LAB-7

Normal Distribution fitting and Plotting

***AIM:** Computing/plotting and visualising the following probability distributions*

About Normal Distribution:-

THE NORMAL DISTRIBUTION :

A random variable X is said to posses normal distribution with mean μ and variance σ^2 , if its probability density function can be expressed of the form,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad , \qquad -\infty < x < +\infty$$

The standard notation used to denote a random variable to follow normal distribution with appropriate mean and variance is, $X \sim N(\mu , \sigma^2)$

STANDARD NORMAL DISTRIBUTION :

If a random variable X follows normal distribution with mean μ and variance σ^2 , its transformation $Z = \frac{X-\mu}{\sigma}$ follows standard normal distribution (mean 0 and unit variance)

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad , \qquad -\infty < z < +\infty$$

The distribution function of the standard normal distribution

$$F(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx$$

R Synatax :-

R has four in built functions to generate normal distribution. They are described below.

```
dnorm(x, mean, sd)
pnorm(x, mean, sd)
qnorm(p, mean, sd)
rnorm(n, mean, sd)
```

Following is the description of the parameters used in above functions:

- ***x** is a vector of numbers.*
- ***p** is a vector of probabilities.*
- ***n** is number of observations(sample size).*
- ***mean** is the mean value of the sample data. It's default value is zero.*
- ***sd** is the standard deviation. It's default value is 1*

(I) Normal distribution computations and graphs

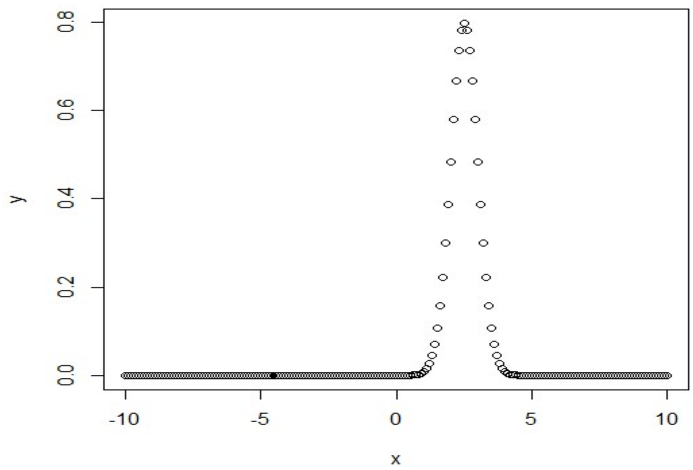
dnorm() :

This function gives height of the probability distribution at each point for a given mean and standard deviation.

CODE:-

```
Create a sequence of numbers between -10 and 10 incrementing by 0.1.
x <-seq(-10,10,by=.1)
x

# Choose the mean as 2.5 and standard deviation as 0.5.
y <-dnorm(x, mean=2.5,sd=0.5)
plot(x,y)
```

OUTPUT:



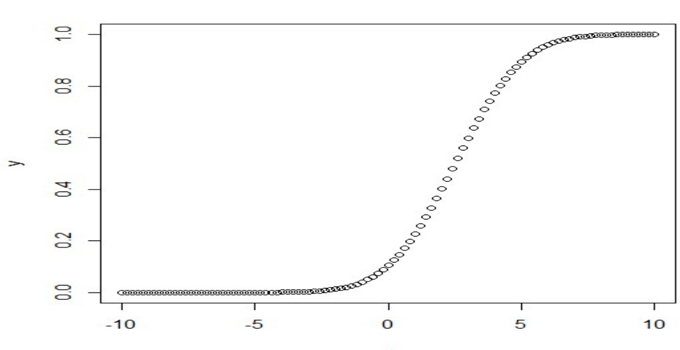
pnorm():

This function gives the probability of a normally distributed random number to be less that the value of a given number. It is also called "Cumulative Distribution Function".

CODE:-

```
# Create a sequence of numbers between -10 and 10 incrementing by 0.2.
x <-seq(-10,10,by=.2)
# Choose the mean as 2.5 and standard deviation as 2.
y <-pnorm(x,mean=2.5,sd =2)
# Plot the graph.
plot(x,y)
```

OUTPUT:



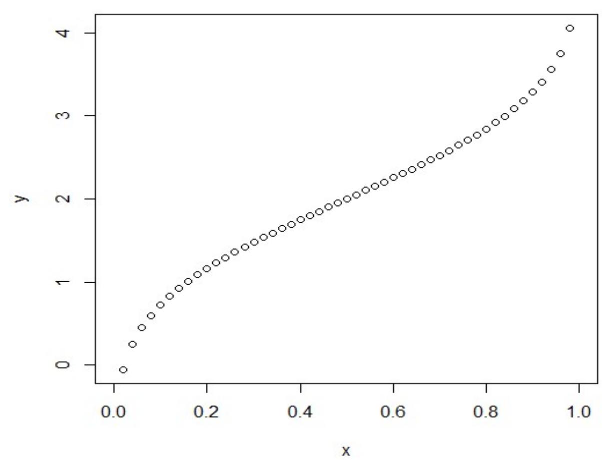
qnorm():-

This function takes the probability value and gives a number whose cumulative value matches the probability value.

