

How To: Apply SPC Techniques to Correlated Variables Using STATGRAPHICS Centurion

by

Dr. Neil W. Polhemus

July 27, 2005

Introduction

When analyzing the quality of manufactured products, it is common to measure more than one variable. For example, two important characteristics of a textile fiber are its tensile strength and its diameter. In applying SPC techniques to samples of data from such a process, it is common to consider each variable separately, even though the variables are likely to be correlated. This can lead to:

1. Erroneous estimates of the percent of the product that is beyond the specification limits.
2. Missed out-of-control signals on control charts.

In such cases, it is important to consider the variables simultaneously. In this guide, we will examine two useful procedures in STATGRAPHICS Centurion: the *Multivariate Capability Analysis* procedure and the *Multivariate T-Squared Control Chart*.

Sample Data

As an example, we will consider data on $n = 200$ fiber specimens. Measurements have been made of two variables: tensile strength and diameter. A portion of the data is shown below:

<i>Specimen</i>	<i>Tensile Strength</i>	<i>Diameter</i>	<i>LSL</i>	<i>Target</i>	<i>USL</i>
1	125.11	1.111	124	125	126
2	124.94	1.091	1.05	1.10	1.15
3	124.79	1.097			
4	124.99	1.105			
5	125.09	1.103			
6	124.86	1.088			
7	125.17	1.114			
8	125.18	1.115			

Figure 1: First 8 Rows of Sample Data

The specification limits for the fiber are:

$$124 \leq \text{tensile strength} \leq 126$$

$$1.05 \leq \text{diameter} \leq 1.15$$

The specification limits have been entered into additional columns in the data file.

Step 1: Plot the Data

The first step in analyzing any new set of data is to plot it. With two variables, a simple *X-Y Plot* will suffice:

Procedure: X-Y Scatterplot

To create an X-Y plot in STATGRAPHICS Centurion, simply press the X-Y Scatterplot  button on the main toolbar. On the data input dialog box, indicate the name of the columns containing the data:

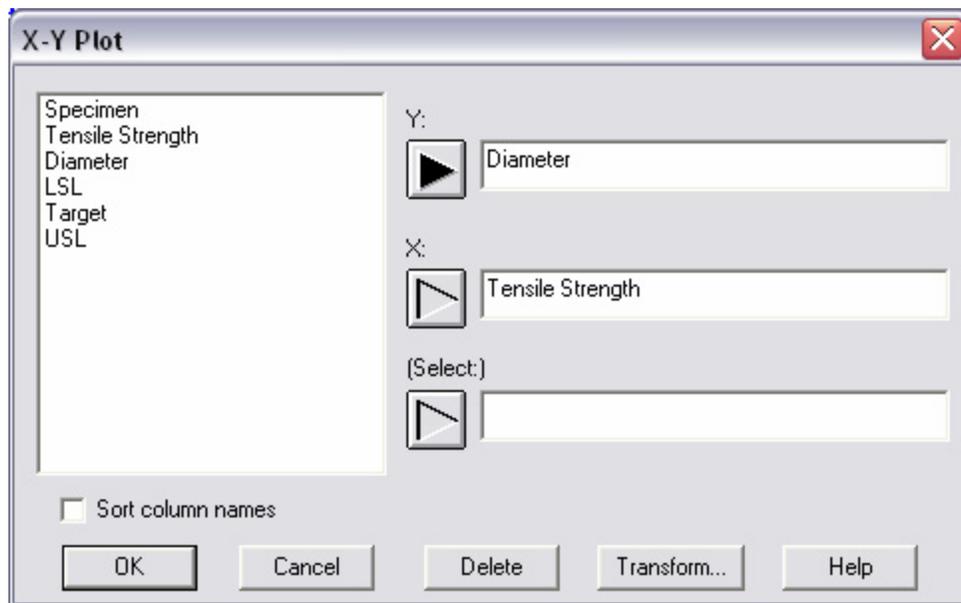


Figure 2: Data Input Dialog Box for X-Y Scatterplot

The resulting plot displays each of the 200 observations in the dataset:

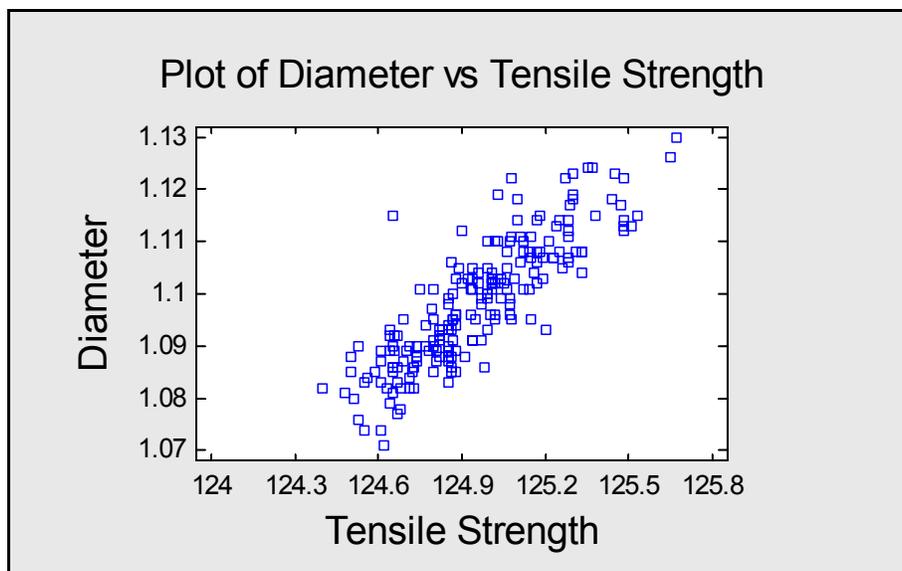


Figure 3: X-Y Scatterplot of Fiber Data

Notice the strong positive correlation between tensile strength and diameter.

