Maternal Smoking and Infant deaths

Smoking by pregnant women may result in fetal injury, premature birth, and low birth weight.

- Is this warning to be taken seriously ?
- Dataset: CHDS - Berkeley, California.
- Taken entirely from Chapter 1 of the book
  Stat Labs: Mathematical Statistics Through Applications by Deborah Nolan and Terry P. Speed
- Check website: https://www.stat.berkeley.edu/users/statlabs/

Check: Normal Distribution

- Do 68-95-99.7 first check to see if data is like normal or not.

  Compute

  $$\text{Skewness} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^3, \qquad \text{Kurtosis} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^4.$$

  - Skewness 0 indicates that the distribution is symmetric
  - Kurtosis is a measure of the peak of the distribution.
  - For standard normal Kurtosis is 3.

  If Kurtosis and Skewness coefficients are *far* from 3 and 0 we can conclude that the data is not normal.

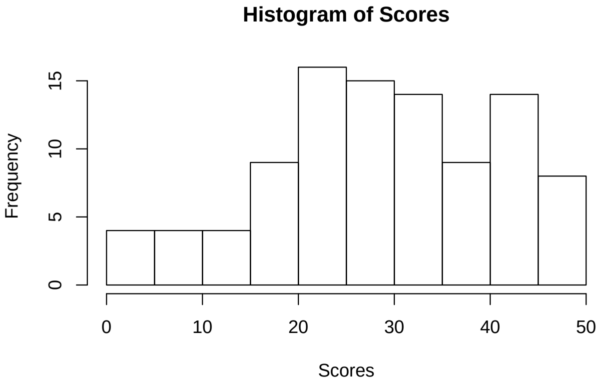
- Final check whether data is normal is to plot

  $$\left(z_{\frac{k}{n+1}}, x_{(k)}\right)$$

  - if the plot is a straight line then it indicates that data is normal

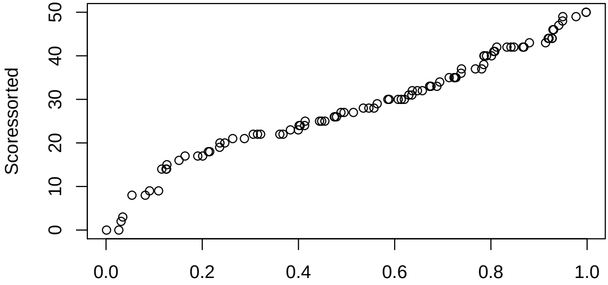
Are the Scores uniform ?

```
> Scores = scan("Scores")
> hist(Scores)
```



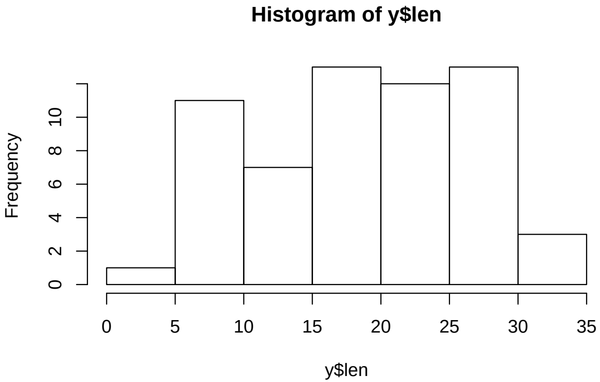
Can try the uniform-quantile plot

```
> Scores = scan("Scores")
> Scoressorted = sort(Scores)
> u = runif(97, 0,1)
> usorted  = sort(u)
> plot(usorted, Scoressorted)
```



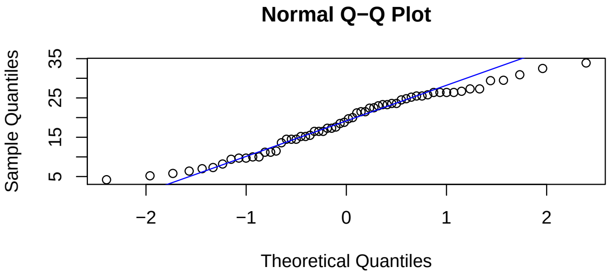
Symmetric, Skewed- Left and Right

```
> y = ToothGrowth
> hist(y$len)
```



