Quality Corner

Determining Outliers

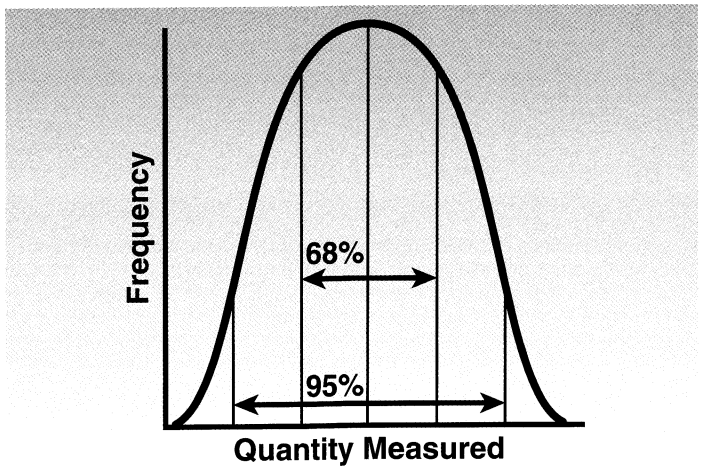
The last *Quality Corner* article explored the construction of a "Box Plot" to determine outlier data. It is important to exclude extreme measurements from a data set to eliminate bias from statistical evaluations. Z-scores provide another tool for determining outliers (Z-scores are measures of location). Extreme or atypical values often are referred to as outliers because of their location outside the normal distribution curve.

Normal distribution curve

When data measurements are normally distributed, we can make the following statements about the measurements. About 68% of the measurements will be within one standard deviation of the mean. About 95% of the measurements will be within two standard deviations of the mean. And all or most all of the measurements will be within three standard deviations of the mean.

Using Z-scores to detect outliers

Z-scores can be calculated for large sample sizes (greater than or equal to 30 data points). Data may not be normally distributed in smaller sample sizes. In general, a Z-score for a specific data point that is greater than 4 or less than -4 is a suspected outlier.



Calculating a Z-score

The formula for calculating a Z-score is:

$$Z = (X - \overline{X})/s$$

where:

Z = Z - score
X = individual measurement
$\overline{X}$ = the mean or average of the measurements
s = the standard deviation

Example: 60 measurements are made giving a mean value of 64.8 and a standard deviation of 13.75.

$$n = 60 \qquad \overline{X} = 64.8 \qquad s = 13.75$$

If one of the 60 measurement points is 125, the Z-score for that data point is calculated as:

$$Z = (125\text{-}64.8)/13.75$$
$$Z = 60.2/13.75$$
$$Z = 4.38$$

This calculated Z-score of 4.38 would indicate the measurement of 125 is an outlier of the normal distribution for that data set.

HACH COMPANY
P.O. Box 389
Loveland, Colorado 80539
Telephone: 970-669-3050
Toll-Free: 800-227-4224
Fax: 970-669-2932
Telex: 160840

Lit. No. 9545
Printed in USA

Reprinted from July 1996 *News & Notes for the Analyst.*