CODE:-

```
# Create a sequence of probability values incrementing by 0.02.
x <-seq(0,1,by=0.02)
# Choose the mean as 2 and standard deviation as 3.
y <-qnorm(x,mean=2,sd=1)
# Plot the graph.
plot(x,y)
```

OUTPUT:-



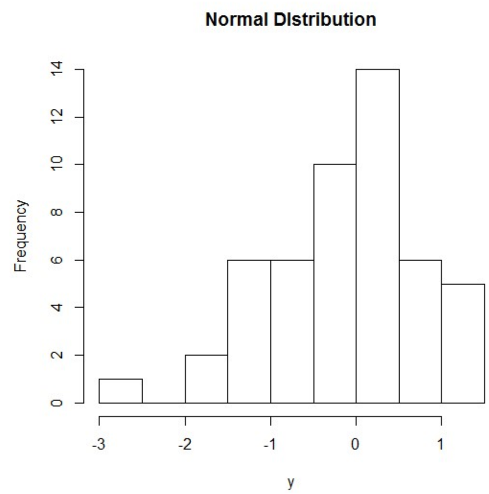
rnorm()

This function is used to generate random numbers whose distribution is normal. It takes the sample size as input and generates that many random numbers. We draw a histogram to show the distribution of the generated numbers.

CODE:-

```
# Create a sample of 50 numbers which are normally distributed.
y <-rnorm(50)
# Plot the histogram for this sample.
hist(y, main ="Normal DIstribution")
```

OUTPUT:-



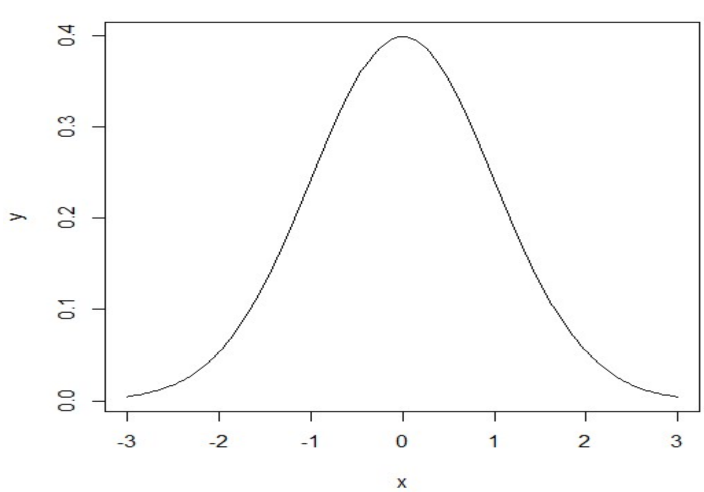
(II) Standard Normal Probability Distribution Plotting and Finding the Area:

CODE 1 :-

 # create a sequence of 200 numbers, beginning at x=-3 and ending at x=3

```
>x=seq(-3,3,length=200)
# use the dnorm command to compute the y-values of the standard normal
# probability density function (mean=0, standard deviation=1)
>y=dnorm(x,mean=0,sd=1)
>plot(x,y)
>plot(x,y,type="l")
```

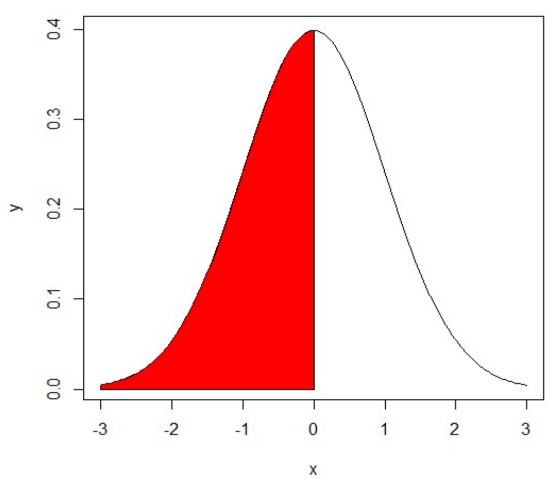
OUTPUT:-



CODE 2:

```
# Draw another normal curve, use a mean=50 and a standard deviation=10.

>x=seq(20,80,length=200)
>y=dnorm(x,mean=50,sd=10)
>plot(x,y,type="l")

# Find the area under the curve to left of the mean

>x=seq(-3,3,length=200)
>y=dnorm(x,mean=0,sd=1)
>plot(x,y,type="l")
>x=seq(-3,0,length=100)
>y=dnorm(x,mean=0,sd=1)
>polygon(c(-3,x,0),c(0,y,0),col="red")

# Find the area to the left of mean=0 (it should be 0.5)

>pnorm(0,mean=0,sd=1)
```

OUTPUT:-

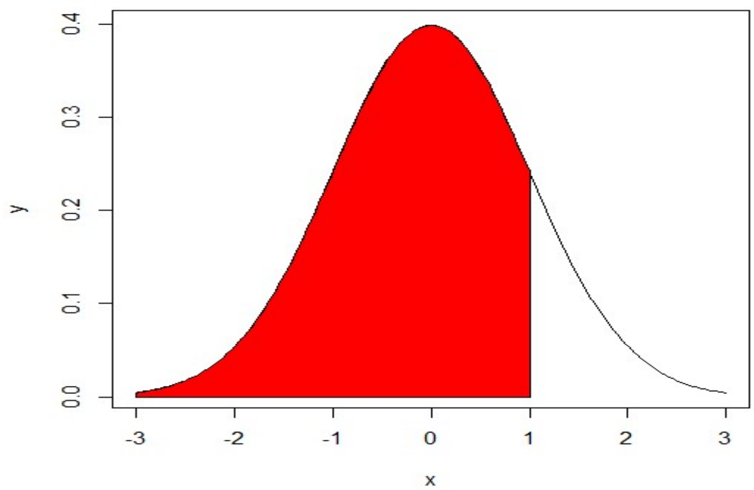


CODE 3:

