Multivariate Tolerance Limits

statgraphics 18®
centurion

Revised: 10/11/2017

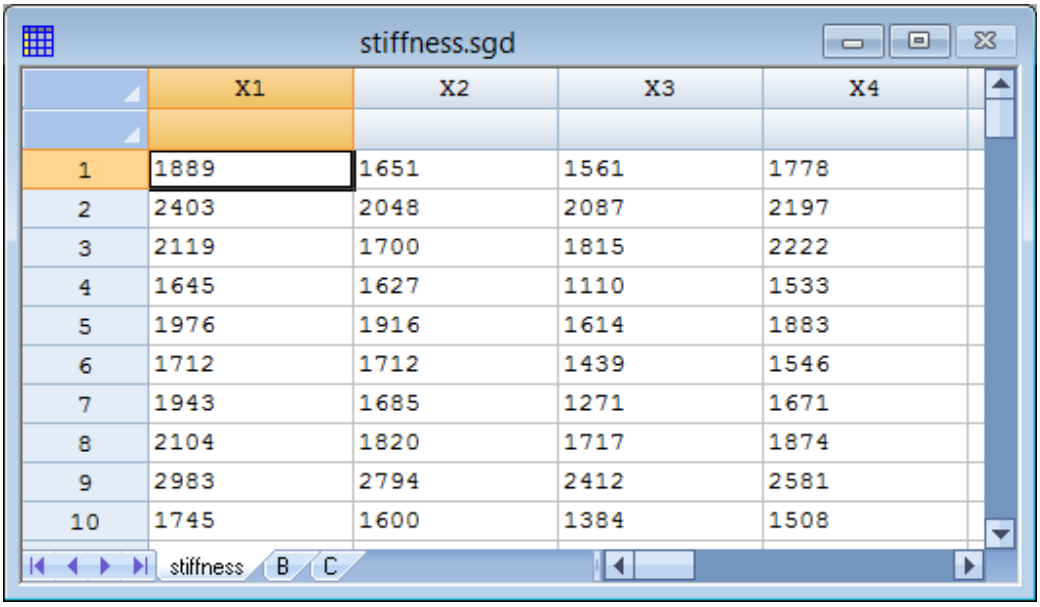
Summary

The *Multivariate Tolerance Limits* procedure creates statistical tolerance limits for data consisting of more than one variable. It includes a tolerance region that bounds a selected p% of the population with 100(1-α)% confidence. It also includes joint simultaneous tolerance limits for each of the variables using a Bonferroni approach. The data are assumed to be a random sample from a multivariate normal distribution. Multivariate tolerance limits are often compared to specifications for multiple variables to determine whether or not most of the population is within spec.

Sample StatFolio: *mvtolerance.sgp*

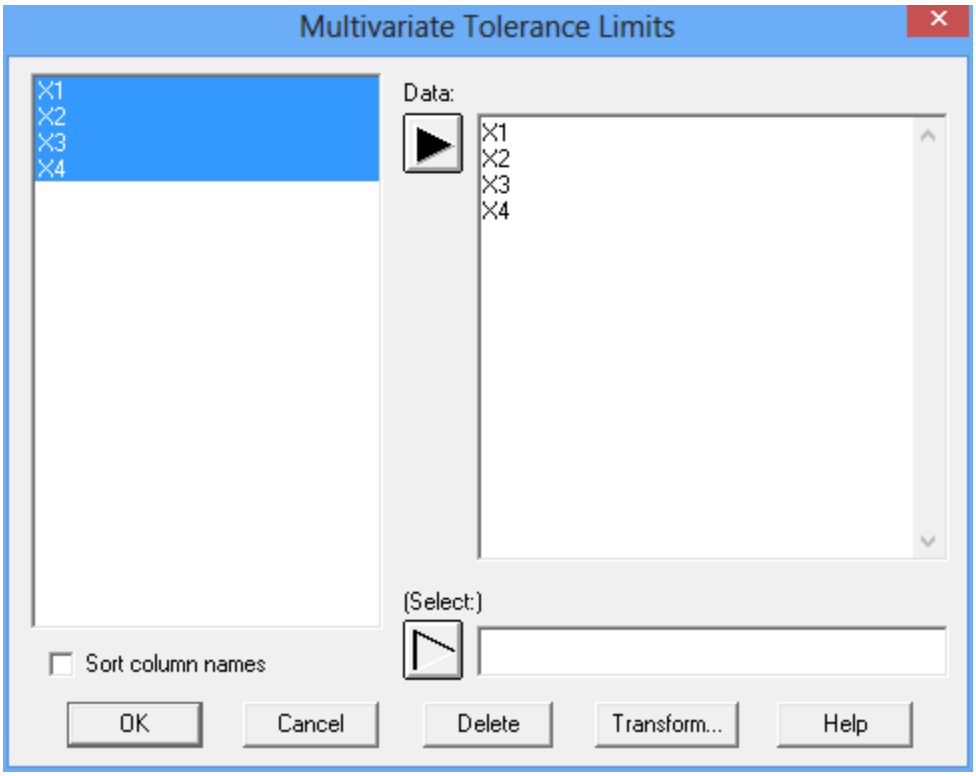
Sample Data:

