ARIMA Charts

Summary

The **ARIMA (AutoRegressive Integrated Moving Average) Charts** procedure creates control charts for a single numeric variable where the data have been collected either individually or in subgroups. In contrast to other control charts, the ARIMA charts do not assume that successive observations are independent. Instead, a statistical model is constructed to describe the serial correlation between observations close together in time. Out-of-control signals are then based on the deviations of the process from this dynamic time series model.

Many continuous processes, if sampled at intervals close together in time, will exhibit the type of autocorrelation that ARIMA charts are designed to handle. Standard control charts would tend to give too many false alarms in such cases.

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

## Sample StatFolio: *ARIMA charts.sgp*

Sample Data:

The file *ARIMA charts.sgd* contains a sample of $n = 120$ concentration measurements taken once every two hours from a chemical process. The data are from Box, Jenkins, and Reinsel (1994). A partial list of the data in that file is shown below:

Sample	Concentration
1	17.0
2	16.6
3	16.3
4	16.1
5	17.1
6	16.9
7	16.8
8	17.4
9	17.1
10	17.0
11	16.7
12	17.4
13	17.2
14	17.4
15	17.4

A standard individuals chart applied to this data shows many out-of-control signals:

X Chart for Concentration
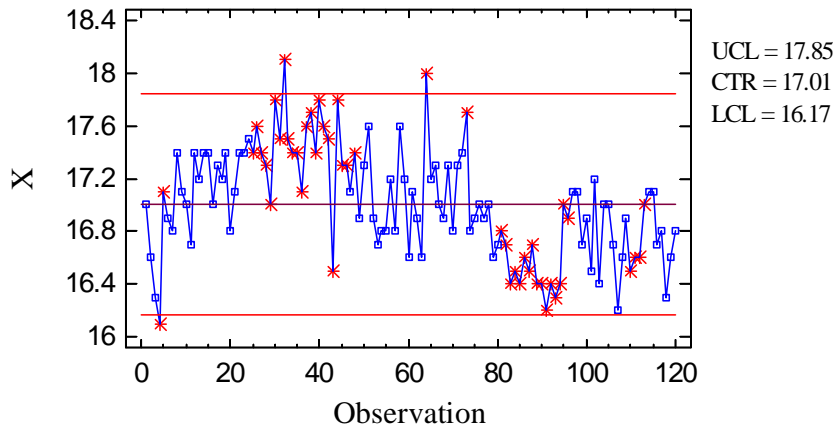
On an hour to hour basis, the process clearly changes. Yet in a long-term sense, the process may actually be "in control", varying around a stable mean with constant variance.

ARIMA Models

Most control charts rely on a fairly simple statistical model. The data are assumed to vary randomly around a fixed mean μ according to:

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

where the a_j are assumed to be random samples from a normal distribution with mean 0 and constant variance σ_a^2 (called "white noise").

The ARIMA charts are built on a more complicated class of models defined by:

$$z_j = \theta_0 + \phi_1 z_{j-1} + \phi_2 z_{j-2} + ... + \phi_p z_{j-p} + a_j - \theta_1 a_{j-1} - \theta_2 a_{j-2} - ... - \theta_q a_{j-q} \tag{2}$$

where

$$z_j = \nabla^d x_j \tag{3}$$

In this model, the measurement at period j, x_j, is related to the measurements at the p previous periods, a random shock to the process a_j at period j, and random shocks at the q previous periods, after applying a difference of order d if necessary. The shocks represent random effects to which the process is subjected each period, which are assumed to be generated from a normal distribution with mean 0 and standard deviation σ_a.

The parameters that define this model are:

θ_0 = constant

$\phi_1, \phi_2, ..., \phi_p$ = autoregressive parameters

$\theta_1, \theta_2, \ldots \theta_q$ = moving average parameters

The value z_j may equal any of several quantities, depending on the order of differencing d:

$$z_j = x_j \qquad\qquad \text{for d = 0 (no differencing)} \qquad (4)$$

$$z_j = (x_j - x_{j-1}) \qquad\qquad \text{for d = 1 (first differencing)} \qquad (5)$$

$$z_j = (x_j - x_{j-1}) - (x_{j-1} - x_{j-2}) \qquad \text{for d = 2 (second differencing) .} \qquad (6)$$

