Statistical Computations using R Software

II B.Sc Statistics

Solving system of equations using inverse matrix

R coding

```
matrix(c(6,4,2,1,-2,8,1,5,7),3,3)
m1<-solve(m)
m1
d<-det(m)
```

Output:

```
> options(digits=1)
> m<-matrix(c(6,4,2,1,-2,8,1,5,7),3,3)
> m1<-solve(m)
> m1
      [,1]   [,2]  [,3]
[1,]  0.18 -0.003 -0.02
[2,]  0.06 -0.131  0.08
[3,] -0.12  0.150  0.05
```

d= - 306

Fitting of Linear model

```
> speed<-c(4,4,7,7,8,9)
> dist<-c(2,10,4,22,16,10)
> linearmod<-lm(dist~speed)
> print(linearmod)
```

Output

```
Call:
lm(formula = dist ~ speed)

Coefficients:
(Intercept)       speed
     0.9922      1.4884
```

$\boldsymbol{dist = Intercept + (\beta * speed)}$

$\Rightarrow dist = 0.9922 + 1.4884 * speed$

Fitting of Quadratic Model

R-Coding

```
> speed<-c(15,20,25,30,35,40,45,50,55,60,65,70,75)
> mileage<-c(22.3,25.5,27.5,29,28.8,30,29.9,30.2,30.4,28.8,27.4,25.3,23.3)
> quadratic.model<-lm(mileage~speed)
> summary(quadratic.model)
```

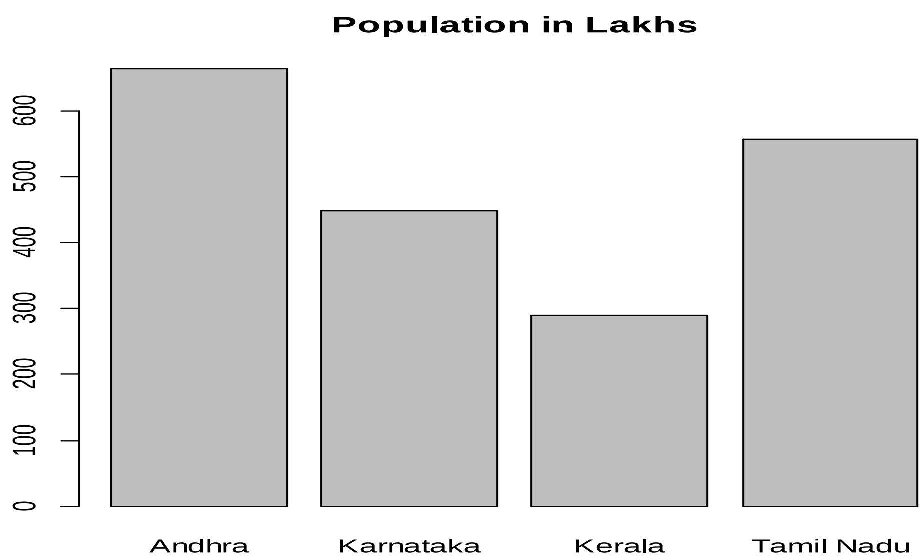
Output

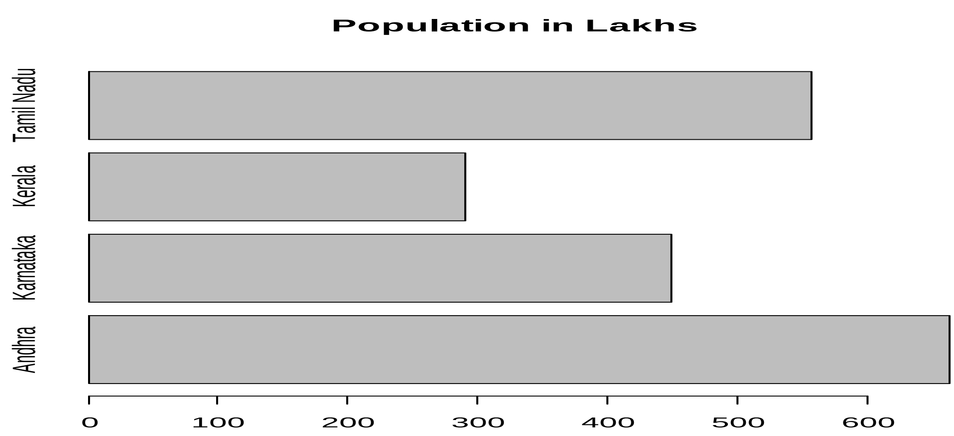
```
Call:
lm(formula = mileage ~ speed)
Residuals:
   Min     1Q Median     3Q    Max
-5.025 -1.866  1.109  2.331  2.749

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 27.203297   2.026072  13.427 3.64e-08 ***
speed        0.008132   0.041574   0.196    0.848
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.804 on 11 degrees of freedom
Multiple R-squared:  0.003466,  Adjusted R-squared:  -0.08713
F-statistic: 0.03826 on 1 and 11 DF,  p-value: 0.8485
```

Simple Bar Diagram

R-coding

```
> population<-c(663,448,290,556)
> state<-c("Andhra","Karnataka","Kerala","Tamil Nadu")
> barplot(population,names.arg=state,main="Population in Lakhs",horiz=TRUE)
> barplot(population,names.arg=state,main="Population in Lakhs",vertical=TRUE)
```

