The concepts of Statistical Process Control (SPC) were initially developed by **Dr. Walter Shewhart** of Bell Laboratories in the 1920's, and were expanded upon by **Dr. W. Edwards Deming**, who introduced SPC to Japanese industry after WWII. After early successful adoption by Japanese firms, Statistical Process Control has now been incorporated by organizations around the world as a primary tool to improve product quality by reducing process variation.

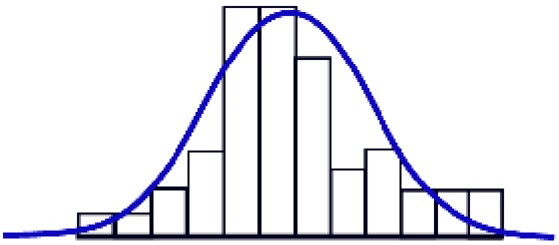
Dr. Shewhart identified two sources of process variation: **Chance** variation that is inherent in process, and stable over time, and **Assignable**, or **Uncontrolled** variation, which is unstable over time **-** the result of specific events outside the system. Dr. Deming relabeled chance variation as **Common Cause** variation, and assignable variation as **Special Cause** variation.

Based on experience with many types of process data, and supported by the laws of statistics and probability, Dr. Shewhart devised control charts used to plot data over time and identify both Common Cause variation and Special Cause variation.

This tutorial provides a brief conceptual background to the practice of SPC, as well as the necessary formulas and techniques to apply it.

Process Variability

If you have reviewed the discussion of frequency distributions in the Histogram module, you will recall that many histograms will approximate a Normal Distribution, as shown below (please note that **control charts do not require normally distributed data in order to work** - they will work with any process distribution **-** we use a normal distribution in this example for ease of representation):



In order to work with any distribution, it is important to have a measure of the data dispersion, or spread. This can be expressed by the range (highest less lowest), but is better captured by the standard deviation (sigma). The standard deviation can be easily calculated from a group of numbers using many calculators, or a spreadsheet or statistics program.

Example

Consider a sample of 5 data points: 6.5, 7.5, 8.0 , 7.2, 6.8

The **Range** is the highest less the lowest, or 8.0 – 6.5 = 1.5

The average ($\bar{X}$) is 7.2

The **Standard Deviation** (s) is:

$$s = \sqrt{\frac{[(6.5-7.2)^2+(7.5-7.2)^2+(8.0-7.2)^2+(7.2-7.2)^2+(6.8-7.2)^2]}{5-1}}$$

$$s = 0.59$$

Why Is Dispersion So Important?

Often we focus on average values, but understanding dispersion is critical to the management of industrial processes. Consider two examples:

- If you put one foot in a bucket of ice water (33 degrees F) and one foot in a bucket of scalding water (127 degrees F), on average you'll feel fine (80 degrees F), but you won't actually be very comfortable!
- If you are asked to walk through a river and are told that the average water depth is 3 feet you might want more information. If you are then told that the range is from zero to 15 feet, you might want to re-evaluate the trip.