

## Multivariate T-Squared Control Chart

### Summary

The **Multivariate T-Squared Control Chart** procedure creates control charts for two or more numeric variables. Examining the variables in a multivariate sense is extremely important when the variables are highly correlated, since joint out-of-control conditions can occur without any individual variable violating its control limits when plotted separately.

The procedure creates a T-Squared chart to monitor the means of each variable and a generalized variance control chart to monitor the variances and covariances.

**Sample StatFolio:** *mvchart.sgp*

### Sample Data:

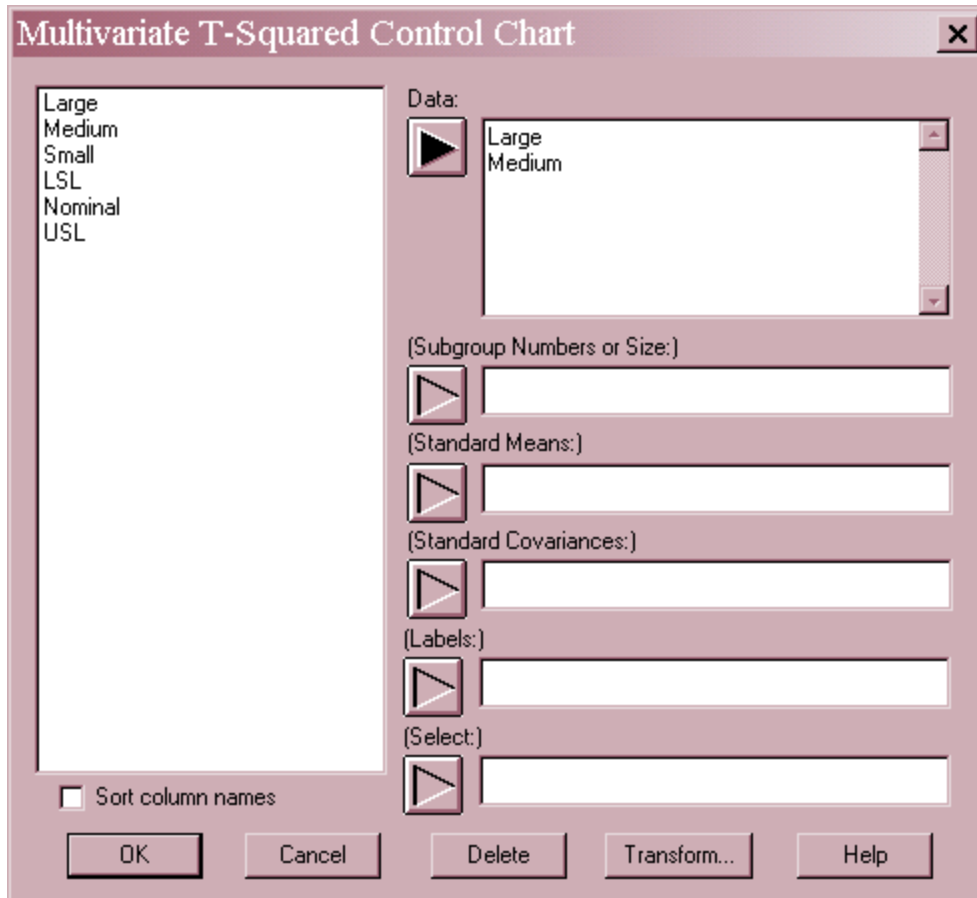
The file *grit.sf3* contains measurements made on  $n = 56$  batches of “grit”, from Holmes and Mergen (1993). The data represent the percentages of large, medium, and small particles in the grit. The table below shows a partial list of the data in that file:

<i>Large</i>	<i>Medium</i>	<i>Small</i>
5.4	93.6	1.0
3.2	92.6	4.2
5.2	91.7	3.1
3.5	86.9	9.6
2.9	90.4	6.7
4.6	92.1	3.3
4.4	91.5	4.1
5.0	90.3	4.7
8.4	85.1	6.5
4.2	89.7	6.1
3.8	92.5	3.7
4.3	91.8	3.9
3.7	91.7	4.6
3.8	90.3	5.9
2.6	94.5	2.9

Since the percentages in the three columns sum to 100% in each row, it is only necessary to create a chart based on the first 2 columns.