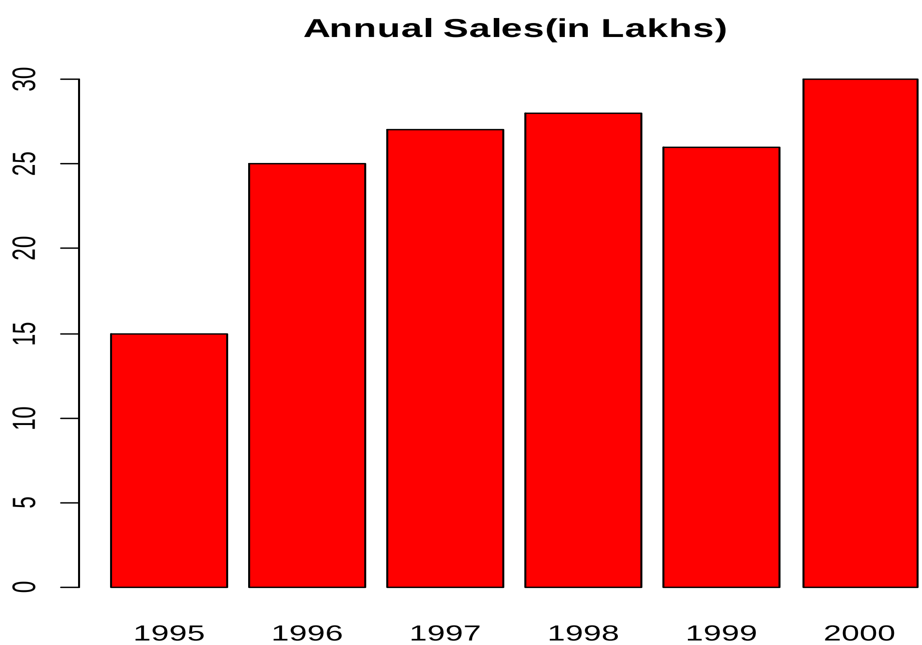




Construction of Simple bar diagram

R coding

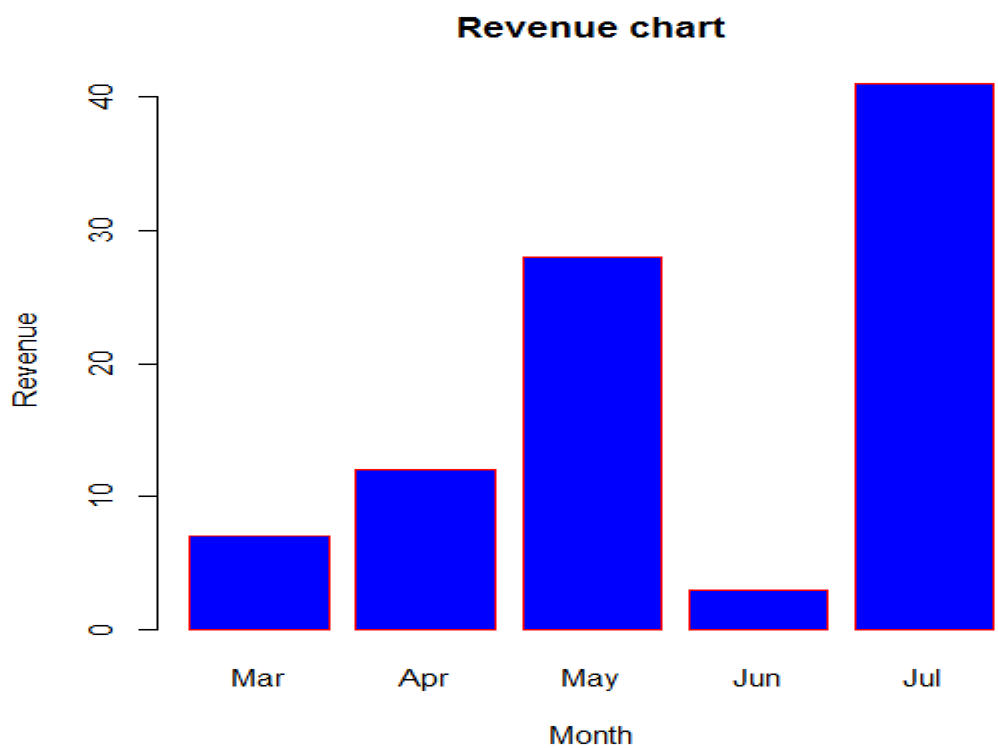
```
> sales<-c(15,25,27,28,26,30)
> year<-c("1995","1996","1997","1998","1999","2000")
> barplot(sales,names.arg=year,main="Annual Sales(in Lakhs)",col="red")
```



Construction of Simple bar diagram

R coding

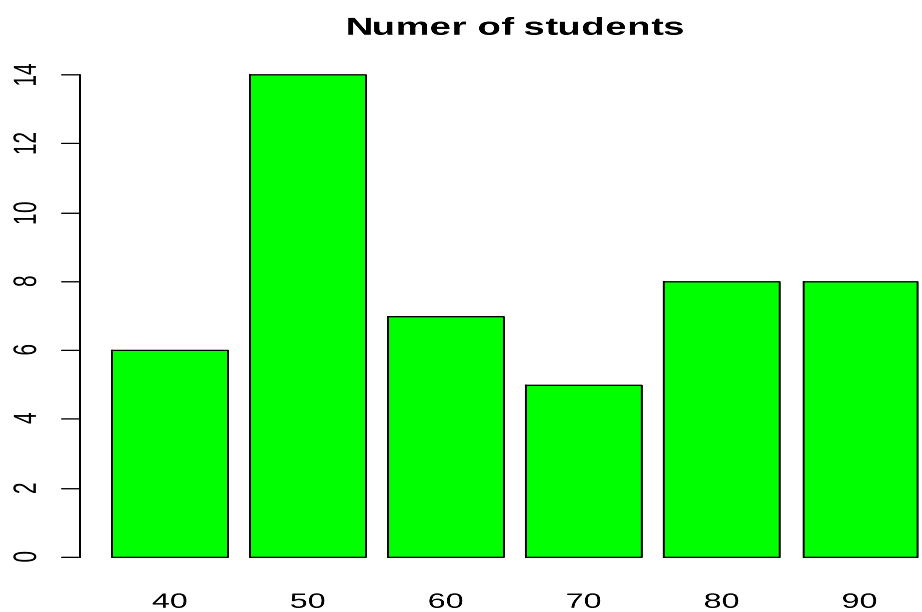
```
H <- c(7,12,28,3,41)
M <- c("Mar","Apr","May","Jun","Jul")
barplot(H,names.arg = M,xlab = "Month",ylab = "Revenue",col = "blue",
main = "Revenue chart",border = "red")
```

