100, 190, 210, 160, 150, 160, 190, 200, 170, 152.
Calculate arithmetic mean, median, Q1, D3, P42, Q3 and print those values.

Solution:

```
> x=c(120, 115, 130, 140, 180, 210, 180, 120, 130, 150, 100, 190, 210, 160, 150, 160, 190, 200,
170, 152)
> n=length(x)
> am=mean(x)
> lx=log10(x)
> gm=10^mean(lx)
> hm=n/sum(1/x)
> me=median(x)
> q1=quantile(x,0.25);d3=quantile(x,0.3);p42=quantile(x,0.42);q3=quantile(x,0.75)
> cat("arithmetic mean=",am)
arithmetic mean= 157.85>                              #Output
> cat("geometric mean=",gm)
geometric mean= 154.5132>                             #Output
> cat("harmonic  mean=",hm)
harmonic  mean= 151.1194>                             #Output
> cat("median=",me)
median= 156>                                          #Output
> cat("Q1=",q1)
Q1= 130>                                              #Output
```

```
> cat("D3=",d3)
D3= 137>                                    #Output
> cat("P42=",p42)
P42= 150>                                   #Output
> cat("Q3=",q3)
Q3= 182.5>                                  #Output
```

Q.2) For the following frequency distribution. Calculate arithmetic mean, mode, median, Q3 and variance.

X	1	2	3	4	5	6	7	8	9	10
F	7	11	10	8	7	5	4	3	2	1

Solution:

```
> x=1:10
> f=c(7,11,10,8,7,5,4,3,2,1)
> n=sum(f)
> y=rep(x,f);am=mean(y)
> m=which(f==max(f));mo=x[m];me=median(y);q3=quantile(y,0.75)
> v=var(y)
> cat("Arithemtic mean=",am)
Arithemtic mean= 4.068966>                  #Output
> cat("mode=",mo)
mode= 2>                                    #Output
> cat("Q3=",q3)
Q3= 5.75>                                   #Output
> cat("variance=",v)
variance= 5.503932>                         #Output
> cat("median=",me)
median= 4>                                  #Output
> cat("variance=",v)
variance= 5.503932>                         #Output
```

Q.3**) Find arithmetic mean, median and mode of the following data.**

Classes	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	6	8	15	24	19	14	12	7

Solution: R code is

```
> lb=seq(0,70,10);ub=seq(10,80,10);h=10
> f=c(6,8,15,24,19,14,12,7)
> x=(lb+ub)/2
> am=sum(f*x)/n;
> lcf=cumsum(f)
> mc=min(which(lcf>=n/2))
> me=lb[mc]+(n/2-lcf[mc-1])*h/f[mc]
> moc=which(f==max(f))
> mo=lb[moc]+h*((f[moc]-f[moc-1])/(2*f[moc]-f[moc-1]-f[moc+1]))
> cat("Aruthmetic mean=",am)
Aruthmetic mean= 74.05172>                    #Output
> cat("medain=",me)
medain= 30>                                   #Output
> cat("mode=",mo)
mode= 36.42857>                               #Output
```

Q.4) **The score in test is given for 20 students is given below:**
20, 15, 19, 17, 21, 16, 15, 22, 24, 12, 17, 13, 12, 12, 18, 17, 18, 19, 16, 13, 23, 10.
Find (1) Range and coefficient of range. (2) Quartile deviation and coefficient of quartile deviation, (3) Coefficient of variation.

Solution: R code is

```
> x=c(20,15,19,17,21,16,15,22,24,12,17,13,12,18,17,19,16,23,10)
>  n=length(x);m=mean(x)
> min=min(x);max=max(x);r=max-min;cr=r/(max+min)
> q1=quantile(x,0.25);
> q3=quantile(x,0.75)
> qd=(q3-q1)/2;cqd=(q3-q1)/(q3+q1)
> v1=var(x)
> v=((n-1)/n)*v1
> sd=v^0.5;cv=sd*100/m
> cat("Aritmetic mean=",m)
Aritmetic mean= 17.15789>                     #Output
> cat("Range=",r)
Range= 14>                                    #Output
> cat("coefficient of range=",cr)
coefficient of range= 0.4117647>              #Output
> cat("Quartile deviation",qd)
Quartile deviation 2.25>                      #Output
> cat("coefficient of variation=",cv)
coefficient of variation= 21.91052>           #Output
```

PRACTICAL 04: MOMENT AND MEASURES OF SKEWNESS AND KURTOSIS

Q.1) For the following data calculate first four central moments.
82,66,76,95,45,30,57,86,78,64,89,92,58,74,64,85,48,72,65,80.