## Data Input

The data to be analyzed consist of  $p = 2$  or more numeric columns containing the variables of interest.



- **Data:** 2 or more numeric columns containing the  $n$  samples, one sample per row.
- **Subgroup numbers or size:** If the data were obtained as individuals, leave this field blank or enter “1”. If the data were collected in subgroups, each of size  $m$ , enter the single value  $m$ . In such a case, each consecutive  $m$  rows in the file will be considered to form a subgroup. If the subgroup sizes are not equal, enter the name of an additional numeric or non-numeric column containing group identifiers. The program will scan this column and place all sequential rows with identical codes into the same group.
- **Standard Means:** For an “initial study” or Phase I analysis where the data will be used to determine the control limits, leave this field blank. For a “control-to-standard” or Phase II analysis, enter the name of a column contain  $p$  means.
- **Standard Covariances:** For an “initial study” or Phase I analysis, leave this field blank. For a “control-to-standard” or Phase II analysis, enter the name of a column contain the  $p^2$  variances and covariances. In entering the values in a covariance matrix, enter the values in the first row of the matrix, then the values in the second row, and so forth. Note: if you select *Save Results* after performing a Phase I analysis, the covariance matrix will be saved in this exact format.

- **Labels:** optional labels for each subgroup. The labels will be applied in sequence to the subgroups when plotting the control charts.
- **Select:** subset selection.

### Analysis Summary

The *Analysis Summary* shows the number of observations included in the analysis and the location of the control limits on the control charts.

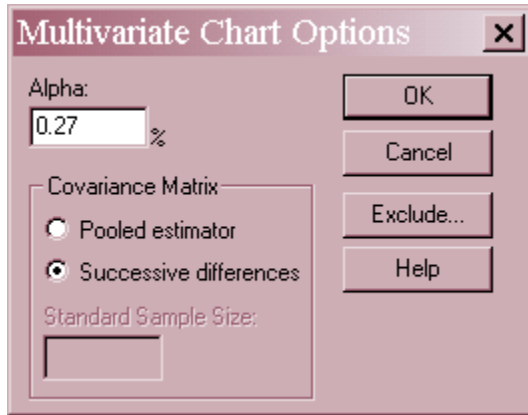
<u>Multivariate T-Squared Control Chart</u>				
Data variables:				
Large				
Medium				
Number of observations included = 56				
Number of observations excluded = 0				
<b>Phase 1 - covariance estimated from current data using successive differences</b>				
<i>Chart</i>	<i>Alpha</i>	<i>LCL</i>	<i>UCL</i>	<i>Beyond limits</i>
T-Squared	0.0027	0.0	10.8055	3

Included in the table are:

- **Phase:** If Phase 1, the method for estimating the covariance matrix is displayed. If Phase 2, the assumptions about the input covariance matrix are shown.
- **Chart:** the type of chart. For individuals data, only a T-Squared chart is created. For grouped data, a generalized variance chart is included if the subgroup size exceeds the number of variables.
- **Alpha:** the false alarm probability of the chart, specified using *Analysis Options*. For standard 3-sigma control charts,  $\alpha = 0.0027$ .
- **LCL:** the lower control limit.
- **UCL:** the upper control limit.
- **Beyond limits:** the number of points on the control chart that are beyond the control limits.

In the example, the  $n = 56$  samples showed 3 points beyond the upper control limit on the T-squared chart when estimating the covariances using the method of successive differences.