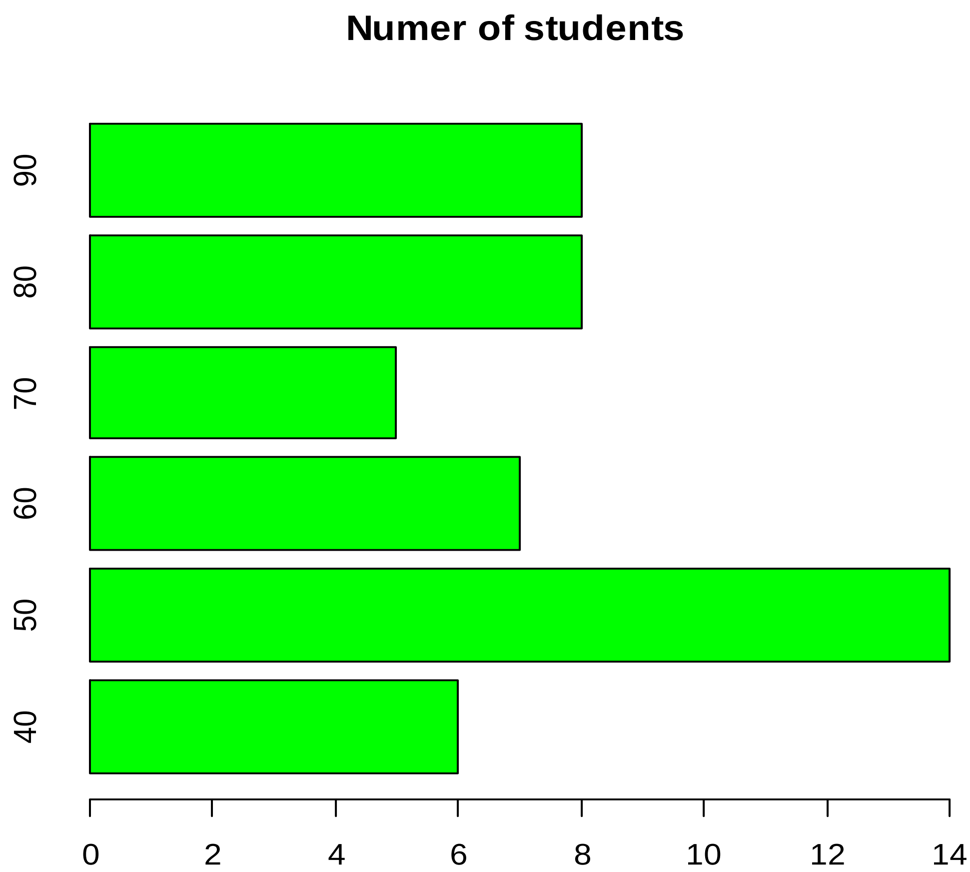


Construction of Simple bar diagram

R coding

```
> marks<-c("40","50","60","70","80","90")

> students<-c(6,14,7,5,8,8)

> barplot(students,names.arg=marks,main="Numer of students",col="green")

> barplot(students,names.arg=marks,main="Numer of students",col="green",horiz=TRUE)
```

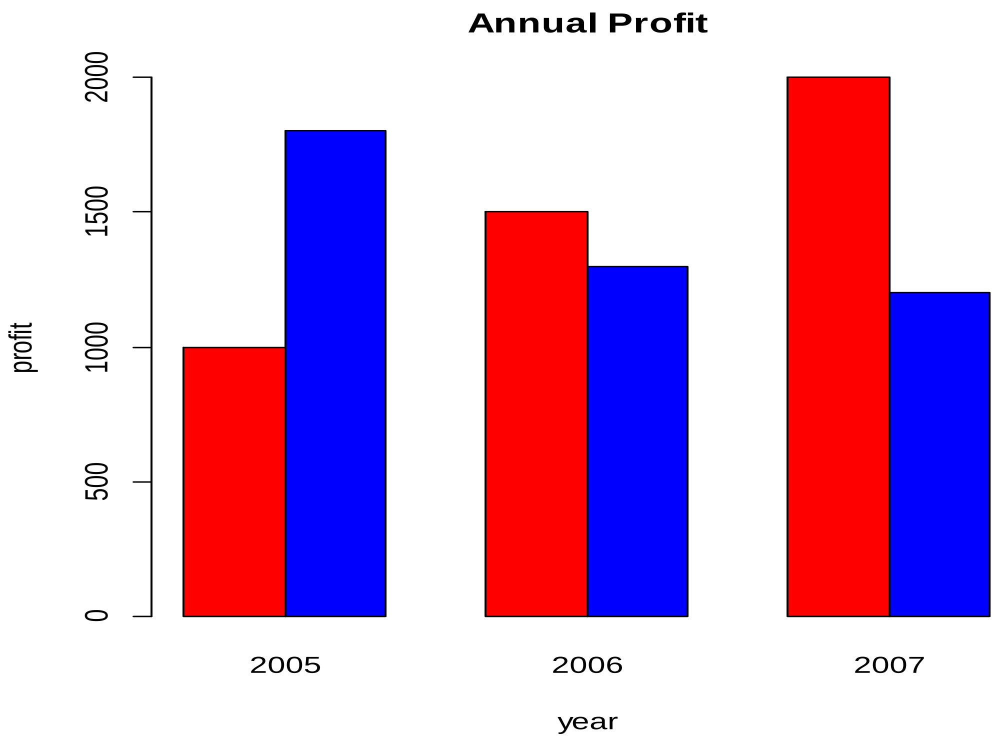



Numer of students
90
80
70
60
50
40
0
2
4
6
8
10
12
14

Construction of Multiple Bar Diagram

R-coding

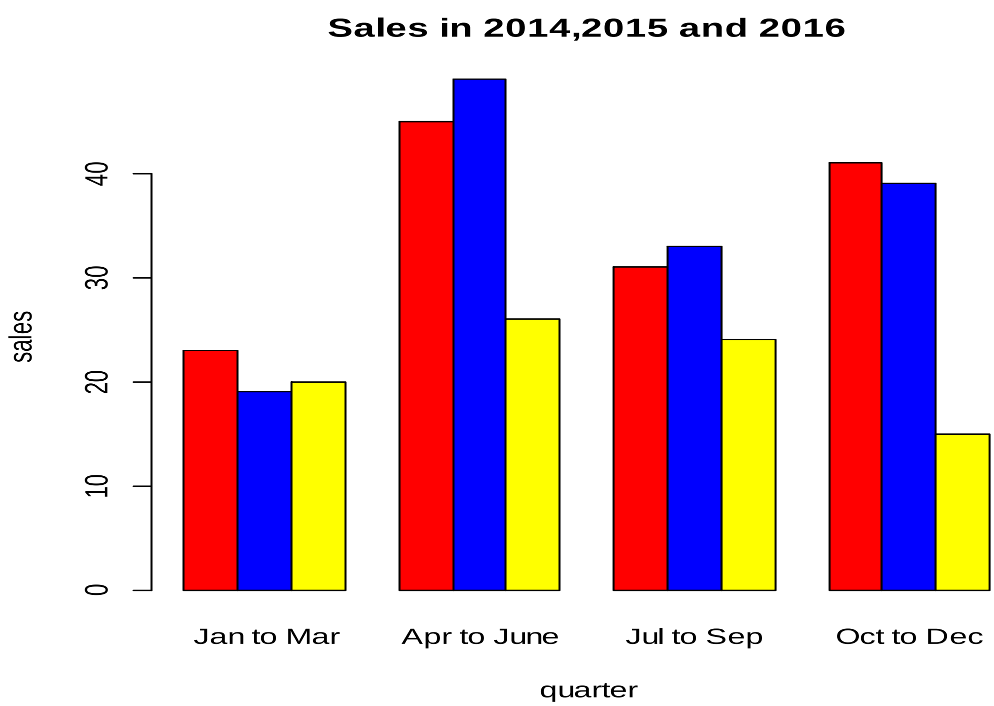
```
> year<-c("2005","2006","2007")

> color<-c("red","blue")

> profit=matrix(c(1000,1500,2000,1800,1300,1200),nrow=2,ncol=3,byrow=T)

>barplot(profit,names.arg=year,xlab="year",ylab="profit",col=color,main="Annual Profit",beside=T)
```

