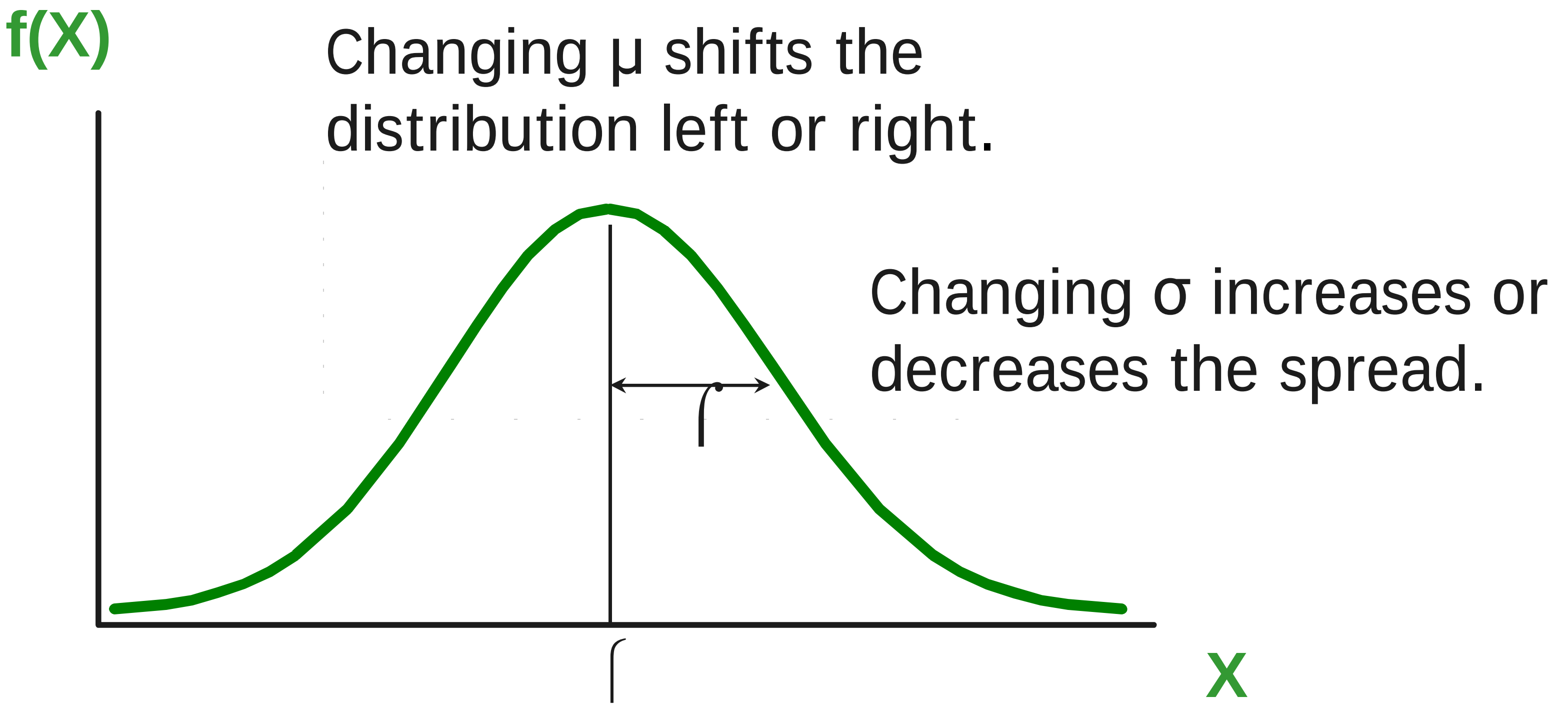
Normal Distribution

MAT2001-Statistics for Engineers

Introduction

- The normal distribution is the most important distribution. It describes well the distribution of random variables that arise in practice, such as the heights or weights of people, the total annual sales of a firm, exam scores etc. Also, it is important for the central limit theorem, the approximation of other distributions such as the binomial, etc.

The Normal Distribution



The Normal Distribution: as mathematical function (pdf)

$$f(x)=\frac{1}{\boxed{\sigma}\sqrt{2\pi}}\cdot e^{-\frac{1}{2}\left(\frac{x-\boxed{\mu}}{\sigma}\right)^2}$$

This is a bell shaped curve with different centers and spreads depending on μ and σ

<u>Note constants:</u>

π=3.14159

e=2.71828

The Normal PDF

It’s a probability function, so no matter what the values of μ and σ, must integrate to 1!

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1$$

Normal distribution is defined by its mean and standard dev.

E(X)=μ = $\boxed{\int_{-\infty}^{+\infty} x \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx}$