The file *stiffness.sgd* contains four different measurements of the stiffness of $n = 30$ pieces of lumber (Johnson and Wichern, 2002). A portion of the data is shown below:



Data Input

To create a multivariate tolerance region for the sample data, choose *Multivariate Tolerance Limits* from the main menu. The data input dialog box is shown below:

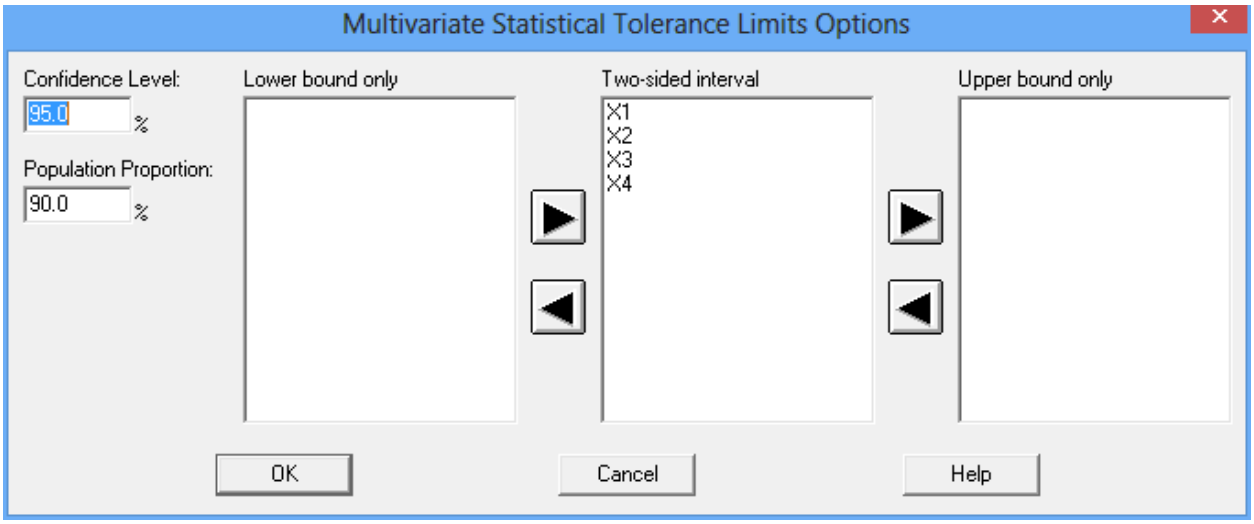


- **Data:** the names of 2 or more numeric columns containing the data.
- **Select:** subset selection.

The data for each of the *m* variables should be placed in a separate column. Each row corresponds to a single multivariate observation.

Analysis Options

The *Analysis Options* dialog box is used to specify the type of tolerance limits desired:



- **Confidence Level:** the confidence level for the tolerance limits.
- **Population Proportion:** the proportion of the population to be contained within the limits.
- **Lower bound only:** list of variables for which a lower bound is desired.
- **Two-sided interval:** list of variables for which both lower and upper limits are desired.
- **Upper bound only:** list of variables for which an upper bound is desired.

Select a variable and click on an arrow button to move it from one list to another.