## Analysis Options



- **Alpha:** the false alarm probability for positioning the control limits. For the equivalent of a standard 3-sigma control chart, set  $\alpha = 0.27\%$ .
- **Covariance Matrix:** details about the covariance matrix. There are 4 possibilities:
  1. If the *Standard Covariances* field was left blank on the data input dialog box and the data are in subgroups, no entry is necessary since the covariances will be estimated from the pooled within-group variability.
  2. If the *Standard Covariances* field was left blank on the data input dialog box and the data are individuals, select *Pooled estimator* to estimate the covariance between variables  $j$  and  $k$  using the usual estimator

$$s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \quad (1)$$

Select *Successive differences* to estimate the covariance using

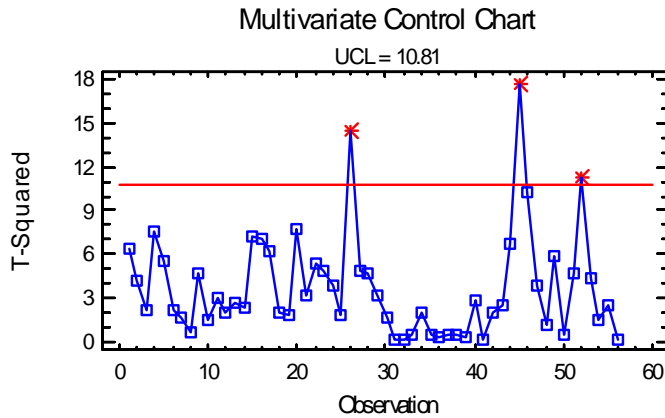
$$s_{jk} = \frac{1}{2(n-1)} \sum_{i=1}^n (x_{ij} - x_{i-1,j})(x_{ik} - x_{i-1,k}) \quad (2)$$

The second estimator is a more “local” estimator, in the sense that it captures only short-term variability, in a similar manner to the way in which a moving range is used to estimate variability for a standard individuals chart.

3. If an entry was made in the *Standard Covariances* field and the estimates provided were obtained from a previous sample, enter the size of that previous sample (if grouped, the number of subgroups) in the *Standard Sample Size* field.
4. If an entry was made in the *Standard Covariances* field and the covariances are assumed to be known, leave the *Standard Sample Size* field empty.

## T-Squared Chart

The *T-Squared Chart* shows the value of  $T^2$  for each data value or subgroup:



The  $i$ -th value of T-squared is calculated from

$$T_i^2 = (x_i - \bar{x})' S^{-1} (x_i - \bar{x}) \tag{3}$$

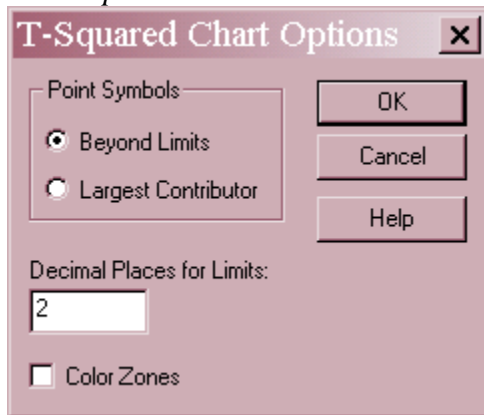
for individuals data and from

$$T_i^2 = (\bar{x}_i - \bar{x})' S^{-1} (\bar{x}_i - \bar{x}) \tag{4}$$

for grouped data, where  $x_i$  is the vector of observed values corresponding to the  $i$ -th sample,  $\bar{x}$  is the vector of sample means, and  $S$  is the covariance matrix. In simple terms,  $T^2$  measures the distance of each observation or subgroup from the centroid of the data, in a standardized scale that compensates for the difference in measurement units and variability. A point above the upper control limits is a signal of a potential out-of-control condition.

For the sample data, 3 out-of-control signals are given by the chart.