 # Find the area to the left of 1. First, draw an image, then compute

```
>x=seq(-3,3,length=200)
>y=dnorm(x,mean=0,sd=1)
>plot(x,y,type="l")
>x=seq(-3,1,length=100)
>y=dnorm(x,mean=0,sd=1)
>polygon(c(-3,x,1),c(0,y,0),col="red")
>pnorm(1,mean=0,sd=1)
```

OUTPUT:

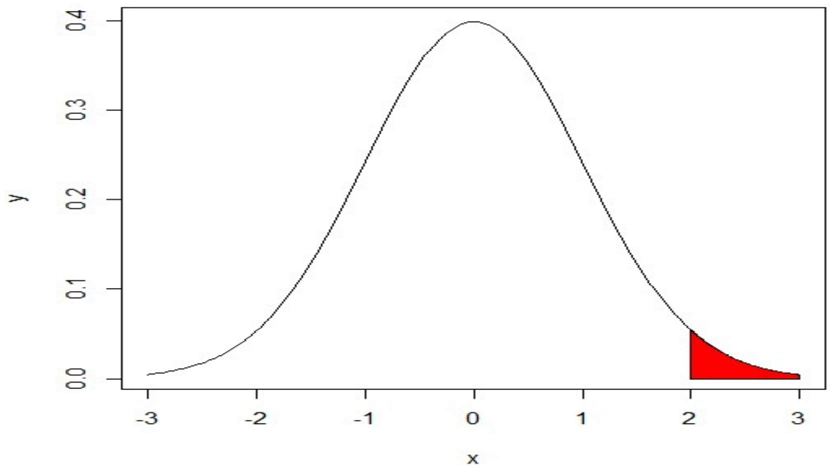


CODE 4:-

Get the area to the right of 2. First, draw an image, then compute

```
>x=seq(-3,3,length=200)
>y=dnorm(x,mean=0,sd=1)
>plot(x,y,type="l")
>x=seq(2,3,length=100)
>y=dnorm(x,mean=0,sd=1)
>polygon(c(2,x,3),c(0,y,0),col="red")
>1-pnorm(2,mean=0,sd=1)
```

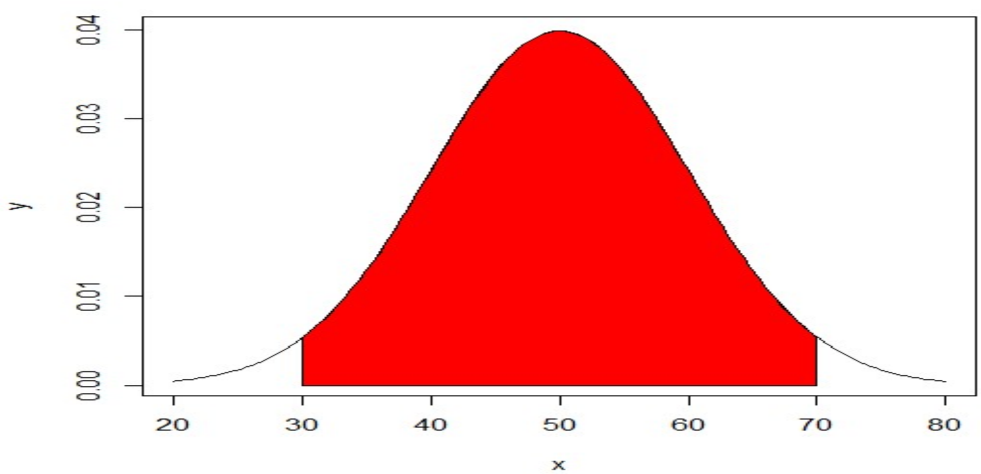
OUTPUT:-



CODE 5:

```
# Use the pnorm command to find areas under the normal density curve,
# regardless of the mean and standard deviation values.
# Example, mean=50 and standard deviation=10.

>x=seq(20,80,length=200)
>y=dnorm(x,mean=50,sd=10)
>plot(x,y,type="l")
>x=seq(30,70,length=100)
>y=dnorm(x,mean=50,sd=10)
>polygon(c(30,x,70),c(0,y,0),col="red")
>pnorm(70,mean=50,sd=10)-pnorm(30,mean=50,sd=10)
```

OUTPUT:

