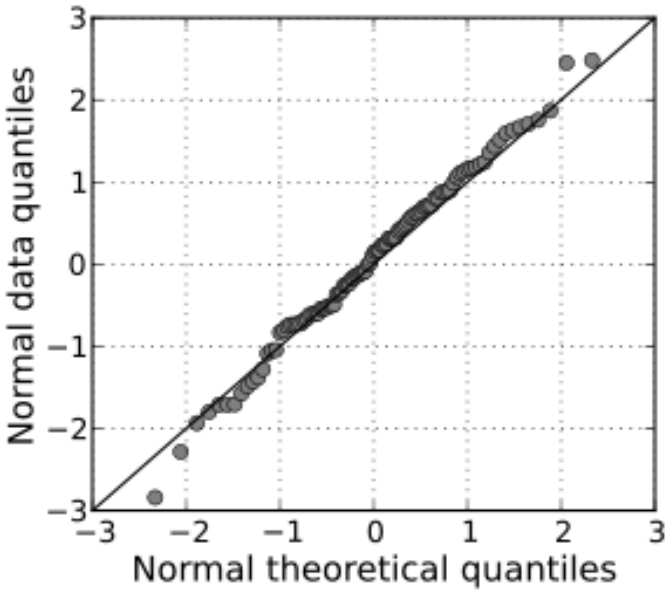
# Checking for Normality

- A normal distribution is often a reasonable model for continuous data. However, without inspecting the data, it is risky to assume a normal distribution.
- There are a number of graphical methods, for example histogram and boxplots, that can be used to indicate deviations of the data from the normal distribution.




## The Q-Q Plot

- The most useful tool for assessing normality is a quantile quantile or QQ plot. This is a scatterplot with the quantiles of the scores on the horizontal axis and the expected normal scores on the vertical axis.
- The steps in constructing a QQ plot are as follows: First, we sort the data from smallest to largest. A plot of these scores against the expected normal scores should reveal a straight line.
- The expected normal scores are calculated by taking the z-scores where I is the rank in increasing order.
- Curvature of the points indicates departures of normality.
- This plot is also useful for detecting outliers. The outliers appear as points that are far away from the overall pattern op points

### 0.0.1 Graphical Methods

- The quantile-quantile (QQ) plot is an excellent way to see whether the data deviate from normal (the plot can be set up to see if the data deviate from other distributions as well but here we are only interested in the normal distribution). The process SPSS goes through for creating a QQ plot involves determining what proportion of the 'observed' scores fall below any one score, then the z score that would fit that proportion if the data were normally distributed is calculated, and finally that z score that would cut off that proportion (the 'expected normal value') is translated back into the original metric to see what raw score that would be.
- A scatter plot is then created that shows the relationship between the actual 'observed' values and what those values would be 'expected' to be if the data were normally distributed.
- This is all quite complicated but the 'bottom line' is quite easy, if the data are normally distributed then the circles on the resulting plot (each circle representing a score) will form a straight line. Use the following examples to learn how to interpret QQ plots, be aware that some stat programs switch the axes around from the way it is set up in SPSS.

## 0.1 Normal Probability Plots

- A normal distribution is often a reasonable model for the data. Without inspecting the data, however, it is risky to assume a normal distribution.
- There are a number of graphs that can be used to check the deviations of the data from the normal distribution.

### 0.1.1 Normal Probability Plots

- A histogram is an example of a graph that can be used to check normality.
- Here, the histogram should reveal a bell shaped curve.

- The most useful tool for assessing normality is a ***quantile -quantile*** or **Q-Q** plot.
- This is a scatterplot with the quantiles of the scores on the horizontal axis and the equivalent values expected when assuming the normal distribution, on the vertical axis.
- The steps in constructing a QQ plot are as follows: First, we sort the data from smallest to largest.
- A plot of these scores against the expected normal scores should reveal a straight line.
- The expected normal scores are calculated by taking the z-scores where I is the rank in increasing order.
- Curvature of the points indicates departures of normality. This plot is also useful for detecting outliers. The outliers appear as points that are far away from the overall pattern op points

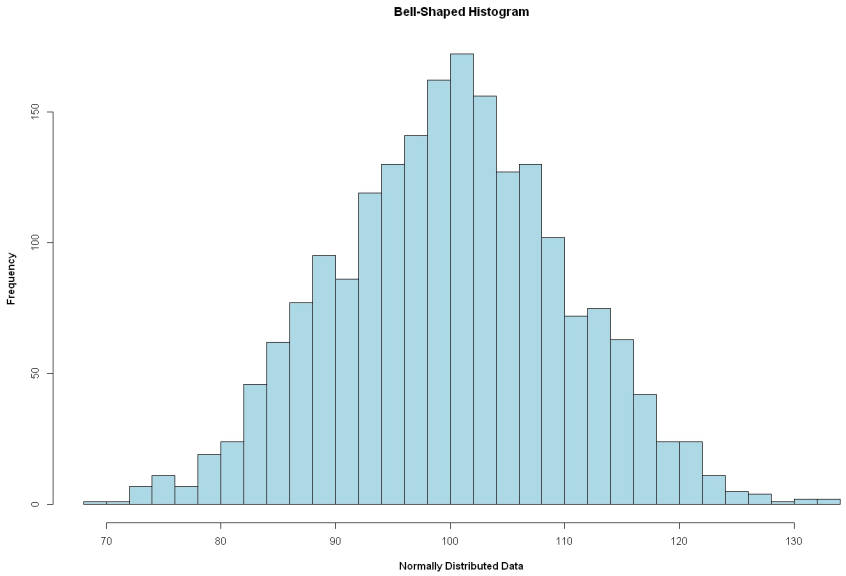
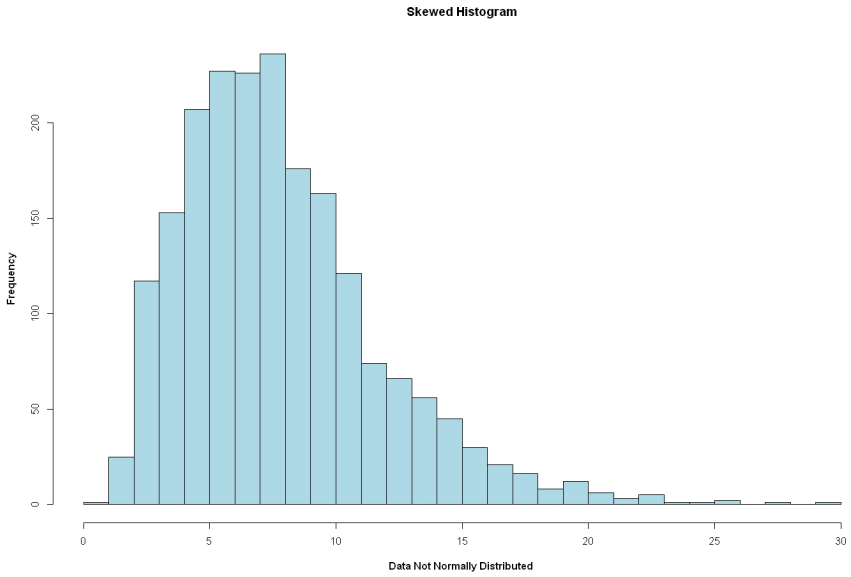

Bell-Shaped Histogram
Frequency
0
50
100
150
70
80
90
100
110
120
130
Normally Distributed Data

Figure 1:


Skewed Histogram
Frequency
0
50
100
150
200
0
5
10
15
20
25
30
Data Not Normally Distributed