### Pane Options

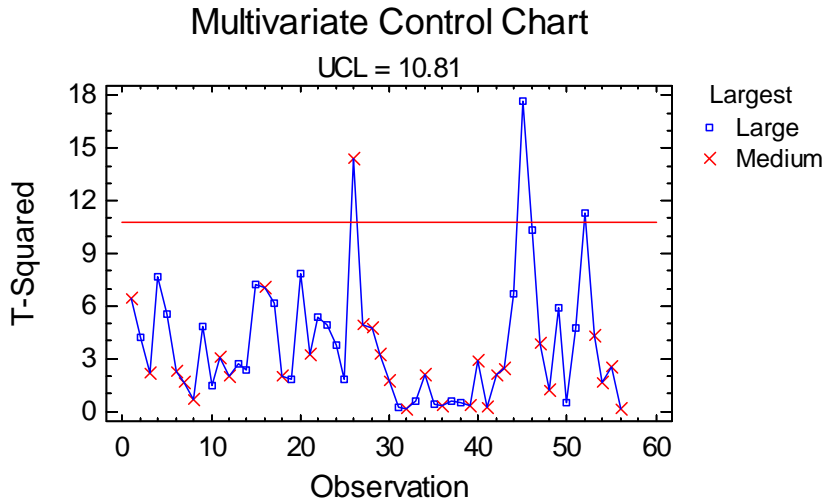


- **Point Symbols:** select *Beyond Limits* to draw special point symbols only for points falling above the control limit. Select *Largest Contributor* to identify the variable that contributes most to each value of  $T^2$ .

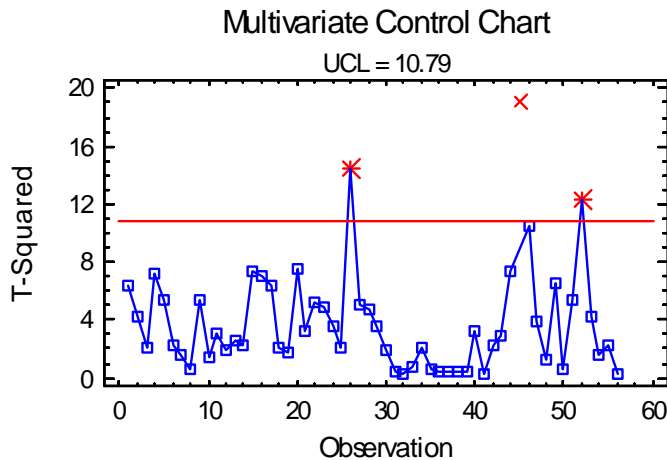
- **Decimal Places for Limits:** number of decimal places for displaying the control limit.
- **Color Zones:** check this box to display green and red zones.

Example: Identifying Largest Contributor

If *Largest Contributor* is selected, the chart will take the following form:



Each point on the chart is coded according to the variable that contribute the most to the value of  $T^2$ . For example, the biggest contributor to the problems in the latter half of the data appears to be the percentage of *Large* particles. If you wish to exclude points from the analysis, you can click on a point and then press the *Exclude/Include* button on the analysis toolbar. The chart will be recalculated without those points.



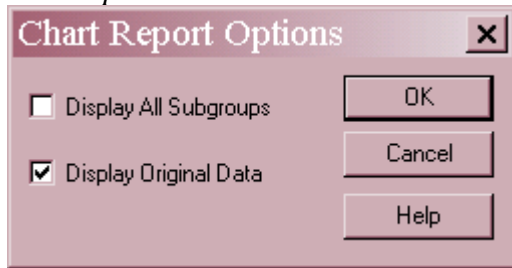
**Multivariate Control Chart Report**

This pane tabulates the points on the control chart:

Multivariate Chart Report			
Observation	T-Squared	Large	Medium
26	* 14.3721	7.3	79.0
45	* 17.6655	10.9	82.2
52	* 11.2594	9.6	80.6

An asterisk indicates any value beyond the control limits. Any points excluded from the analysis using the *Exclude* button are indicated with an X.

