Cuscore Charts

Summary

The **Cuscore (Cumulative Score) Charts** procedure creates control charts designed for directed process monitoring, where it is feared that a specific type of pattern might be present in the data, indicating a process disturbance of known cause. For example, a sine wave of known frequency might occur when a specific device begins to fail. A cuscore could be constructed to detect that specific frequency. The charts are typically used as supplements to standard control charts that look for more general out-of-control situations.

The cuscore charts can handle data that have been collected either individually or in subgroups. Like the ARIMA charts procedure, they can deal with random noise or noise that follows a specific parametric time series model. The user is responsible for specifying both the type of disturbance that is feared and the noise model.

The procedure creates both a cuscore chart and an R chart, S chart, or MR(2) chart. Although the charts may be constructed in an *Initial Study* (Phase 1) mode, where the current data determine the control limits, they are primarily intended for *Control to Standard* (Phase 2) uses, where the limits come from either a known standard or from prior data

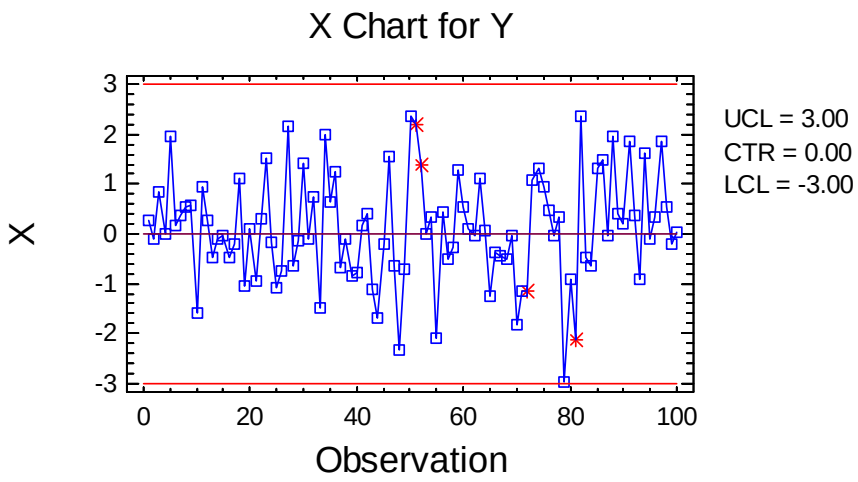
Sample StatFolio: *cuscorechart.sgp*

Sample Data:

The file *sineplusnoise.sgd* contains a sample of $n = 144$ observations, similar to an example in Box and Luceño (1997). A partial list of the data is shown below:

Sample	*Concentration*
1	0.262
2	-0.109
3	0.849
4	0.001
5	1.959
6	0.173
7	0.385
8	0.534
9	0.569
10	-1.583
11	0.932
12	0.263
13	-0.460
14	-0.112
15	-0.017

The data consist of a set of random numbers generated from a normal distribution with $\mu = 0$ and $\sigma = 1$, plus a sine wave of period 12 and amplitude 0.5 added to the random numbers between observations 48 and 96. An X-Chart of the data is shown below:



The sine wave is pretty much invisible, generating no points beyond the control limits, although there are a few runs rules violations.

Cuscore Charts

In constructing a cuscore chart, one begins by specifying a **null model** that describes the behavior of the variable being monitored when the process is under control. Often, the model is as simple as random deviations around a fixed mean μ:

$$y_j = \mu + a_j \tag{1}$$

where the a_j are random samples from a normal distribution with mean 0 and constant variance σ_a^2 (called "white noise"). In other cases, a more complicated ARIMA model might be postulated. For example, a commonly used model for nonstationary data that do not have a fixed mean is the IMA(1,1) model given by:

$$y_j = y_{j-1} + a_j - \theta a_{j-1} \tag{2}$$

This model expresses the observation at time *j* as being equal to the previous observation plus a random shock at the current period minus a fraction of the random shock at the previous period. For a general discussion of ARIMA models, see the *ARIMA Charts* documentation.

In addition to the null model, a **discrepancy model** is also postulated. The discrepancy model specifies the type of disturbance that the chart is designed to detect. For example, the disturbance might be a step change in the mean. When added to the null model of random deviations around a fixed mean, the discrepancy model in that case becomes:

$$y_j = \mu + a_j + \delta \tag{3}$$

Given both models, a "detector" series r_t can be created that is specifically designed to be sensitive to the postulated disturbance. The general method for creating a detector is described by Box and Luceño (1997). For the model above, the detector equals simply $r_j = 1$ for all *j*.