Identification of the proper ARIMA model to be used in any instance relies on tools such as the autocorrelation function, found in the *Descriptive Methods* procedure under *Time Series Analysis* on the main STATGRAPHICS menu:

Estimated Partial Autocorrelations for Concentration
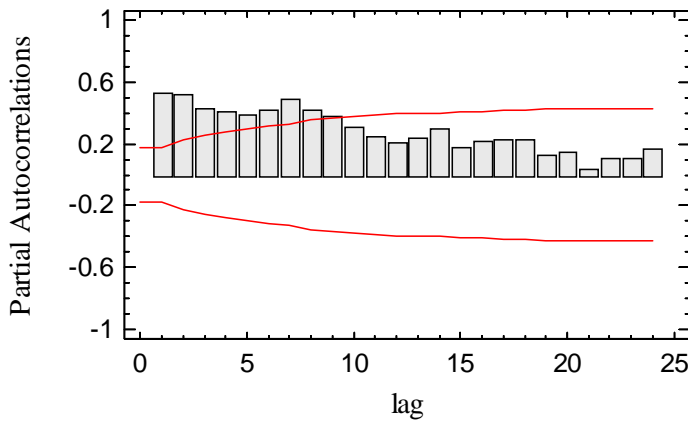
The autocorrelation function shows how the correlation between pairs of values (x_j, x_{j-k}) varies as a function of the lag or number of periods k between them. From the shape of the autocorrelation function and related partial autocorrelation function, an experienced analyst can select a good model for a particular set of data. For a full discussion of ARIMA model identification and fitting, see Box, Jenkins and Reinsel (1994).

Luckily, although the general form of the ARIMA model is very complicated, simple models often do a good job in modeling real-life processes. The default model used by STATGRAPHICS is a second order autoregressive model AR(2) which has the rather simple form

$$x_j = \theta_0 + \phi_1 x_{j-1} + \phi_2 x_{j-2} + a_j \qquad (7)$$

This model states that the average measurement at period j equals a constant, plus a multiple of the averages at the two previous time periods, plus a random shock. For processes that vary around a constant mean (*stationary* processes), such a model is often sufficient. For a discussion of other useful ARIMA models, see Montgomery (2005).

Data Input

There are two menu selections that create ARIMA 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.
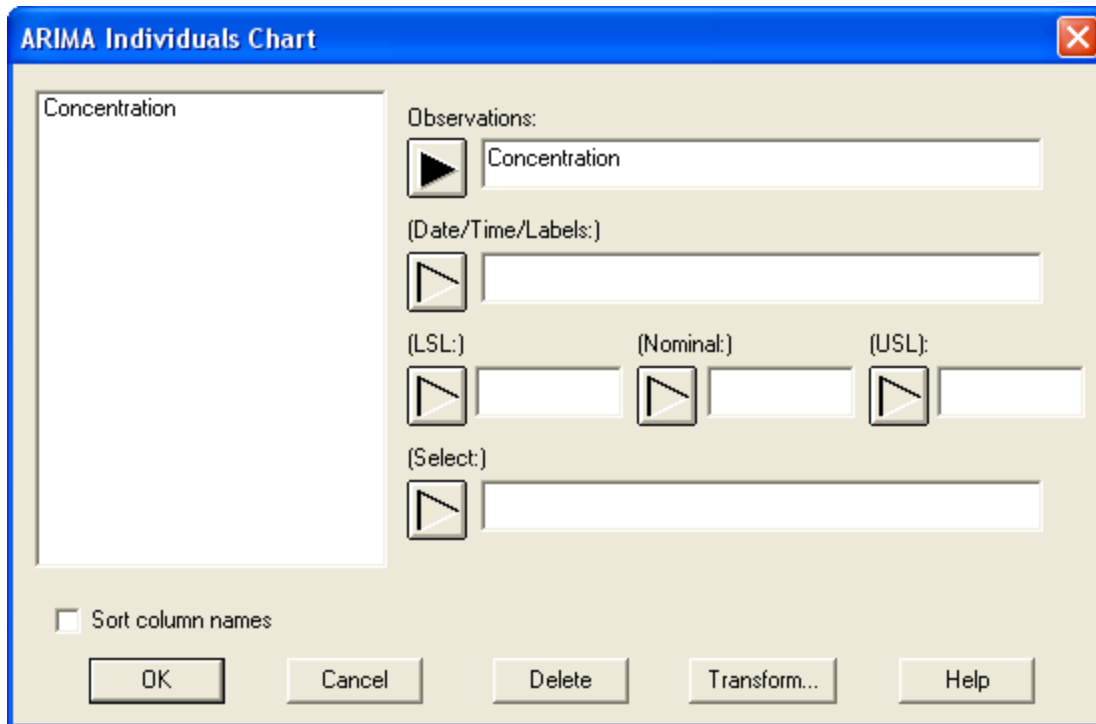
- **Observations:** numeric column containing the data to be analyzed.

