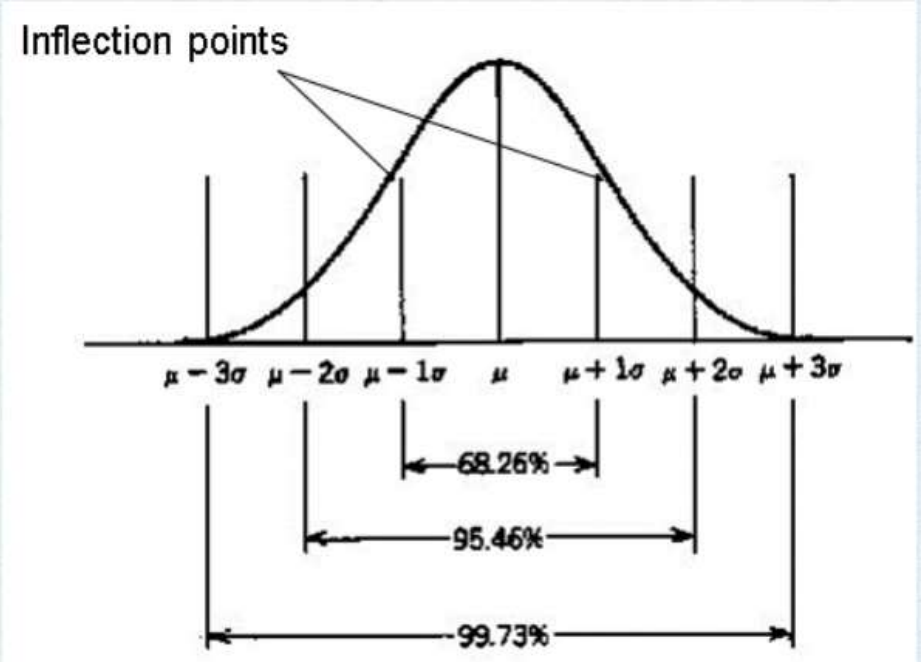
# Exp 3b- Normal Distribution

In a random collection of data from independent sources, it is generally observed that the distribution of data is normal. Which means, on plotting a graph with the value of the variable in the horizontal axis and the count of the values in the vertical axis we get a bell shape curve. The centre of the curve represents the mean of the data set. In the graph, fifty percent of values lie to the left of the mean and the other fifty percent lie to the right of the graph. This is referred as normal distribution in statistics.

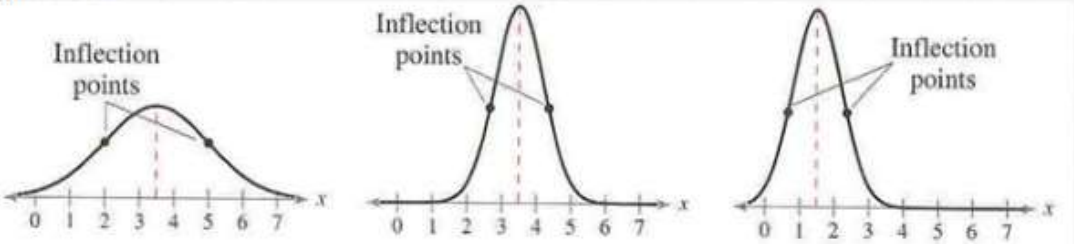
**Properties:**

A normal distribution is a continuous probability distribution for a random variable, x. The graph of a normal distribution is called the normal curve. A normal distribution has the following properties.

1. The mean, median and mode are equal.
2. The normal curve is bell-shaped and is symmetric about the mean.
3. The total area under the normal curve is equal to 1.
4. The normal curve approaches, but never touches the x-axis as it extends farther and farther away from the mean.
5. Between $\mu - \sigma$ and $\mu + \sigma$ (in the center of the curve) the graph curves downward. The graph curves upward to the left of $\mu - \sigma$ and to the right of $\mu + \sigma$. The points at which the curve changes from curving upward to curving downward are called **inflection points.**



6. A normal distribution can have any mean and any positive standard deviation. These two parameters, $\mu$ and $\sigma$ completely determine the shape of a normal curve. The mean gives the location of the line of symmetry and the standard deviation describes how much the data are spread out.