Solution: R code is

```
> x=c(82,66,76,95,45,30,57,86,78,64,89,92,58,74,64,85,48,72,65,80)
> m=mean(x);n=length(x)
> m1=sum((x-m)^1)/n
> m2=sum((x-m)^2)/n;m3=sum((x-m)^3)/n;m4=sum((x-m)^4)/n
> cat("mean=",m)
mean= 70.3>                                    #Output
cat("first central moment",m1)
first central moment 2.842171e-15>                    #Output
cat("second central moment",m2)
second central moment 272.41>                  #Output
cat("third central moment",m3)
third central moment -2779.596>                #Output
cat("fourth central moment=",m4)
fourth central moment= 210459.1>               #Output
```

Q.2) For the following data, find (1) Karl Pearson's coefficient of skewness, (2) Bowley's

coefficient of skewness, (3) Coefficient of skewness based on moments, (4) Measure of kurtosis.
182, 166, 176, 195, 145, 130, 157, 106, 198, 124, 199, 197, 128, 104, 94, 115, 128, 102, 125, 130.

Solution: R code is

```
> x=c(182, 166, 176, 195, 145, 130, 157, 106, 198, 124, 199, 197, 128, 104, 94, 115, 128, 102,
125, 130)
> m=mean(x); n=length(x)
> m1=sum((x-m)^1)/n; m2=sum((x-m)^2)/n;m3=sum((x-m)^3)/n;m4=sum((x-m)^4)/n
> s=m2^0.5;q1=quantile(x,0.25);q3=quantile(x,0.75);me=quantile(x,0.5)
> sk=3*(m-me)/s;sb=(q3+q1-2*me)/(q3-q1)
> ms=m3^2/m2^3;mk=m4/m2^2;
> cat("karl pearson cofficient of skewness=",sk)
karl pearson cofficient of skewness= 1.296233>                    #Output
> cat("Bowley's coefficient of skewness",sb)
Bowley's coefficient of skewness 0.7040359>                       #Output
> cat("coefficient of skewness based on moments=",ms)
coefficient of skewness based on moments= 0.09354859>             #Output
> cat("measure of kurtosis=",mk)
measure of kurtosis= 1.687262>                                    #Output
```

Q.3) **For the following data, find (1) Karl Pearson's coefficient of skewness, (2) Bowley's coefficient of skewness, (3) Coefficient of skewness based on moments, (4) Measure of kurtosis.**

X	**2**	**3**	**4**	**5**	**6**	**7**
Frequency	**5**	**10**	**15**	**20**	**15**	**10**

Solution: R code is

```
> x=2:7; f=c(5,10,15,20,15,10);n=sum(f)
> m=sum(f*x)/n;m1=sum(f*(x-m)^1)/n;m2=sum(f*(x-m)^2)/n;m3=sum(f*(x-m)^3)/n;
> m4=sum(f*(x-m)^4)/n;
> s=m2^0.5;mc=which(f==max(f));mo=x[mc]
> sk=(m-mo)/s;
> y=rep(x*f);q1=quantile(y,0.25);q2=quantile(y,0.5);q3=quantile(y,0.75)
> sb=(q3+q1-2*q2)/(q3-q1)
> ms=m3^2/m2^3;mk=m4/m2^2
> cat("karl pearson coefficient of skewness=",sk)
karl pearson coefficient of skewness= -0.1404879>                 #Output
> cat("Bowley's coefficent of skewness=",sb)
Bowley's coefficent of skewness= -0.1578947>                      #Output
> cat("coefficient of skewness based on moments=",ms)
coefficient of skewness based on moments= 0.03985639>             #Output
> cat("measure of kurtosis=",mk)
```