To help determine whether the relationship between the variables is linear, press the *Smooth/Rotate* button  on the analysis toolbar. When the dialog box appears, request a *Robust LOWESS* smoother:

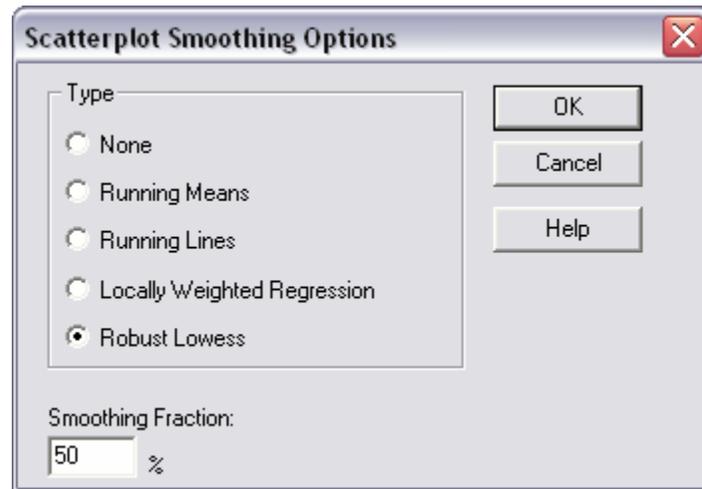


Figure 4: Dialog Box for Scatterplot Smoothing

LOWESS stands for *Locally Weighted Scatterplot Smoothing* and is a technique that can be applied to any X-Y scatterplot to help visualize the relationship between the variables plotted on each axis. It is made “robust” or resistant to outliers by smoothing the data twice, down-weighting points far from the first smooth when the second smooth is performed. The result is shown below:

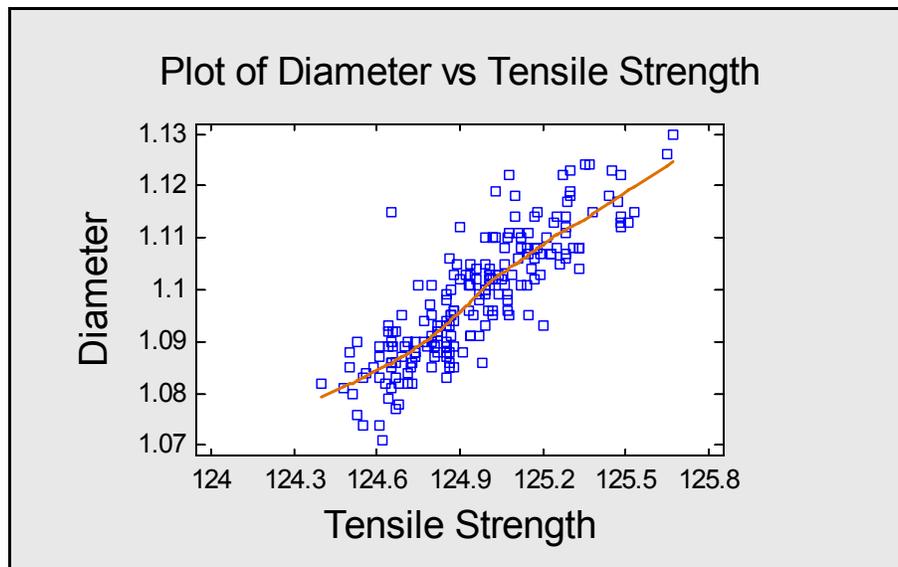


Figure 5: X-Y Plot with LOWESS Smooth

The relationship does appear to be nearly linear.

Step 2: Create Control Charts

To determine whether the process from which the data were obtained is in a state of statistical control, control charts may be generated.

Procedure: Individuals Charts

Assuming that the samples were taken one at a time over some sampling period, the standard approach is to construct an X chart and an MR(2) chart for each variable. In STATGRAPHICS Centurion, this is done by:

- If using the Classic menu, select: *SPC – Control Charts – Basic Variables Charts - Individuals*.
- If using the Six Sigma menu, select: *Control – Variables Control Charts – Basic Control Charts – Individuals*.

Complete the data input dialog box as shown below:

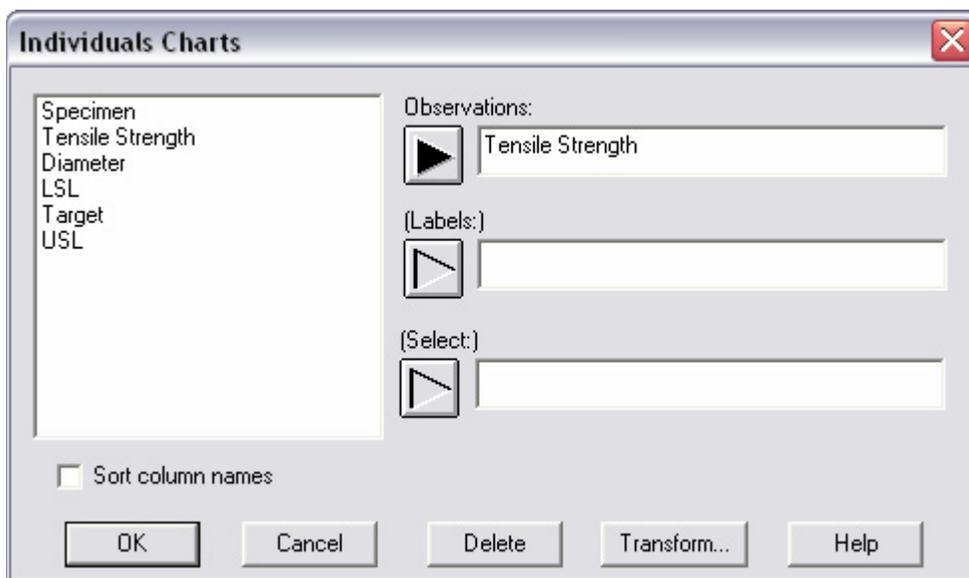


Figure 6: Data Input Dialog Box for Individuals Charts

The procedure creates an X-Chart and an MR(2) moving range chart. The X-Chart is shown below:

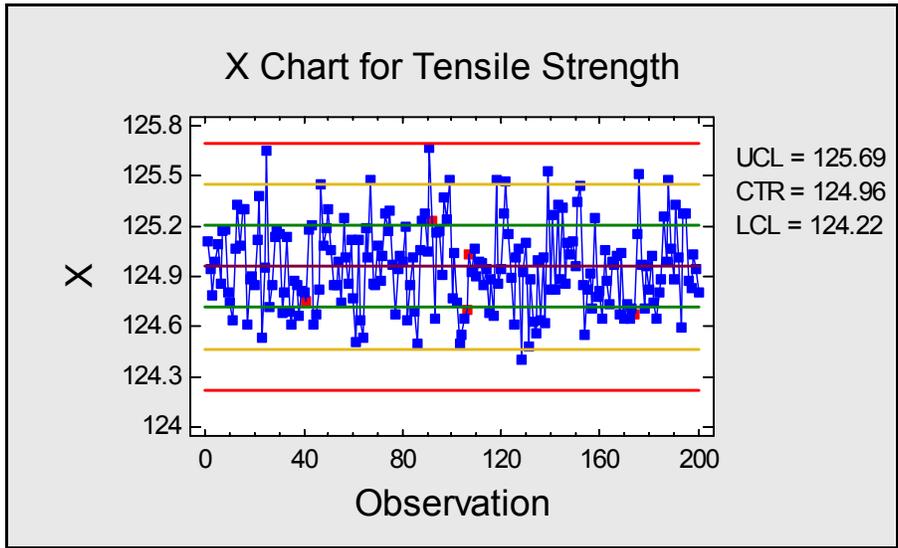


Figure 7: X-Chart for Tensile Strength

There are no observations beyond the 3-sigma control limits, although there are some runs rules violations that seem to correspond to groups of similar values.

To plot *Diameter*, press the *Input dialog* button  on the analysis toolbar and change the data variable:

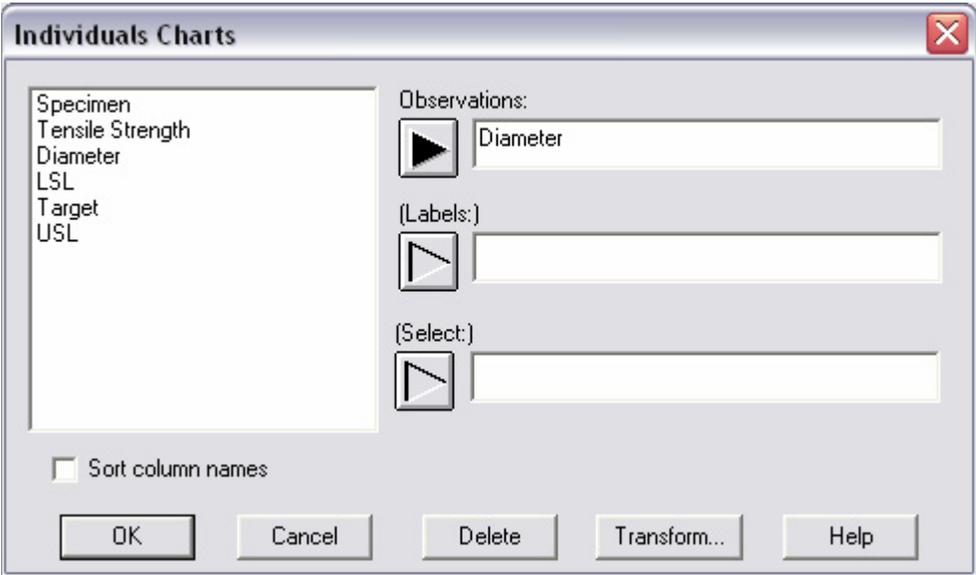


Figure 8: Modified Data Input Dialog Box

The resulting X-Chart again shows no observations beyond the 3-sigma limits:

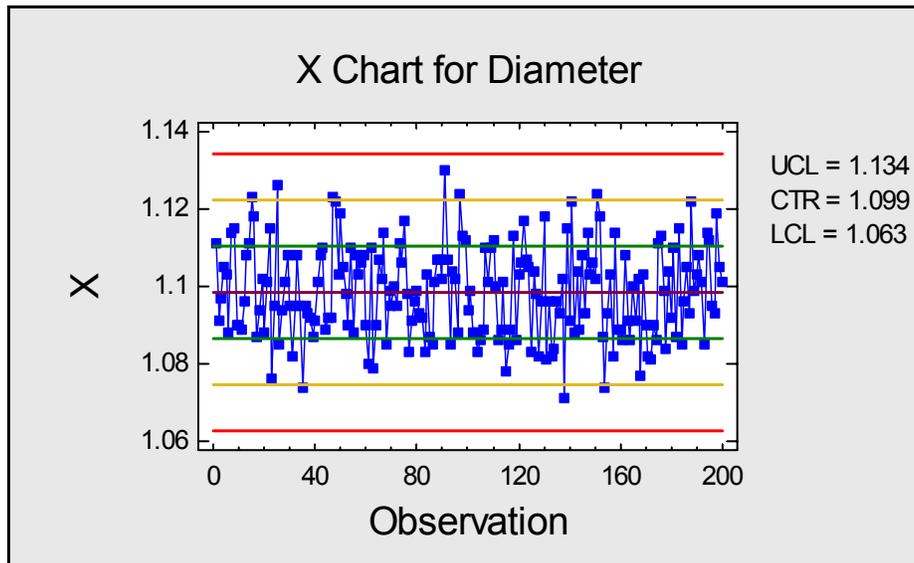


Figure 9: X Chart for Diameter

Again, there are no points beyond the 3-sigma control limits.

It should be noted that the control charts have been created in *Initial Studies* or Phase I mode, allowing the data to determine the location of the centerline and control limits. As such, the absence of points beyond the control limits leads to the conclusion that neither variable deviated unexpectedly far from the average of the 200 samples.