Symmetric, Skewed- Left and Right

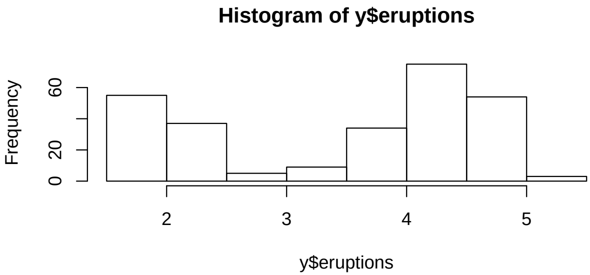
```
> y = ToothGrowth
> qqnorm(y$len)
> qqline(y$len, col="blue") # adds a reference line
```



Notice that we did not scale or center. slope and intercept

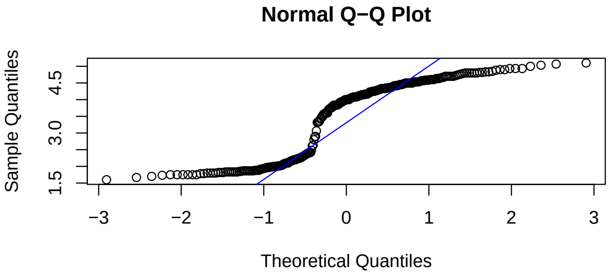
Symmetric, Skewed- Left and Right

```
> y = faithful
> hist(y$eruptions)
```



Bi-Modal

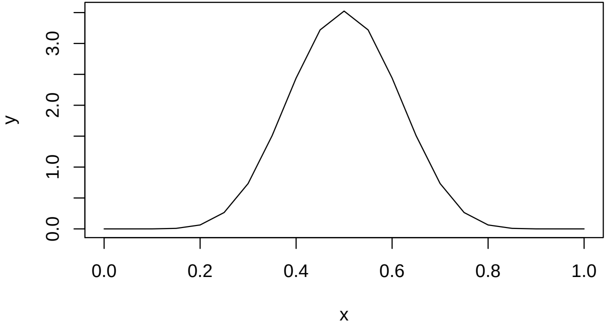
```
> y = faithful
> qqnorm(y$eruptions)
> qqline(y$eruptions, col="blue") # adds a reference line
```



Notice that we did not scale or center. slope and intercept

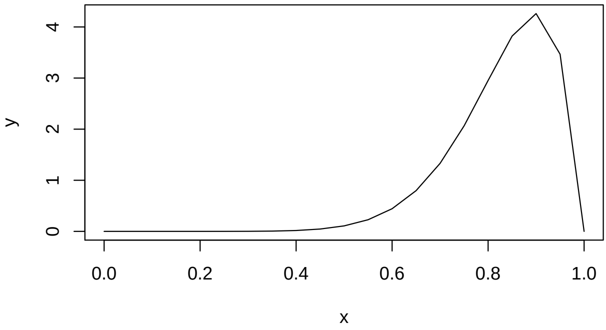
Beta-distribution

```
> x = seq(0,1, by=0.05)
> y = dbeta(x, 10,10)
> plot(x,y, type="l")
```



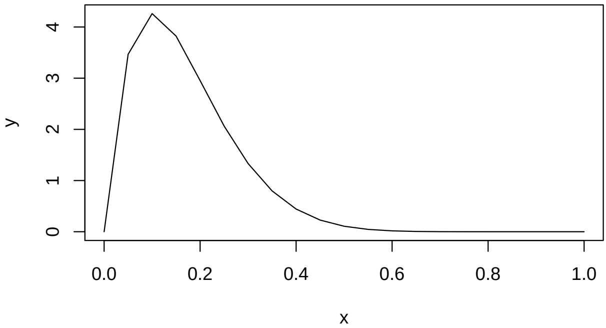
Beta-distribution

```
> x = seq(0,1, by=0.05)
> y = dbeta(x, 10,2)
> plot(x,y, type="l")
```