measure of kurtosis= 2.219875> #Output

PRACTICAL 05: CORRELATION AND REGRESSION ANALYSIS

Q.1) Following are the marks obtained by 10 students in 2 subjects X & Y

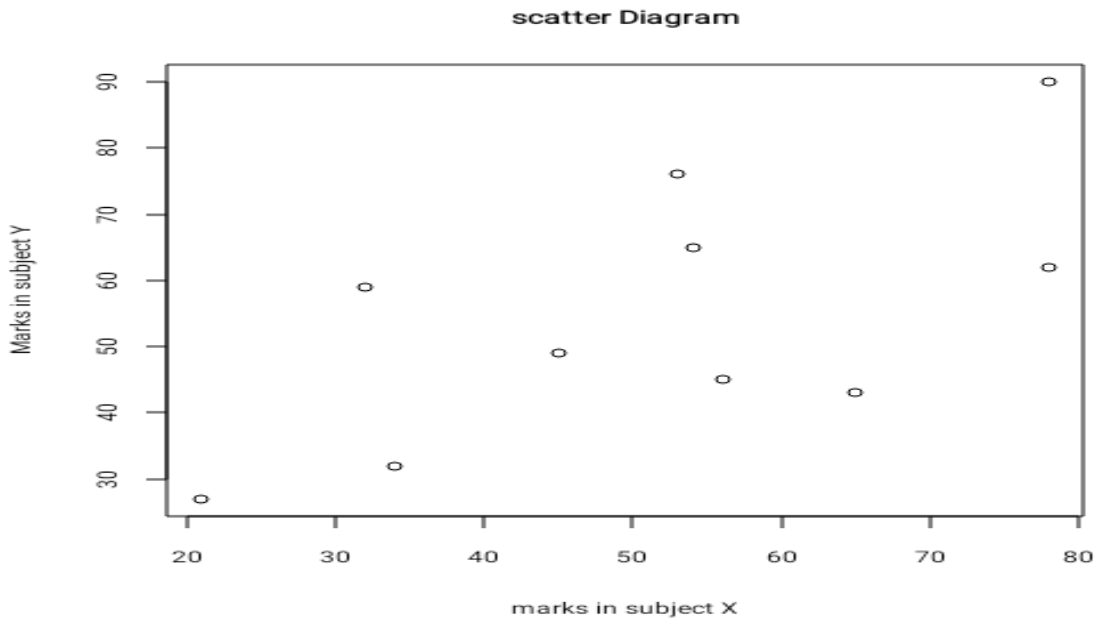
X	**34**	**56**	**78**	**21**	**45**	**32**	**78**	**53**	**65**	**54**
Y	**32**	**45**	**62**	**27**	**49**	**59**	**90**	**76**	**43**	**65**

1) **Draw a scatter diagram**
2) **Find the correlation coefficient between X and Y.**

Solution: R code is

```
> x=c(34,56,78,21,45,32,78,53,65,54)
> y=c(32,45,62,27,49,59,90,76,43,65)
> plot(x,y,main="scatter Diagram",xlab="marks in subject X",ylab="Marks in subject Y")
> r=cor(x,y)
> cat("correlation coefficient=",r)
correlation coefficient= 0.6511619>
```

Output:



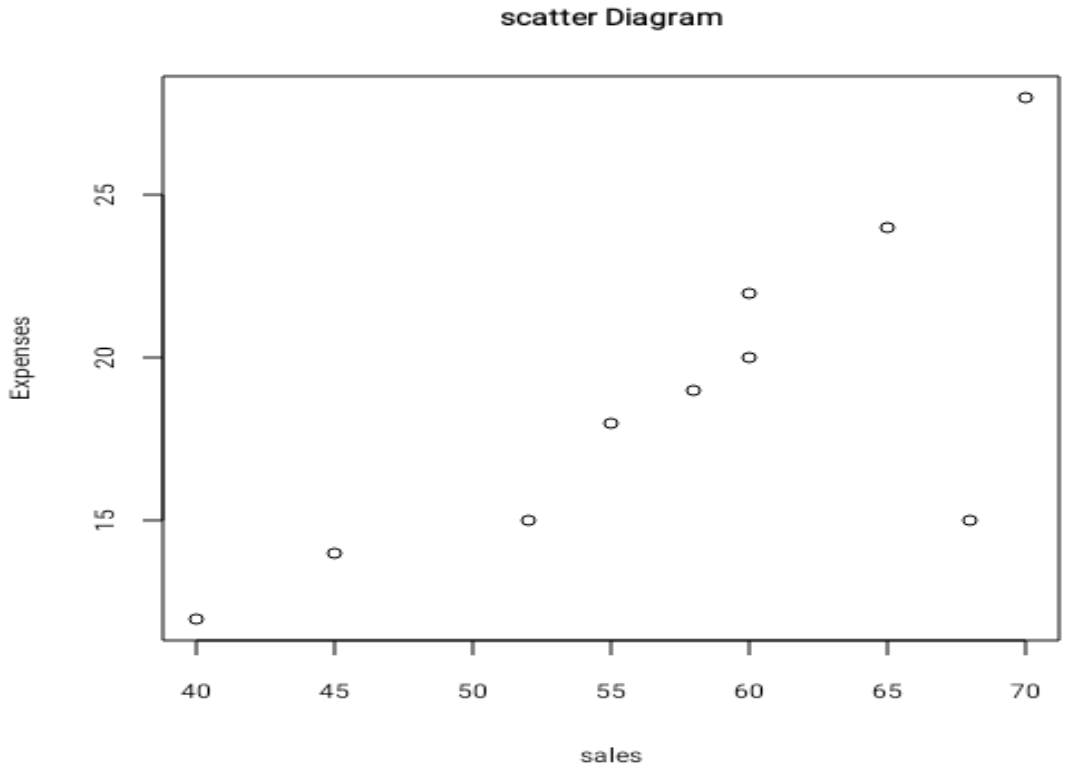
Q.2) For the following data write R program to calculate correlation coefficient and plot scatter diagram. Print given values of sales and expense (figure are given in '000 Rs.)

