R LAUGUAGE

DATA LOADING(TYPE 2)

```
a1=c(64,78,48,11,47,50,47,06,63,34)

a2=c (22,43,77,76,66,39,44,34,84,85)

a3=c (24,66,18,20,10,45,62,96,09,44)

a=cbind (a1, a2, a3)

summary (a)

boxplot (a)
```

DATA LOADING(TYPE 3)

MATRIAL STATUS DATA

```
status=matrix(c(180,210,70,40), nrow=1,ncol=4, byrow=T)

rownames (status)=c("Marital Status")

colnames (status)=c("Never married", "Married","Divorced","Widowed")

status

barplot (status, beside=T,main="Graphical Representation",

legend.text=rownames (status),col='green')
```

FREQUENCY COUNT

```
a=c(1,1,1,3,4,3,5,5,6,6,6,2,1,3,7,7,7,4,3,5,7,1,2,3,4,8)

b=table (a)

b

cbind (b)

plot(b,col='red',xlab='Numbers',ylab='Frequency',lwd=3)
```

HISTROGRAM

TYPE 1

```
a=rep(c(149.55,154.55,159.55,164.55,169.55,174.55,179.55,184.55),c(1,3,24,58,60,27,2,2))

H=hist (a, breaks=c(144.55,149.55,154.55,159.55,164.55,169.55,174.55,179.55,184.55))

H$counts= H$counts/(5*sum (H$counts))

plot (H, xlab="Height",ylab="Relative Frequency Density",main="Hisogram",ylim=c(0,0.1))

# With normal approximation

curve (dnorm (x,mean(a), sd(a)),col="blue", add=TRUE)
```

TYPE 2

```
a=rep(c(149.55,154.55,159.55,164.55,169.55,174.55,179.55, 184.55),c(1,3,24,58,60,27,2,2))

H=hist (a, breaks=c(144.55,149.55,154.55,159.55,164.55,169.55,174.55,179.55, 184.55))

H$counts= H$counts/5

plot (H,xlab="Height",ylab="Frequency Density", main="Hisogram",ylim=c(0,15))

# With normal approximation

b=length(a)

curve (b*dnorm(x,mean(a), sd(a)),col="blue", add=TRUE)
```

TYPE 3

```
a=rep(c(149.55,154.55,159.55,164.55,169.55,174.55,179.55,184.55),c(1,3,24,58,60,27,2,2))

H=hist(a,breaks=c(144.55,149.55,154.55,159.55,164.55,169.55,174.55,179.55,184.55))

plot(H,xlab="Height",ylab="Frequency",main="Hisogram",ylim=c(0,70))

b=length(a)

b1=min(diff (H$breaks))

curve(b*b1*dnorm(x,mean(a),sd(a)),col="blue", add=TRUE)
```

BOX PLOT

TYPE 1

```
a=c(31,71,69,59,70,74,41,87,37,46,56,34,62,15,53,81,76,24,21,8)

summary (a)

boxplot (a)

boxplot (a, col = "red")
```

TYPE 2

```
a=c(31,71,69,59,70,74,41,87,37,46,56,34,62,15,53,81,76,24,21,8)

b=c(10,23,55,60,62,68,53,78,60,87,78,93,94,90,65,88,76,96,86,66)

dat=cbind (a,b)

colnames(dat) =c("Marks A","Marks B")

boxplot (dat)

boxplot(dat,col =c("red","blue"))
```

STEM- LEAF PLOT

```
a=c(11,13,18,16,21,29,24,30,39,34,33)

stem (a)

b=c(11,13,18,16,21,29,24,30,39,34,33,48, 47,42,50,55,53,52,61,69,65,62)

stem (b)

stem (b, scale=2)
```

FREQUENCY POLYGON

```
a=c(0,1,2,3,4,5,6,7,8)

b=c (0,4,33,76,50,26,8,1,0)

plot (a,b,type="o",xlab="Number of Peas",ylab="Frequency")
```

PIE CHART

TYPE 2

```
states=c("Punjab","Haryana", "Tamil Nadu","Andhra Pradesh", "Uttar Pradesh", "Others")

procurement=c(5486,1248,589,3987,1296, 1654)

rice.proc=data.frame(states,procurement)

rice.proc=transform(rice.proc,percentage=(rice.proc$procurement/14260)*100)

rice.proc

attach(rice.proc)

pie(procurement,labels=c("Punjab(38.47%)", "Haryana (8.75%)", "Tamil Nadu(4.13%)",

"Andhra Pradesh (27.96%)", "Uttar Pradesh(9.09%)","Others(11.60%)"),

main="Procurement of Rice('000 tonnes)\nduring oct. 1993 to Sept. 1994",

col=c("royalblue4","orchid","chocolate4","olivedrab3","orange3","purple 1","violetred4"))

detach(rice.proc)
```