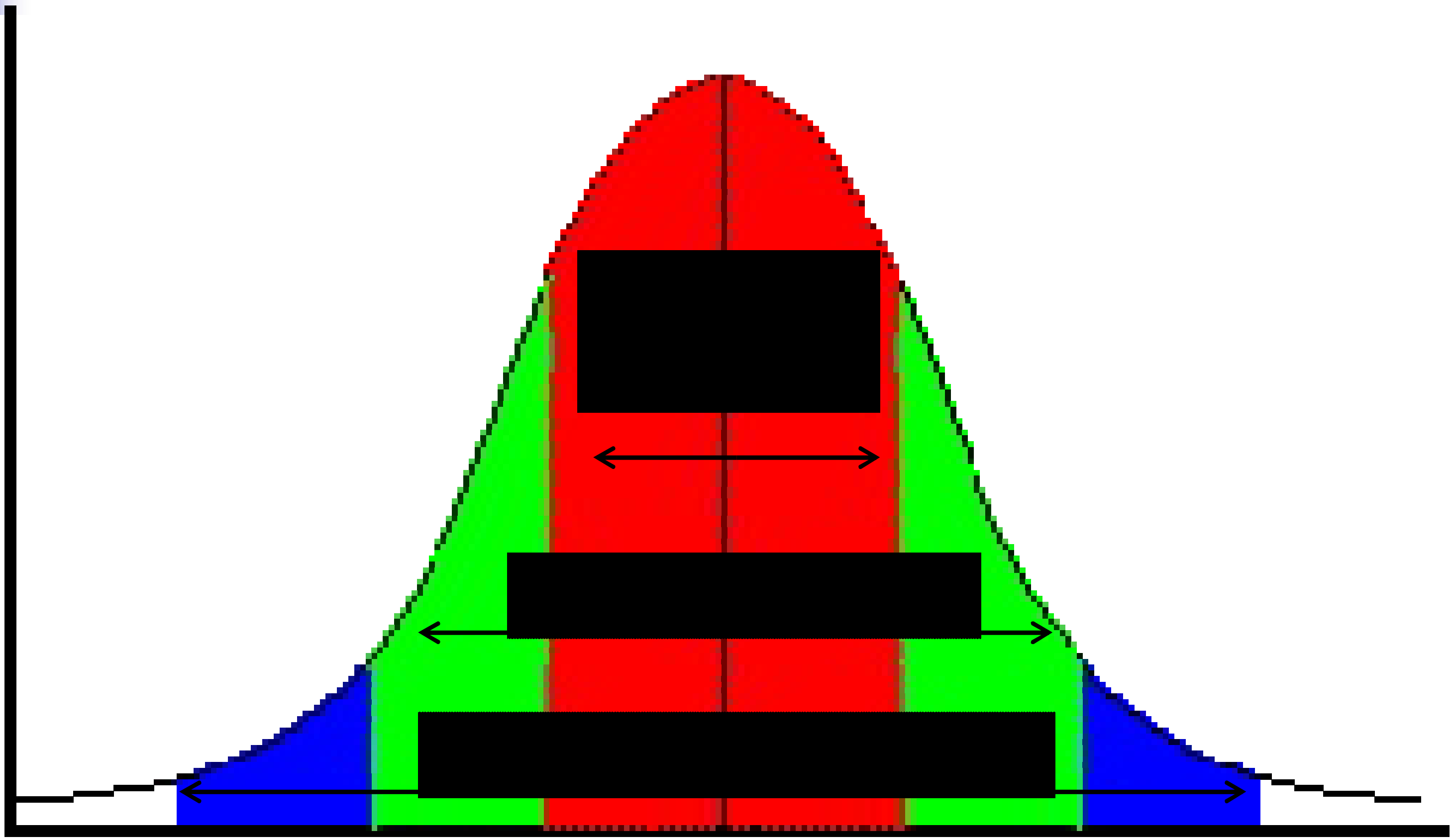
Var(X)=σ² = $\boxed{\int_{-\infty}^{+\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx) - \mu^2}$

Standard Deviation(X)=σ

**The beauty of the normal curve:

No matter what μ and σ are, the area between $\mu-\sigma$ and $\mu+\sigma$ is about 68%; the area between $\mu-2\sigma$ and $\mu+2\sigma$ is about 95%; and the area between $\mu-3\sigma$ and $\mu+3\sigma$ is about 99.7%. Almost all values fall within 3 standard deviations.

68-95-99.7 Rule



68-95-99.7 Rule in Math terms…

$$\int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{\sigma\sqrt{2\pi}} \bullet e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx = .68$$

$$\int_{\mu-2\sigma}^{\mu+2\sigma} \frac{1}{\sigma\sqrt{2\pi}} \bullet e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx = .95$$

$$\int_{\mu-3\sigma}^{\mu+3\sigma} \frac{1}{\sigma\sqrt{2\pi}} \bullet e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx = .997$$

R Syntax

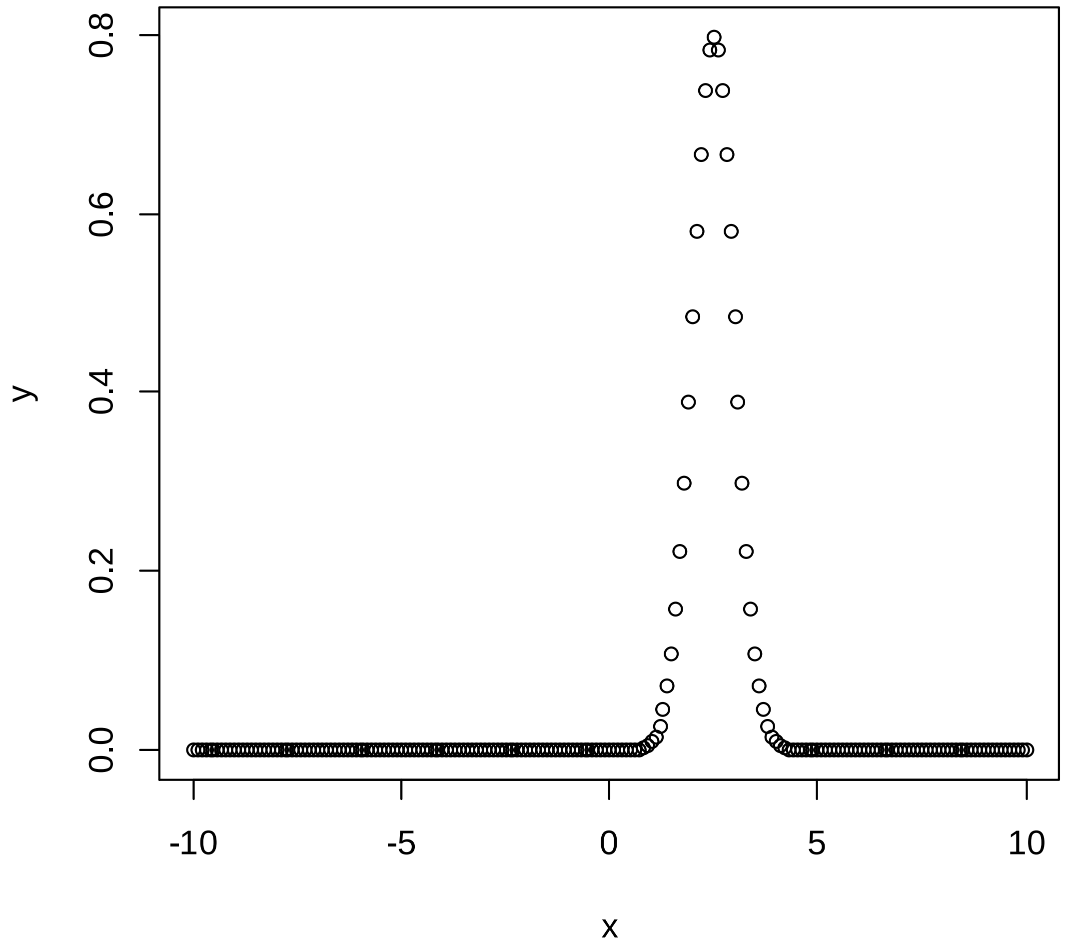
- R has four in built functions to generate normal distribution. They are described below.