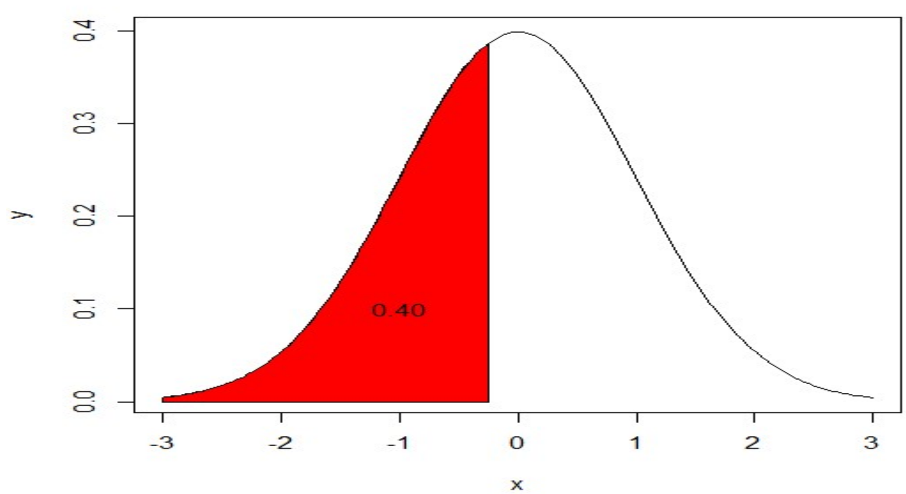


CODE 5:

Find the Quantile (Percentile) - i.e., reverse the process.
That is, given the area, find the value of x.

```
>x=seq(-3,3,length=200)
>y=dnorm(x,mean=0,sd=1)
>plot(x,y,type="l")
>x=seq(-3,-0.2533,length=100)
>y=dnorm(x,mean=0,sd=1)
>polygon(c(-3,x,-0.2533),c(0,y,0),col="red")
>text(-1,0.1,"0.40")
>qnorm(0.40,mean=0,sd=1)
```

OUTPUT:-



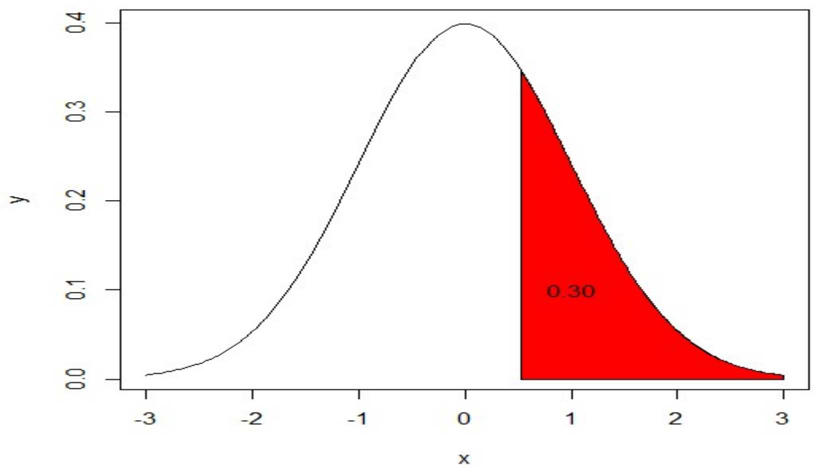
CODE 6:

density curve to the right of x is the given area.

```
>x=seq(-3,3,length=200)
>y=dnorm(x,mean=0,sd=1)
>plot(x,y,type="l")
>x=seq(0.5244,3,length=100)
>y=dnorm(x,mean=0,sd=1)
>polygon(c(0.5244,x,3),c(0,y,0),col="red")
>text(1,0.1,"0.30")
```

>qnorm(0.70,mean=0,sd=1)

OUTPUT:-



CODE 7:

1. Find $P(0 < Z < 1.24)$

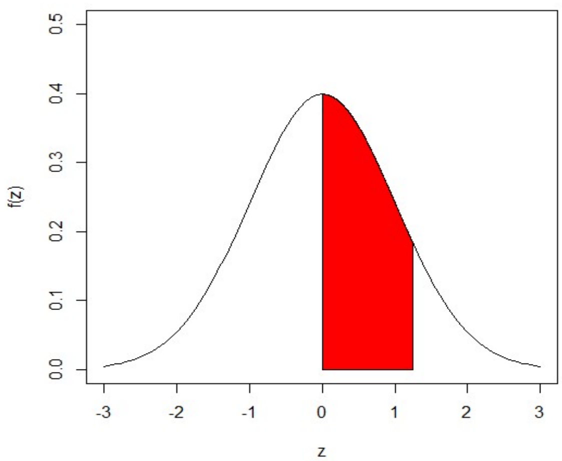
>pnorm(1.24) - pnorm(0)

[1] 0.3925123

Normal Probability Shape :

```
>plot.new()
>curve(dnorm, xlim = c(-3, 3), ylim = c(0, 0.5), xlab = "z", ylab="f(z)")
>zleft = 0
>zright = 1.24
> x = c(zleft, seq(zleft, zright, by=.001), zright)
> y = c(0, dnorm(seq(zleft, zright, by=.001)), 0)
>polygon(x, y, col="red")
```