Analysis Summary

The *Analysis Summary* shows sample statistics for the variables and the estimated tolerance limits:

Multivariate Tolerance Limits
Data variables:
X1
X2
X3
X4

	Mean	*Standard deviation*
X1	1906.1	324.987
X2	1749.53	318.607
X3	1509.13	303.178
X4	1724.97	322.844

Sample Correlations

	X1	X2	X3	X4
X1	1.0	0.913762	0.88593	0.898121
X2	0.913762	1.0	0.788213	0.788103
X3	0.88593	0.788213	1.0	0.923101
X4	0.898121	0.788103	0.923101	1.0

Number of observations = 30

95% Simultaneous Bonferroni Tolerance Limits for 90% of the Population

	Lower Limit	*Upper Limit*
X1	1139.34	2672.86
X2	997.826	2501.24
X3	793.827	2224.44
X4	963.263	2486.67

Observations beyond Bonferroni limits: 1

95% Elliptical Tolerance Region for 90% of the Population: Squared distance <= 13.2206
Observations outside elliptical region: 1

The first table shows the vector of sample means $\bar{\boldsymbol{X}}$ and the vector of sample standard deviations $\boldsymbol{s}$. The second table shows the m by m sample correlation matrix $\boldsymbol{R}$. The element in the jth row and kth column of the correlation matrix is calculated from

$$r_{j,k} = \frac{\sum_{i=1}^{n}\left(x_{j,i} - \bar{x}_j\right)\left(x_{k,i} - \bar{x}_k\right)}{\sqrt{\left[\sum_{i=1}^{n}\left(x_{j,i} - \bar{x}_j\right)^2\right]\left[\sum_{i=1}^{n}\left(x_{k,i} - \bar{x}_k\right)^2\right]}} \qquad (1)$$

The lower section of the output shows two types of tolerance limits.

Simultaneous tolerance limits

The output displays simultaneous tolerance limits for each of the *m* variables using a Bonferroni approach. This approach calculates separate tolerance limits for each of the variables using the standard K-factor

$$\bar{x} \pm Ks \tag{2}$$

where $\bar{x}$ and *s* are the sample mean and sample standard deviation of the selected variable. However, instead of using a K-factor corresponding to the desired level of confidence, it uses a K with a confidence level of

$$CL = 100\ (1 - \alpha/m)\ \% \tag{3}$$

The resulting tolerance limits bound *p*% of the joint distribution of the *m* variables with confidence equal to or greater than 100(1-α)%.

For the sample data, we can state with 95% confidence that 90% of the joint distribution of the 4 variables is within the limits:

$$1139.34 \le X_1 \le 2672.86$$

$$997.826 \le X_2 \le 2501.24$$

$$793.827 \le X_3 \le 2224.44$$

$$963.263 \le X_4 \le 2486.67$$

The Bonferroni limits are somewhat conservative, meaning that they may contain more than the stated population percentage. Krishnamoorthy and Mathew (2009) studied the conservatism of the Bonferroni intervals for *m* = 2 variables and found that the approach is quite satisfactory unless the correlation between the variables is very large. The approach also has the advantage that it is usually easy to determine whether or not the entire tolerance region is within a set of specifications.

Elliptical tolerance region

An exact tolerance region for the *m* variables is also calculated. It takes the form

$$(\boldsymbol{X} - \bar{\boldsymbol{X}})^T \boldsymbol{S}^{-1} (\boldsymbol{X} - \bar{\boldsymbol{X}}) \le c \tag{4}$$

where $\boldsymbol{S}$ is the *m* by *m* sample covariance matrix. This corresponds to an elliptical region in *m* dimensions. A multivariate observation $\boldsymbol{X}_i$ is within the tolerance region if the squared generalized distance from the centroid $\bar{X}$

$$d_i^2 = (X_i - \bar{X})^T \boldsymbol{S}^{-1} (\boldsymbol{X}_i - \bar{X}) \quad (5)$$