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

dnorm()

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

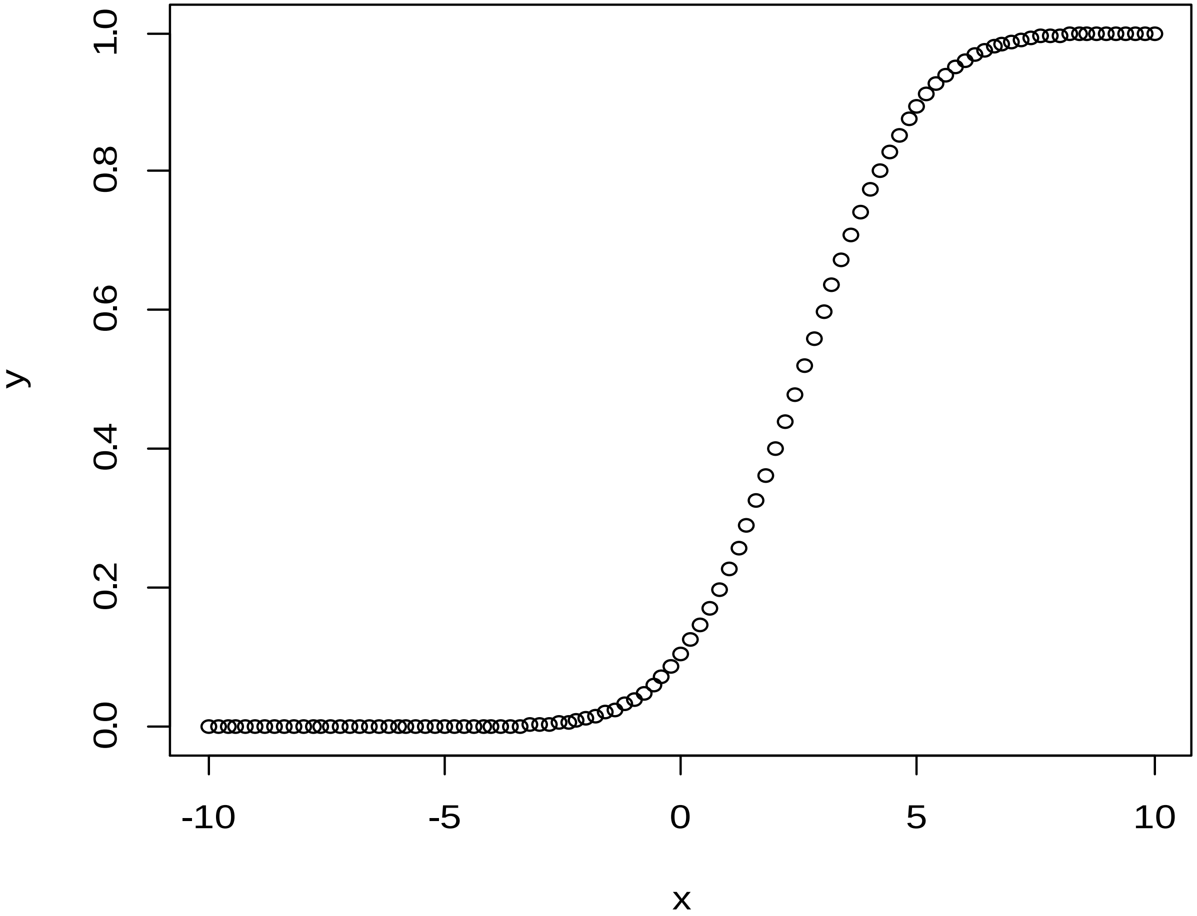
```
> # Create a sequence of numbers between -10 and 10 incrementing by 0.1.
> x <- seq(-10, 10, by = .1)
>
> # Choose the mean as 2.5 and standard deviation as 0.5.
> y <- dnorm(x, mean = 2.5, sd = 0.5)
>
> plot(x,y)
```


0.0
0.2
0.4
0.6
0.8
y
-10
-5
0
5
10
x

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"