PLOT OGIVES

```
Height=c(149.55,154.55,159.55,164.55,169.55,174.55,179.55,184.55)

a = c(1,3,24,58,60,27,2,2)

b=cumsum (a)

plot (Height,b, type="o",ylab="Cummulative Frequency", col="blue",xlim = c(144,185))

a1=rev(a)

b1=cumsum (a1)

b2=rev (b1)
lines (Height-5,b2, type="o",col="red")
```

DESCRIPTIVE STATISTICS

TYPE1

```
x=c(70,48,84,21,84,33,64,36,61,50,27,87,41,65,33)

max (x)-min(x) # Range

var (x) # Variance

sd(x) # Standard Deviation

sum (abs (x-40))/length(x) # Mean deviation about 50

sum (abs (x-mean (x)))/length(x) # Mean deviation about mean

quantile (x) # Quartiles

IQR=67.5-64.5 # Using Quartile Data

(sd(x)/mean (x))*100 # Coefficient of Variation
```

TYPE2

```
x=c (60,39,75,33,69,47,69,32,59,52,49,63)

r=3 # r=1,2,3,4

sum (x^r)/length (x) # Raw Moments

sum ( (x-mean (x))^r)/length(x) # Central Moments
```

CORRELATION

```
x=c(126,157,153,156,152,135,145,132,153,143,165,132,180,161,170,176,165,163,157,180)

y=c(292,367,361,366,343,326,331,307,343,322,363,296,412,377,394,384,370,365,349,403)

z=data.frame(x,y)

colnames (z)=c("Marks in School", "Marks in College")

plot (z,col="blue") # Scatter Plot

cor (x,y) # Correlation Coefficient

cor (z) # Correlation Matrix
```

REGRESSION

TYPE1

```
x=c(126,157,153,156, 152,135,145,132, 153,143,165,132,180,161,170,176,165,163,157,180)

y=c(292,367,361,366,343,326,331,307,343,322,363,296,412,377,394,384,370,365,349,403)

fit=lm(y~x)

summary(fit)

plot (x,y,xlab = 'Marks in School', ylab='Marks in College',col='blue',cex=1.2,

pch=16,main='Regression Fit')

abline(fit,lwd=2,col='green’)

x1 = data.frame (x=seq(min(x),max(x),0.5))

y1=predict (fit, interval="conf",newdata=x1)

y2=predict (fit, interval="pred",newdata=x1)

matlines(x1$x,y1,lty=c(1,2,2),col=c("green", "red","red"),lwd=c(2,1,1))

matlines(x1$x,y2,lty=c(1,3,3),col=c("green", "black","black"),lwd=c (2,1,1))
```

TYPE2

```
x=c(107,110,121,124,129,132,131,153,142,151,158,152,164,152,162,156,175,178,172,177,174,183,
172,176)

y=c(99,82,102,81,89,87,59,112,74,91,116,91,96,85,131,136, 162,121,137,82,165,159,91,193)

fit=lm(y~x + I(x^2)) # Polnomial Fit

summary (fit)

plot (x,y,xlab ='X',ylab='Y',ylim=c(20,200), col='blue',main='Polynomial Fit',

cex=1.2,pch=16)

x1 = data.frame(x=seq(min(x), max (x),0.5))

lines(x1$x,predict(fit,newdata=x1),col="green",lwd=2)

y1=predict (fit,interval="conf",newdata=x1)

y2=predict (fit,interval="pred",newdata=x1)

matlines(x1$x,y1,lty=c(1,2,2),col=c("green", "red", "red"),lwd=c(2,1,1))

matlines(x1$x,y2,lty=c(1,3,3),col=c("green", "black", "black"),lwd=c(2,1,1))
```

ESTIMATE

```
x1=rnorm(100, 0, 1) # Generate 100 Random Samples from N(0,1)

xbar1=mean (x1)

sd1=sd(x1) # var (sd) is sample variance(sd) in R with denominator n-1

x2=rnorm(500, 0, 1) # Generate 500 Random Samples from N(0,1)

xbar2=mean (x2)

sd2=sd(x2)

x3=rnorm(5000, 0, 1) # Generate 5000 Random Samples from N(o,1)

xbar3=mean (x3)

sd3=sd(x3)

hist(x3,col='green',main="Histogram") # Histogram for sample size 5000
```