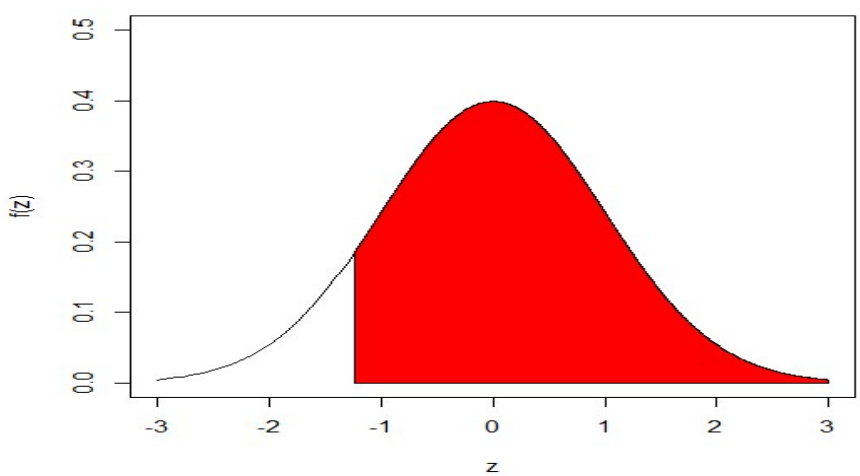
Output:-



2. Find $P(Z > -1.24)$

>1 - pnorm(-1.24)
>plot.new()
>curve(dnorm, xlim = c(-3, 3), ylim = c(0, 0.5), xlab = "z", ylab="f(z)")
>z = -1.24
>x = c(z, seq(z, 3, by=.001), 3)
>y = c(0, dnorm(seq(z, 3, by=.001)), 0)
>polygon(x, y, col="red")

Output:-

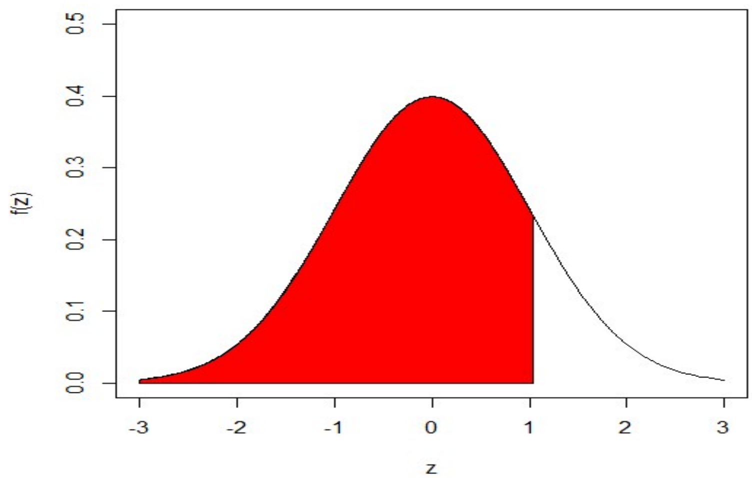


3. Find P_{85}, the 85_{th} percentile of the standard normal "z" distribution.

>qnorm(0.85)
[1] 1.036433
>plot.new()
>curve(dnorm, xlim = c(-3, 3), ylim = c(0, 0.5), xlab = "z", ylab="f(z)")

```
>prob = 0.85
> x = c(-3, seq(-3, qnorm(prob), by=.001), qnorm(prob))
> y = c(0, dnorm(seq(-3, qnorm(prob), by=.001)), 0)
>polygon(x, y, col="red")
```

