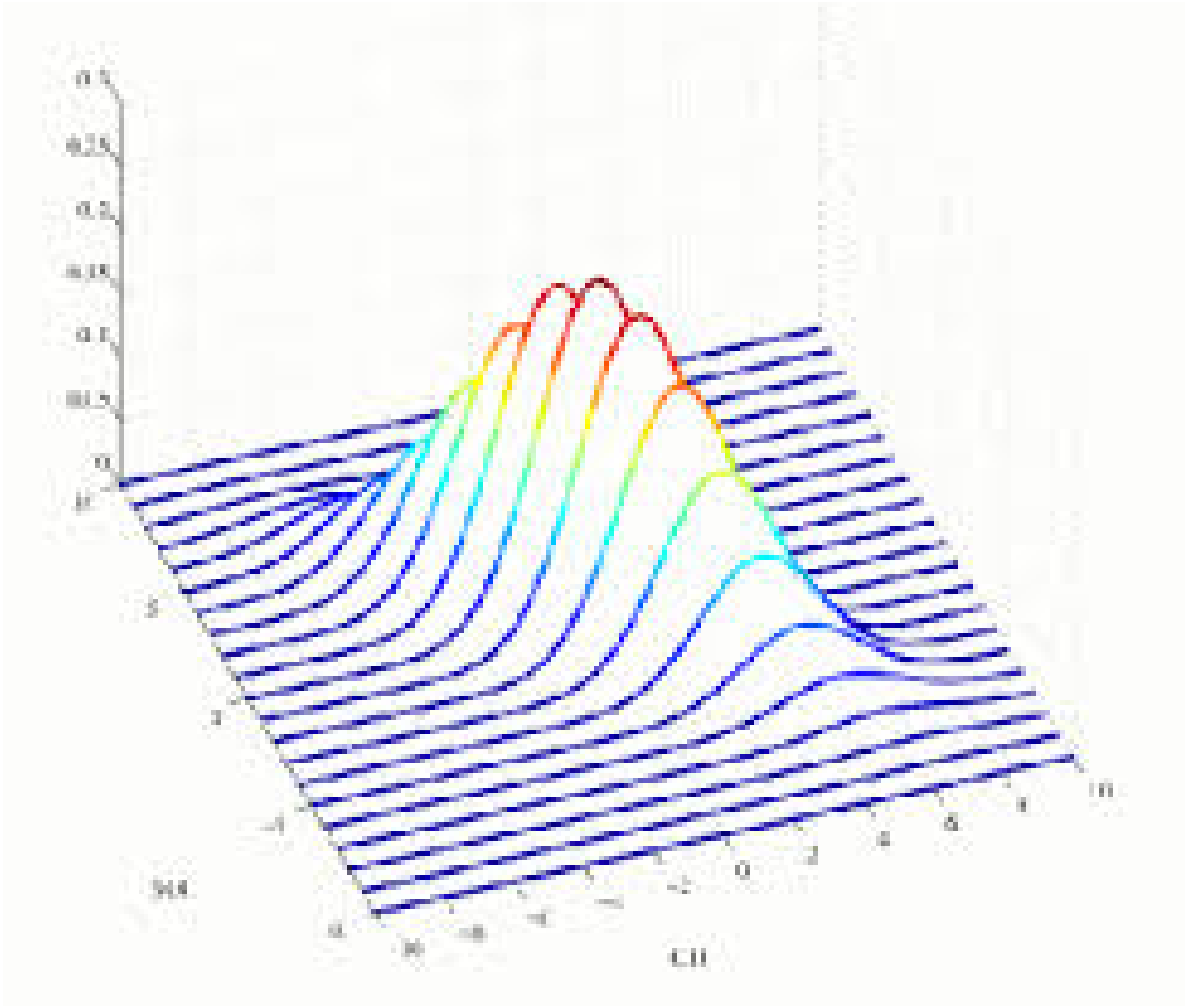
Multivariate Normal

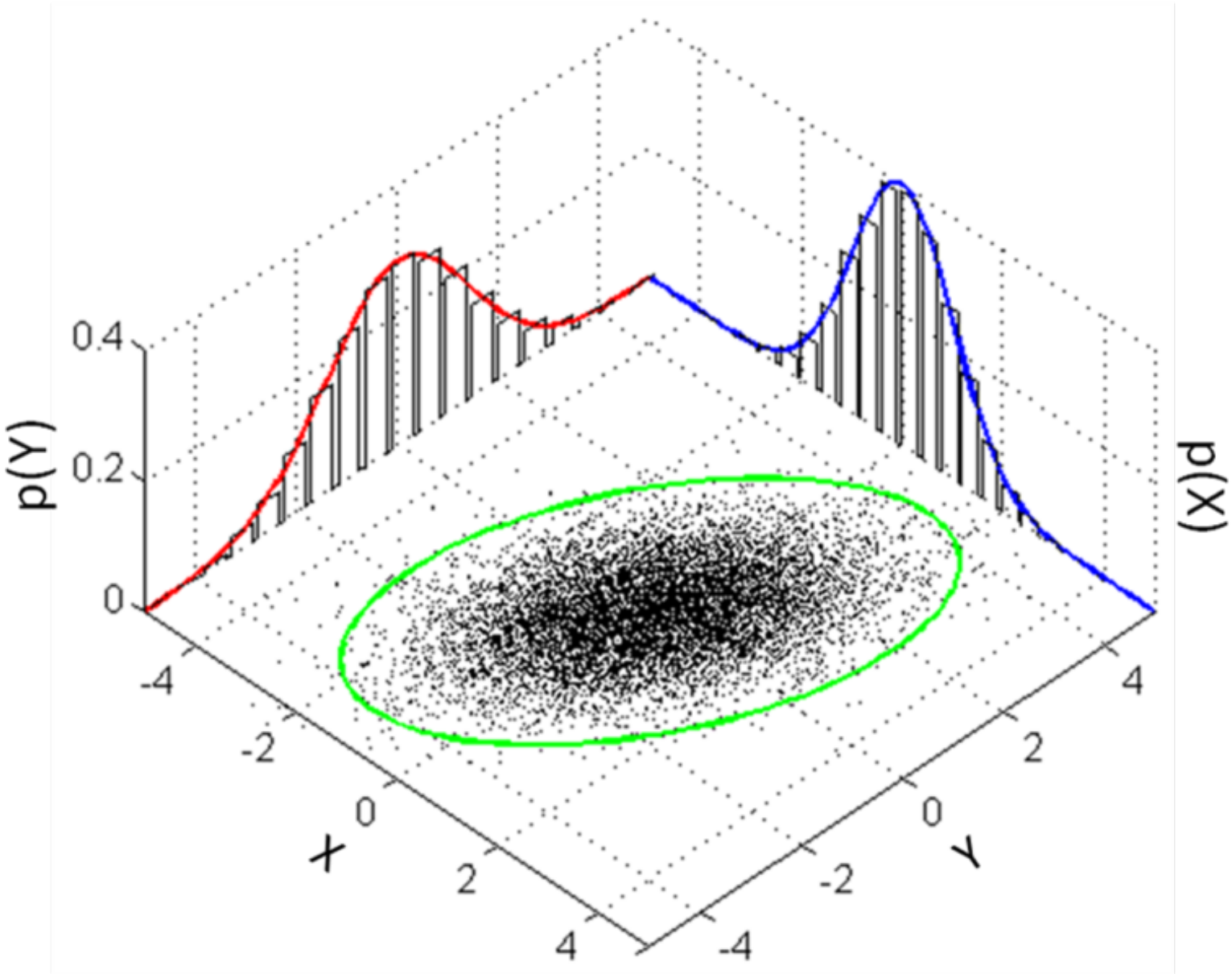
- The multivariate normal distribution (or multivariate Gaussian distribution), is a generalization of the one-dimensional (univariate) normal distribution to higher dimensions.
- The multivariate normal distribution is often used to describe, at least approximately, any set of (possibly) correlated real-valued random variables each of which clusters around a mean value.

Bivariate Normality



Bivariate describes the case of two variables.