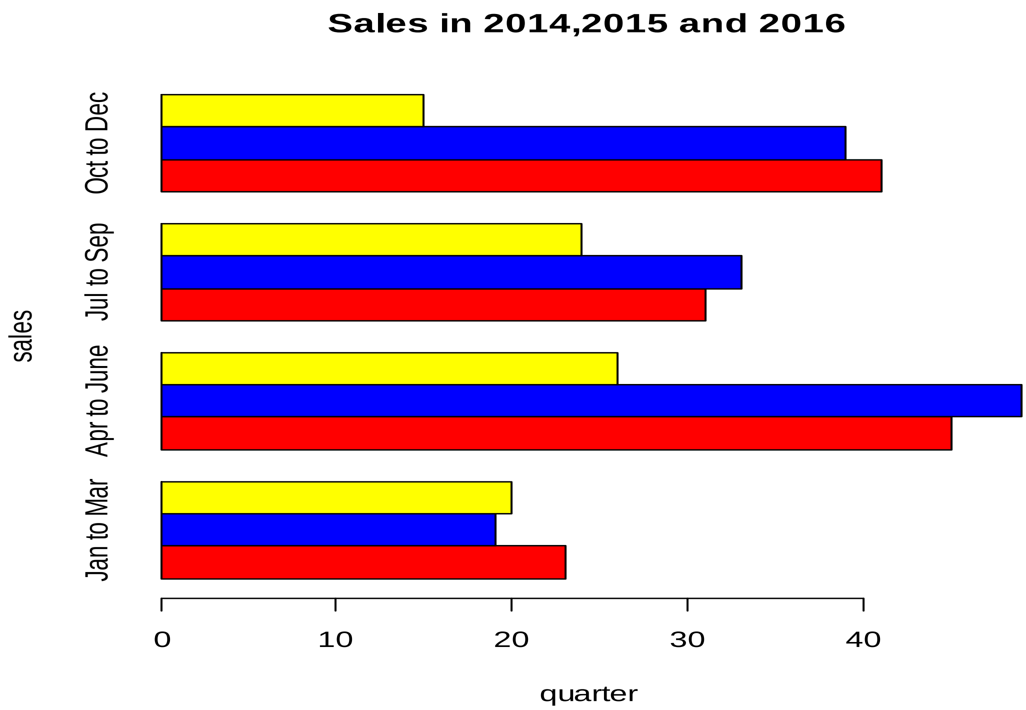


Construction of Multiple Bar Diagram

R-code

```
> quarter<-c("Jan to Mar","Apr to June","Jul to Sep","Oct to Dec")

> color<-c("red","blue","yellow")

> sales=matrix(c(23,45,31,41,19,49,33,39,20,26,24,15),nrow=3,ncol=4,byrow=T)

>barplot(sales,names.arg=quarter,col=color,xlab="quarter",ylab="sales",main="Sales in 2014,2015 and 2016",beside=T)

>barplot(sales,names.arg=quarter,col=color,xlab="quarter",ylab="sales",main="Sales in 2014,2015 and 2016",horiz=TRUE,beside=T)
```