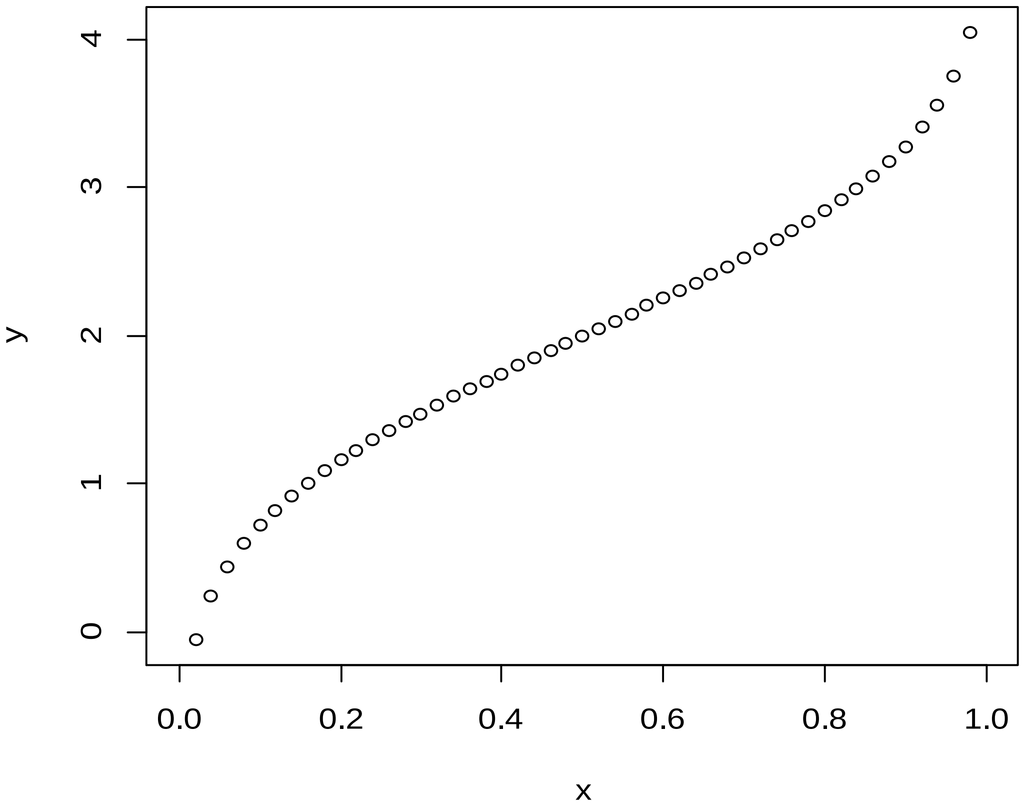
- > # Create a sequence of numbers between -10 and 10 incrementing by 0.2.
- > x <- seq(-10,10,by = .2)
- > # Choose the mean as 2.5 and standard deviation as 2.
- > y <- pnorm(x, mean = 2.5, sd = 2)
- > # Plot the graph.
- > plot(x,y)


1.0
0.8
0.6
0.4
0.2
0.0
y
-10
-5
0
5
10
x

qnorm()

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

- \# 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)



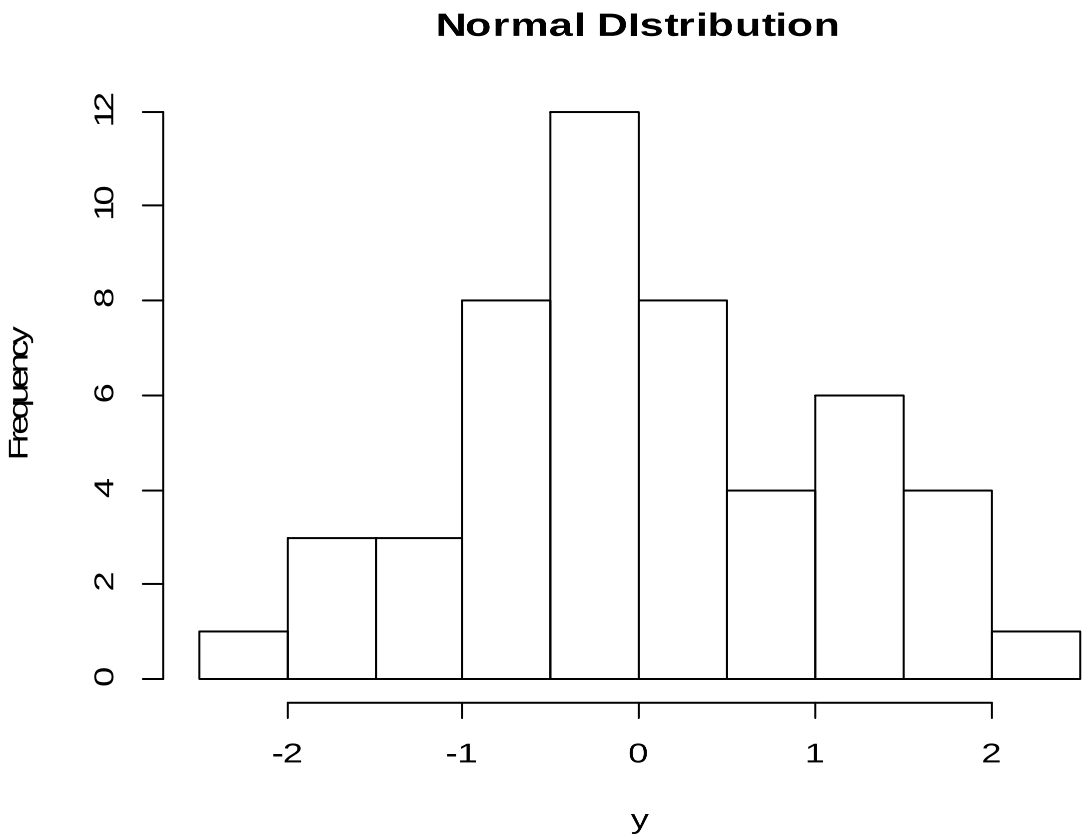
0
1
2
3
4
y
0.0
0.2
0.4
0.6
0.8
1.0
x

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

.

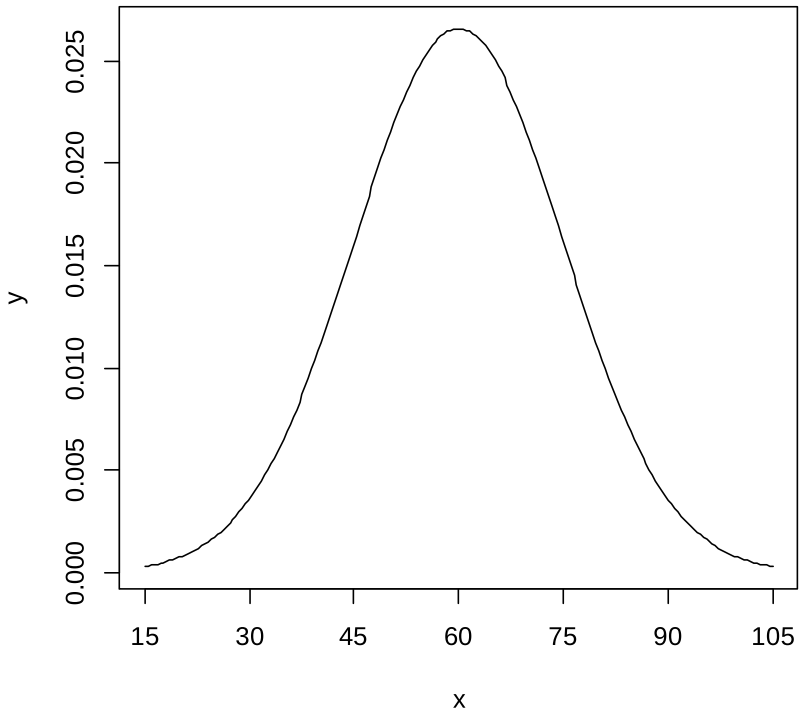
- > # Create a sample of 50 numbers which are normally distributed.
- > y <- rnorm(50)
- > # Plot the histogram for this sample.
- > hist(y, main = "Normal DIstribution")