Output:-

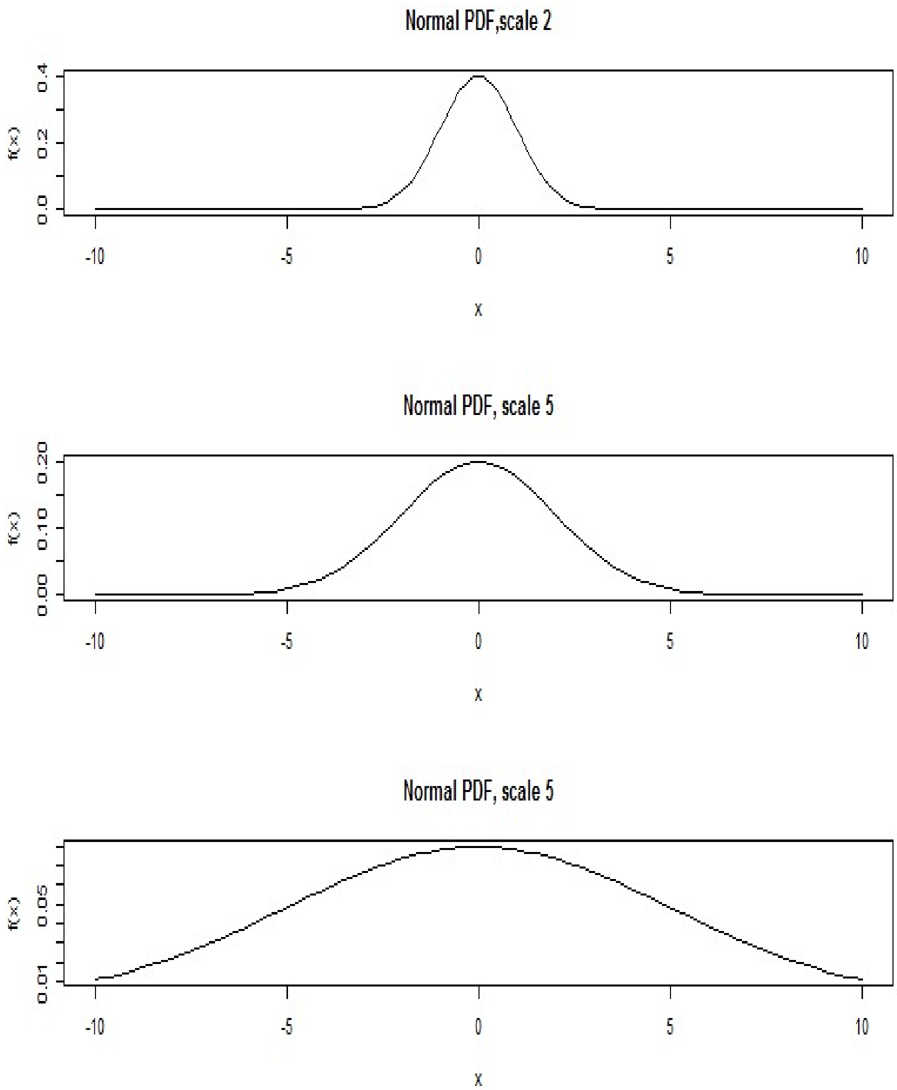


CODE 8:

(a) Plot a normal density for a range of x from –10 to 10 with mean 0 and standard deviation 1:{This Problem Explains types of kurtosis by changing standard deviation}

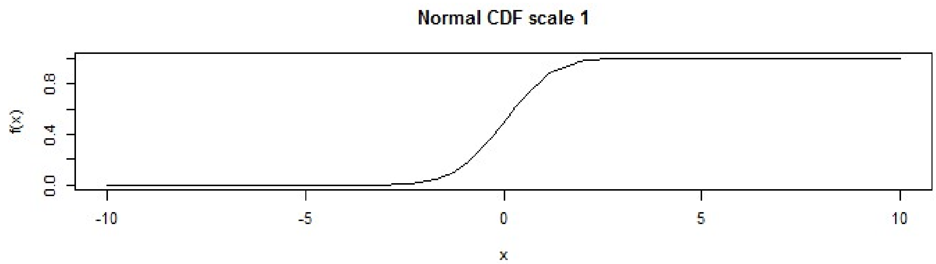
```
> x<-seq(-10,10,length=100)
>plot(x,dnorm(x,0,1),xlab="x", ylab="f(x)", type='l', main="Normal PDF")
>par(mfrow=c(3,1))
>plot(x,dnorm(x,0,1),xlab="x",ylab="f(x)", type='l', main="Normal PDF,scale 2")
>plot(x,dnorm(x,0,2),xlab="x",ylab="f(x)", type='l',main="Normal PDF, scale5")
>plot(x,dnorm(x,0,5),xlab="x",ylab="f(x)", type='l',main="Normal PDF, scale5")
```

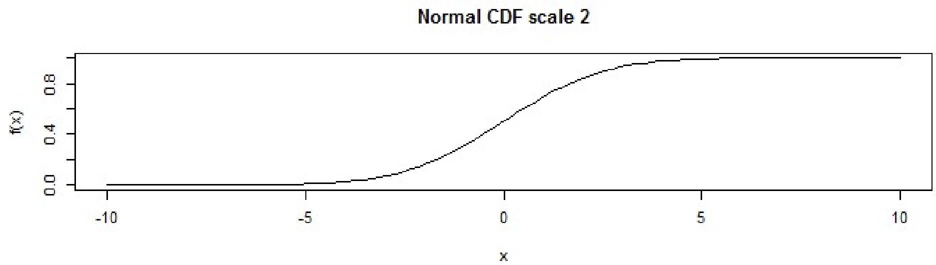
OUTPUT:-

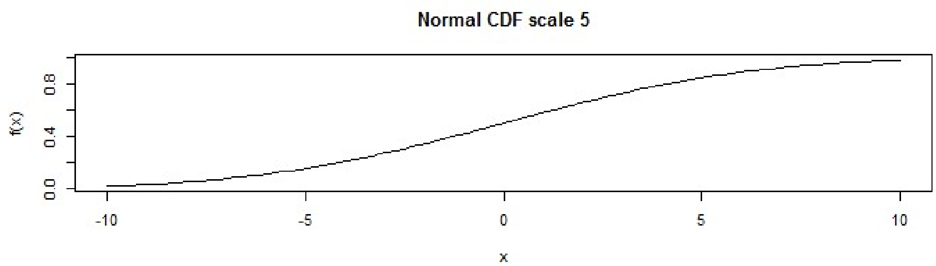


(b) Normal distribution Cumulative Distribution Function with different scale parameters

```
>par(mfrow=c(3,1))
>plot(x,pnorm(x,0,1),xlab="x",ylab="f(x)", type='l', main="Normal CDF scale 1")
>plot(x,pnorm(x,0,2),xlab="x", ylab="f(x)", type='l', main="Normal CDF scale 2")
>plot(x,pnorm(x,0,5),xlab="x", ylab="f(x)", type='l', main="Normal CDF scale 5")
```







CODE 9:

Problem :In a photographic process the developing times of prints may be looked upon as a random variable having the normal distribution with a mean of 16.28 seconds and a standard deviation 0.12 second. Find the probability that it will take

(i) Atleast 16.20 seconds to develop one of the prints;
(ii) atmost 16.35 seconds to develop one of the prints

Ans) *Atleast 16.20 seconds to develop one of the prints;*
Print developing time : $X \sim N(16.28, (0.12)^2)$