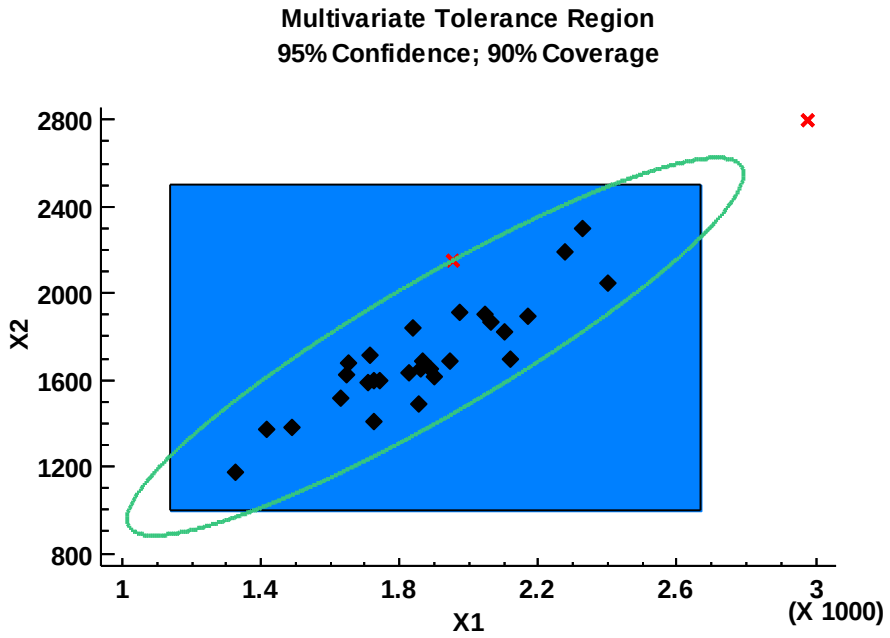
is no greater than *c*. The critical distance *c* depends on the number of variables *m*, the sample size *n*, the coverage percentage *p*, and the confidence level 100(1-α)%.

Since there is no theoretical way to calculate *c* and the available approximations are not satisfactory for all combinations of *m*, *n*, *p* and α, it is necessary to use a Monte Carlo simulation to obtain the value of *c*. Statgraphics does so using Algorithm 9.2 described by Krishnamoorthy and Mathew (2009). Whenever a value of *c* is needed, it is obtained using that algorithm with 100,000 repetitions. Tests have indicated that the value of *c* obtained is quite stable with that many repetitions, and the values obtained are very close to those tabulated by the authors in their textbook.

As displayed in the Analysis Summary, a 95% elliptical tolerance region for 90% of the population is given by equation (4) with *c* = 13.2206.

Multivariate Tolerance Plot

This option displays the data and calculated tolerance regions for any pair of observations:

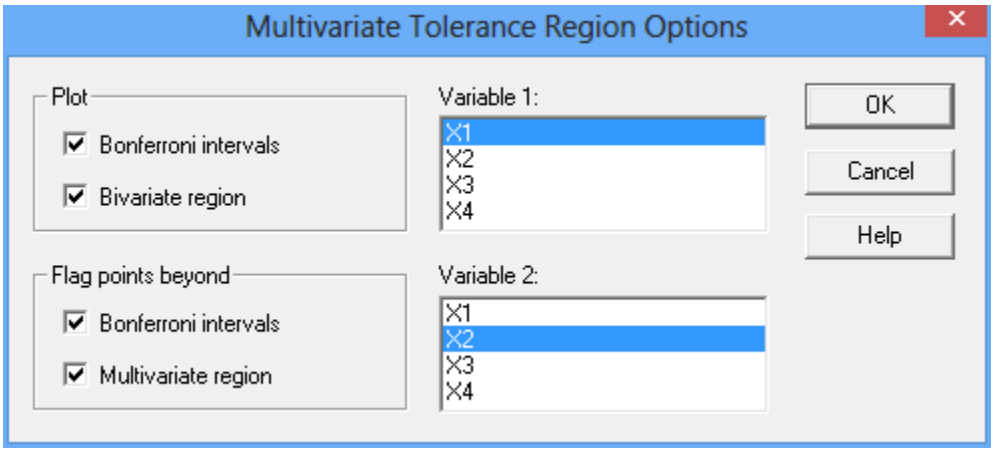


By default, it displays both the Bonferroni limits and the elliptical tolerance region. The Bonferroni limits are calculated using equation (3) with *m* equal to the total number of variables.

The elliptical tolerance region is calculated using the equation for $m = 2$ based on only the 2 variables being plotted.

By default, points outside either of the tolerance limits are shown using a red X.

Pane Options



- **Plot:** the regions to plot on the graph.
- **Flag points beyond:** the regions used to determine which points are plotted using a red X.
- **Variable 1:** the variable displayed on the horizontal axis.
- **Variable 2:** the variable displayed on the vertical axis.

Data Table

This table displays each of the observations and the calculated squared distance d_i^2 from equation (5), setting *m* equal to the total number of variables.