Bivariate Normality



Testing for Normality

Hypothesis Tests for Univariate Data

Graphical Methods

- Histograms (with Kerney Density Estimation Line)
- Normal Probability Plots

Formal Hypothesis Tests

- Shapiro-Wilk Test (inbuilt with R)
- Anderson-Darling Test (nortest package)
- D'Agostino Test (MSQC package)

Testing for Normality

Hypothesis Tests for Multivariate Data

- Mardia Test (MSQC package)
- Henze and Zirkler (MSQC package)
- Royston Test (MSQC package)

Testing for Normality

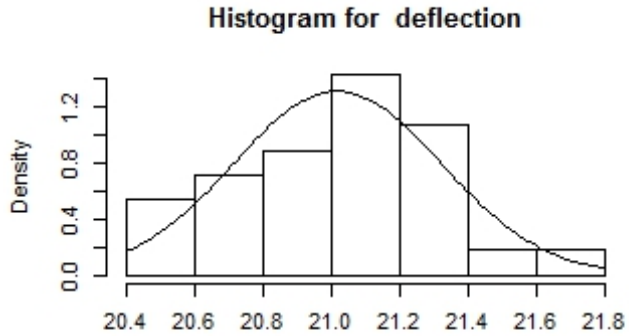
The bimetal data set (MSQC package)

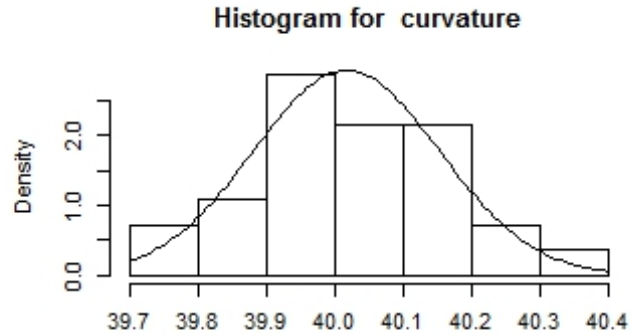
- Bimetal thermostat has innumerable practical uses. These types of thermostats hold a bimetallic strip composed by two strips of different metals that convert the changing of temperature in mechanical displacement due to the difference in thermal expansion.
- Certain type of strip composed of brass and steel is analyzed in a quality laboratory by testing the deflection, curvature, resistivity, and hardness in low and high expansion sides.

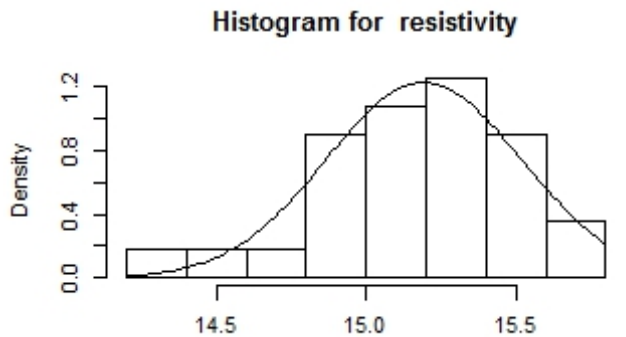
Testing for Normality

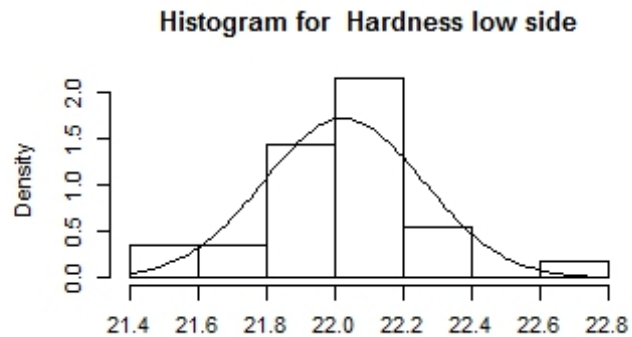
```
> tail(bimetal1)
deflection curvature resistivity Hardness low side Hard
[23,]      20.76     39.98        14.98                22.2
[24,]      21.00     40.11        15.17                22.0
[25,]      20.57     39.73        14.35                22.0
[26,]      20.78     39.83        15.27                21.6
[27,]      20.96     40.03        15.26                21.9
[28,]      21.14     39.93        14.98                21.8
```