Procedure: Multivariate Control Charts

Plotting separate control charts for each variable is common practice and will detect cases in which a single variable is far from its expected value. When the variables are correlated, however, this does not tell the whole story. An additional chart should also be created which considers the *joint properties* of the variables. Such a chart would show any unusual samples with respect to the variables when considered together.

STATGRAPHICS Centurion provides two control charts for plotting multivariate data: a chart based on Hotelling's T-Squared statistic, and a multivariate EWMA chart. To construct the T-squared chart:

- If using the Classic menu, select: *SPC – Control Charts – Multivariate Control Charts – T-Squared Chart.*
- If using the Six Sigma menu, select: *Control – Variables Control Charts – Multivariate Control Charts – T-Squared Chart.*

Complete the data input dialog box as shown below:

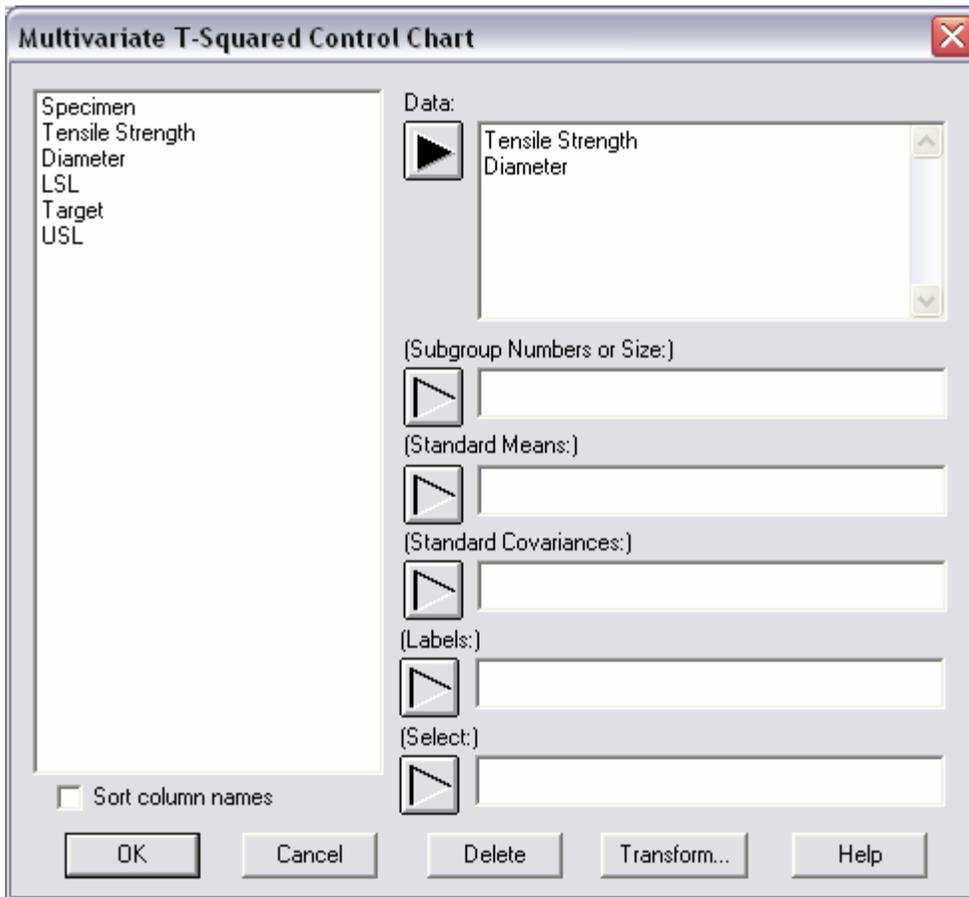


Figure 10: Multivariate T-Squared Control Chart Dialog Box

To create a Phase I chart, we need only enter the names of the variables to be considered. The procedure will then calculate the means, variances, and covariances for the set of variables and display the following chart:

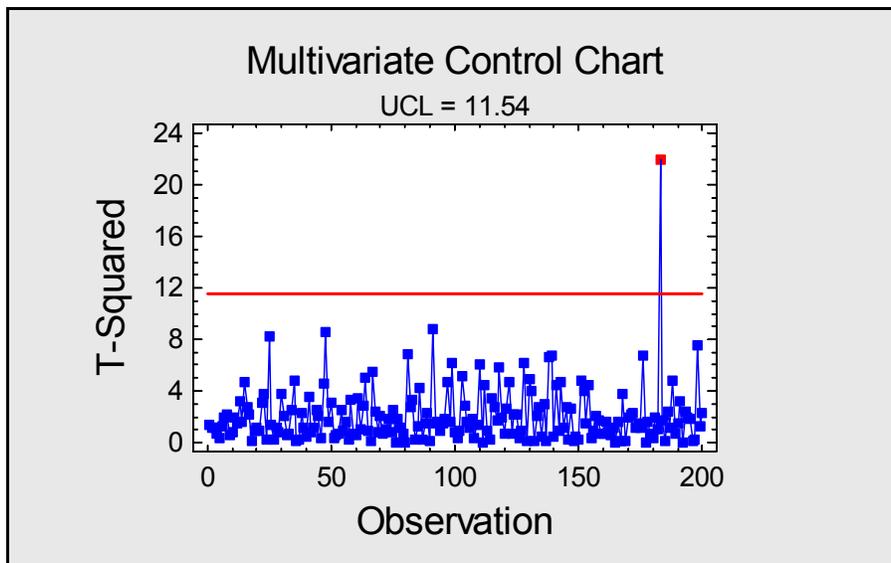


Figure 11: Multivariate T-Squared Control Chart

A point is plotted for each of the $n = 200$ samples, with an upper control limit. For the fiber data, one sample is way beyond the upper limit. Clicking on the red point with the mouse indicates that this sample is specimen #183.