(Solution).(i) Required event : $[X \geq 16.20]$

$$P[X \geq 16.20] = P\left[\frac{X - 16.28}{0.12} \geq \frac{16.20 - 16.28}{0.12}\right] = P[Z \geq -0.6667]$$

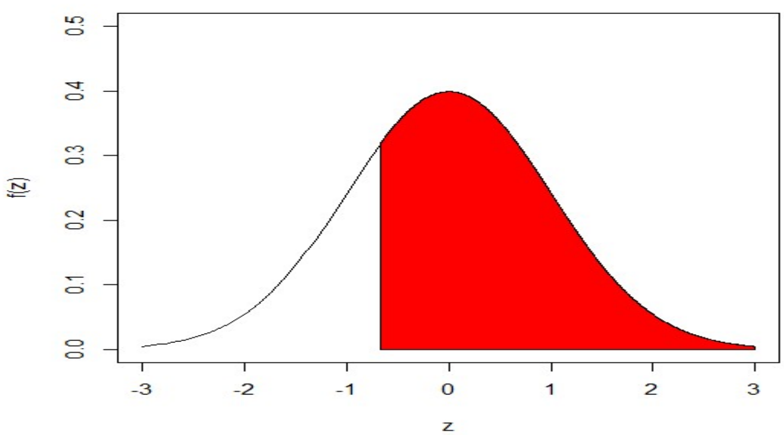
$$P[X \geq 16.20] = 0.7486 \text{(Dotted area)}$$

R Code:-

(i). $P[Z \geq -0.6667]$

```
>(1-pnorm(-0.6667))
[1] 0.7475181
```

```
> 1 - pnorm((-0.6667))
>plot.new()
>curve (dnorm, xlim = c(-3, 3), ylim = c(0, 0.5), xlab = "z", ylab="f(z)")
>z = -0.6667
>x = c(z, seq(z, 3, by=.001), 3)
>y = c(0, dnorm(seq(z, 3, by=.001)), 0)
>polygon(x, y, col="red")
```



(ii) *Required event : [X ≤ 16.35]*

$$P[X \leq 16.20] = P\left[Z \leq \frac{16.35 - 16.28}{0.12}\right]$$

$$= P[X \leq 0.5833] \text{(dotted area)}$$

= 0.5 + 0.2190 = 0.7190

$$P[X \leq 16.35] = 0.7190$$

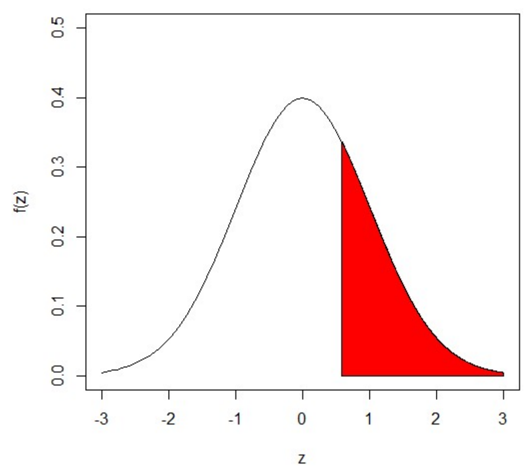
R code:-

>pnorm(0.5833)

[1] 0.7201543

```
>pnorm(0.5833)
>plot.new()
>curve (dnorm, xlim = c(-3, 3), ylim = c(0, 0.5), xlab = "z", ylab="f(z)")
>z = 0.5833
>x = c(z, seq(z, 3, by=.001), 3)
>y = c(0, dnorm(seq(z, 3, by=.001)), 0)
>polygon(x, y, col="red")
```

Output:-



(III) GENERAL NORMAL PROBABILITY DISTRIBUTIONS

1. *Suppose X is normal with mean 527 and standard deviation 105. Compute P(X ≤310).*

   ```
   >pnorm(310,527,105)
   [1] 0.01938279
   ```

2. *If* $X \sim N$ (μ = 100 pts., σ= 15 pts.)

(i) Find $P(80 \text{ pts.} < X < 95 \text{ pts.})$

```
>pnorm(95, mean=100, sd=15) - pnorm(80, mean=100, sd=15)
[1] 0.2782301
```

(ii) Find $P(X > 125 \text{ pts.})$.

```
> 1 - pnorm(125, mean=100, sd=15)
[1] 0.04779035
```

(iii) Find P_{75}, the 75_{th} percentile of the above distribution. This is the same as the 0.75 quantile.

```
>qnorm(0.75, mean=100, sd=15)
[1] 110.1173
```

3. *The weekly wages of 1000 workmen are normally distributed around a mean of Rs. 70 with S.D of Rs 5.Estimate the number of workers whose weekly wages will be*
   - *(i) Between Rs 69 and Rs 72*
   - *(ii) Less than Rs 69*
   - *(iii) More than Rs 72*

(i) Between Rs 69 and Rs 72

```
>(pnorm(72, mean=70, sd=5) - pnorm(69, mean=70, sd=5))*1000
[1] 234.6815
```

The number of workers whose wages lies between Rs.69 and Rs.72 is 234

(ii) Less than Rs 69

```
> (pnorm(69, mean=70, sd=5))*1000
[1] 420.7403
```

The number of workers whose wages is less than Rs.69 is 421

(iii) More than Rs 72

```
> (1 - pnorm(72, mean=70, sd=5))*1000
[1] 344.5783
```