- **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	16.62	0.187	5
2	17.04	0.053	5
3	17.22	0.092	5
4	17.14	0.058	5
5	17.36	0.023	5
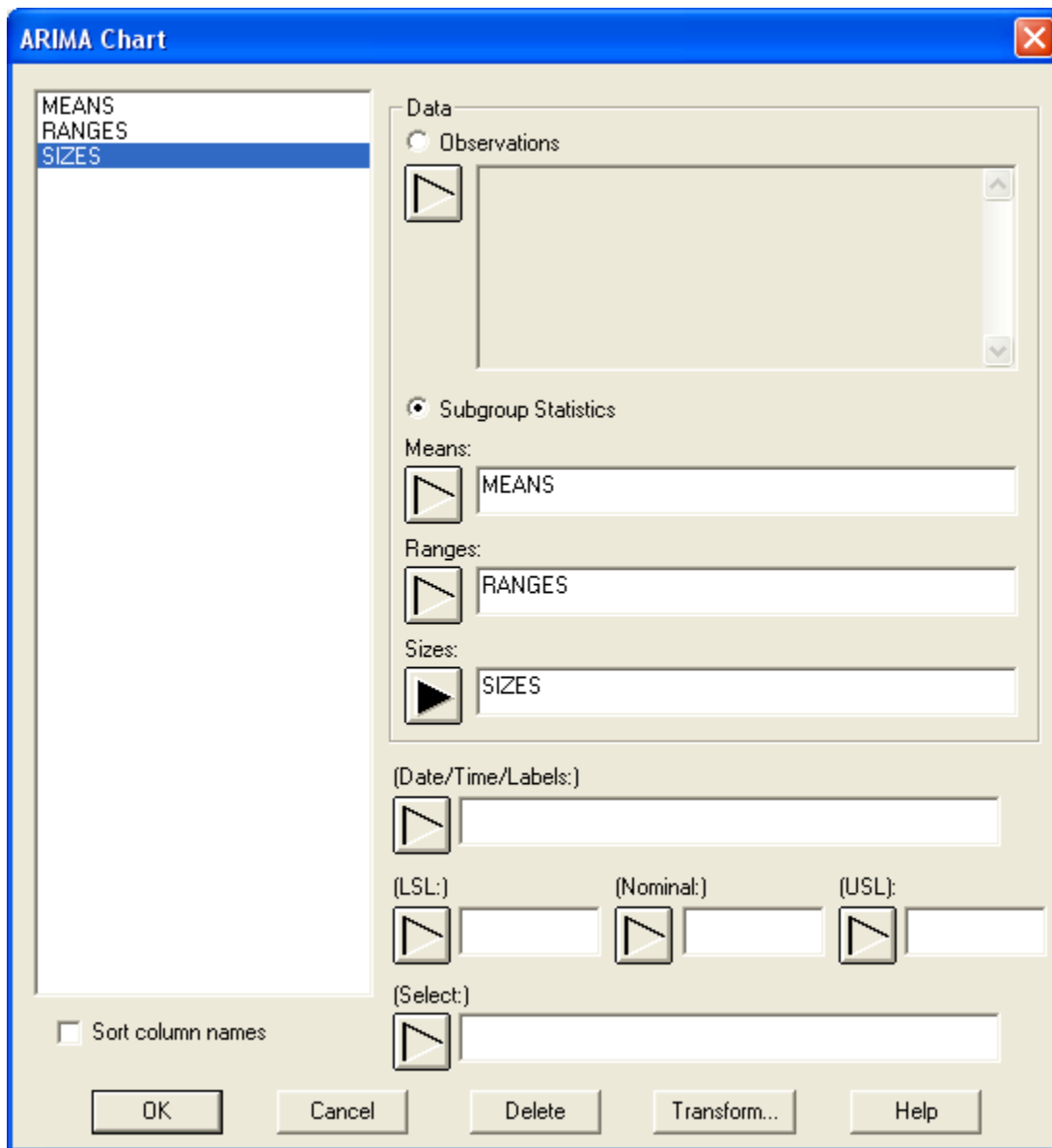
- **Subgroup Statistics**: the names of the columns containing the subgroup means, subgroup ranges, and subgroup sizes.