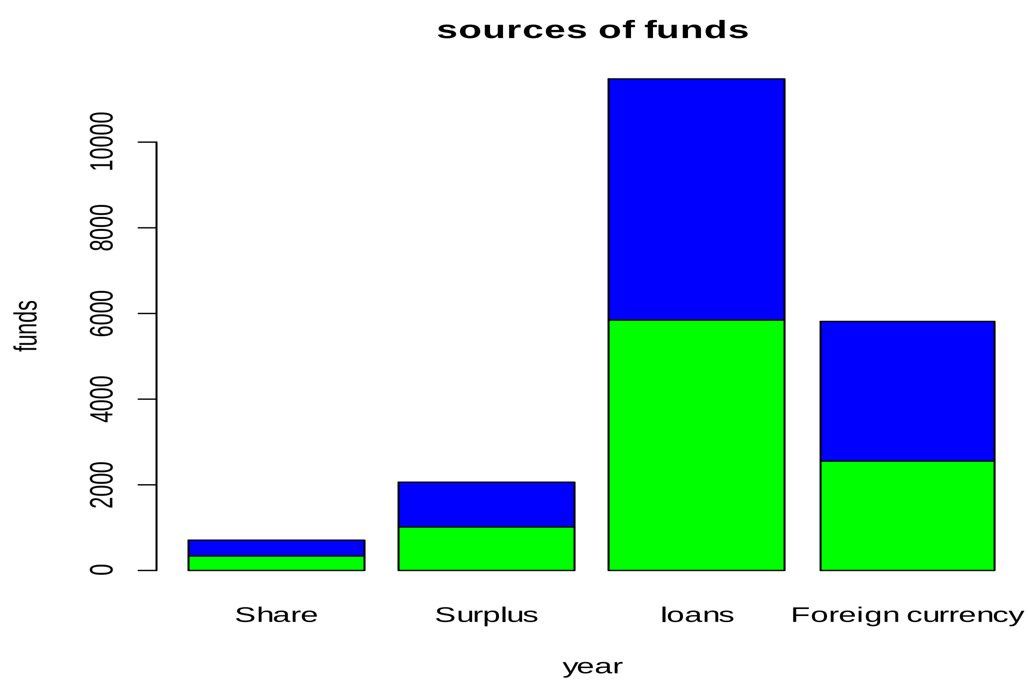


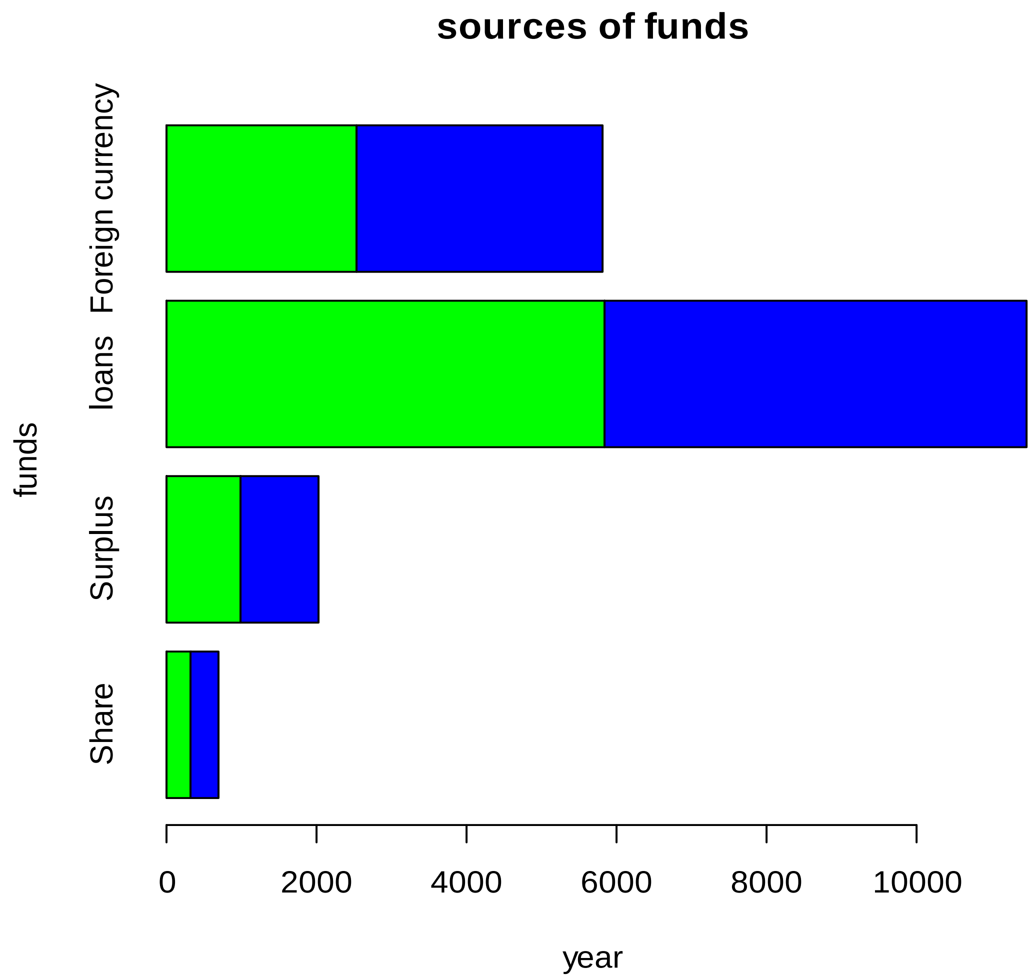
Sales in 2014,2015 and 2016
sales
Oct to Dec
Jul to Sep
Apr to June
Jan to Mar
0
10
20
30
40
quarter

Construction of Sub divided Bar Diagram

R-code

```
> funds<-c("Share","Surplus","loans","Foreign currency")

> colors<-c("green","blue")

> values<-
matrix(c(339,998,5843,2552,352,1043,5614,3262),nrow=2,ncol=4,byrow=TRUE)

> barplot(values,names.arg=funds,xlab="year",ylab="funds",main="sources of
funds",col=colors)

> barplot(values,names.arg=funds,xlab="year",ylab="funds",main="sources of
funds",col=colors)

> barplot(values,names.arg=funds,xlab="year",ylab="funds",main="sources of
funds",col=colors,horiz=TRUE)
```