The Q score statistic is constructed from the cumulative sum of the detector times the residuals obtained when fitting the null model:

$$Q_j = \sum_{i=1}^{j} \hat{a}_j r_j \qquad (4)$$

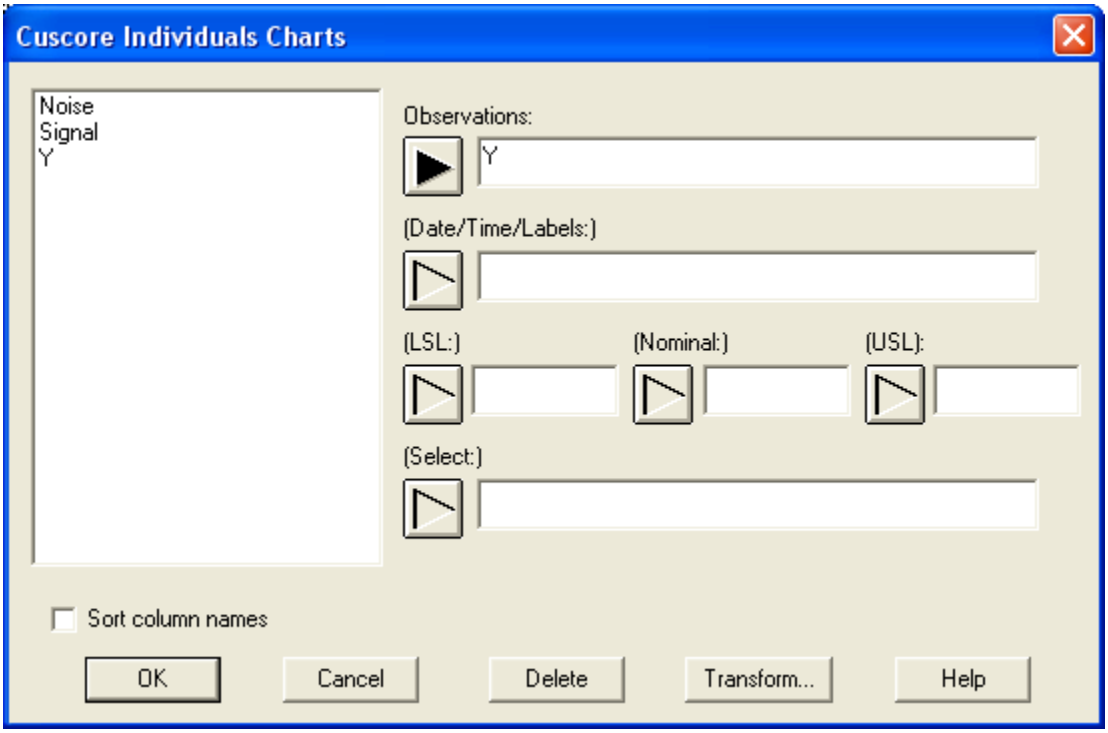
To determine whether the feared disturbance has occurred, the Q's are plotted on a "cuscore" chart with upper and lower control limits.

Data Input

There are two menu selections that create cuscore charts, one for individuals data and one for grouped data. In the case of grouped data, the original observations may be entered, or subgroup statistics may be entered instead.

Case #1: Individuals

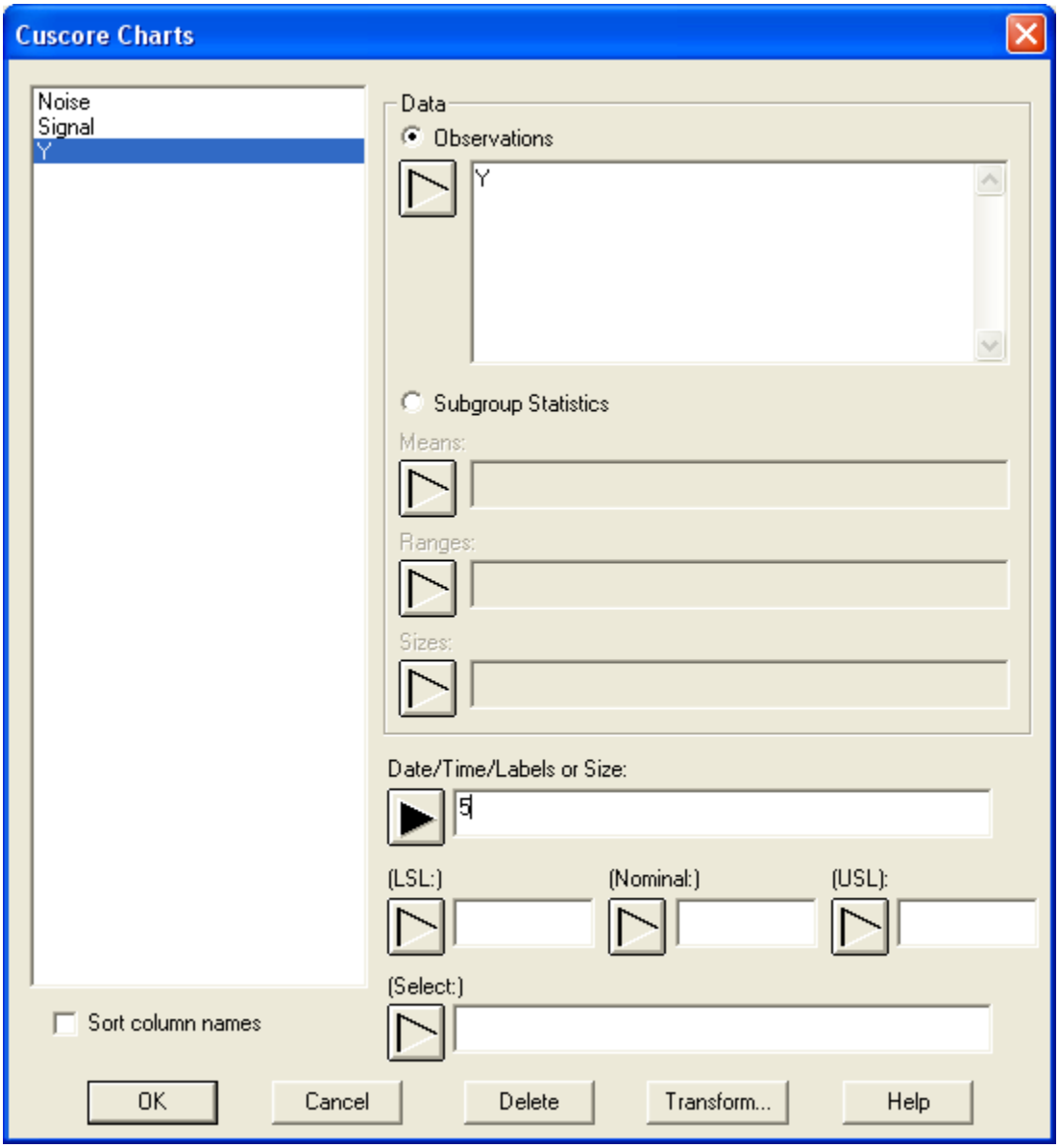
The data to be analyzed consist of a single numeric column containing *n* observations. The data are assumed to have been taken one at a time.