Mean: $\mu = 3.5$
Standard deviation: $\sigma = 1.5$

Mean: $\mu = 3.5$
Standard deviation: $\sigma = 0.7$

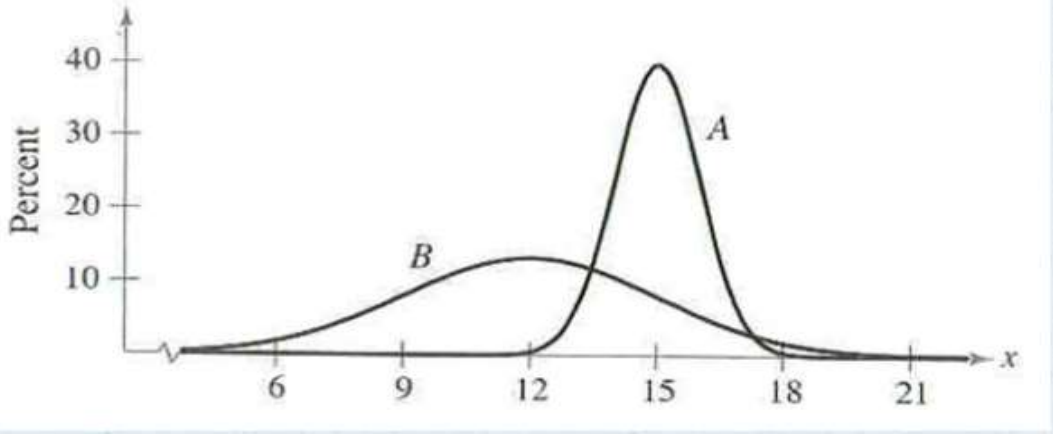
Mean: $\mu = 1.5$
Standard deviation: $\sigma = 0.7$

See the line of symmetry for each? That's the mean. However, if it is fatter, then the standard deviation is greater. That's the difference.

**Understanding Mean & Standard Deviation**

Which normal curve has a greater mean?

Which normal curve has a greater standard deviation



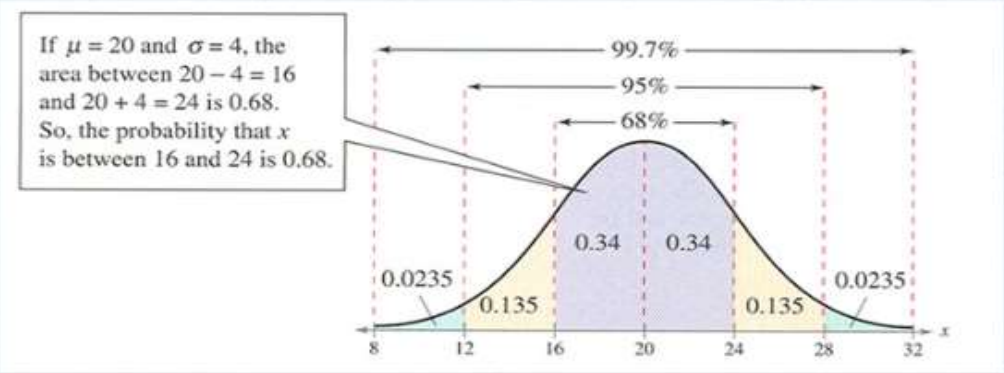

The line of symmetry of curve A occurs at x = 15. The line of symmetry of curve B occurs at x = 12. So, curve A has a greater mean.

Curve B is more spread out than curve A, so curve B has a greater standard deviation.