Sales	**40**	**45**	**52**	**60**	**58**	**60**	**65**	**68**	**70**	**55**
Expenses	**12**	**14**	**15**	**20**	**19**	**2**	**24**	**15**	**28**	**18**

Solution: R code is

```
> sales=c(40,45,52,60,58,60,65,68,70,55)
> expen=c(12,14,15,20,19,22,24,15,28,18)
> plot(sales,expen,main="scatter Diagram",xlab="sales",ylab="Expenses")
> d=data.frame(sales,expen)
> r=cor(sales,expen)
> print(d)
   sales expen                                    #Output
1     40    12
2     45    14
3     52    15
4     60    20
5     58    19
6     60    22
7     65    24
8     68    15
9     70    28
10    55    18
> cat("corrleation coefficient is=",r)
corrleation coefficient is= 0.7595066>
```

Output:



Q.3) The sales of a company (in million Rs) for each year are shown in the table below.

X (year)	2005	2006	2007	2008	2009	2010
Y (sales)	12	19	29	37	45	54

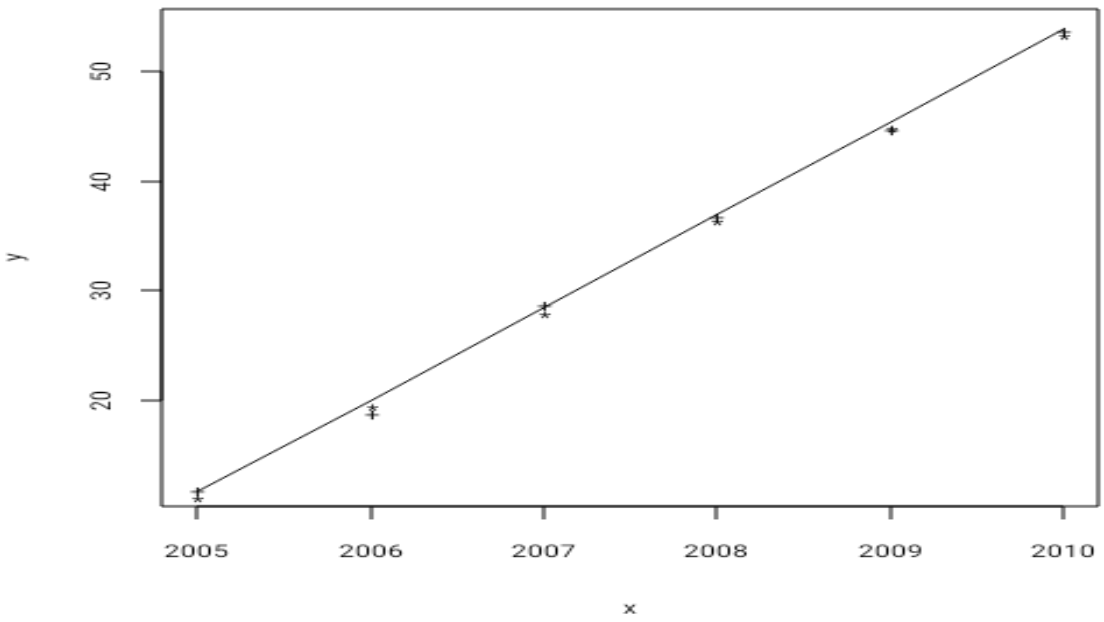
(1) Find equation of line of regression of Y on X.
(2) Find estimated values corresponding to given values and print those along with given data.
(3) Estimate Y when X=2016.
(4) Plot observed and estimated values along the regression line.