Normal DIstribution
Frequency
0
2
4
6
8
10
12
-2
-1
0
1
2
y

Program-1

Draw a normal distribution with a mean=60 and a standard deviation=15.

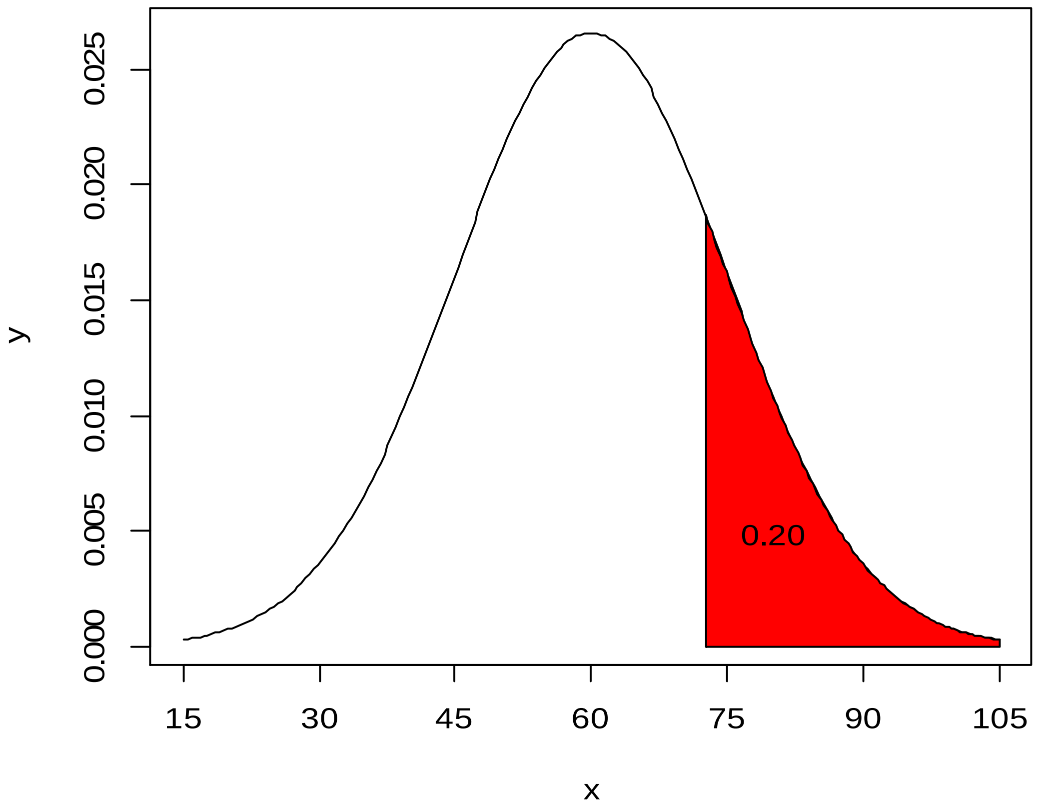
- >x=seq(15,105,length=200)
- >y=dnorm(x,mean=60,sd=15)
- >plot(x,y,type="l",xaxt="n")
- >axis(1,at=c(15,30,45,60,75,90,105))


0.000
0.005
0.010
0.015
0.020
0.025
y
15
30
45
60
75
90
105
x

Programme2:

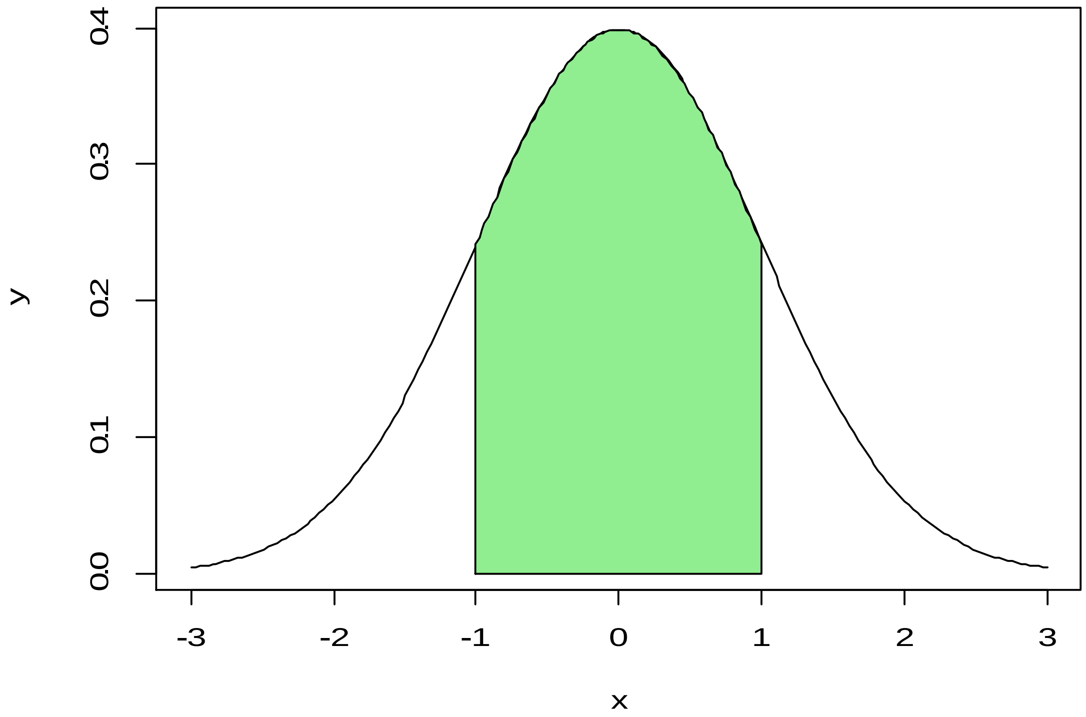
#shade the top 20% of the area under the normal density curve

```
>x=seq(15,105,length=200)
>y=dnorm(x,mean=60,sd=15)
>plot(x,y,type="l",xaxt="n")
>axis(1,at=c(15,30,45,60,75,90,105))
>x=seq(72.62,105,length=100)
>y=dnorm(x,mean=60,sd=15)
>polygon(c(72.62,x,105),c(0,y,0),col="red")
>text(80,0.005,"0.20")
```