A second very informative plot can be created by pressing the *Graphs* button  and selecting *Control Ellipse*:

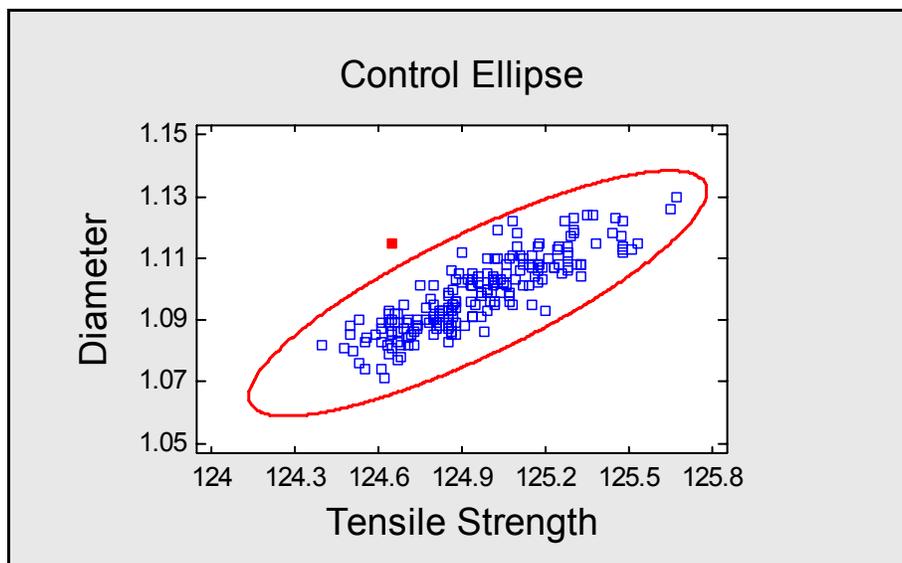


Figure 12: 3-Sigma Control Ellipse

This plot shows the measurements for each of the 200 specimens, together with a control ellipse. The control ellipse is similar to a 3-sigma limit, in the sense that it bounds 99.73% of the expected distribution of the samples. It is based on a multivariate normal distribution, with its centroid located at the sample means and the length and orientation of its axes calculated from the sample covariance matrix.

Notice that the unusual specimen is well within the overall limits for both *Tensile Strength* and *Diameter*. However, it is unusually weak for a specimen *with its diameter*. This sample would only be identified as unusual by considering both variables together. An attempt to determine the cause of the unusual specimen might well provide important information about the process.

Step 3: Calculate Process Capability

Except for the one unusual specimen, the process from which the textile specimens were obtained appears to be fairly stable. It thus makes sense to perform a capability analysis. The goal of such an analysis is to estimate the percentage of specimens that meet (or do not meet) the specification limits for both tensile strength and diameter.

Procedure: Process Capability Analysis

If the variables were not correlated, a separate capability analysis could be done on each variable. To perform a capability analysis on a single variable:

- If using the Classic menu, select: *SPC – Capability Analysis – Variables - Individuals*.
- If using the Six Sigma menu, select: *Analyze – Variable Data – Capability Analysis – Individuals*.

Beginning with *Tensile Strength*, the dialog input dialog box should be completed as shown below:

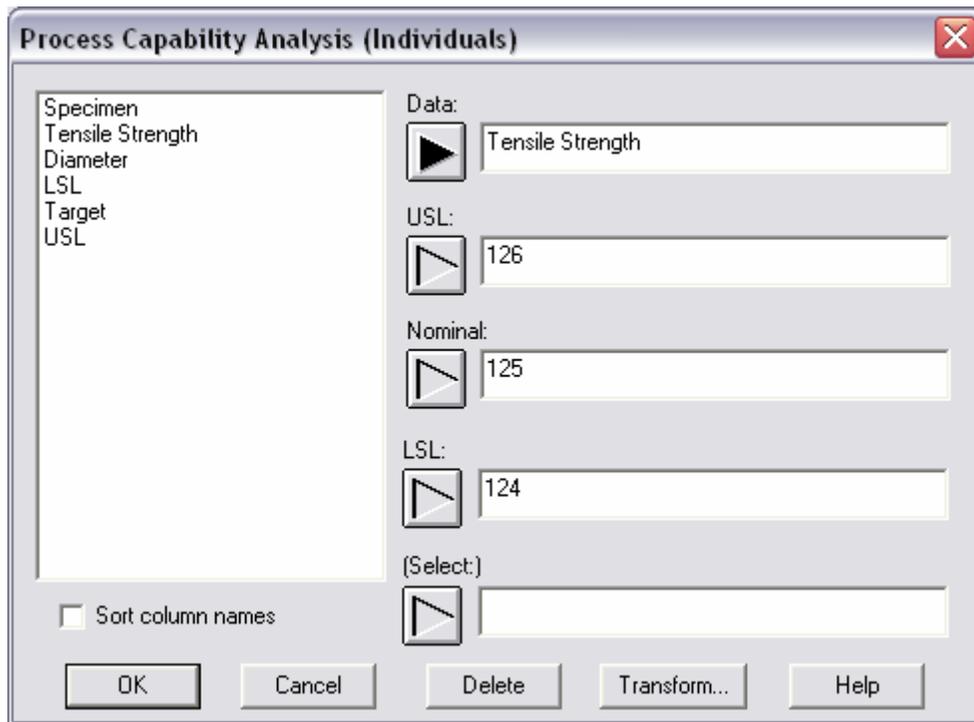


Figure 13: Data Input Dialog Box for Process Capability Analysis

The procedure creates a plot displaying a histogram of the sample data, a fitted normal distribution, and several common capability indices:

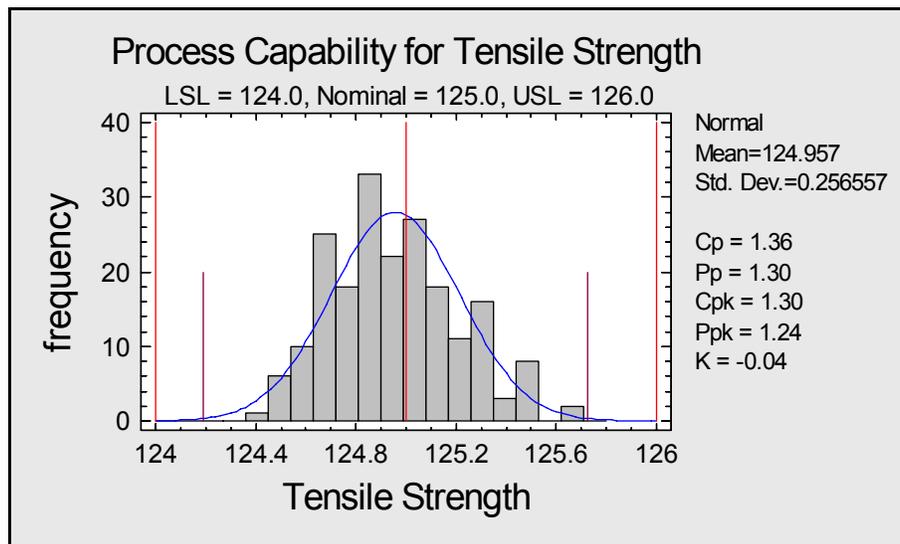


Figure 14: Process Capability Plot for Tensile Strength

Also displayed is a table of capability indices:

Capability Indices for Tensile Strength

Specifications
 USL = 126.0
 Nom = 125.0
 LSL = 124.0

	Short-Term	Long-Term
	Capability	Performance
Sigma	0.24542	0.256557
Cp/Pp	1.35821	1.29926
Zusl	4.24842	4.06401
Zlsl	3.90086	3.73153
Zmin	3.90086	3.73153
Cpk/Ppk	1.30029	1.24384
K		-0.04265
DPM	58.7163	119.321
Sigma Quality Level	5.35143	5.17416

Based on 6 sigma limits. Short-term sigma estimated from average moving range. The Sigma Quality Level includes a 1.5 sigma drift in the mean.

Figure 15: Estimated Capability Indices for Tensile Strength

Of major interest are:

1. The one-sided capability indices C_{pk} and P_{pk} . These compare the distance from the sample mean to the nearer specification limit, divided by $3\hat{\sigma}$. Indices are shown using a “short-term” estimate of sigma obtained from the moving range, and a “long-term” estimate based on the sample standard deviation. The higher the value of these indices, the better. Most companies have a goal of at least 1.33 for both indices.
2. The long-term estimate leads to an estimated defects per million (DPM) of 119.3. This represents the estimated number of specimens in every million that would fall outside the specification limits.
3. The Sigma Quality Level (SQL) is 5.17. In a typical Six Sigma program, the goal is to achieve an SQL of 6, which equates to 3.4 DPM.

Repeating the same analysis on *Diameter* yields:

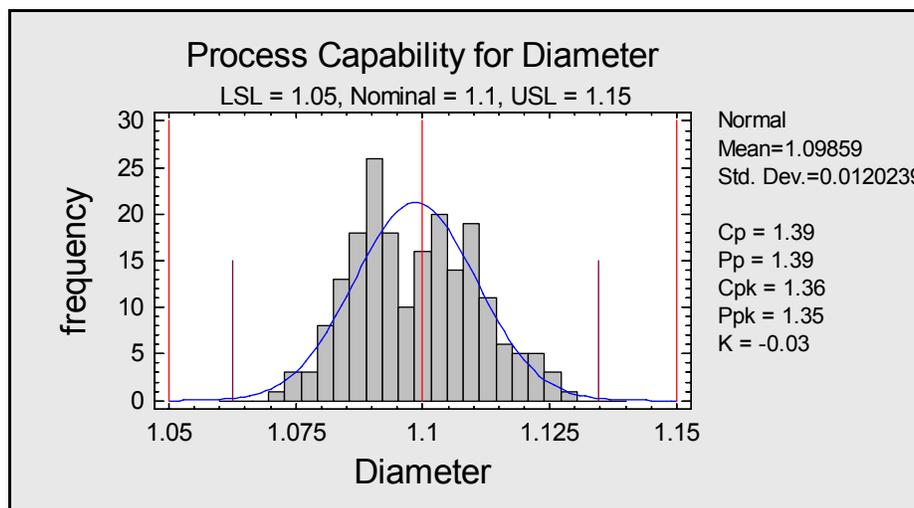


Figure 16: Capability Plot for Diameter

The calculated capability indices are:

Capability Indices for Diameter		
Specifications		
USL = 1.15		
Nom = 1.1		
LSL = 1.05		
	<i>Short-Term</i>	<i>Long-Term</i>
	<i>Capability</i>	<i>Performance</i>
Sigma	0.011948	0.0120239
Cp/Pp	1.39493	1.38613
Zusl	4.3028	4.27564
Zlsl	4.06678	4.04111
Zmin	4.06678	4.04111
Cpk/Ppk	1.35559	1.34704
K		-0.0282
DPM	32.285	36.1486
Sigma Quality Level	5.49546	5.4686

Based on 6 sigma limits. Short-term sigma estimated from average moving range. The Sigma Quality Level includes a 1.5 sigma drift in the mean.

Figure 17: Capability Plot for Diameter

The long-term estimate of DPM for diameter is 36.1.