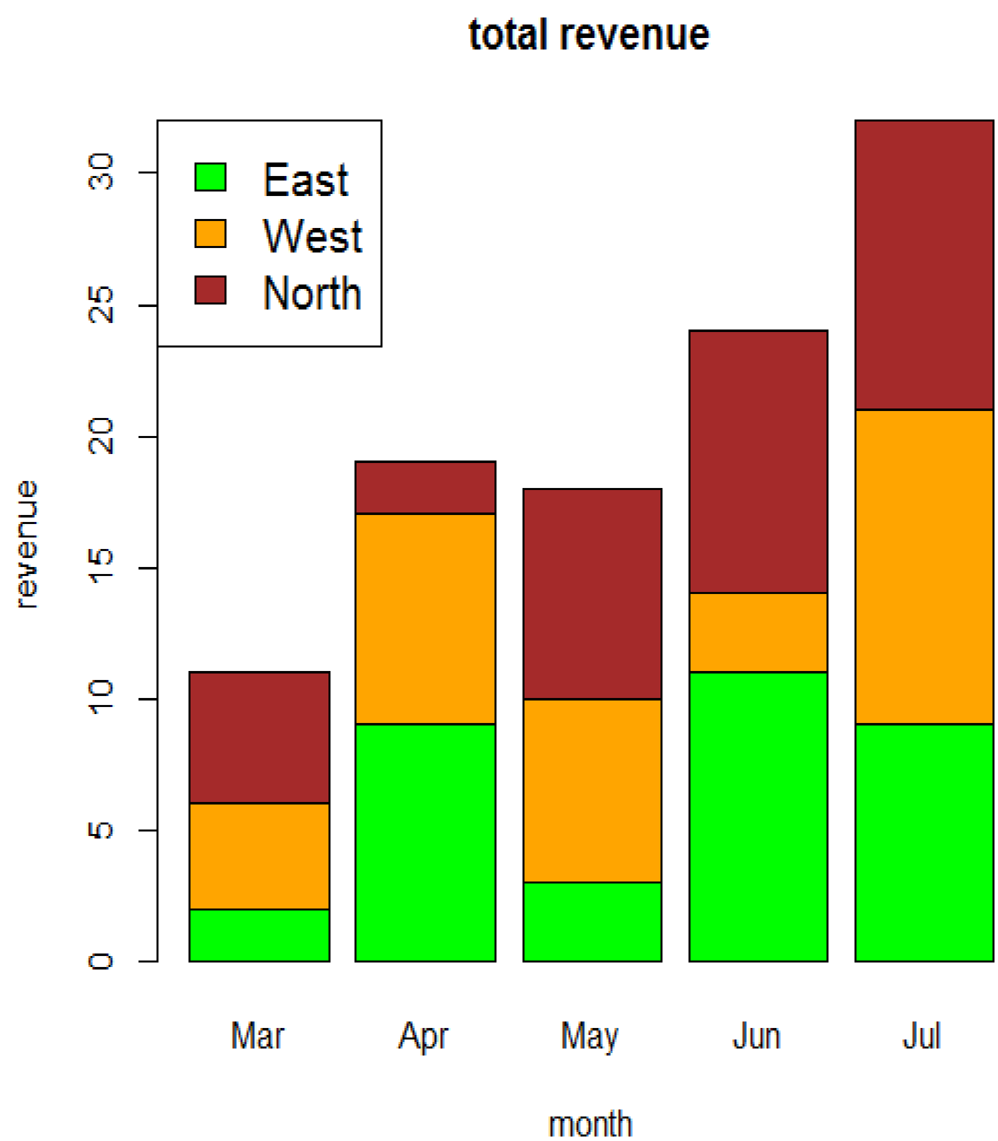



sources of funds
Foreign currency
loans
Surplus
Share
funds
0
2000
4000
6000
8000
10000
year

Construction of Sub divided bar diagram

R-code

```
colors <- c("green","orange","brown")

months <- c("Mar","Apr","May","Jun","Jul")

regions <- c("East","West","North")

Values <- matrix(c(2,9,3,11,9,4,8,7,3,12,5,2,8,10,11),nrow = 3,ncol = 5,byrow =
TRUE)

barplot(Values,main = "total revenue",names.arg = months,xlab = "month",ylab =
"revenue", col = colors )
```

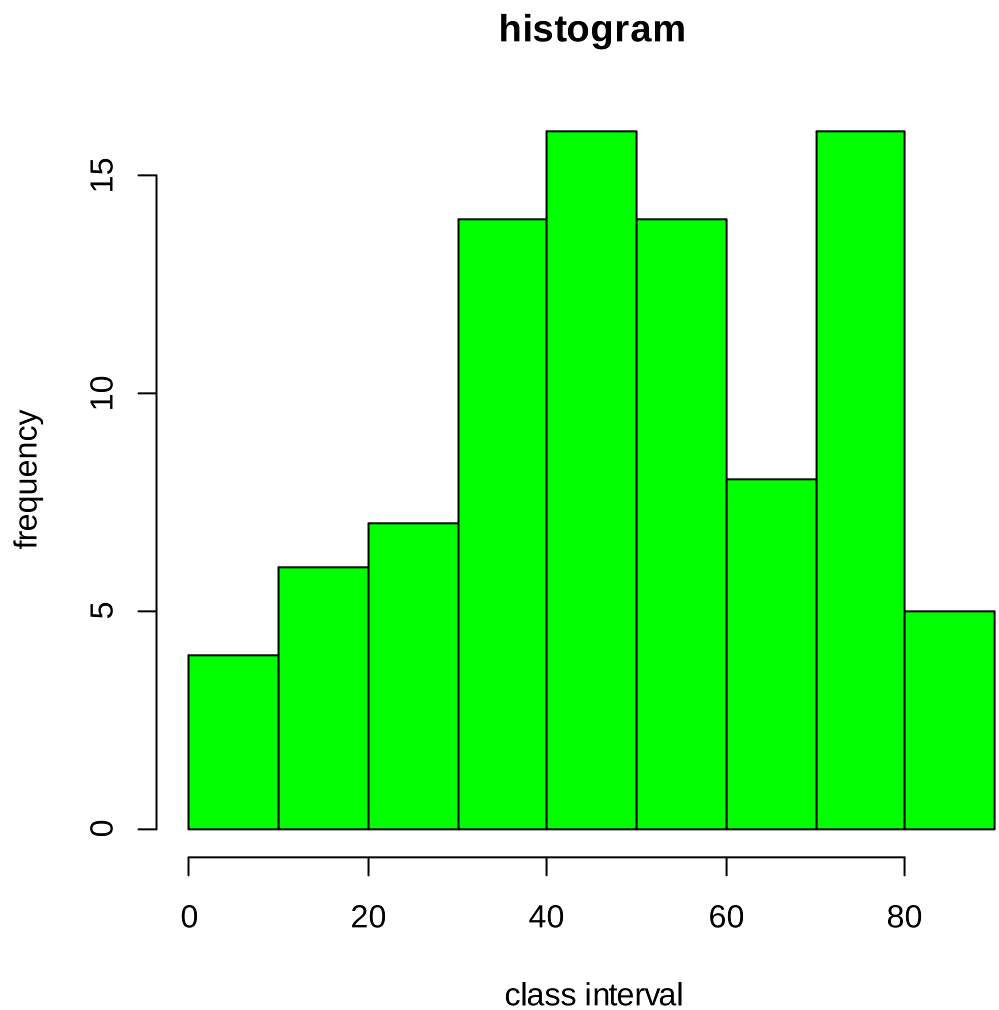


total revenue
East
West
North
0
5
10
15
20
25
30
revenue
Mar
Apr
May
Jun
Jul
month

Construction of Histogram

R-code

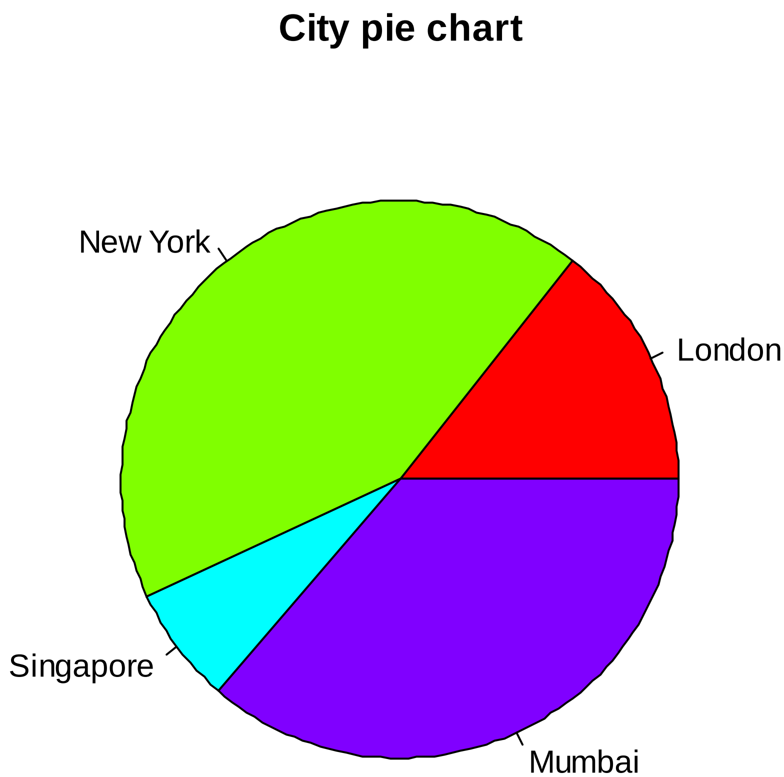
```
> x<-c(5,15,25,35,45,55,65,75,85)
> f<-c(4,6,7,14,16,14,8,16,5)
> a<-rep(x,f)
> brk=seq(0,90,by=10)
> hist(a,brk,xlab="class
interval",ylab="frequency",col="green",main="histogram")
```