Exponential curve fitting

```
x=c(0,.01,.03,.05,.07,.09,.11,.13,.15,.17,.19,.21)

y=c(1,1.03,1.06,1.38,2.09,3.54,6.41,12.6,22.1,1.39,65.32,99.78)

y1=log(y)

fit=lm(y1~x)

a=exp(fit$coefficients[1])

b=fit$coefficients[2]

summary(fit)

plot(x,y,main = "Exp Fit")

x1=seq(min(x),max(x)+0.1,0.01)

y1=a*exp(b*x1)

lines(x1,y1,lty=1,col="blue",lwd=2)
```

NORMAL DISTRIBUTION GRAPH

```
x=seq(-5,5,0.1)

f=dnorm(x,mean = 0.4,sd=1.2)

plot(x,f,type="l",xlim = c(min(x),max(x)),ylim = c(0,1),lwd=2,col="blue",ylab="f(x)" )

x=seq(-5,5,0.1)

f=dnorm(x,mean = 0.4,sd=1.2)

f1=dnorm(x,mean = 0.4,sd=0.6)

f2=dnorm(x,mean = 0.4,sd=1.8)

plot(x,f,type="l",xlim = c(min(x),max(x)),ylim = c(0,1),lwd=2,col="blue",ylab="f(x)" )

lines(x,f1,type="l",xlim = c(min(x),max(x)),ylim = c(0,1),lwd=2,col="red")

lines(x,f2,type="l",xlim = c(min(x),max(x)),ylim = c(0,1),lwd=2,col="green")

legend(3,0.8,legend =
c("N(0.4,1.2)","N(0.4,0.6)","N(0.4,1.8)",col=c("blue","red","green"),lty=c(1,1,1)))
```

BINOMIAL CURVE FITTING

```
x=c(0,1,2,3,4,5,6,7,8,9)

f=c(7,33,54,38,35,15,7,4,1,1)

n=sum(f)

dat=rep(x,f)

hist(dat,col="blue",xlab="x",main="Histogram") #Idea about the distribution

lam=mean(dat) #estimate of lambda=bar(x)

f1=f*0 #Initialize expected frequency

f1[1]=exp(-lam) #Initial value of e^(-lam)

for(i in 1:8)

{

  f1[i+1]=(lam/i)*f1[i]

}

f1[10]=1-sum(f1) #adjusting total probability

Ef=n*f1 #expected frequency

out=data.frame(x,Ef,f)

colnames(out) = c("|x|","|Expected Frequency|","|Observed Frequency|")

format(out,scientific=F,digit=3)
```

BINOMIAL PLOT DISTRIBUTION

```
n=10

p=0.2

x=0:n

f=dbinom(x, n, p, log = FALSE)

plot (x,f,type="h",xlim =

c(min(x),max(x)),ylim =

c(0,1),lwd=2, col="blue",ylab = "p(X=x")

points(x,f,pch=16,cex=1,col="dark red")
```

NORMAL CURVE FITTING

```
x=seq (146.5, 181.5,5) # Midvalues

f=c(1,3,24,58,60,27,2,2)

n=sum (f)

dat=rep(x,f)

hist(dat,xlab="Height",main="Histogram",ylim=c(0,max(f)+3)) # ylim needed for curve

xbar=mean (dat) # Estimate of mu = mean (x)

s=sd(dat)

curve (n*5*dnorm(x,xbar,s),col="blue", add=T) # Add curve

xi=seq(144,184,5) # Intervalvalues

x1= (xi-xbar)/s # Normalize

y1=dnorm(x1) # Standard normal value

F1=pnorm(x1) # Df values

F2=c (0,F1,1) # Extended DF, including -Inf, Inf

F3= diff (F2) # Actual Probability values

Ef=n+F3 # Expected frequency

x1=c(-Inf,x, Inf) # Extended x, including -Inf, Inf

f1=c (0,f,0) # Extended f, including -Inf, Inf

Out=data.frame (x1,Ef,f1)

 colnames (Out) =c("Ix(Mid) |", "|Expected Frequency ", "IObserved Frequency ")

format (Out, scientific = F,digit =3)
```

CIRCULAR SYSTEMATIC SAMPLING

```
N=90

n=17

r=sample(c(1:90),size=1) #One sample using SRS

r

k=N/n

k=5 #Check the above k and vrite the nearest integer

samp=seq(r,r+k*n,k) #Generate systematic sample

samp=samp%%90*1 #Use nodulo function i.e ramainder
```