**The Empirical Rule**

In a normal distribution with mean μ and standard deviation σ, you can approximate areas under the normal curve as follows:

1. About 68% of the area lies between μ - σ and μ + σ
2. About 95% of the area lies between μ - 2σ and μ + 2σ
3. About 99.7% of the area lies between μ - 3σ and μ + 3σ




## Normal Distribution

A random variable X is said to possess normal distribution with mean μ and variance $\sigma^2$, 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 $\sigma^2$ , its transformation $Z = \frac{X-\mu}{\sigma}$ follows standard normal distribution (mean 0 and unit variance)

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

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 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 −

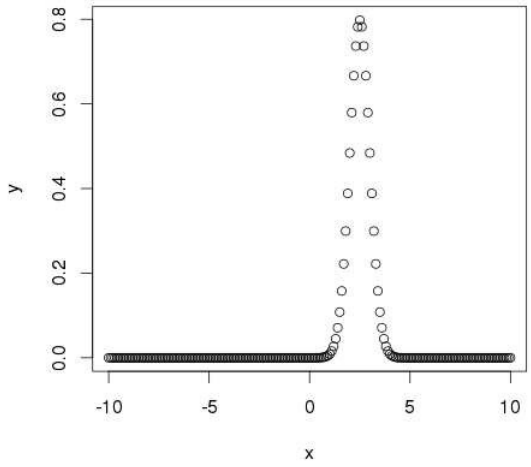
- **x** is a vector of numbers.
- **p** is a vector of probabilities.
- **n** is number of observations (sample size).
- **mean** is the mean value of the sample data. Its default value is zero.
- **sd** is the standard deviation. Its default value is 1.

**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)

# Give the chart file a name.
png(file = "dnorm.png")
plot(x,y)

# Save the file.
dev.off()
```

When we execute the above code, it produces the following result −

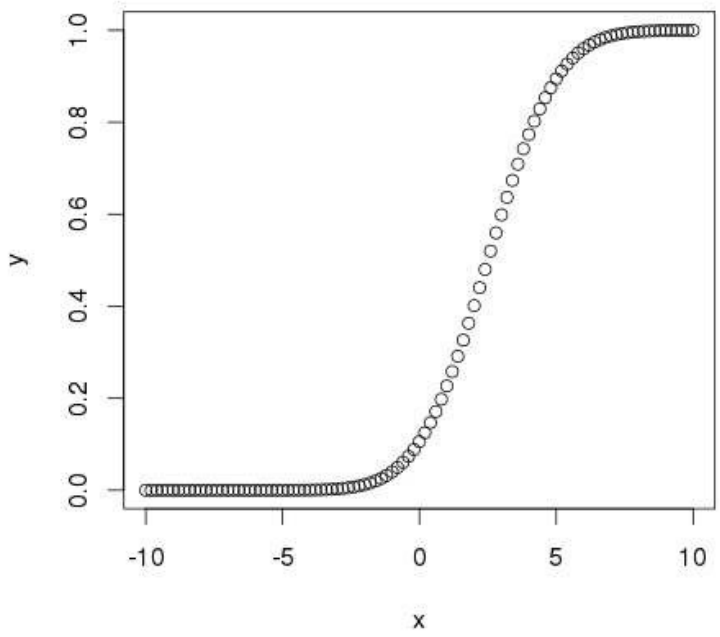



**pnorm()**

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)

# Give the chart file a name.
png(file = "pnorm.png")

# Plot the graph.
plot(x,y)

# Save the file.
dev.off()
```