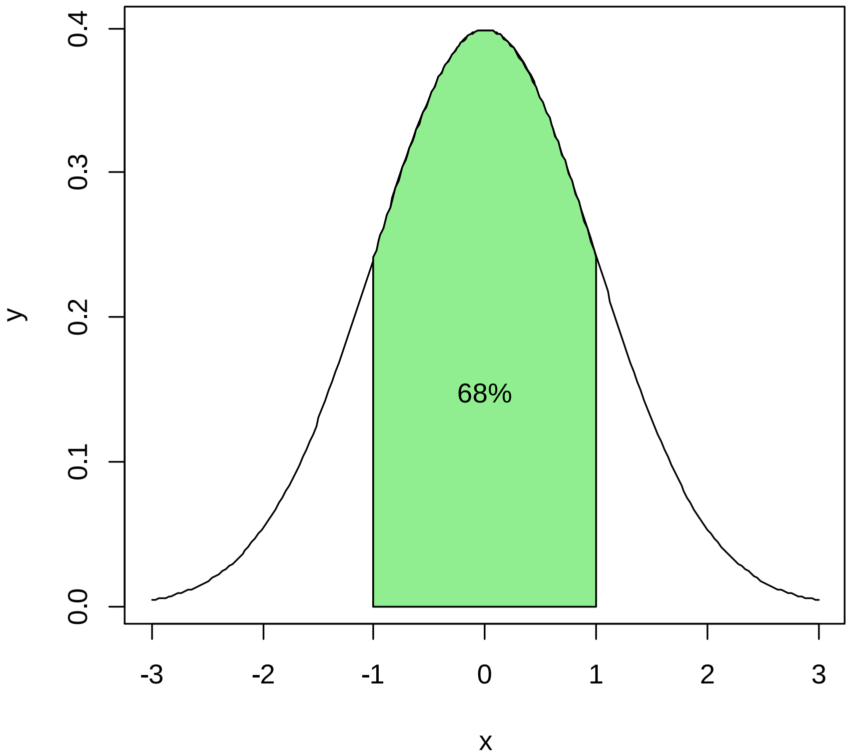

0.20
0.000
0.005
0.010
0.015
0.020
0.025
y
15
30
45
60
75
90
105
x

Programme 3
Simulate a standard normal density curve (mean=0 and standard deviation=1)

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

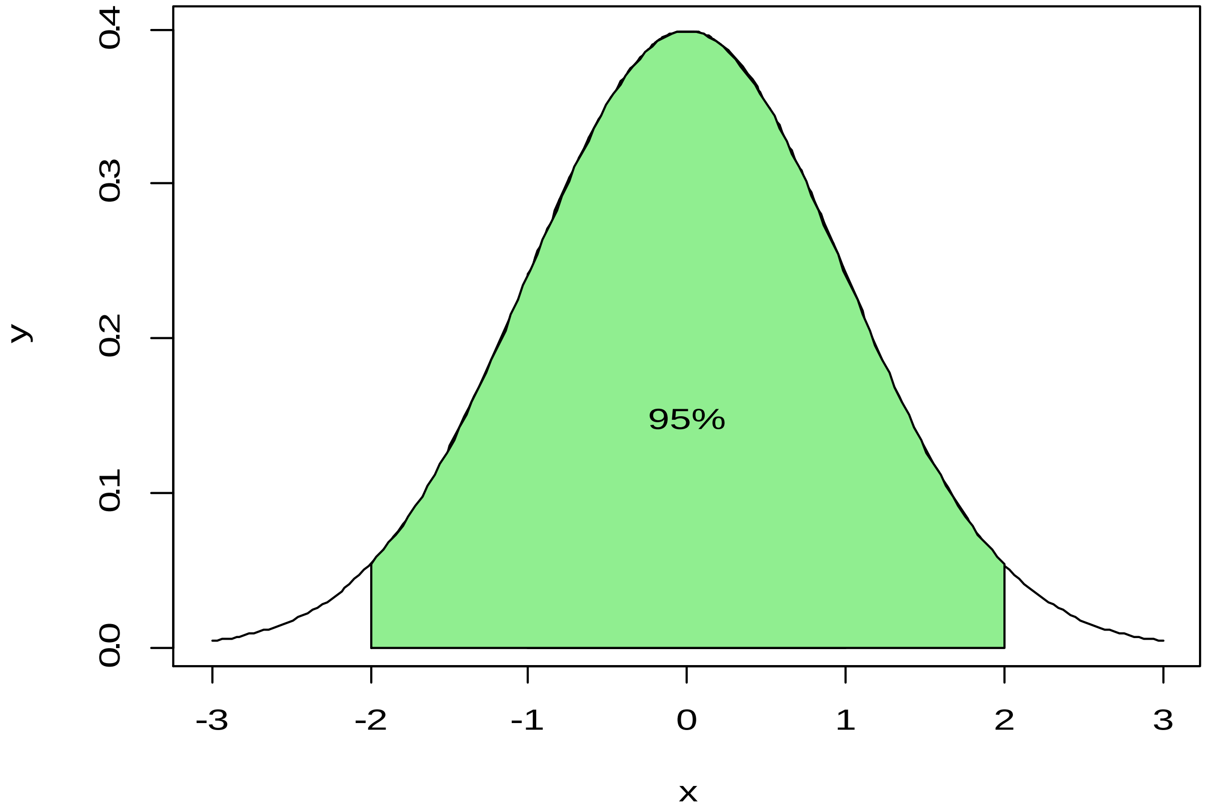

0.0
0.1
0.2
0.3
0.4
y
-3
-2
-1
0
1
2
3
x

- text(0,0.15,"68%")