Procedure: Multivariate Capability Analysis

We have now estimated that about 119 of every million fibers will be out of spec with respect to tensile strength and about 36 of every million out of spec with respect to diameter. If the variables were independent, the total DPM would be very nearly the sum of the two, i.e., DPM = 155. Although the chance of being out of spec on both variables at the same time exists, it is very small *if the variables are independent*:

$$\text{Prob}(\text{out of spec on both if independent}) = \frac{119}{1,000,000} \times \frac{36}{1,000,000} = \frac{0.004}{1,000,000}$$

To calculate the Sigma Quality Level of the combined variables, we can use the STATGRAPHICS Centurion *Six Sigma Calculator*, found under *Tools – Six Sigma Calculator*. Enter the combined DPM in the top section and press the *Calculate* button:

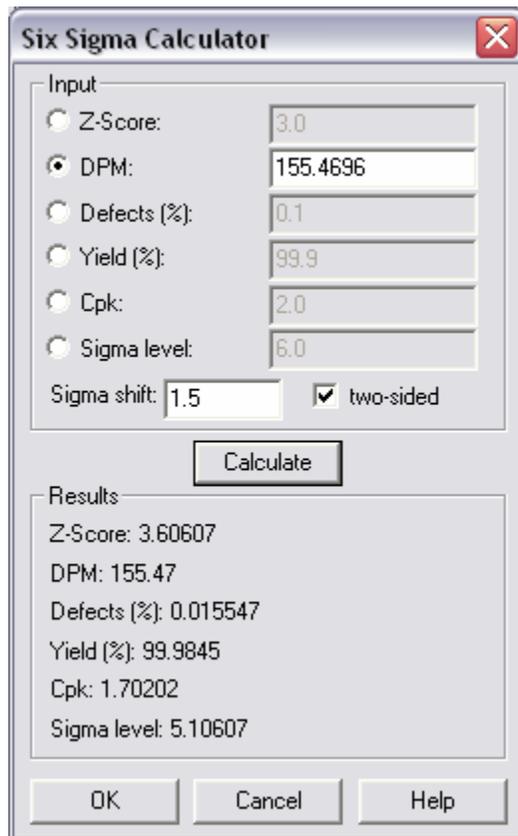


Figure 18: Six Sigma Calculator

The calculator shows that, considering the combined DPM for the two variables, the process is operating at an SQL of approximately 5.11.

In fact, tensile strength and diameter are not independent, so it is incorrect to sum the DPM of the two variables. Instead, the *Multivariate Capability Analysis* procedure should be used:

- If using the Classic menu, select: *SPC – Capability Analysis – Variables – Multivariate Capability*.
- If using the Six Sigma menu, select: *Analyze – Variable Data – Capability Analysis – Multivariate Capability*.

Enter the names of the variables in the study, together with the names of columns containing the specification limits and target values:

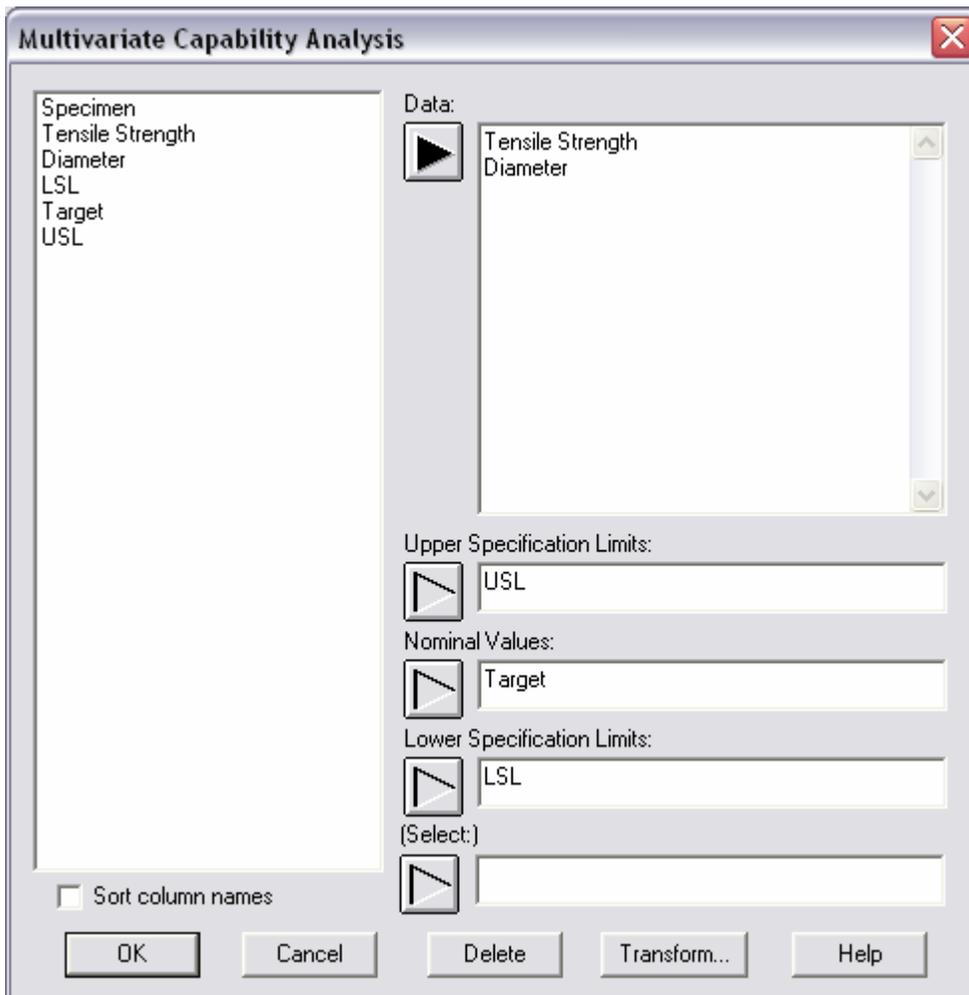


Figure 19: Data Input Dialog Box for Multivariate Capability Analysis

The latter columns contain one entry for each variable, as shown earlier in Figure 1. The *Analysis Summary* produces a table that summarizes the observed and estimated DPM:

Multivariate Capability Analysis

Data variables:
 Tensile Strength
 Diameter

Number of complete cases: 200

	<i>Sample</i>	<i>Sample</i>			
<i>Variable</i>	<i>Mean</i>	<i>Std. Dev.</i>	<i>LSL</i>	<i>Nominal</i>	<i>USL</i>
Tensile Strength	124.957	0.256557	124.0	125.0	126.0
Diameter	1.09859	0.0120239	1.05	1.1	1.15

	<i>Observed</i>	<i>Estimated</i>	<i>Estimated</i>
<i>Variable</i>	<i>Beyond Spec.</i>	<i>Beyond Spec.</i>	<i>DPM</i>
Tensile Strength	0.0%	0.0119321%	119.321
Diameter	0.0%	0.00361486%	36.1486
Joint	0.0%	0.0149248%	149.248

The StatAdvisor
 This procedure determines the percentage of items beyond a set of multivariate specification limits. In this case, the estimated frequency of non-conformities with respect to at least one of the 2 variables equals 149.248 per million.