When we execute the above code, it produces the following result −

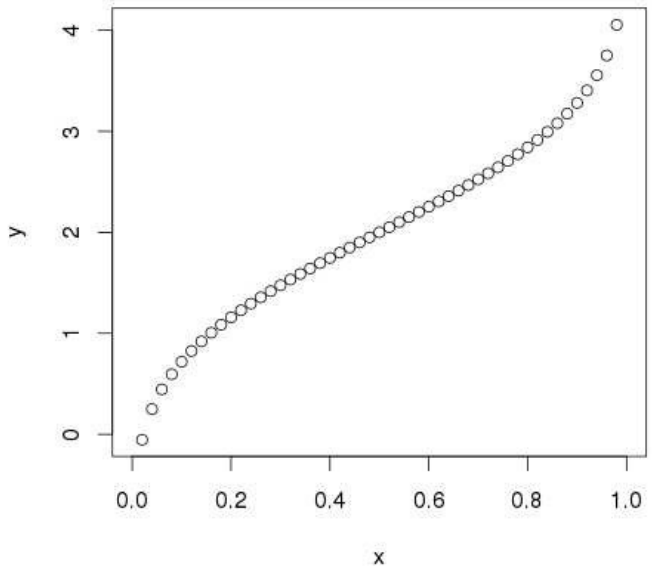



**qnorm()**

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)

# Give the chart file a name.
png(file = "qnorm.png")

# Plot the graph.
plot(x,y)

# Save the file.
dev.off()
```

When we execute the above code, it produces the following result −

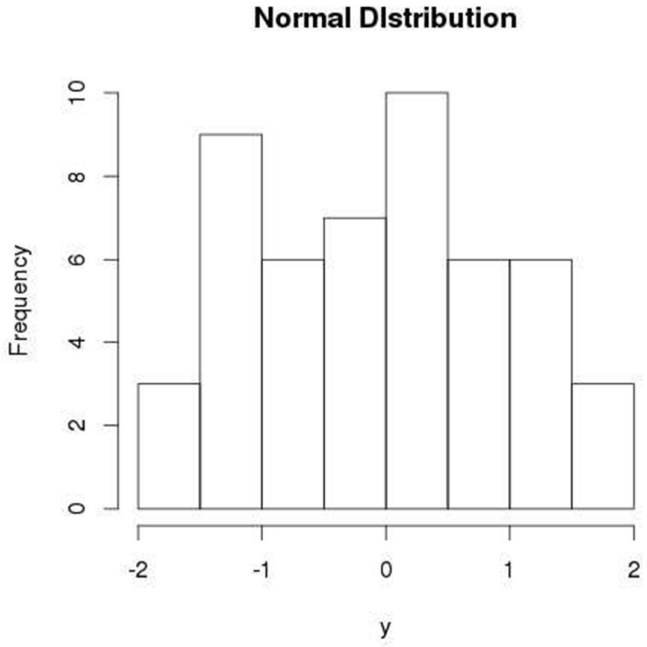



**rnorm()**

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)

# Give the chart file a name.
png(file = "rnorm.png")

# Plot the histogram for this sample.
hist(y, main = "Normal DIstribution")

# Save the file.
dev.off()
```

When we execute the above code, it produces the following result −

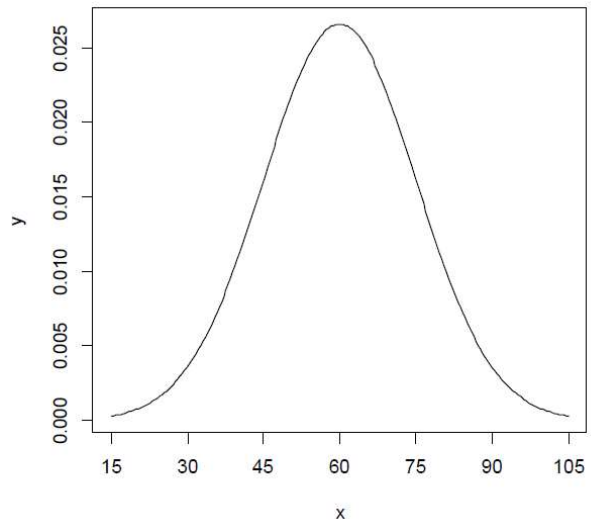



**1.** The weekly wages of 1000 workmen are normally distributed around a mean of Rs. 70 with S.D of Rs 5.Estimate the number of workers whose weekly wages will be
(i) Between Rs 69 and Rs 72
(ii) Less than Rs 69
(iii) More than Rs 72

```
> #(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
```