- **Observations:** numeric column containing the data to be analyzed.
- **Date/Time/Labels:** optional labels for each observation.
- **LSL, Nominal, USL**: optional lower specification limit, nominal (target) value, and upper specification limit.
- **Select:** subset selection.

Case #2: Grouped Data – Original Observations

The data to be analyzed consist of one or more numeric columns. The data are assumed to have been taken in groups, in sequential order by rows.



- **Observations:** one or more numeric columns. If more than one column is entered, each row of the file is assumed to represent a subgroup with subgroup size *m* equal to the number of columns entered. If only one column is entered, then the *Date/Time/Labels or Size* field is used to form the groups.

- **Date/Time/Labels or Size**: If each set of *m* rows represents a group, enter the single value *m*. For example, entering a 5 as in the example above implies that the data in rows 1-5 form the first group, rows 6-10 form the second group, and so on. 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 sequential rows with identical codes into the same group.

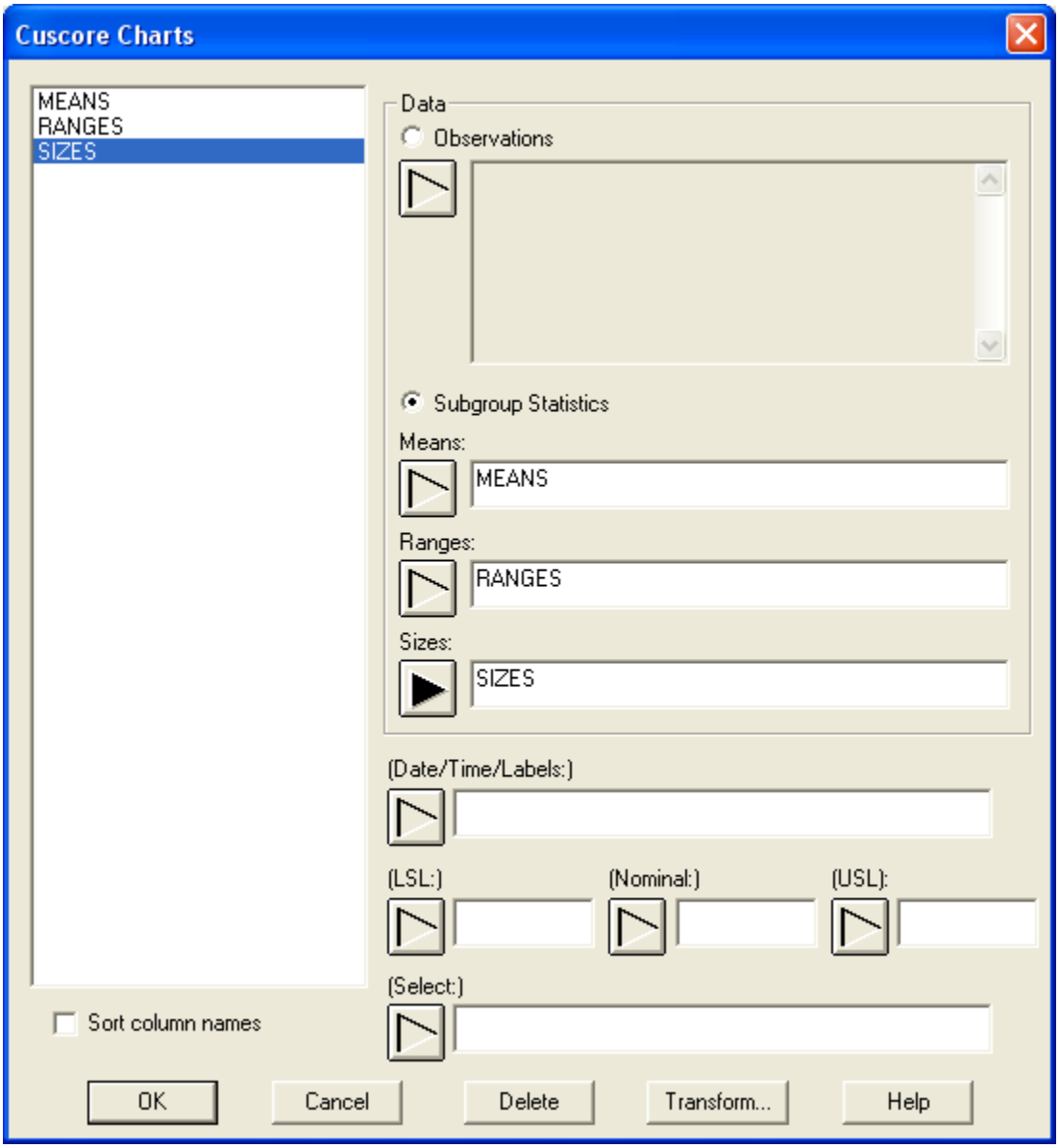
- **LSL, Nominal, USL**: optional lower specification limit, nominal (target) value, and upper specification limit.