The number of workers whose wages is More than Rs.72 is 345

4. *In a test on 2000 Electric bulbs ,it was found that the life of particular make, was normally distributed with an average life of 2040 hours and S.D of 60 hours. Estimate the number of bulbs likely to burn for*
   - *(i) More than 2150 hours*
   - *(ii) Less than 1950 hours*
   - *(iii) More than 1920 hours but less than 2160 hours*
   - *(iv) More than 2150 hours*

R CODE:-

```
>(1 - pnorm(2150, mean=2040, sd=60))*2000
 [1] 66.75302
```

The number of bulbs expected to burn for more than 2150 hours is 67 (approximately)

(i) *Less than 1950 hours*

```
>(pnorm(1950, mean=2040, sd=60))*2000

[1] 133.6144
```

The number of bulbs expected to burn for less than 1950 hours is 134 (approximately)

(ii) *More than 1920 hours but less than 2160 hours*

```
>(pnorm(2160, mean=2040, sd=60) - pnorm(1920, mean=2040,
sd=60))*2000
[1] 1908.999
```

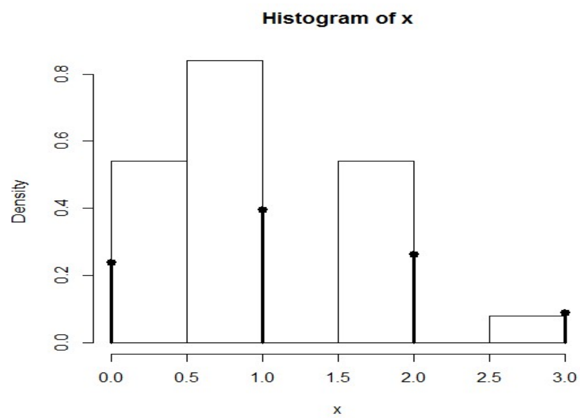
The number of bulbs expected to burn more than 1920 hours but less than 2160is 1909 (approximately)

(IV) BINOMIAL DISTRIBUTION TENDS TO NORMAL DISTRIBUTION AS 'n' TENDS TO INFINITY:

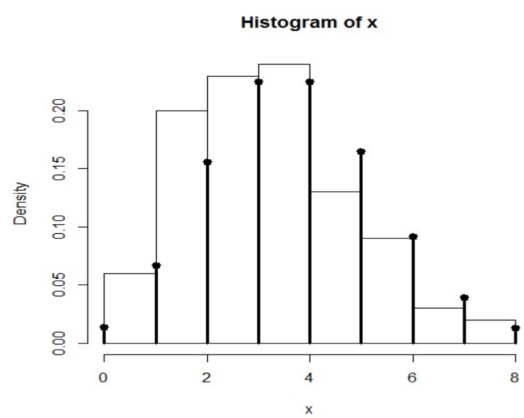
Binomial distribution tends to Normal distribution :------

```
> n=5;p=.25
> x=rbinom(100,n,p)
>hist(x,probability=TRUE,)
>## use points, not curve as dbinom wants integers only for x
>xvals=0:n;points(xvals,dbinom(xvals,n,p),type="h",lwd=3)
>points(xvals,dbinom(xvals,n,p),type="p",lwd=3)
```

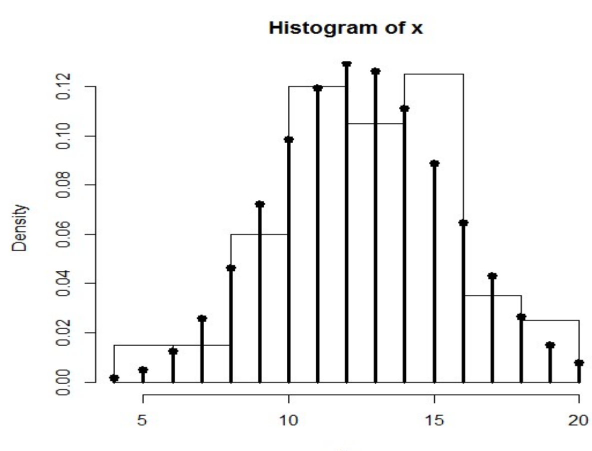
Output:-



n=15



n=50



(Note ;- In Statistics any probability distribution follows normal distribution as 'n' tends to 'infinity' with certain assumptions)

Practice Experiments:-

Standard Normal distribution:-

1. *Find (i)* $P(0.8 \le Z \le 1.5)$ *(ii)* $P(Z \le 2)$ *(iii)* $P(Z \ge 1)$
   Find These probability values and Plot the graph .

   General Normal Distribution.

2. *If mean=70 and Standard deviation is 16*

   i) $P(38 \le X \le 46)$ *ii)* $P(82 \le X \le 94)$ *iii)* $P(62 \le X \le 86)$

   Find the Probability values and Plot the graph with text.

   (***Standard Normal distribution or General Normal Distribution.)***

3. *1000 students had Written an examination the mean of test is 35 and standard deviation is 5.Assumning the to be normal find*
   i) How many students Marks Lie between 25 and 40
   ii) How many students get more than 40
   iii) How many students get below 20
   iv) How many students get 50