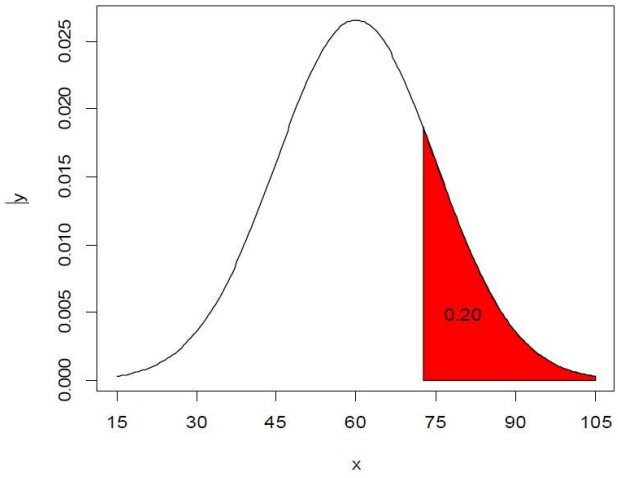
2. Draw a normal distribution with a mean=60 and a standard deviation=15.
>x=seq(15,105,length=200)
>y=dnorm(x,mean=60,sd=15)

```
>plot(x,y,type="l",xaxt="n")
>axis(1,at=c(15,30,45,60,75,90,105))
```




3.Shade the top 20% of the area under the normal density curve

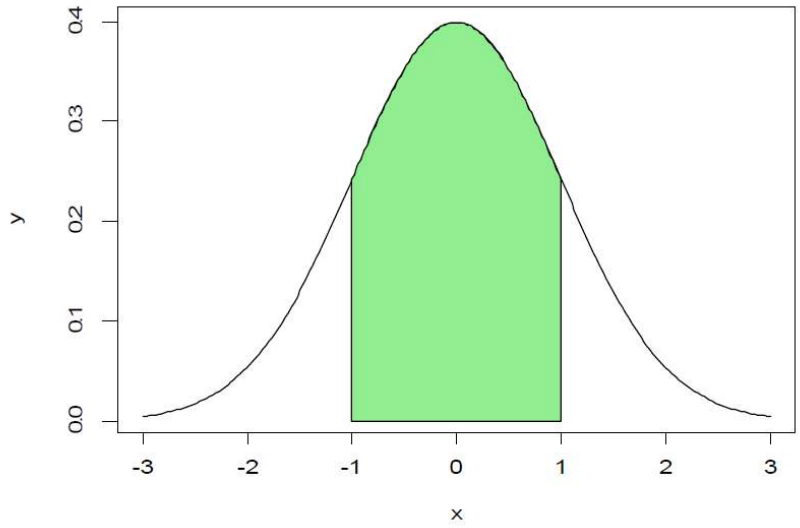
```
>x=seq(15,105,length=200)
>y=dnorm(x,mean=60,sd=15)
>plot(x,y,type="l",xaxt="n")
>axis(1,at=c(15,30,45,60,75,90,105))
>x=seq(72.62,105,length=100)
>y=dnorm(x,mean=60,sd=15)
>polygon(c(72.62,x,105),c(0,y,0),col="red")
>text(80,0.005,"0.20")
```