- **Select:** subset selection.

Case #3: Grouped Data – Subgroup Statistics

In this case, the statistics for each subgroup have been computed elsewhere and entered into the datasheet, as in the table below:

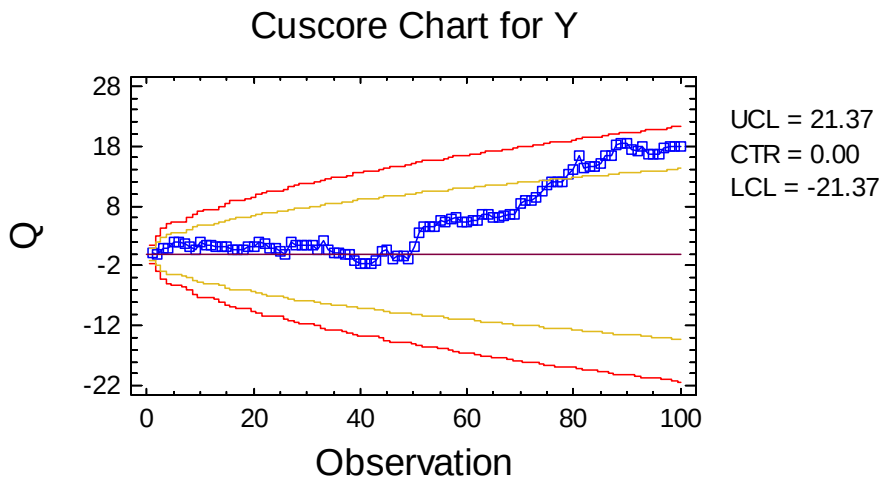
Sample	*Means*	*Ranges*	*Sizes*
1	0.5924	2.068	5
2	0.0156	2.152	5
3	0.1212	1.392	5
4	-0.1044	2.141	5
5	-0.0764	2.600	5



- **Subgroup Statistics**: the names of the column containing the subgroup means, subgroup ranges, and subgroup sizes.
- **Date/Time/Labels:** optional labels for each subgroup.
- **LSL, Nominal, USL**: optional lower specification limit, nominal (target) value, and upper specification limit.
- **Select:** subset selection.

Cuscore Chart

The *Cuscore Chart* plots the calculated scores for each observation or subgroup.



In this chart, the points plotted are the scores Q_j, calculated based on the null and discrepancy models specified on the *Analysis Options* dialog box. The centerline is drawn at 0, with control limits at

$$\pm k \sum_{i=1}^{j} r_j^2 \hat{\sigma}_a^2 \tag{5}$$

where *k* is the multiple (usually equal to 3) specified on the *Control Charts* tab of the *Preferences* dialog box, accessible from the *Edit* menu. The standard deviation of the random shocks σ_a may be estimated using either the mean squared error (MSE) of the fitted ARIMA model or using the subgroup ranges or standard deviations, as described below.

In the current example, the null model was selected to be random deviations from a normal distribution with $\mu = 0$ and $\sigma = 1$, since that model was used to generate the random numbers:

$$\textbf{Null model: } y_j = \mu + a_j \tag{6}$$

The discrepancy model included a sinusoidal disturbance of period 12, which had been hidden in the noise:

$$\textbf{Discrepancy model: } y_j = \mu + \delta \sin\left(\frac{2\pi j}{12}\right) + a_j \tag{7}$$

