1 Assumption of Normality

1.1 Multivariate Normal

- The multivariate normal distribution (or multivariate Gaussian distribution), is a generalization of the one-dimensional (univariate) normal distribution to higher dimensions.
- One possible definition is that a random vector is said to be k-variate normally distributed if every linear combination of its k components has a univariate normal distribution.
- 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.

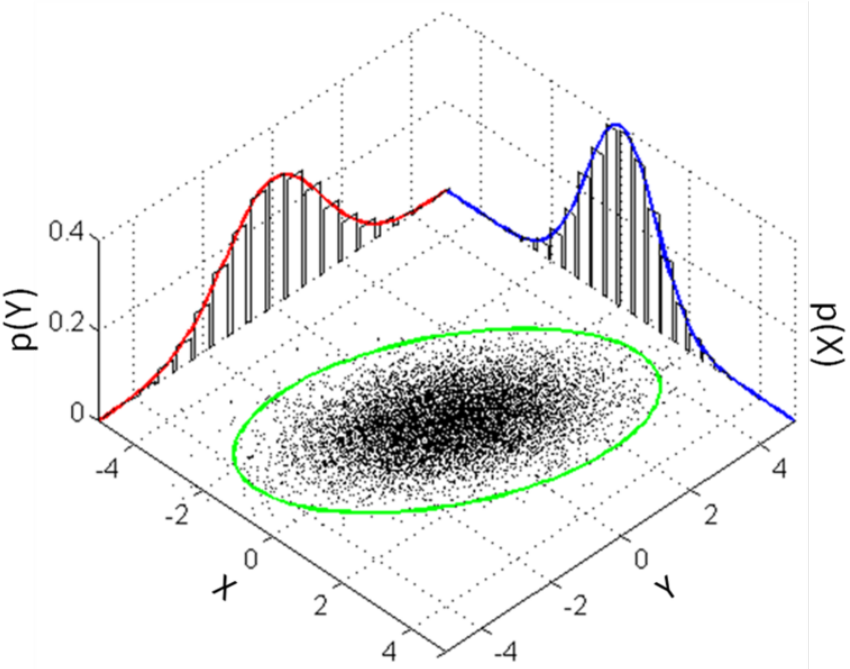

Figure 1

1.2 Testing for Normality

Graphical Methods

- Histograms
- Normal Probability Plots

Hypothesis Tests for Univariate Data

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

Hypothesis Tests for Multivariate Data

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

1.2.1 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.

```
> tail(bimetal1)
      deflection curvature resistivity Hardness low side Hardness high side
[23,]      20.76     39.98       14.98             22.29              26.03
[24,]      21.00     40.11       15.17             22.04              25.99
[25,]      20.57     39.73       14.35             22.02              25.80
[26,]      20.78     39.83       15.27             21.60              25.89
[27,]      20.96     40.03       15.26             21.98              25.94
[28,]      21.14     39.93       14.98             21.84              25.98
```

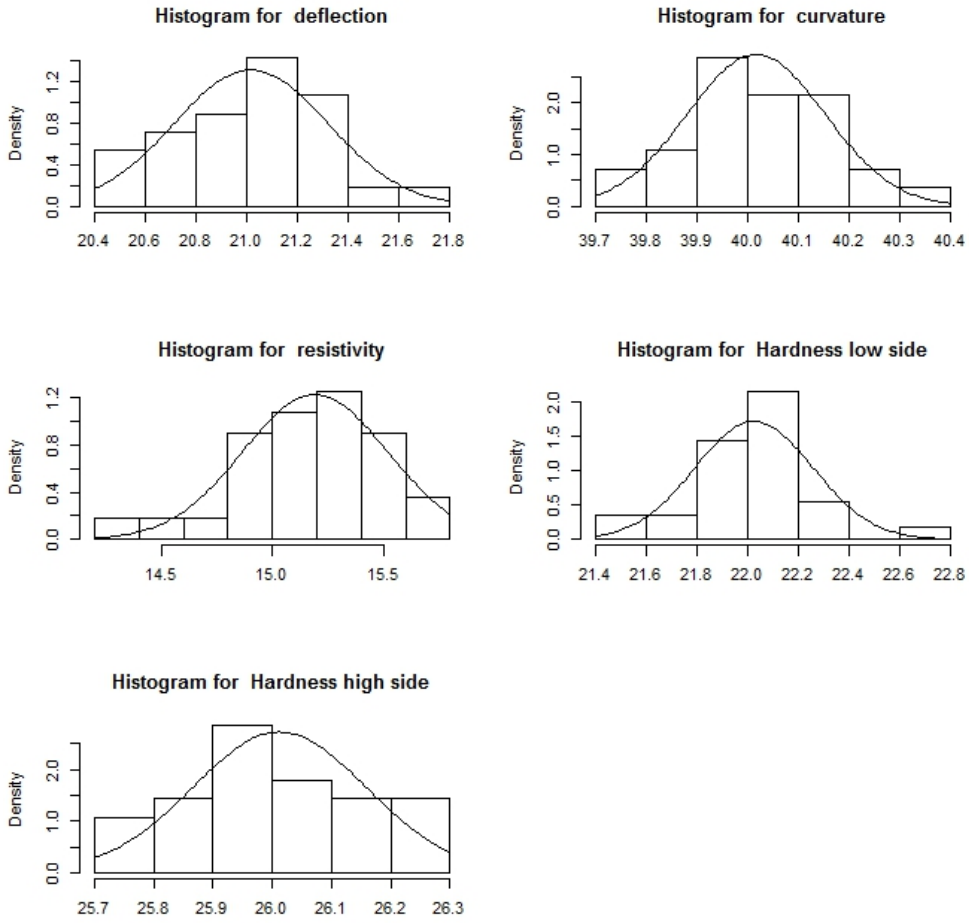
Histogram for deflection
Density
0.0 0.4 0.8 1.2
20.4 20.6 20.8 21.0 21.2 21.4 21.6 21.8
Histogram for curvature
Density
0.0 1.0 2.0
39.7 39.8 39.9 40.0 40.1 40.2 40.3 40.4
Histogram for resistivity
Density
0.0 0.4 0.8 1.2
14.5 15.0 15.5
Histogram for Hardness low side
Density
0.0 0.5 1.0 1.5 2.0
21.4 21.6 21.8 22.0 22.2 22.4 22.6 22.8
Histogram for Hardness high side
Density
0.0 1.0 2.0
25.7 25.8 25.9 26.0 26.1 26.2 26.3