Data Table

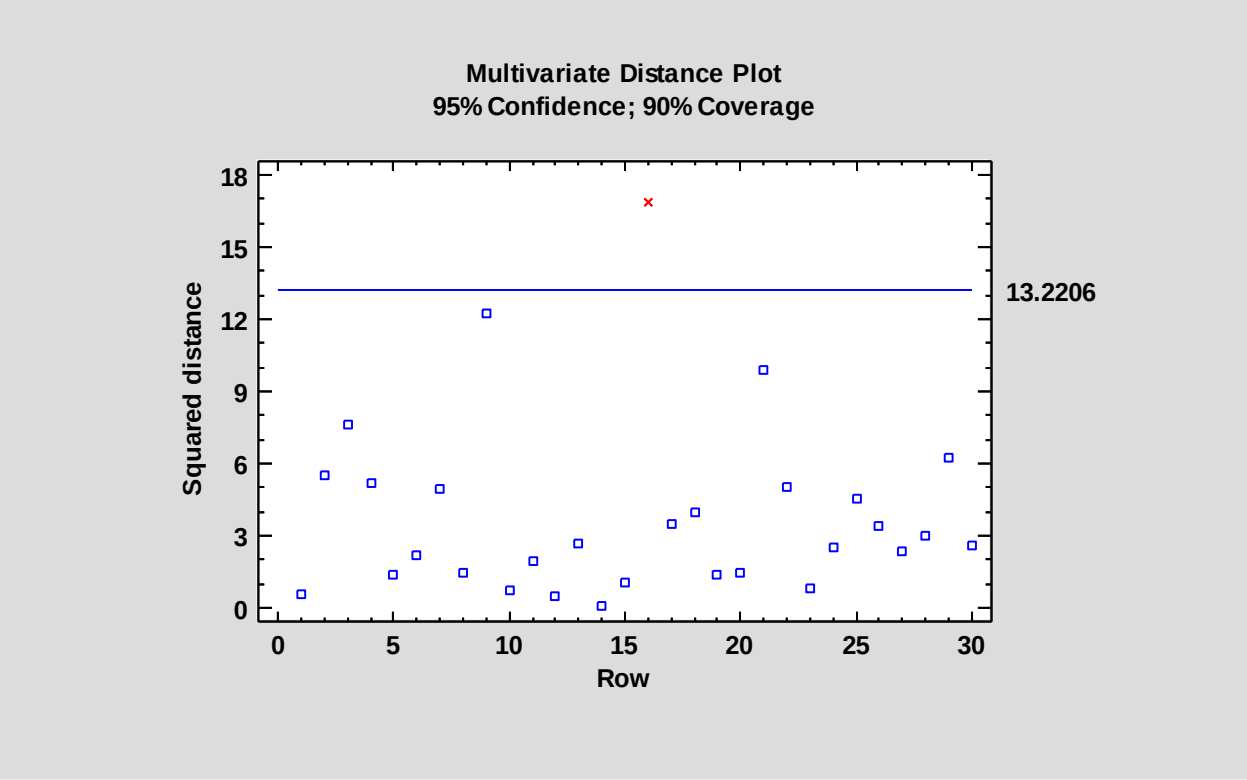
E:Outside elliptical region B:Beyond Bonferroni limits

Row	*X1*	*X2*	*X3*	*X4*	*Squared distance*	*Beyond limits*
1	1889.0	1651.0	1561.0	1778.0	0.600013	
2	2403.0	2048.0	2087.0	2197.0	5.47702	
3	2119.0	1700.0	1815.0	2222.0	7.61664	
4	1645.0	1627.0	1110.0	1533.0	5.20761	
5	1976.0	1916.0	1614.0	1883.0	1.39808	
6	1712.0	1712.0	1439.0	1546.0	2.21914	
7	1943.0	1685.0	1271.0	1671.0	4.98835	
8	2104.0	1820.0	1717.0	1874.0	1.48766	
9	2983.0	2794.0	2412.0	2581.0	12.2648	B
10	1745.0	1600.0	1384.0	1508.0	0.76654	
11	1710.0	1591.0	1518.0	1667.0	1.93078	
12	2046.0	1907.0	1627.0	1898.0	0.463516	
13	1840.0	1841.0	1595.0	1741.0	2.6959	
14	1867.0	1685.0	1493.0	1678.0	0.129571	
15	1859.0	1649.0	1389.0	1714.0	1.07925	
16	1954.0	2149.0	1180.0	1281.0	16.8474	E
17	1325.0	1170.0	1002.0	1176.0	3.50183	
18	1419.0	1371.0	1252.0	1308.0	3.99006	
19	1828.0	1634.0	1602.0	1755.0	1.36321	
20	1725.0	1594.0	1313.0	1646.0	1.46499	
21	2276.0	2189.0	1547.0	2111.0	9.89804	
22	1899.0	1614.0	1422.0	1477.0	5.05574	
23	1633.0	1513.0	1290.0	1516.0	0.79621	
24	2061.0	1867.0	1646.0	2037.0	2.53856	
25	1856.0	1493.0	1356.0	1533.0	4.57679	
26	1727.0	1412.0	1238.0	1469.0	3.39798	
27	2168.0	1896.0	1701.0	1834.0	2.3816	
28	1655.0	1675.0	1414.0	1597.0	2.99518	
29	2326.0	2301.0	2065.0	2234.0	6.28376	
30	1490.0	1382.0	1214.0	1284.0	2.58382	