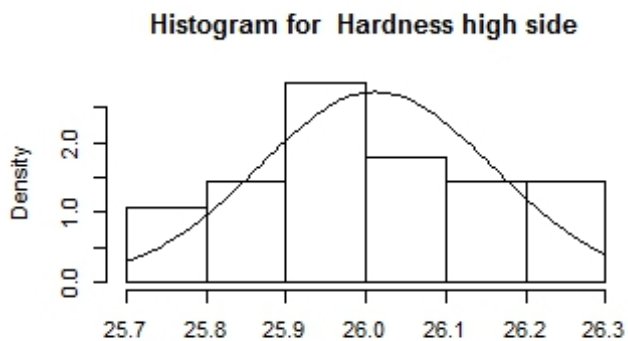
Testing for Normality



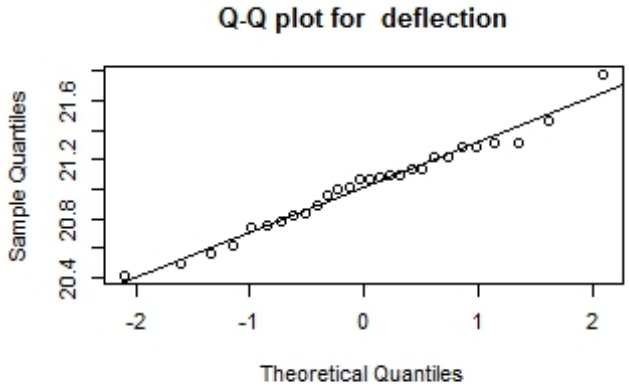


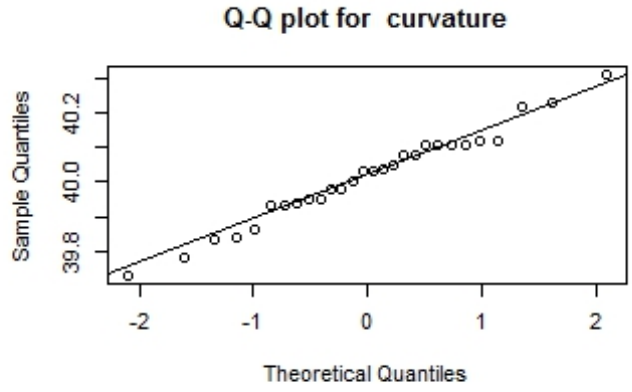


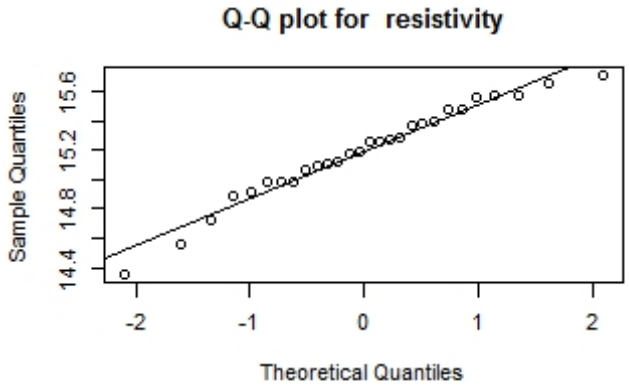


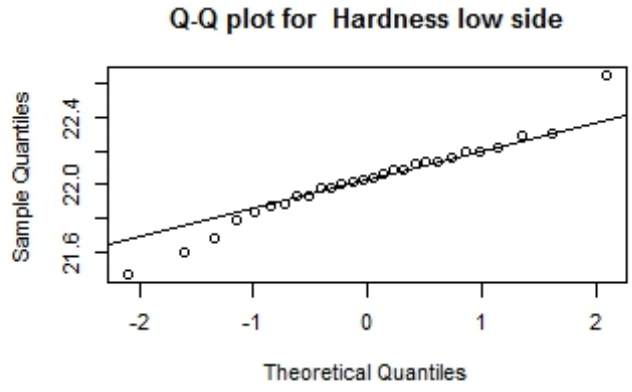


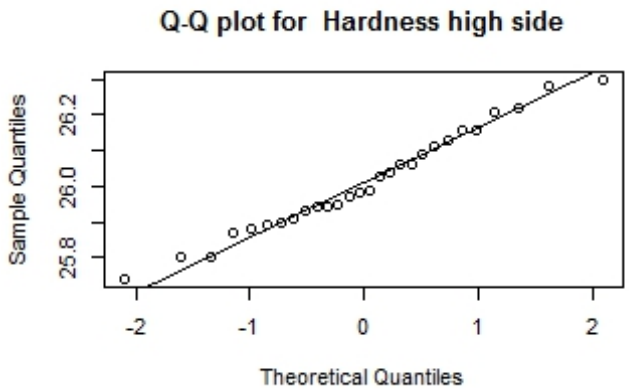
Testing for Normality











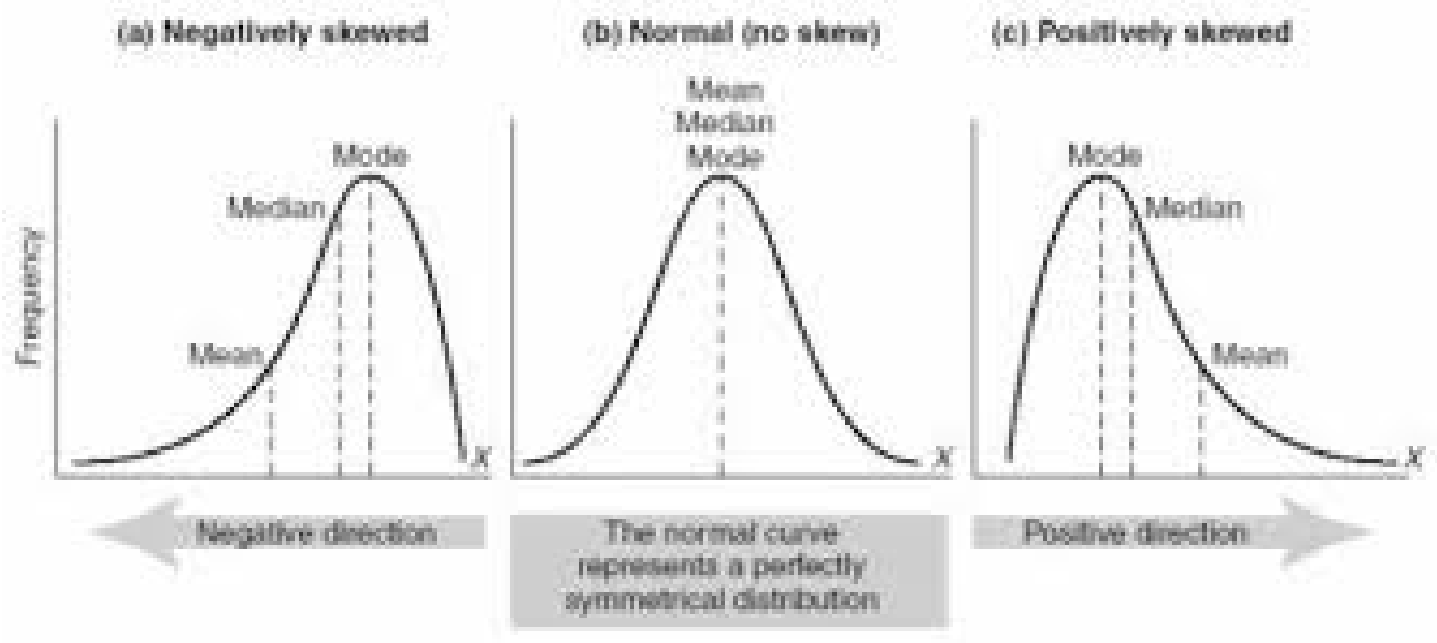
Skewness and Kurtosis

- Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point.
- Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. That is, data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. Data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak.

Skewness and Kurtosis

- The normal distribution is a symmetric distribution with well-behaved tails.
- Theoretically a Normal Distibuted Random Variable will have a skewness of 0.00.
- Also The kurtosis of 2.96 is near the expected value of 3.

Skewness



- Negatively Skewed - Skewness is negative number
- Symmetric - Skewness is zero
- Positively Skewed - Skewness is positive number