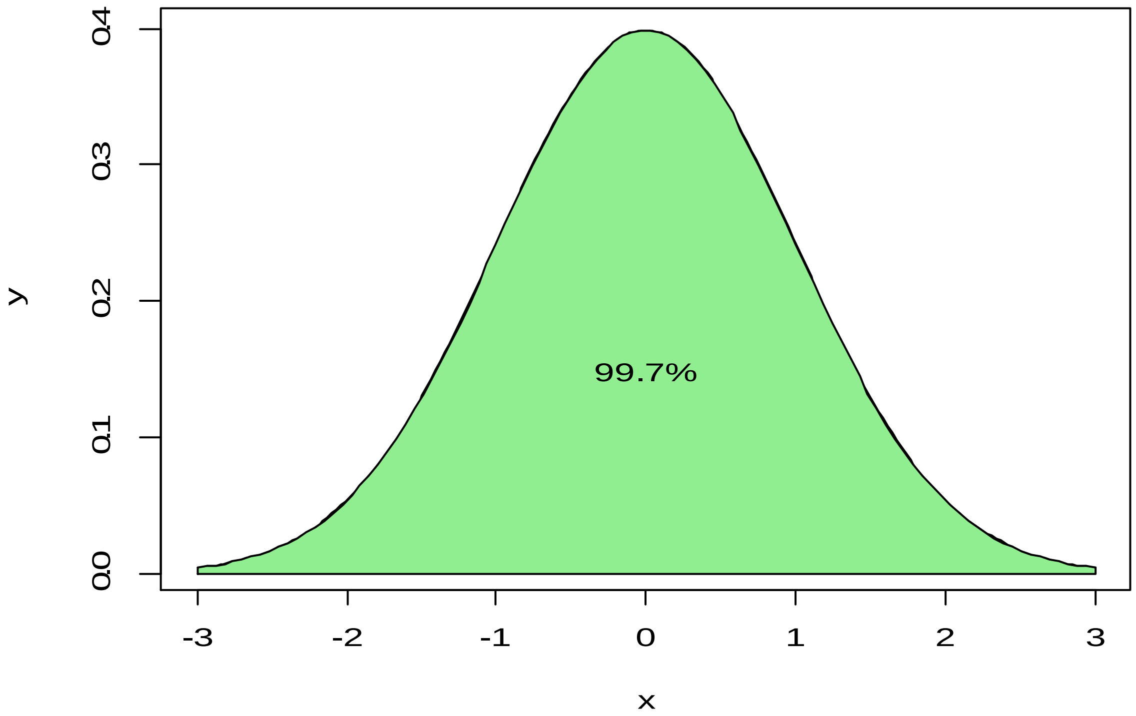
#Hence,approximately 95% of the area lies within two standard deviations of the mean.

```
>x=seq(-3,3,length=200)
>y=dnorm(x) plot(x,y,type="l")
>x=seq(-2,2,length=100)
>y=dnorm(x)
>polygon(c(-2,x,2),c(0,y,0),col="lightgreen")
>text(0,0.15,"95%")
```


95%
y
x
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3

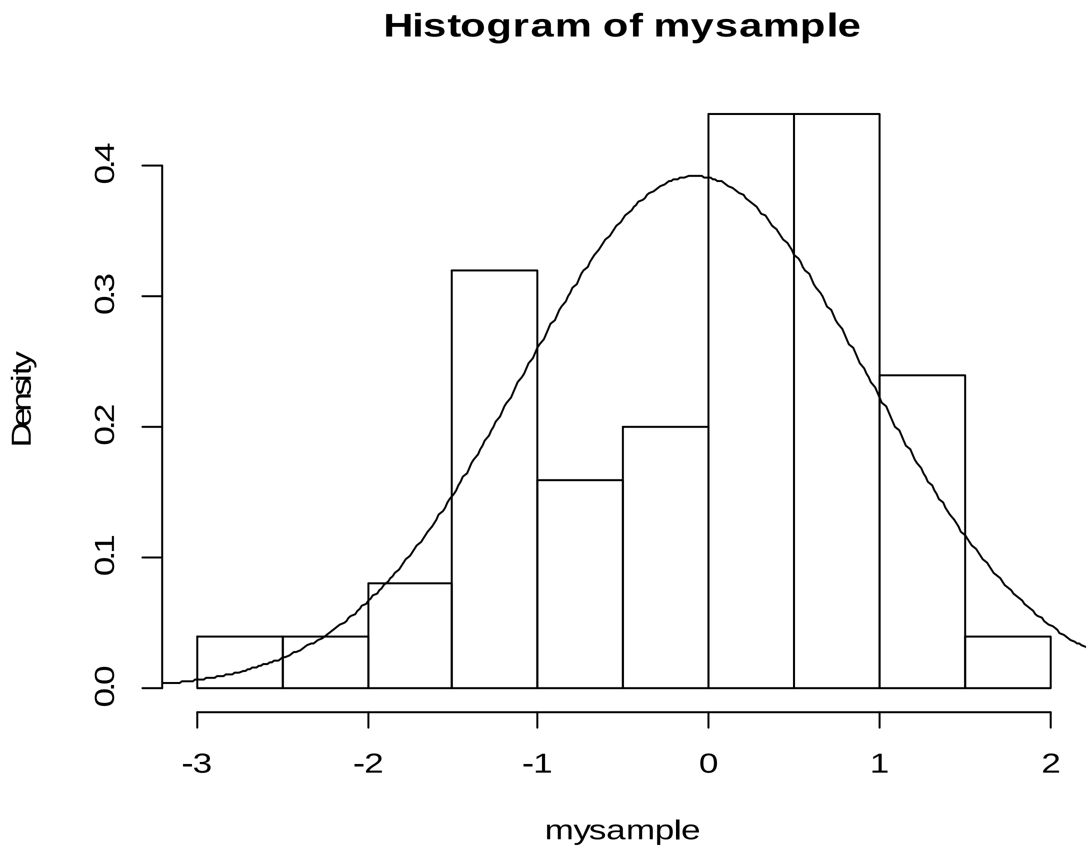
#Hence, approximately 99.7% of the area lies within three standard deviations of the mean.

```
> x=seq(-3,3,length=200)
> y=dnorm(x)
> plot(x,y,type="l")
> x=seq(-3,3,length=100)
> y=dnorm(x)
> polygon(c(-3,x,3),c(0,y,0),col="lightgreen")
> text(0,0.15,"99.7%")
```