In this case, the cuscore statistic is

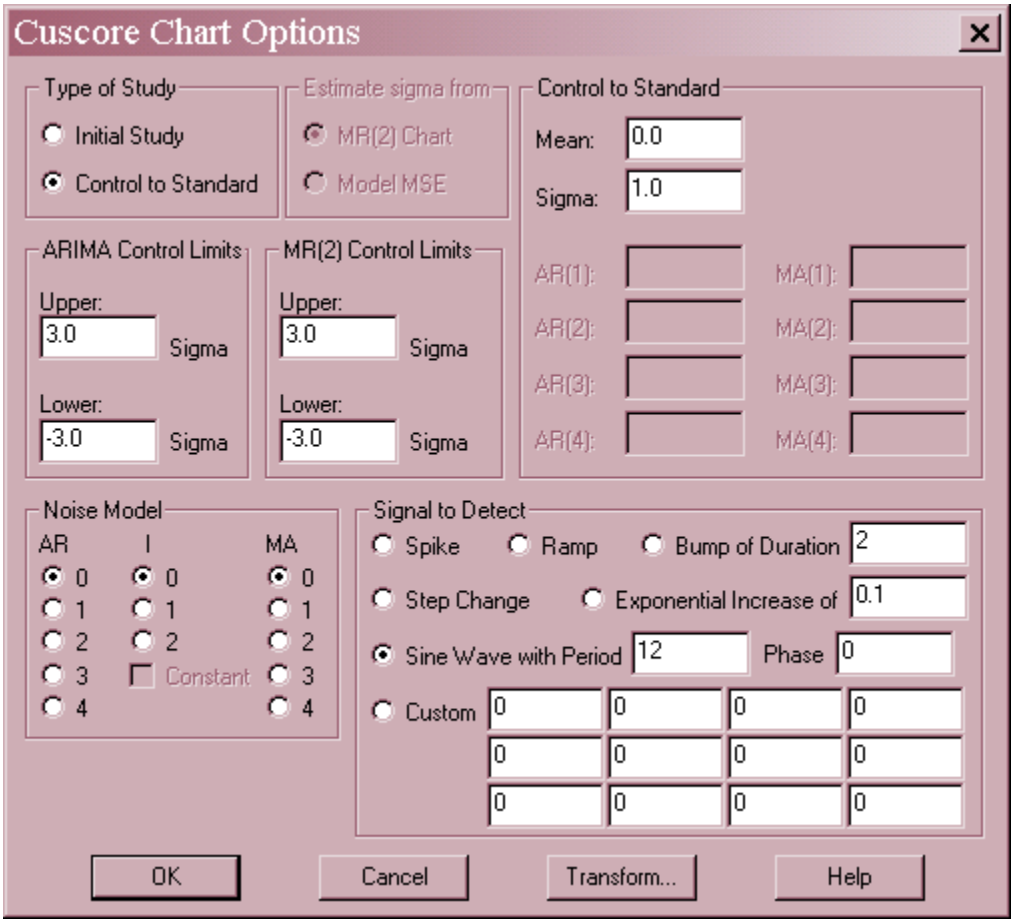
$$Q_j = \sum_{i=1}^{j} \hat{a}_j \sin\left(\frac{2\pi i}{12}\right) \tag{8}$$

where the noise estimates are simply the observed data values (minus $\mu = 0$):

$$\hat{a}_j = y_j \tag{9}$$

In the plot, the scores vary around 0 until approximately $j = 50$, when the detector begins to pick up the buried sine wave. The scores rise steadily until about $j = 85$, when the scores begin to oscillate around a new level. Although the signal was not strong enough to cause the scores to cross the 3-sigma limits, it does come very close.

Analysis Options



- **Type of Study:** determines how the control limits are set. For an *Initial Study* (Phase 1) chart, the limits are estimated from the current data. For a *Control to Standard* (Phase 2) chart, the control limits are determined from the information in the *Control to Standard* section of the dialog box.

- **Estimate sigma from**: specifies whether the standard deviation of the random shocks σ_a should be estimated from the dispersion chart, or whether it should be estimated from the residual mean squared error of the fitted ARIMA model.

- **Cuscore Control Limits:** specify the multiple k to use in determining the upper and lower control limits on the cuscore chart. To suppress a limit completely, enter 0.

- **MR(2) Control Limits:** specify the multiple *k* to use in determining the upper and lower control limits on the MR(2), R, or S chart. To suppress a limit completely, enter 0.

- **Control to Standard:** To perform a Phase 2 analysis, select *Control to Standard* for the *Type of Study* and then enter the established standard process mean and sigma and the ARIMA model parameters. If AR or MA terms are included, "sigma" refers to the standard deviation of the noise σ_a.

- **Model**: specifies the order *p* of the autoregressive (AR) portion of the model, the order of differencing *d* (I), and the order *q* of the moving average (MA) portion of the model. If d > 0, the constant term can be removed.

- **Signal to Detect**: the type of disturbance assumed in the discrepancy model. The chart will be constructed to be sensitive to the indicated type of change. The choices are:

  Spike – a value δ added to the process mean for a single time period.

  Ramp – a linear increase of the form δt.

  Bump – a value δ added to the process mean, lasting for *k* time periods, where $k \geq 2$.

  Step Change – a value δ added to the process mean, lasting indefinitely.

  Exponential Increase - an increase of the form $\exp(\lambda t)$ where $0 < \lambda < 1$.

  Sine Wave – a periodic disturbance of the form $\delta \sin(2\pi t/P+\phi)$ where *P* is the number of observations in a cycle (the *Period*) and ϕ specifies the phase.