Construction Pie Diagram

R-code

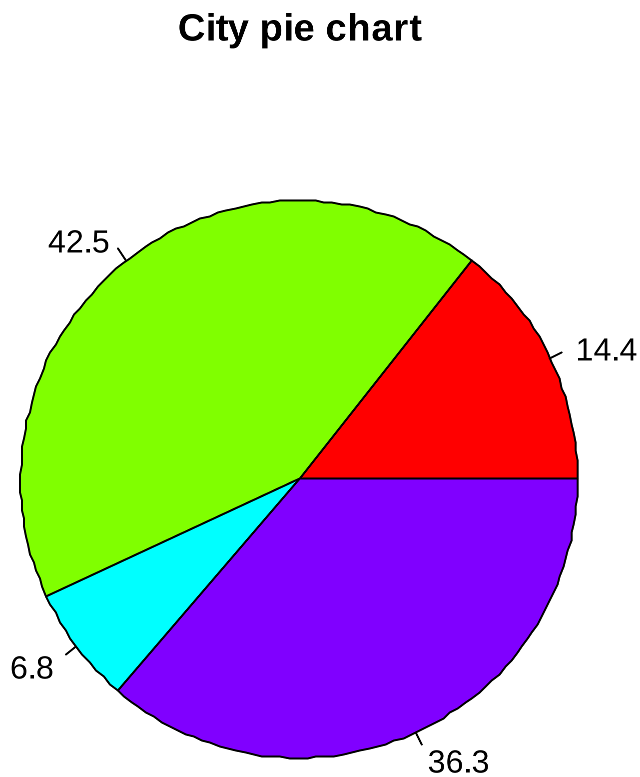
```
> x <- c(21, 62, 10, 53)
> labels <- c("London", "New York", "Singapore", "Mumbai")
> pie(x, labels, main = "City pie chart", col = rainbow(length(x)))
```



Construction Pie Diagram

R code

```
> x <-  c(21, 62, 10,53)
> labels <-  c("London","New York","Singapore","Mumbai")
> piepercent<- round(100*x/sum(x), 1)
> pie(x, labels = piepercent, main = "City pie chart",col = rainbow(length(x)))
```



Correlation coefficient

R-code

```
x<-c(10,12,18,8,13,20,22,15,5,17)
y<-c(88,90,94,86,87,92,96,94,88,85)

cor(x,y,method="pearson")
```

Output

```
0.6369544
```

Correlation coefficient

R-code

```
 Aptitude_Score<-c(57,58,59,59,60,61,62,64)
 Productivity_Index<-c(67,68,65,68,72,72,69,71)
cor(Aptitude_Score,Productivity_Index,method="pearson")
```

Output

```
 0.6030227
```

Regression Coefficient

R-code

```
> Y<-c(88,90,94,86,87,92,96,94,88,85)
> X<-c(10,12,18,8,13,20,22,15,5,17)
> simple.fit = lm(Y~X)
> summary(simple.fit)
```

Output

```
Call:
lm(formula = Y ~ X)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.3409 -1.1591  0.3409  2.1648  3.5530 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  83.7424     2.8523  29.360 1.96e-09 ***
X             0.4470     0.1913   2.337   0.0476 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.108 on 8 degrees of freedom
Multiple R-squared:  0.4057,    Adjusted R-squared:  0.3314 
F-statistic: 5.461 on 1 and 8 DF,  p-value: 0.04764
```

Regression Coefficient

R-code

```
> inde<-c(10,12,13,12,16,15)
> depen<-c(40,38,43,45,37,43)
> simple.fit = lm(depen~inde)
> summary(simple.fit)
```

Output

```
Call:
lm(formula = depen ~ inde)

Residuals:
    1     2     3     4     5     6
-1.75 -3.25  2.00  3.75 -3.25  2.50

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  44.2500     9.3489   4.733  0.00908 **
inde         -0.2500     0.7108  -0.352  0.74279
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.482 on 4 degrees of freedom
Multiple R-squared:   0.03,     Adjusted R-squared:  -0.2125
F-statistic: 0.1237 on 1 and 4 DF,  p-value: 0.7428
```

One way table

R-code

```
> x<-c(10,12,5,6,7,12,7,6,5,12,14,14,15,17,18,17,16,20)
> table(x)
```

output

```
x
 5  6  7 10 12 14 15 16 17 18 20
 2  2  2  1  3  2  1  1  2  1  1
> transform(table(x))
    x Freq
1   5    2
2   6    2
3   7    2
4  10    1
5  12    3
6  14    2
7  15    1
8  16    1
9  17    2
10 18    1
11 20    1
```

Two way table

R-code

```
> gender<-c(1,2,1,2,1,2,1,2,2,1,2,1,2,1,2,1,2,1,1,1)
> relign<-c(1,2,3,2,1,1,2,1,3,2,1,2,1,2,1,2,3,2,3,2)
> tble=table(gender,religien)
> tble
```

Output

```
      religien
gender 1 2 3
     1 2 7 2
     2 5 2 2
```

Measures of Central Tendency

Arithmetic mean

R-code

```
> Family<-c("A","B","C","D","E","F","G","H","I","J")
> Expenditure<-c(30,70,10,75,500,8,42,250,40,36)
> mean(Expenditure)
```

output

mean= 106.1

R-code

```
> persons<-c(2,3,4,5,6)
> house<-c(10,25,30,25,10)
> fx=sum(persons*house)
> fx
[1] 400
> f=sum(house)
> f
[1] 100
> fxx=(fx/f)
> fxx
```