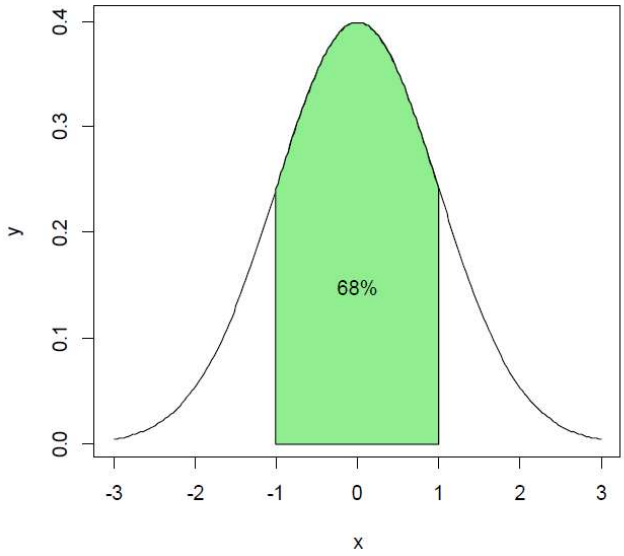

3. Simulate a standard normal density curve (mean=0 and standard deviation=1)

```
>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")
```

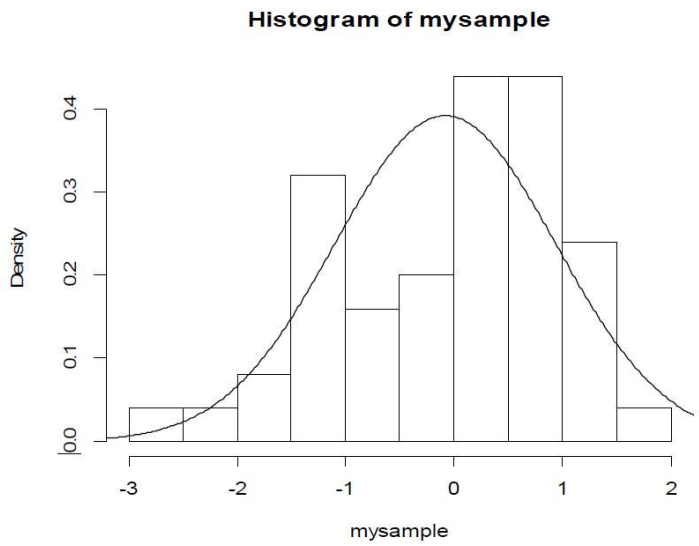



```
>text(0,0.15,"68%")
```



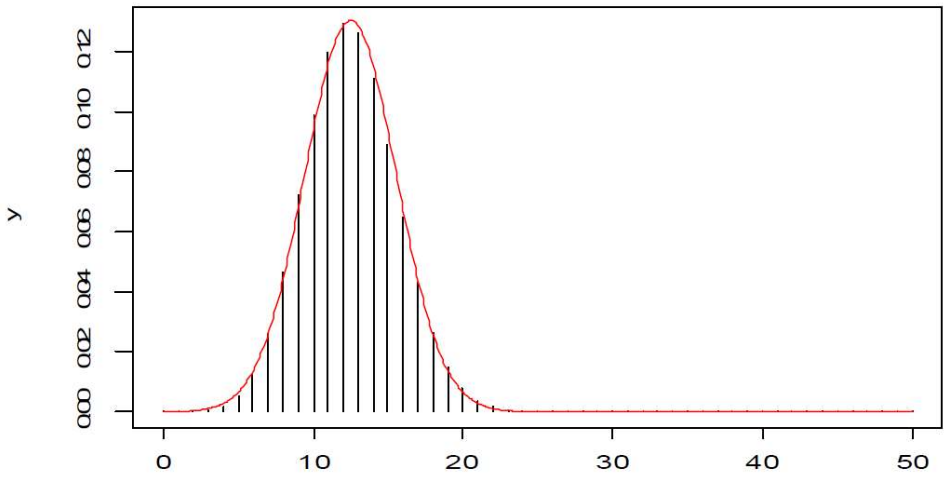

4. Generate 50 (standard) normally distributed random numbers and to display them as a histogram.

```
>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)
```




5. 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")
```




**Practice:-**

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

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

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

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

3. 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

# References


Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S Language*. Wadsworth & Brooks/Cole.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995) *Continuous Univariate Distributions*, volume 1, chapter 13. Wiley, New York.