- **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.

ARIMA Chart

The ARIMA charts can be drawn in several different ways, depending on the settings in *Analysis Options*. For stationary models in which the data vary around a fixed mean µ, a standard type of Shewhart chart can be created.

ARIMA Chart for Concentration
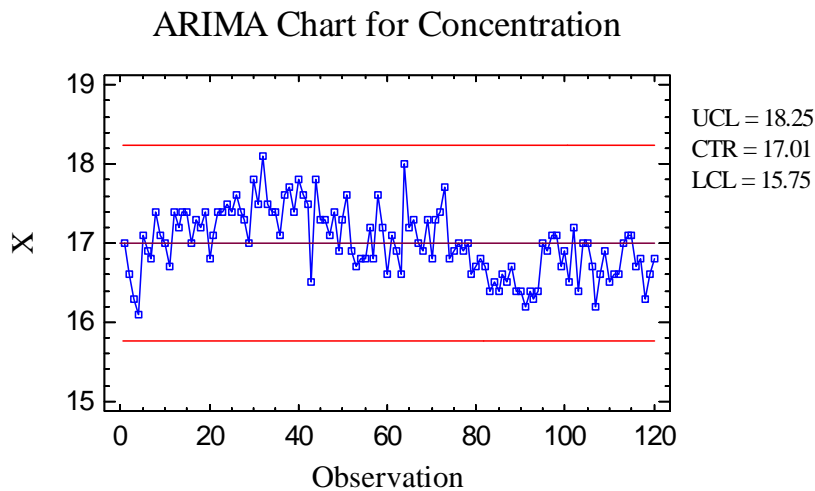
In this chart, the data are drawn around a centerline located at µ with control limits at

$$\hat{\mu} \pm k\hat{\sigma}_z \tag{8}$$

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 mean and standard deviation depend on the ARIMA model that is selected. For an AR(2) model,

$$\mu = \frac{\theta_0}{\left(1 - \phi_1 - \phi_2\right)} \tag{9}$$

and

$$\sigma_z = \sqrt{\left(\frac{1 - \phi_2}{1 + \phi_2}\right) \frac{\sigma_a^2}{\left\{(1 - \phi_2)^2 - \phi_1^2\right\}}} \tag{10}$$

In *Initial Studies* mode, the parameters of the ARIMA model are estimated using an unconditional maximum likelihood estimation procedure with full backforecasting. 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 through the average range, as described below.

Notice that the control limits for the chemical concentrations are considerably wider than on the X chart. The autocorrelation between successive measurements represents a type of dynamic behavior that causes the process to vary around its mean in a pseudo-periodic manner. As shown below, the standard deviation of the measurements is consequently about 25% larger than the standard deviation of the random shocks.

Analysis Summary

This pane displays the fitted ARIMA model and summarizes the control charts.

ARIMA Individuals Chart - Concentration
Number of observations = 120
0 observations excluded

Distribution: Normal
Transformation: none

ARIMA Chart

Period	#1-120
UCL: +3.0 sigma	18.2484
Centerline	17.0067
LCL: -3.0 sigma	15.753

0 beyond limits

MR(2) Chart

Period	#1-120
UCL: +3.0 sigma	1.22844
Centerline	0.375981
LCL: -3.0 sigma	0.0

2 beyond limits

Estimates

Period	#1-120
Process mean	17.0007
Process sigma	0.415913
Mean residual MR(2)	0.375981
Residual sigma	0.333316

Sigma estimated from average moving range of residuals

ARIMA Model Summary

Parameter	Estimate	Stnd. Error	t	P-value
AR(1)	0.349114	0.0868699	4.01881	0.0001
AR(2)	0.335688	0.0869563	3.86043	0.0002
Mean	17.0007	0.0954675	178.079	0.0000
Constant	5.35859			

Backforecasting: yes
Estimated white noise variance = 0.116811 with 117 degrees of freedom
Estimated white noise standard deviation = 0.341777

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.