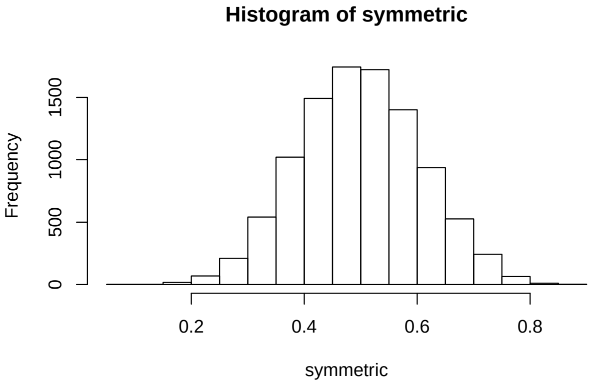
Beta-distribution

```
> x = seq(0,1, by=0.05)
> y = dbeta(x, 2,10)
> plot(x,y, type="l")
```



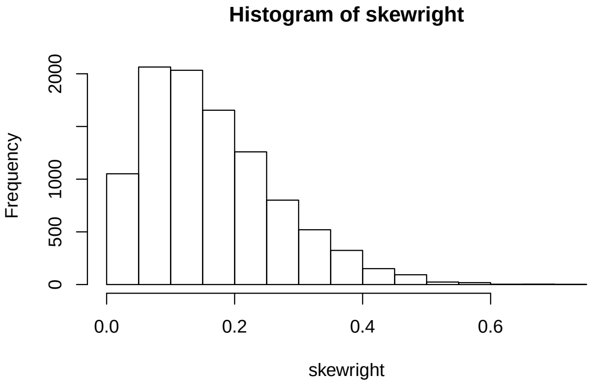
Symmetric, Skewed- Left and Right

```
> symmetric = rbeta(10000,10,10)

> hist(symmetric)
```



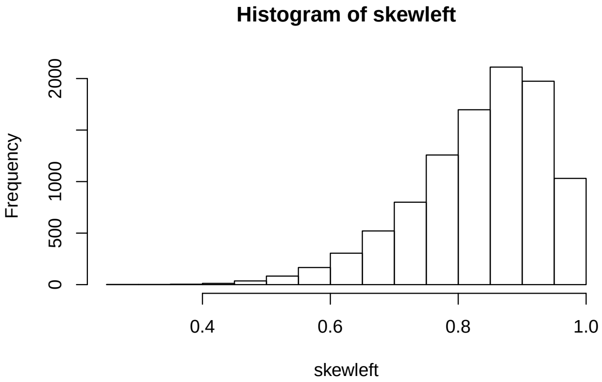
Symmetric, Skewed- Left and Right

```
> skewright= rbeta(10000,2,10)

> hist(skewright)
```



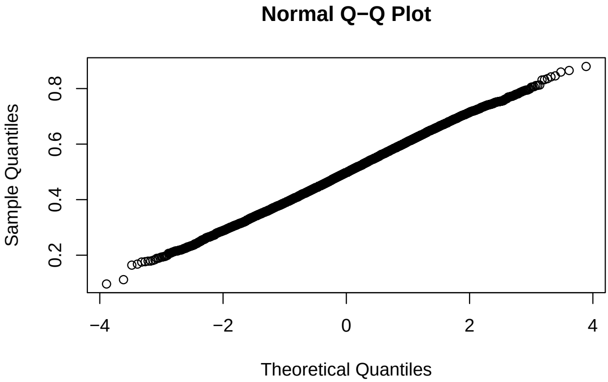
Symmetric, Skewed- Left and Right

```
> skewleft= rbeta(10000,10,2)

> hist(skewleft)
```