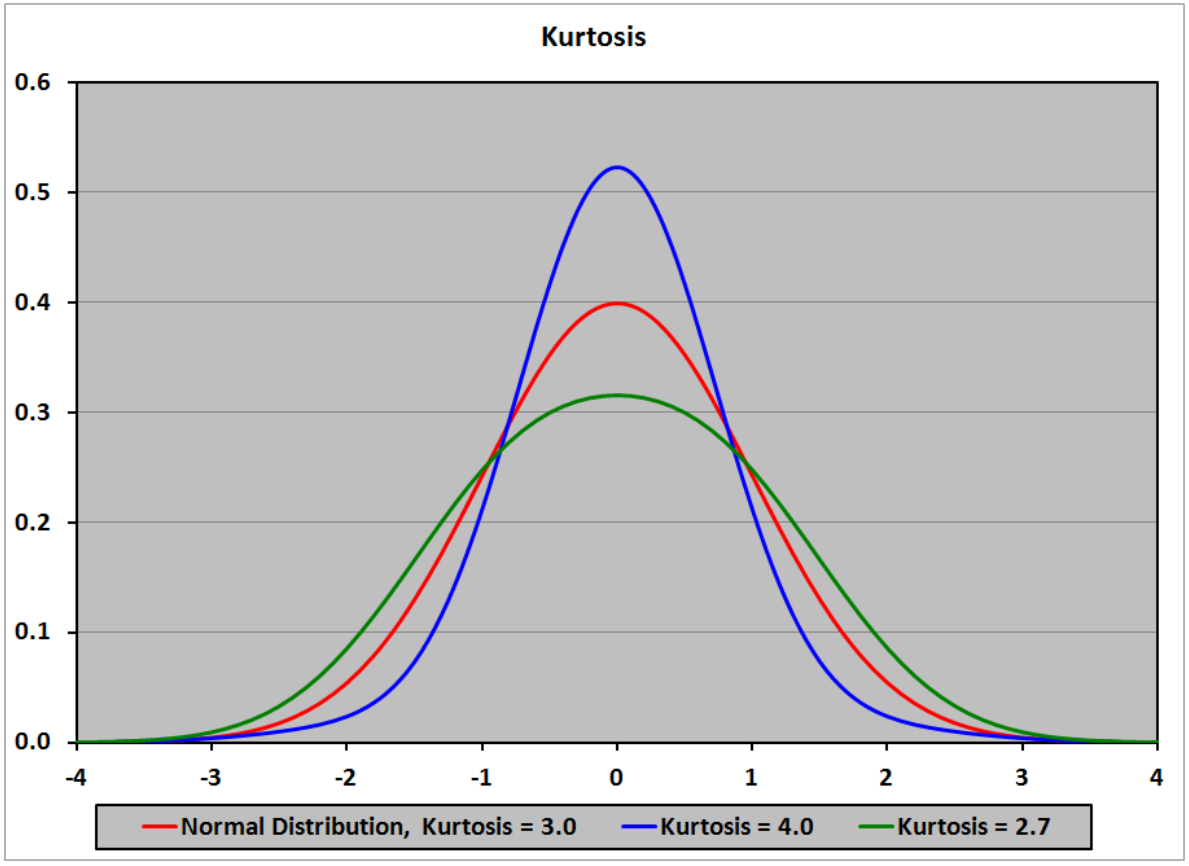

Kurtosis
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-4
-3
-2
-1
0
1
2
3
4
Normal Distribution, Kurtosis = 3.0
Kurtosis = 4.0
Kurtosis = 2.7

Figure:

D'Agostino Test

- DAgostinos K^2 test, named for Ralph D'Agostino, is a goodness-of-fit measure of departure from normality, that is the test aims to establish whether or not the given sample comes from a normally distributed population.
- The test is based on transformations of the **sample kurtosis** and **skewness**, and has power only against the alternatives that the distribution is skewed and/or kurtic.

Testing for Normality

D'Agostino Test (MSQC Package)

- Using the bimetal1 data set in MSQC package
- Run the procedure on the first variable only.

```
library(MSQC)
DAGOSTINO(bimetal1[,1])
```

Output on next two slides

Testing for Normality

```
D'Agostino Test

Skewness
Skewness coefficient: 0.0831225
Statistics: 0.2117358
p-value: 0.8323131

Kurtosis
The kurtosis coefficient: 3.0422
Statistics: 0.591983
p-value: 0.553862
```

Testing for Normality

```
....
Omnibus Test
Chi-squared: 0.3952759
Degree of freedom: 2
p-value: 0.8206669
```

Testing for Normality

Some Multivariate (MSQC Pacakge)

```
> MardiaTest(bimetal1)
$skewness
[1] 6.982112

$p.value
[1] 0.585327

$kurtosis
[1] 33.77373

$p.value
[1] 0.3490892
```

Testing for Normality

```
>
> HZ.test(bimetal1)
[1] 0.6068650 0.7709586
>
```

Testing for Normality

```
> Royston.test(bimetal1)
test.statistic          p.value
1.1814742       0.9364221
```

Testing for Normality

Box Cox Transformation

- The Box-Cox transforms non-normally distributed data to a set of data that has approximately normal distribution.