- **ARIMA Chart:** a summary of the centerline and control limits for the ARIMA 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 ARIMA model. For the chemical concentration measurements, the model is:

$$x_j = 5.359 + 0.349\ x_{j-1} + 0.336\ x_{j-2} + a_j \tag{11}$$

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, as with all the coefficients in the current model, then that coefficient is significantly different from 0 at the 5% significance level. Also shown are the estimated process mean $\hat{\mu} = 17.00$ and the estimated standard deviation of the random shocks $\hat{\sigma}_a = 0.342$.

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.

- **ARIMA Control Limits:** specify the multiple k to use in determining the upper and lower control limits on the ARIMA 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 (or other parameters if not assuming a normal distribution) and the ARIMA model parameters.

- **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 θ_0 can be removed.

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

- **Chart type**: The ARIMA chart may be constructed in any of 4 ways:

1. **Data with long-term limits** - plots the subgroup averages with limits defined by

$$\hat{\mu} \pm k\hat{\sigma}_z \tag{12}$$

2. **Data with One-Step Limits** - plots the subgroup averages with limits defined by

$$\hat{z}_j(j-1) \pm k\hat{\sigma}_a \tag{13}$$

where $\hat{z}_j(j-1)$ is the conditional expectation of z at period j given all information through period j - 1. For example, the conditional expectation for the observation at period j using the fitted AR(2) model for concentration is

$$\hat{z}_j(j-1) = 5.359 + 0.349 z_{j-1} + 0.336 z_{j-2} \tag{14}$$

3. **Residuals** - plots the residuals defined by

$$\hat{a}_j = z_j - \hat{z}_j(j-1) \tag{15}$$

4. **Normalized Residuals** - plots standardized residuals defined by

$$\hat{a}_j / \hat{\sigma}_a \tag{16}$$

- **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*.

Example – Data with One-Step Limits

ARIMA Chart for Concentration
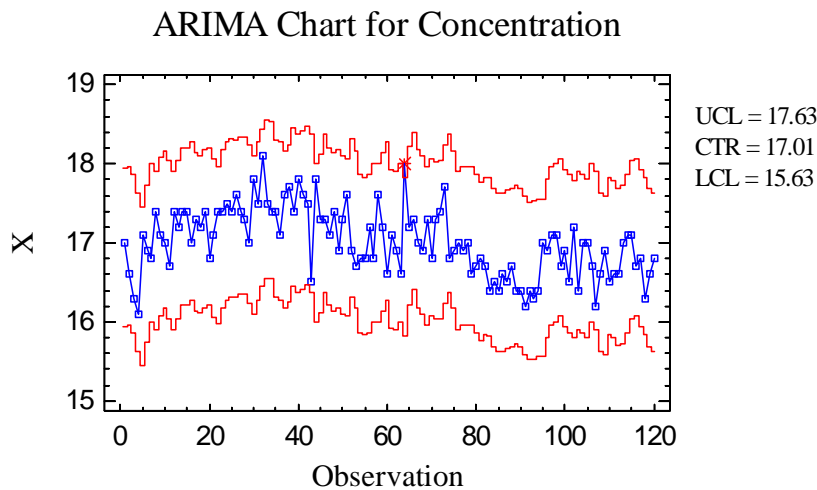
This form of the chart puts sigma limits around the expected observation at time *j*, given all of the information through time *j* - 1. It is useful for determining whether an unusual event has

occurred at time period *j*. In this format, you are looking at the local shocks to the system, rather than taking a more global view of the deviations of the process from its look-term mean.