Figure 20: Analysis Summary for Multivariate Capability Analysis

Considering both variables together, the estimated DPM is 149.2, not 155.5. This is because the strong positive correlation between the variables creates a non-trivial chance that when one variable is out of spec, the other variable will be also.

If you press the *Tables* button  on the analysis toolbar, a table of *Capability Indices* may be displayed:

Capability Indices	
Index	Estimate
MCP	1.21
MCR	82.95
DPM	149.248
Z	3.61661
SQL	5.11661

Based on 6 sigma limits. The Sigma Quality Level includes a 1.5 sigma drift in the mean.

Figure 21: Table of Multivariate Capability Indices

In addition to the DPM, the table shows:

1. *MCP* – a multivariate capability index similar to C_{pk} . It is equal to the value of C_{pk} that yields the observed DPM when considering only one variable.
2. *SQL* – the Sigma Quality Level corresponding to the estimated DPM. A DPM of 149.2 equates to an SQL of approximately 5.12.

It will be noted that the estimated number of defects is about 4% lower when allowing for the correlation between the variables. Depending on circumstances, the difference could be larger or smaller than in this example. The more variables that are considered and the stronger the correlations between them, the more important it becomes to use the multivariate approach.

Also created by default is a *Multivariate Capability Plot*:

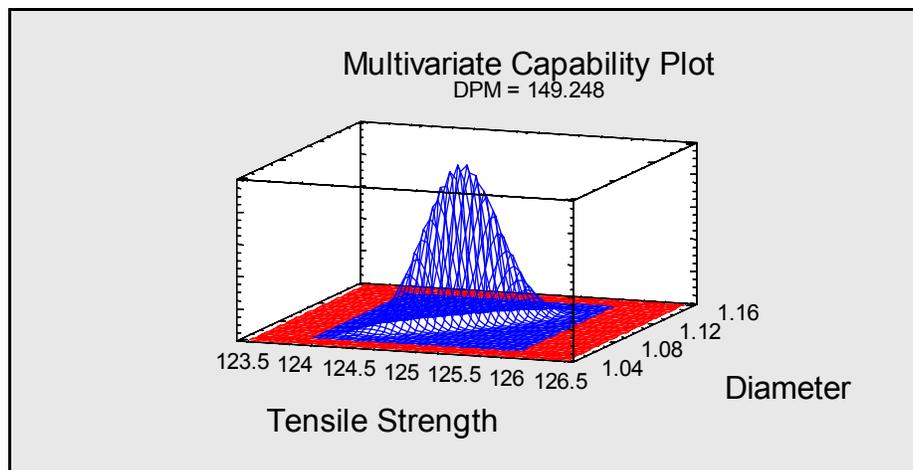


Figure 22: Plot of Fitted Multivariate Normal Distribution

This plot shows the multivariate normal distribution from which the estimate of DPM was obtained. The red area corresponds to values that are outside the specification limits.

Also of interest is the *Capability Ellipse*:

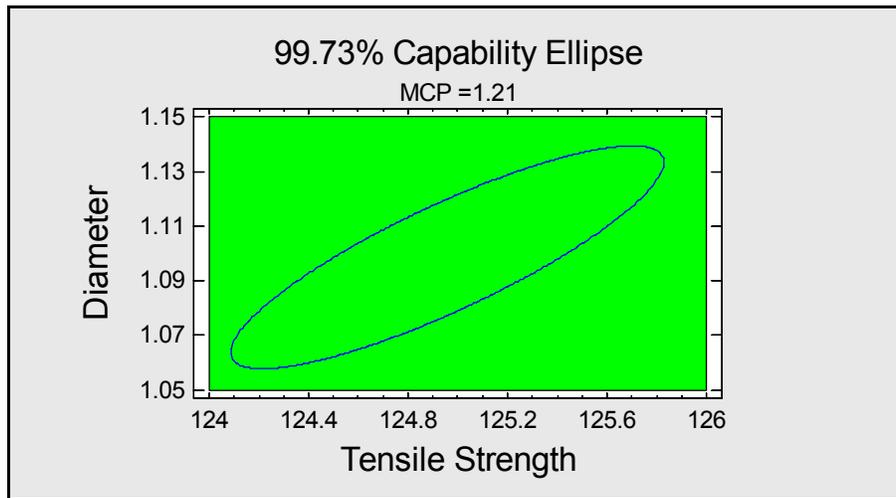


Figure 23: Capability Ellipse

The ellipse contains 99.73% of the multivariate normal distribution. The green rectangle corresponds to the specification limits. A highly capable process will contain the ellipse comfortably within the rectangle.

Conclusion

When applying SPC techniques to multiple variables, it can be important to treat them as a multivariate sample rather than as independent measurements. If a strong correlation exists between the variables, a multivariate control chart will sometimes pick up unusual events that separate control charts for each variable would miss. When estimating the fraction of product outside the specification limits, use of a multivariate normal distribution gives a more accurate estimate of the DPM than simply summing the DPM's for each of the variables.

Although the computations required to treat variables from a multivariate perspective are somewhat involved and can be time-consuming, it is important to consider related variables together. Taking a multivariate approach can show features that might otherwise be missed, which can lead to important process improvements.

Note: The author welcomes comments about this guide. Please address your responses to neil@statgraphics.com.