Figure 2

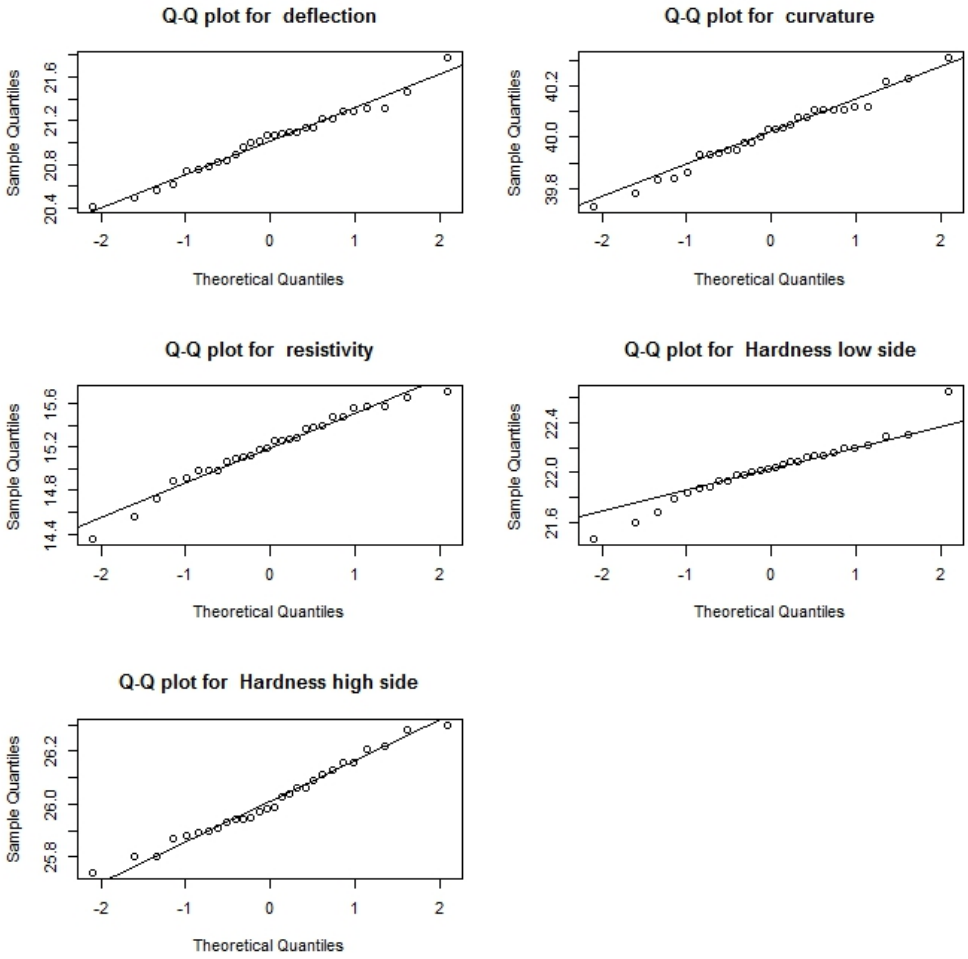
Q-Q plot for deflection
Sample Quantiles
20.4 20.8 21.2 21.6
-2 -1 0 1 2
Theoretical Quantiles
Q-Q plot for curvature
Sample Quantiles
39.8 40.0 40.2
-2 -1 0 1 2
Theoretical Quantiles
Q-Q plot for resistivity
Sample Quantiles
14.4 14.8 15.2 15.6
-2 -1 0 1 2
Theoretical Quantiles
Q-Q plot for Hardness low side
Sample Quantiles
21.6 22.0 22.4
-2 -1 0 1 2
Theoretical Quantiles
Q-Q plot for Hardness high side
Sample Quantiles
25.8 26.0 26.2
-2 -1 0 1 2
Theoretical Quantiles

Figure 3

1.2.2 D'Agostino Test (MSQC Package)

- Using the bimetal1 data set in MSQC package

```
> for (i in 1 : 5){
+  DAGOSTINO(bimetal1[,i])
+  }
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
    Omnibus Test
      Chi-squared: 0.3952759
      Degree of freedom: 2
      p-value: 0.8206669
....
....
D'Agostino Test
    Skewness
      Skewness coefficient: -0.04173762
      Statistics: -0.1063873
      p-value: 0.9152751
    Kurtosis
      The kurtosis coefficient: 4.162062
      Statistics: 1.675258
      p-value: 0.09388364
    Omnibus Test
      Chi-squared: 2.817807
      Degree of freedom: 2
      p-value: 0.2444111
```

1.2.3 Some Multivariate (MSQC Pacakge)

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

$p.value
[1] 0.585327

$kurtosis
[1] 33.77373

$p.value
[1] 0.3490892

>
>
>
> HZ.test(bimetal1)
[1] 0.6068650 0.7709586
>
>
> Royston.test(bimetal1)
test.statistic        p.value
     1.1814742      0.9364221
```

1.2.4 Box Cox Transformation

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