99.7%
y
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3
x

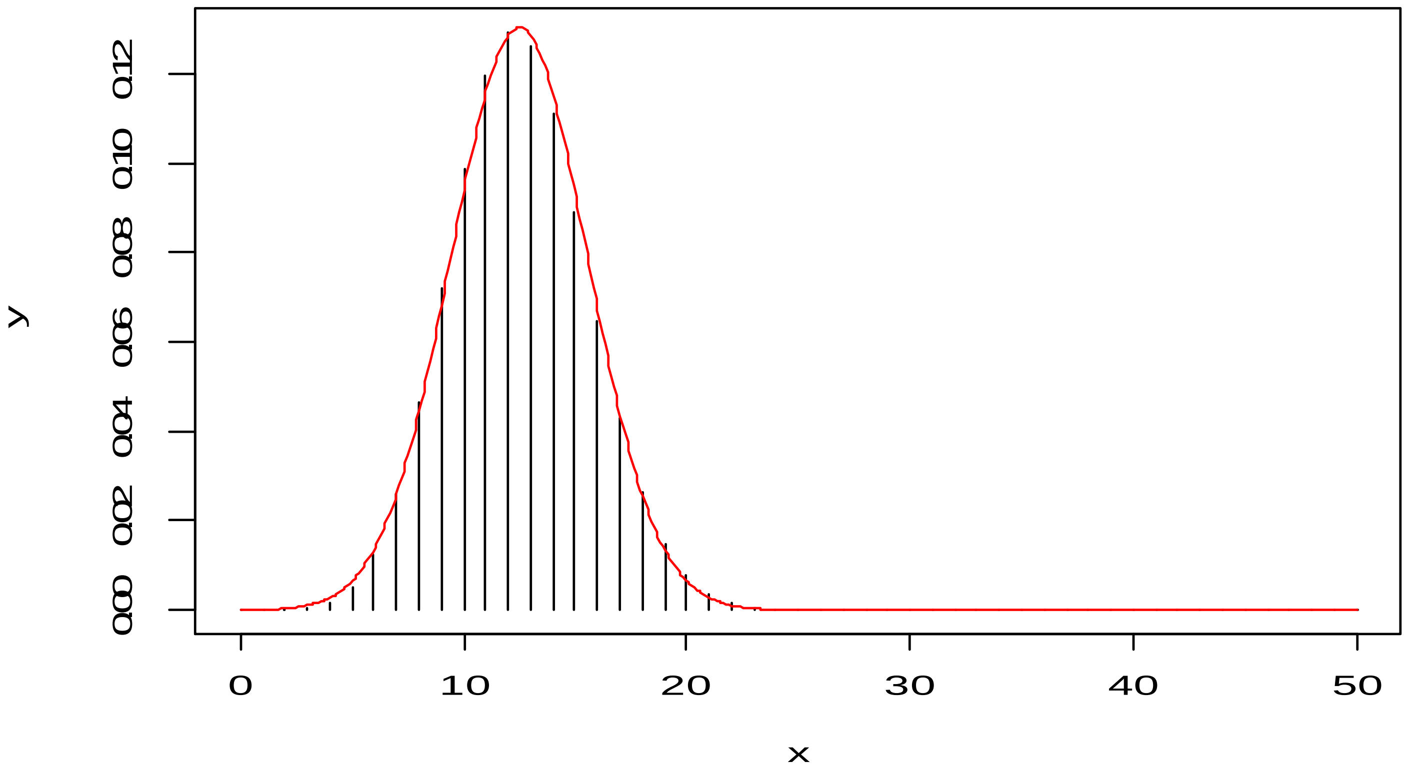
generate 50 (standard) normally distributed random numbers and to display them as a histogram.

```
>mysample <- rnorm(50)
>hist(mysample, prob = TRUE)
>mu <- mean(mysample)
>sigma <- sd(mysample)
>x <- seq(-4, 4, length = 500)
>y <- dnorm(x, mu, sigma)
>lines(x,y)
```


Histogram of mysample
Density
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
mysample

approximation of the binomial distribution with the normal distribution

```
> x <- 0:50
> y <- dbinom(x, 50, 0.25)
> plot(x, y, type="h")
> x2 <- seq(0, 50, length = 500)
> y2 <- dnorm(x2, 50*0.25, sqrt(50*0.25*(1-0.25)))
> lines(x2, y2, col = "red")
```



0.00
0.02
0.04
0.06
0.08
0.10
0.12
y
0
10
20
30
40
50
x

Practice:-

(i) Suppose X is normal with mean 527 and standard deviation 105. Compute P(X >=310)

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

(ii) Find P(80 pts< x<95 pts.)

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

- 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
```

- 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

```
> (1 - pnorm(2150, mean=2040, sd=60))*2000
[1] 66.75302
> (pnorm(1950, mean=2040, sd=60))*2000
[1] 133.6144
> ( pnorm(2160, mean=2040, sd=60)-pnorm(1920,mean=2040,sd=60))*2000
[1] 1908.999
```

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

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

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