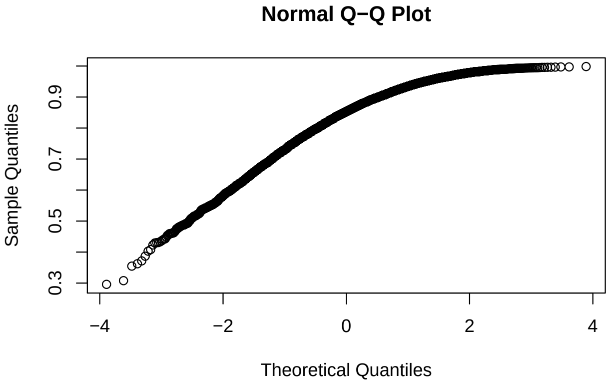
Symmetric, Skewed- Left and Right

```
> qqnorm(symmetric)
```



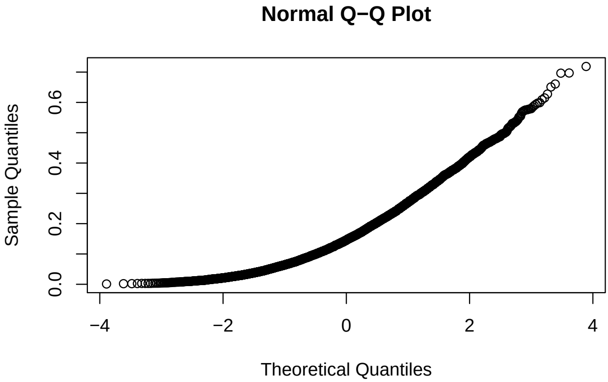
Symmetric, Skewed- Left and Right

```
> qqnorm(skewleft)
```



Symmetric, Skewed- Left and Right

```
> qqnorm(skewright)
```



Symmetric, Skewed- Left and Right

$$\text{Skewness} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^3.$$

- Skewness is a measure of symmetry.
- Negative skewness will imply that the mean of the data is less than the median, and the data distribution is left-skewed.
- Positive skewness will imply that the mean of the data values is larger than the median, and the data distribution is right-skewed.

Symmetric, Skewed- Left and Right

$$\text{Skewness} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^3.$$

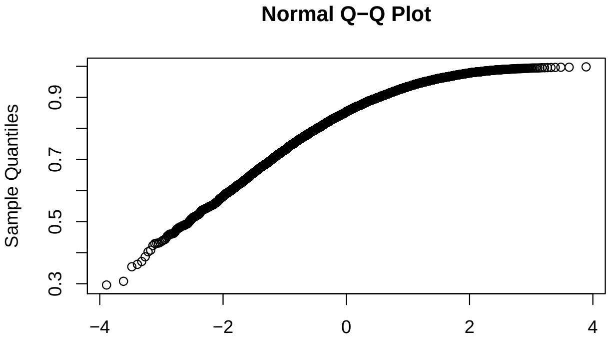
- In a histogram setting we can infer from mode:
  - Left Skew. Find the mode (the highest point of the distribution). The right of the mode should be shorter than the left of the mode.
  - Right Skew. Find the mode (the highest point of the distribution). The right of the mode should be longer than the left of the mode

Symmetric, Skewed- Left and Right

$$\text{Skewness} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^3.$$

- From Normal-Q plot

```
> qqnorm(skewleft)
```

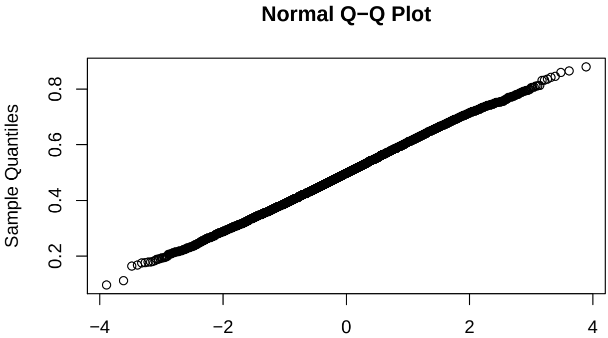


Symmetric, Skewed- Left and Right

$$\text{Skewness} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^3 .$$

- From Normal-Q plot

```
> qqnorm(symmetric)
```



Symmetric, Skewed- Left and Right

$$\text{Skewness} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^3.$$

- From Normal-Q plot

```
> qqnorm(skewright)
```