  Custom – a user-specified disturbance lasting up to 12 periods. For example, a diminishing sequence of oscillations might be represented by the sequence (1.0, -0.9, 0.8, -0.7, 0.6, -0.5, 0.4, -0.3, 0.2, -0.1, 0, 0).

- **Transform Button**: Use this button to specify a transformation or non-normal distribution.

For a discussion of the *Transform* feature, see the documentation for *Individuals Control Charts*.

Analysis Summary

This pane summarizes the control charts.

Cuscore Individuals Charts - Y
Number of subgroups = 100
Subgroup size = 1.0
0 subgroups excluded

Distribution: Normal
Transformation: none

Cuscore Chart

Period	*#1-100*
UCL: +3.0 sigma	21.3717
Centerline	0.0
LCL: -3.0 sigma	-21.3717

0 beyond limits

MR(2) Chart

Period	*#1-100*
UCL: +3.0 sigma	3.6855
Centerline	1.128
LCL: -3.0 sigma	0.0

1 beyond limits

Estimates

Period	*#1-100*	*Standard*
Process mean	0.0	0.0
Process sigma	1.0	1.0
Mean residual MR(2)	1.19637	
Residual sigma	1.06062	

Sigma estimated from average moving range of residuals

ARIMA Model Summary

Parameter	*Value*
Mean	0.0
Constant	0.0

White noise standard deviation = 1.0

Included in the table are:

- **Subgroup Information**: the number of observations or subgroups and the average subgroup size.
- **Distribution:** the assumed distribution for the data. By default, the data are assumed to follow a normal distribution. However, one of 26 other distributions may be selected using *Analysis Options*.
- **Transformation:** any transformation that has been applied to the data. Using *Analysis Options*, you may elect to transform the data using either a common transformation such as a square root or optimize the transformation using the Box-Cox method.
- **Cuscore Chart:** a summary of the centerline and control limits for the cuscore chart, together with a count of any points beyond the control limits.
- **MR(2)/R/S Chart:** a summary of the centerline and control limits for the dispersion chart.
- **Estimates:** estimates of the process mean μ and the process standard deviation σ. The method for estimating the process sigma depends on the settings on the *Analysis Options* dialog box, described below.
- **Mean residual MR(2), Average Range, or Average S:** the average of the values plotted on the dispersion chart.

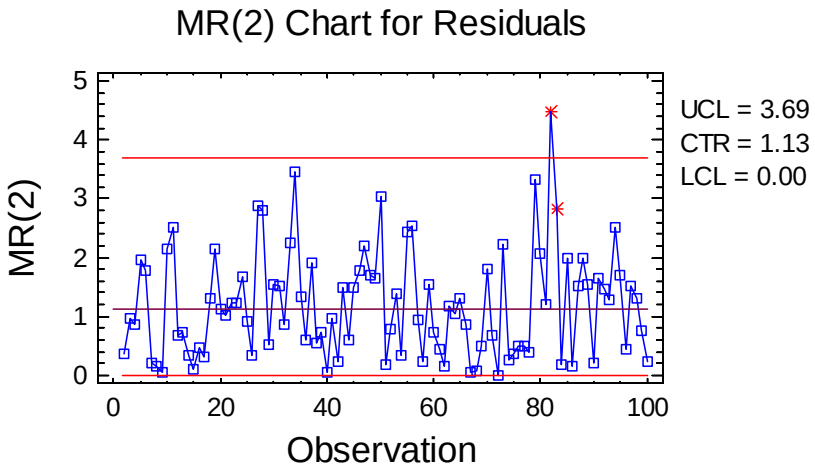
- **ARIMA Model Summary**: summarizes the fitted or specified ARIMA model. For the sample data, the model is:

$$y_j = a_j \tag{10}$$

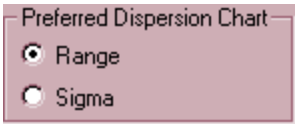
If the model was estimated, each of the estimated model coefficients is shown together with a standard t-test. If the P-value associated with a selected coefficient is less than 0.05, then that coefficient is significantly different from 0 at the 5% significance level. Also shown are the process mean and the estimated standard deviation of the random shocks σ_a.

MR(2)/R/S Chart

A second chart is also included to monitor the process variability.



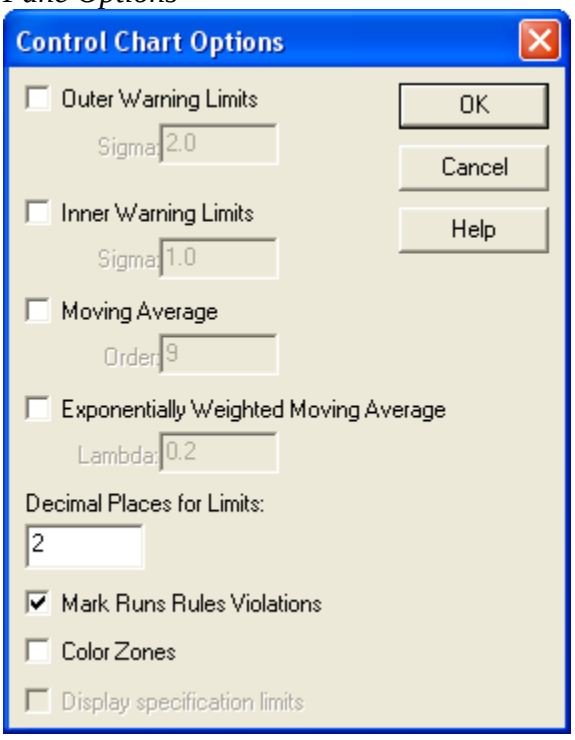
For individuals data, the chart displayed is an MR(2) chart, described in the *Individuals Control Charts* documentation. For grouped data, either an R chart or an S chart is plotted, depending on the setting on the *Control Charts* tab of the *Preferences* dialog box:



These charts are described in the *X-Bar and R Charts* and the *X-Bar and S Charts* documents.

The dispersion chart is always created using the residuals from the fitted model.

Pane Options



- **Outer Warning Limits**: check this box to add warning limits at the specified multiple of sigma, usually at 2 sigma.

- **Inner Warning Limits**: check this box to add warning limits at the specified multiple of sigma, usually at 1 sigma.

- **Moving Average:** check this box to add a moving average smoother to the chart. In addition to the subgroup means, the average of the most recent q points will also be displayed, where q is the *order* of the moving average. The default value $q = 9$ since the 1-sigma inner warning limits for the original subgroup means are equivalent to the 3-sigma control limits for that order moving average.

- **Exponentially Weighted Moving Average:** check this box to add an EWMA smoother to the chart. In addition to the subgroup means, an exponentially weighted moving average of the subgroup means will also be displayed, where λ is the smoothing parameter of the EWMA. The default value $\lambda = 0.2$ since the 1-sigma inner warning limits for the original subgroup means are equivalent to the 3-sigma control limits for that EWMA.

- **Decimal Places for Limits:** the number of decimal places used to display the control limits.

- **Mark Runs Rules Violations:** flags with a special point symbol any unusual sequences or runs. The runs rules applied by default are specified on the *Runs Tests* tab of the *Preferences* dialog box.

- **Color Zones:** check this box to display green, yellow and red zones.

Cuscore Chart Report

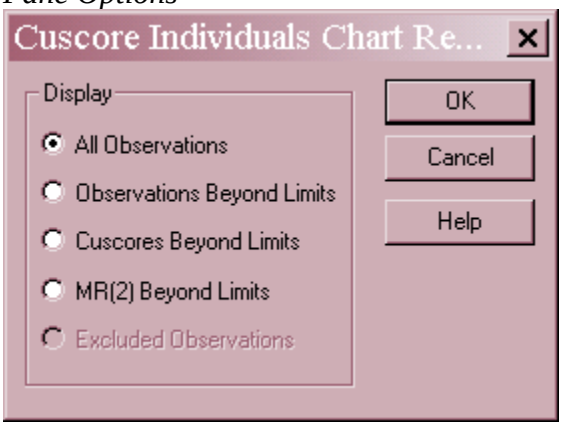
This report displays the values plotted on the control charts.

Cuscore Individuals Chart Report
Observations Beyond Limits
* = Beyond Limits

Observation	*Q*	*MR(2)*
82	14.246	* 4.476

Out-of-control points are indicated by an asterisk. Points excluded from the calculations are indicated by an X.

Pane Options



- **Display:** specify the observations or subgroups to display in the report.

Runs Tests

The *Runs Tests* pane displays the results of standard tests designed to look for unusual sequence of points on the dispersion chart.

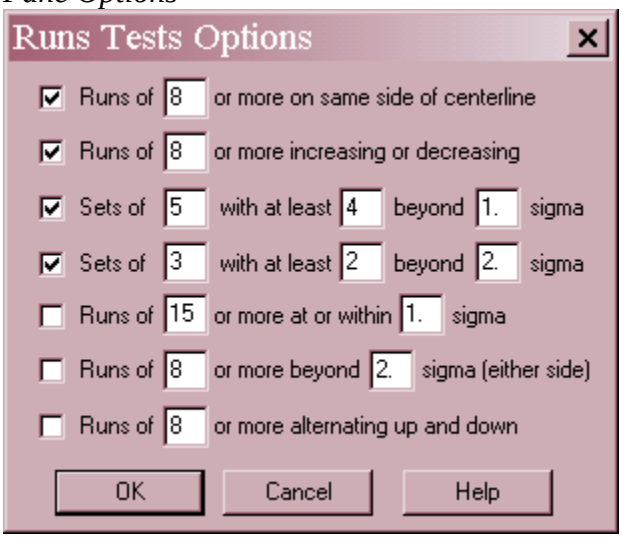
Runs Tests
Rules
(A) runs above or below centerline of length 8 or greater.
(B) runs up or down of length 8 or greater.
(C) sets of 5 observations with at least 4 beyond 1.0 sigma.
(D) sets of 3 observations with at least 2 beyond 2.0 sigma.

Violations

Observation	*MR(2) Chart*
83	C

For a detailed description of these tests, see the *X-Bar and R Charts* documentation.