The *Beyond limits* column indicates any point which is beyond the tolerance limits using each of the two methods.

Distance Plot

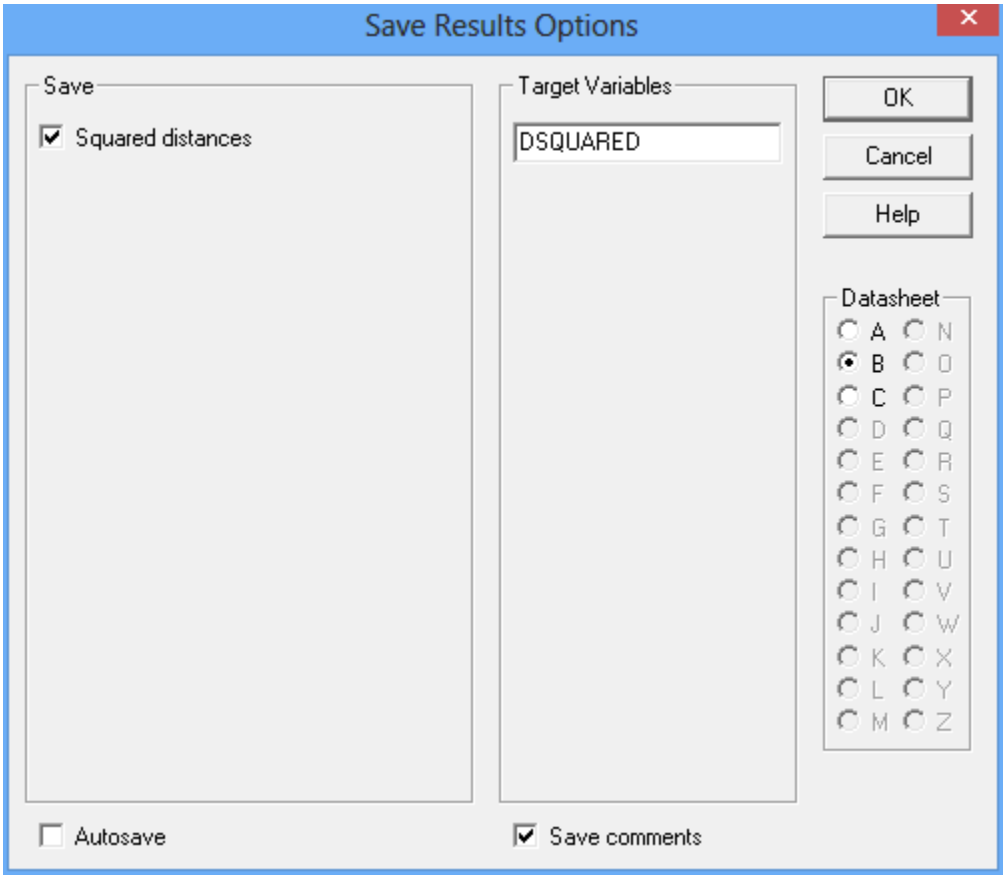
This plot displays the calculated squared distance d_i^2 for each of the n observations, together with the critical distance c.



Any points beyond the critical distance are drawn using a red X.

Save Results

The squared distances may be saved to a datasheet by pressing the *Save Results* button on the analysis toolbar. This displays the following dialog box:



Check the box for *Squared distances* and press *OK*.

References

Johnson, R.A. and Wichern, D. W. (2002). Applied Multivariate Statistical Analysis, fifth edition. Prentice Hall, Upper Saddle River, N.J.

Krishnamoorthy, K. and Mathew, T. (2009). Statistical Tolerance Regions: Theory, Applications, and Computation. John Wiley and Sons, Hoboken, N.J.