Output

Mean= 4

Harmonic mean

R-code

```
> har<-c(6,15,35,40,900,520,300,400,1800,2000)
> aa=(1/har)
> aa
 [1] 0.1666666667 0.0666666667 0.0285714286 0.0250000000 0.0011111111
 [6] 0.0019230769 0.0033333333 0.0025000000 0.0005555556 0.0005000000


> stt=data.frame(har,aa)
> stt
    har       X_data
1     6 0.1666666667
2    15 0.0666666667
3    35 0.0285714286
4    40 0.0250000000
5   900 0.0011111111
6   520 0.0019230769
7   300 0.0033333333
8   400 0.0025000000
9  1800 0.0005555556
10 2000 0.0005000000
> n=length(har)
> n
[1] 10
> sttt=sum(st)
> sttt
[1] 0.2968278
> haa=(n/sttt)
> haa
```

output

```
[1] 33.68956
```

Median

```
> x<-c(57,58,61,42,38,65,72,66)
> median(x)
output
[1] 59.5
```

Measures of Dispersion

Quartile Deviation

R-code

```
> x<-c(391,384,591,407,672,522,777,733,1490,2488)
> quantile(x,0.25)
   25%
435.75
> quantile(x,0.75)
75%
766
> IQR(x)
[1] 330.25
```

Standard Deviation

R-code

```
> x<-c(40,50,60,70,80,90,100)
> sd(x)
```

output

```
[1] 21.60247
```

Coefficient of variation

R-code

```
> y<-c(40,41,45,49,50,51,55,59,60,60)
> mean(y)
[1] 51
> sd(y)
[1] 7.483315
> cv=mean(y)/sd(y)*100
> cv
[1] 681.5162
> cvv=sd(y)/mean(y)*100
```

Output

```
> cvv
[1] 14.67317
```

FITTING OF BINOMIAL DISTRIBUTION

Fit a binomial distribution for the following data

X	0	1	2	3	4	5	6	7
F	0	4	13	28	42	20	6	2

R CODING

```
x<-0:7
f<-c(0,4,13,28,42,20,6,2)
n<-max(x)
N<-sum(f)
smean<-sum(f*x)/sum(f)
p<-smean/n
px<-dbinom(0:6,n,p)
px
p7<-1-sum(px)
p7
px<-c(px,p7)
px<-round(px,7)
px
ex<-px*N
fr.dist<-data.frame(x,f,px,ex)
fr.dist
```

OUTPUT

```
>x<-0:7
> f<-c(0,4,13,28,42,20,6,2)
> n<-max(x)
> N<-sum(f)
> smean<-sum(f*x)/sum(f)
> p<-smean/n
> px<-dbinom(0:6,n,p)
> px

[1] 0.004585542 0.037176082 0.129169442 0.249335117 0.288774183
0.200670961
[7] 0.077470827
> p7<-1-sum(px)
> p7
[1] 0.01281785
> px<-c(px,p7)
> px<-round(px,7)
> px
[1] 0.0045855 0.0371761 0.1291694 0.2493351 0.2887742 0.2006710 0.0774708
[8] 0.0128178
> ex<-px*N
> fr.dist<-data.frame(x,f,px,ex)
> fr.dist

   x   f       px          ex
1  0   0   0.0045855   0.5273325
2  1   4   0.0371761   4.2752515
3  2  13   0.1291694  14.8544810
4  3  28   0.2493351  28.6735365
5  4  42   0.2887742  33.2090330
6  5  20   0.2006710  23.0771650
7  6   6   0.0774708   8.9091420
8  7   2   0.0128178   1.474
```

FITTING OF POISSON DISTRIBUTION

Find the poisson probability for the following data which gives the the frequency of the no.of horse kick in 10 carps per anum for 20 years

X	0	1	2	3	4
Y	109	65	22	3	1

R-coding

```
x<-0:4
f<-c(109,65,22,3,1)
fx<-f*x
smean<-sum(f*x)/sum(f)
px<-dpois(x,smean)
ex<-sum(f)*px
r<-round(ex,digits=0)
fr.dist<-data.frame(x,f,fx,px,ex,r)
fr.dist
```

Output:

	x	f	fx	px	ex	r
1	0	109	0	0.543350869	108.6701738	109
2	1	65	65	0.331444030	66.2888060	66
3	2	22	44	0.101090429	20.2180858	20
4	3	3	9	0.020555054	4.1110108	4
5	4	1	4	0.003134646	0.6269291	1

Fitting of normal distribution for the following data

Class	60-65	65-70	70-75	75-80	80-85	85-90	90-95	95-100
Freq	3	21	150	335	326	135	26	4

R-Coding

```
x1<-c(60,65,70,75,80,85,90,95)

f<-c(3,21,50,335,326,135,26,4)

x<-(x1+(x1+5))/2

fx<-f*x

xx<-x*x

fxx<-f*xx

fr.dist<-data.frame(x,f,fx,xx,fxx)

fr.dist

mean<-(sum(fx))/(sum(f))

mean

sd<-sqrt((sum(fxx)/sum(f))-(mean*mean))

sd

z<-((x1)-mean)/sd

z

A<-pnorm(x1,mean,sd)

A
```

Output

```
> x1<-c(60,65,70,75,80,85,90,95)
> f<-c(3,21,50,335,326,135,26,4)
> x<-(x1+(x1+5))/2
> fx<-f*x
> xx<-x*x
> fxx<-f*xx

> fr.dist<-data.frame(x,f,fx,xx,fxx)
> fr.dist
```

	x	f	fx	xx	fxx
1	62.5	3	187.5	3906.25	11718.75
2	67.5	21	1417.5	4556.25	95681.25
3	72.5	50	3625.0	5256.25	262812.50
4	77.5	335	25962.5	6006.25	2012093.75
5	82.5	326	26895.0	6806.25	2218837.50
6	87.5	135	11812.5	7656.25	1033593.75
7	92.5	26	2405.0	8556.25	222462.50
8	97.5	4	390.0	9506.25	38025.00

```
> mean<-(sum(fx))/(sum(f))
> mean
[1] 80.77222
> sd<-sqrt((sum(fxx)/sum(f))-(mean*mean))
> sd
[1] 5.108632
> z<-((x1)-mean)/sd
> z
[1] -4.0661030 -3.0873673 -2.1086316 -1.1298960 -0.1511603  0.8275754
1.8063110
[8]  2.7850467
> A<-pnorm(x1,mean,sd)
> A
[1] 2.390291e-05 1.009690e-03 1.748820e-02 1.292600e-01 4.399246e-01
[6] 7.960445e-01 9.645651e-01 9.973240e-01
```