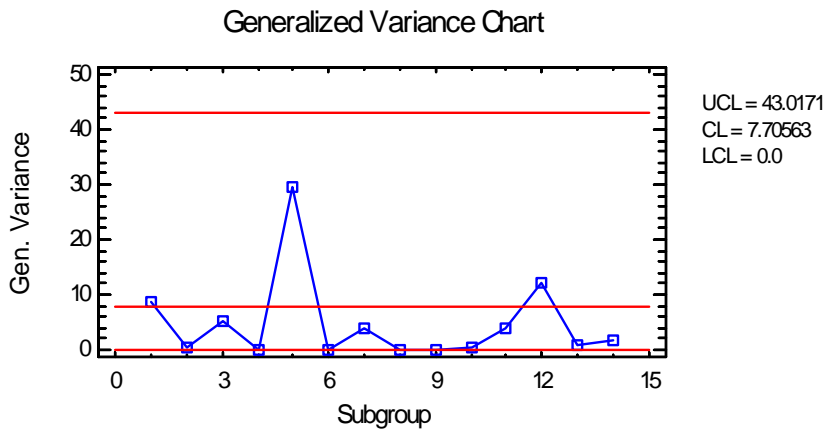
*Pane Options*



- **Display All Subgroups** – if checked, all observations or subgroups will be tabulated. Otherwise, only points beyond the control limits will be included in the table.
- **Display Original Data** – if checked, the values of each variable will be displayed. Otherwise, only the control chart values will be tabulated.

**Generalized Variance Chart**

A T-Squared chart is designed to monitor the means of  $p$  variables. To monitor the variance, a Generalized Variance Chart by sometimes be displayed:

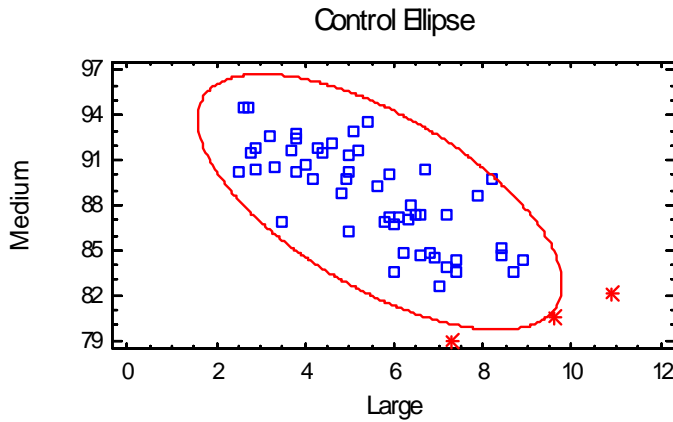


This chart is created only for data arranged in subgroups, and only if the subgroup size is at least  $p + 1$ . The generalized variance  $|S_i|$  for the  $i$ -th subgroup is defined as the determinant of its variance-covariance matrix.

The above chart shows the grit data grouped in subgroups of 4 consecutive observations each. Any point beyond the upper control limit would indicate an unusually large amount of variability within that subgroup. In this case, no such points are present.

## Control Ellipse

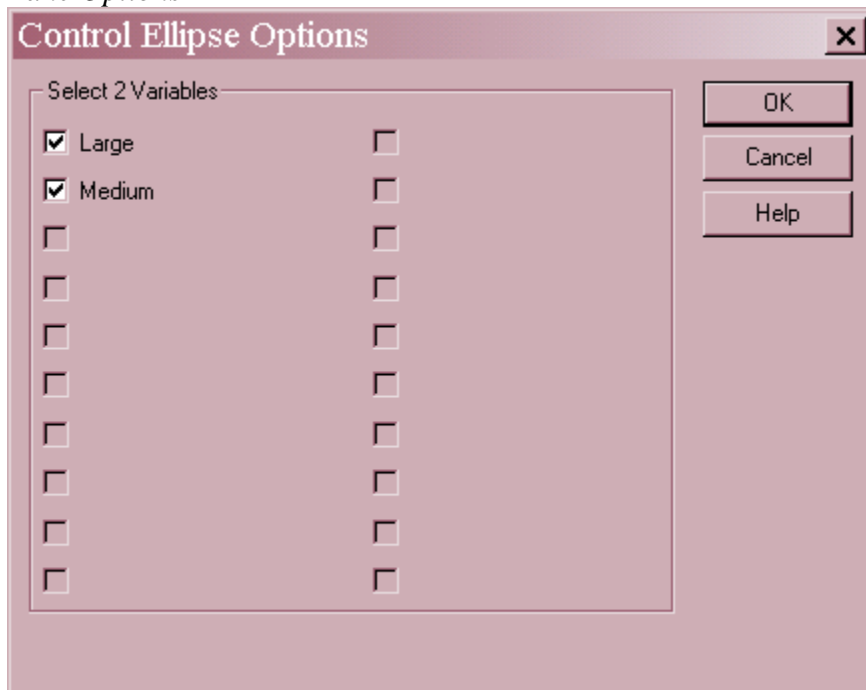
If out-of-control signals are shown on the control chart, it is useful to examine those values in detail. A good chart to use in the case of  $p = 2$  variables is the *Control Ellipse*:



The upper control limit on the T-Squared chart corresponds to an elliptical region in the space of any two of the variables, with the other variables held at a fixed value. For  $p = 2$ , any out-of-control signals will show up as points outside the ellipse.

In the sample data, it will be noticed that the 3 out-of-control signals all correspond to low percentages of medium particles.

### Pane Options



- Select 2 variables:** select any 2 variables to define the control ellipse. The ellipse will be plotted assuming that all other variables are held at their mean levels. Care should be taken in interpreting the plot if  $p > 2$ , since the true elliptical control region for each point depends on the value of the variables that are not plotted.



### T-Squared Decomposition

The  $T^2$  statistic can be decomposed into components attributable to each of the variables. One method for measuring the contribution of the  $j$ -th variable to an out-of-control  $T^2$  value is by looking at how much smaller  $T^2$  would be if the  $j$ -th variable was not included in the analysis. The *T-Squared Decomposition* pane does such a decomposition for each out-of-control signal on the T-Squared chart:

T-Squared Decomposition			
Relative Contribution to T-Squared Signal			
Observation	T-Squared	Large	Medium
26	14.3721	1.72503	12.6968
45	17.6655	12.2741	0.240343
52	11.2594	2.6211	1.43539

Following Runger, Alt and Montgomery (1996), the table shows

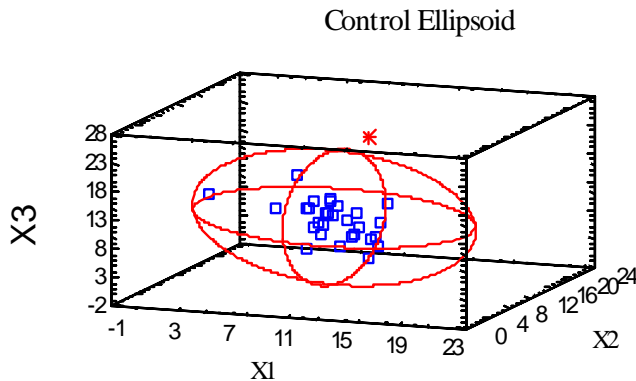
$$d_j = T^2 - T_{(j)}^2 \tag{5}$$

where  $T_{(j)}^2$  is the value of the statistic using all variables except the  $j$ -th.