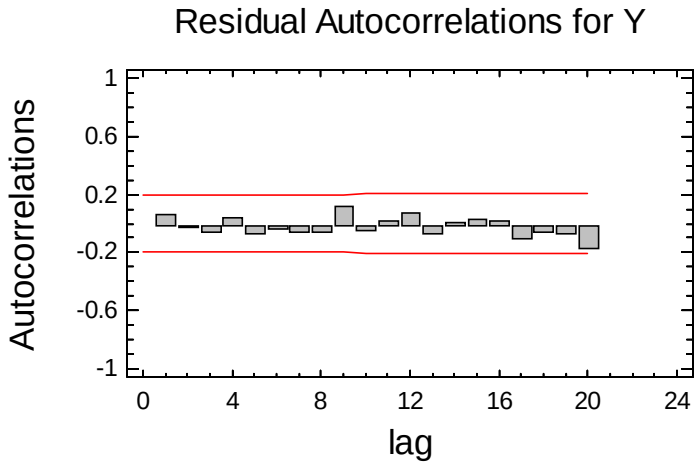
Pane Options



Select the runs tests to be applied and the parameters that define those tests. For example, some practitioners prefer to test for runs of length 7 rather than 8.

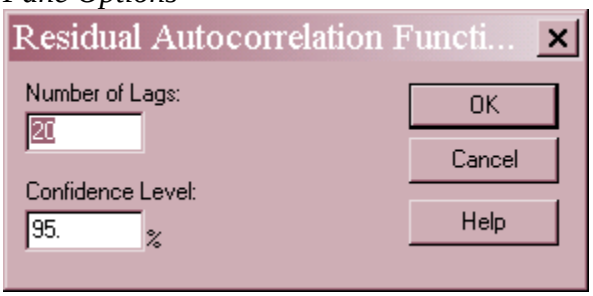
Residual Autocorrelation Function

The *Residual ACF* plots the autocorrelations of $\hat{a}_j$:



The autocorrelation measures the relationship between residuals at a specified separation in time, called the "lag". If the selected ARIMA model fits the data well, the residuals should be random and all the bars should remain within the indicated probability limits. Any bars extending beyond the limits indicate statistically significant autocorrelation between residuals separated by the indicated number of periods.

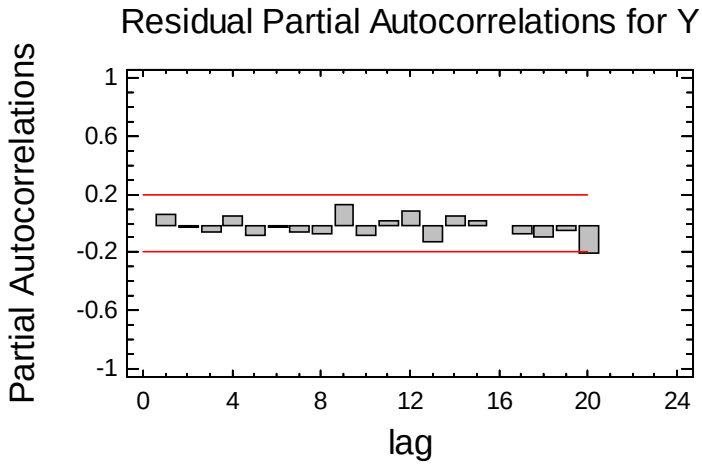
Pane Options



- **Number of Lags**: maximum lag for estimating the autocorrelation.
- **Confidence Level**: level used to calculate the probability limits.

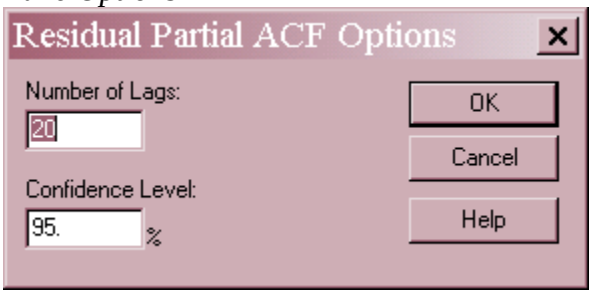
Residual Partial Autocorrelation Function

The residual partial autocorrelation function is used to determine whether additional AR terms should be added to the model.



If the values at very low lags are beyond the probability limits, then the order of the AR model might have to be increased. In this case, the first several spikes are all small, suggesting that the selected ARIMA model is appropriate for the observed data.

Pane Options



- **Number of Lags**: maximum lag for estimating the partial autocorrelation.
- **Confidence Level**: level used to calculate the probability limits.

Save Results

The following results can be saved to the datasheet, depending on whether the data are individuals or grouped:

1. *Q scores* – the values plotted on the cuscore chart.
2. *Ranges, sigmas, or moving ranges* – the values plotted on the dispersion chart.
3. *Sizes* – the subgroup sizes.
4. *Labels* – the subgroup labels.
5. *Process Mean* – the estimated process mean.
6. *Process Sigma* – the estimated process standard deviation.
7. *Residuals* – the residuals from the ARIMA model.

Calculations

For more information on the estimation of ARIMA models, see the *Forecasting* documentation.