Solution: R code is

```
> x=c(2005,2006,2007,2008,2009,2010)
> y=c(12,19,29,37,45,54)
> rl=lm(y~x);co=coef(rl);mco=matrix(co)
> a=mco[1,1];b=mco[2,1]
> esty=fitted(rl)
> x1=2016;ey=a+b*x1
> cat("constant term=",a)
constant term= -16945.05>                                   #Output
> cat("constant term=",b)
constant term= 8.457143>                                    #Output
> cat("estimated value of sale for the year 2016 is",ey)
estimated value of sale for the year 2016 is 104.5524>          #Output
> plot(x,y,pch="+")
> points(x,esty,pch="*");lines(x,esty)
> d=data.frame(x,y,esty)
```

```
> print(d)
    x  y    esty                                    #Output
1 2005 12 11.52381
2 2006 19 19.98095
3 2007 29 28.43810
4 2008 37 36.89524
5 2009 45 45.35238
6 2010 54 53.80952
```

Output:



PRACTICAL 06: PROBABILITY

Q.1) An urn contains 10 white, 6 red and 9 black balls. If 6 balls are drawn at random find the probability that

(1) Four of the balls drawn are white
(2) Two is of each colour
(3) At least one is white

Solution: R Code is

```
> n=choose(25,6)
> m1=choose(10,4)*choose(15,2)
```

```
> p1=m1/n
> m2=choose(10,2)*choose(9,2)*choose(6,2)
> p2=m2/n
> m3=choose(19,4)
> p3=m3/n
> m4=choose(15,6)
> p4=1-m4/n
> cat("Prob four of the balls drawn are white=",p1)
Prob four of the balls drawn are white= 0.1245059>                    #Output
> cat("prob two is of each colour=",p2)
prob two is of each colour= 0.1372106>                                #Output
> cat("prob None is red=",p3)
prob None is red= 0.02188594>                                         #Output
> cat("prob at least one is white=",p4)
prob at least one is white= 0.9717391>                                #Output
```

Q.2) **A stockiest has 40 items in a lot. Out of which 30 are non-defective and 10 are defective. A customer select 6 items from the lot. What is the probability that out of these six items**

1. **All items are non-defective**
2. **Four are non-defective and two is defective**

Solution: R code is

```
> n=choose(40,6)
> m1=choose(30,6)
> m2=choose(30,4)*choose(10,2)
> p1=m1/n;p2=m2/n
> cat("pro all selected are non-defective=",p1)
pro all selected are non-defective= 0.1546942>                                #Output
> cat("Prob among selected 4 are non-defective and 2 defective=",p2)
Prob among selected 4 are non-defective and 2 defective= 0.3212879>           #Output
```

Q.3) A box contains counters numbers 1, 2, ……., 25. A counter is selected at random. E1 denotes number on the counter is divisible by 2: E2 denoted number on the counter is divisible by 3. Calculate

1. **Prob of (E1 and E2)**
2. **Prob of (E1 or E2)**
3. **P(E1/E2)**

Solution: R code is

```
> n=25;
> e1=seq(2,25,2);e2=seq(3,25,3);e12=seq(6,25,6)
> m1=length(e1);m2=length(e2);m12=length(e12)
> pe1=m1/n;pe2=m2/n;pe12=m12/n
> p1=pe12;p2=pe1+pe2-pe12
```

```
> p3=pe12/pe1
> cat("Prob of E1 and E2=",p1)
Prob of E1 and E2= 0.16>                    #Output
> cat("prob of (E1 or E2)=",p2)
prob of (E1 or E2)= 0.64>                   #Output
> cat("Prob of (E1/E2)=",p3)
Prob of (E1/E2)= 0.3333333>                 #Output
```