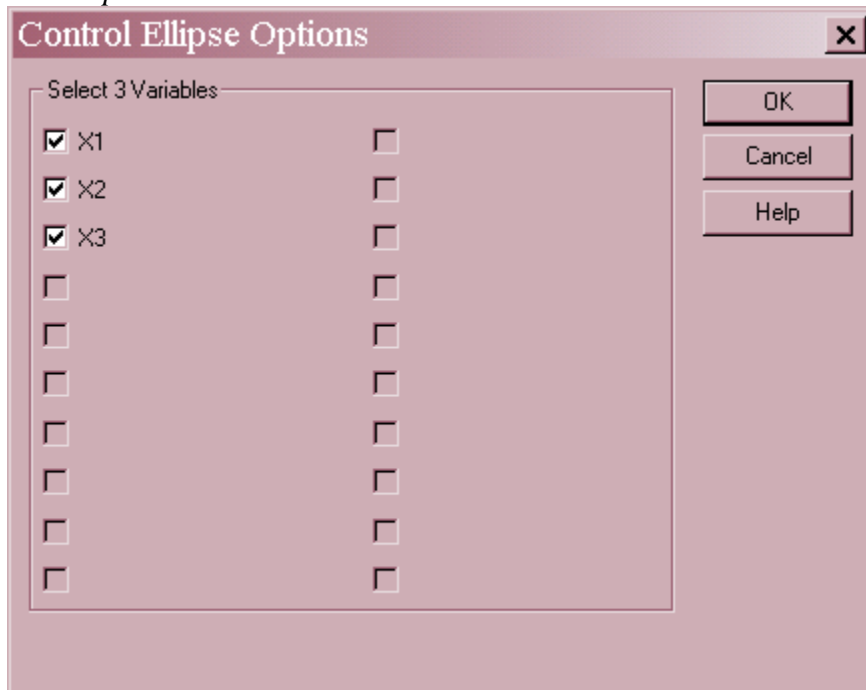
For the current data, *Medium* appears to be the dominant variable for row 26, while *Large* appears to be the dominant variable for row 45. For row 52, neither variable on its own reduces  $T^2$  by much, meaning that both variables contribute to the out-of-control signal.

### 3-D Control Chart

When the data consist of  $p = 3$  variables, a 3-D control chart can be very helpful, since the control region is then an ellipsoid.



The above plot shows the outline of a typical control ellipsoid for 3 variables. One data point is outside the control ellipse. Note: to explore this plot, it is very helpful to use the dynamic rotation button on the analysis toolbar.

*Pane Options*

- **Select 3 variables:** select any 3 variables to define the control ellipsoid. The ellipsoid will be plotted assuming that all other variables are held at their mean levels. Care should be taken in interpreting the plot if  $p > 3$ , since the true control region for each point depends on the value of the variables that are not plotted.

**Save Results**

The following results can be saved to the datasheet:

1. *T-Squared* – the calculated  $T^2$  values for each observation or subgroup.
2. *Means* – the  $p$  variable means.
3. *Covariances* – the  $p^2$  variances and covariances in rowwise order.
4. *Labels* – the labels corresponding to each value of  $T^2$ .
5. *Generalized Variance* - if calculated, the values of  $|S|$ .

Calculations**T-Squared Control Limit if Covariances are Known**

$$UCL = X_{\alpha,p}^2 \quad (6)$$

**T-Squared Control Limit if Covariances are Estimated from  $k$  Previous Samples**

$$UCL = \frac{p(k+1)(k-1)}{k(k-p)} F_{\alpha,p,k-p} \quad (7)$$

**T-Squared Control Limit if Covariances are Estimated from Current Data**

$$UCL = \frac{(n-1)^2}{n} \text{Beta}_{\alpha,p/2,(n-p-1)/2} \quad (8)$$

**Generalized Variance Control Limits**

$$UCL = |\Sigma| (b_1 + 3\sqrt{b_2}) \quad (9)$$

$$CL = b_1 |\Sigma| \quad (10)$$

$$LCL = |\Sigma| (b_1 - 3\sqrt{b_2}) \quad (11)$$

where

$$b_1 = \frac{1}{(n-1)^p} \prod_{i=1}^p (n-i) \quad (12)$$

$$b_2 = \frac{1}{(n-1)^{2p}} \prod_{i=1}^p (n-i) \left[ \prod_{j=1}^p (n-j+2) - \prod_{j=1}^p (n-j) \right] \quad (13)$$

If  $\Sigma$  is not known, it is replaced by the estimate  $|\mathbf{S}|/b_1$ .