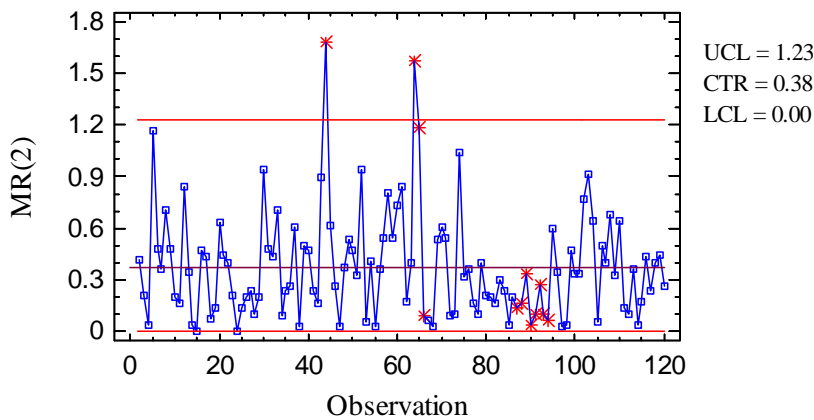Example - Residuals

ARIMA Residual Chart for Concentration
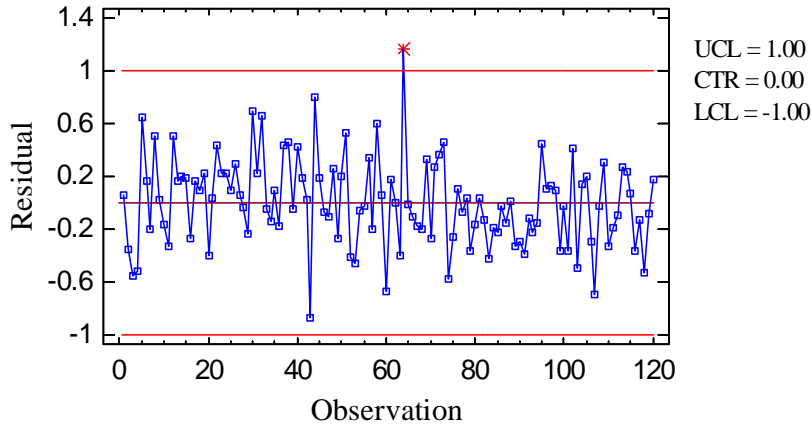
This form of the chart plots the estimated shocks to the system. The out-of-control signals will be the same as for the chart with one-step limits.

MR(2)/R/S Chart

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

MR(2) Chart for Residuals

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:

Preferred Dispersion Chart
- ⦿ Range
- ○ Sigma
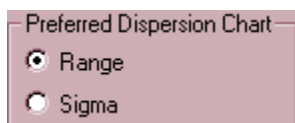
These charts are described in the *X-Bar and R Charts* and the *X-Bar and S Charts* documents.

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

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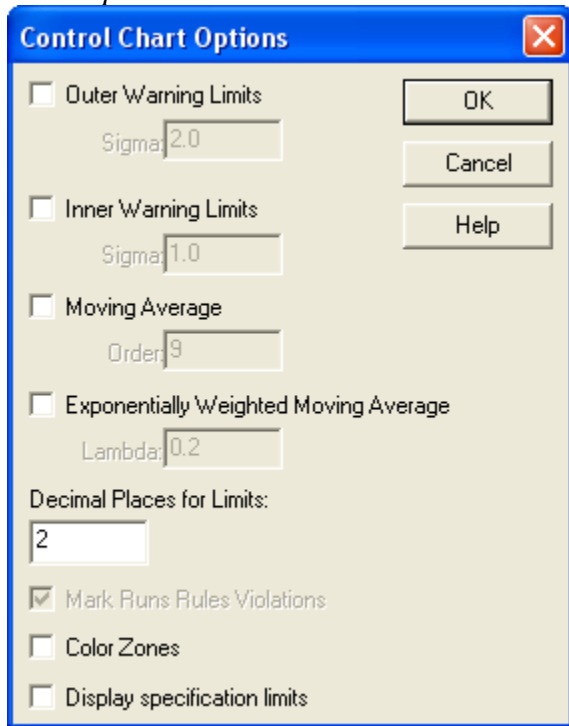
- **Outer Warning Limits**: check this box to add warning limits at the specified multiple of sigma, usually at 2 sigma.

- **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.

- **Display Specification Limits**: whether to add horizontal lines to the chart displaying the location of the specification limits (if any).

ARIMA Chart Report

ARIMA Individuals Chart Report Observations Beyond Limits * = Beyond Limits		
Observation	*X*	*MR(2)*
44	17.8	* 1.68268
64	18.0	* 1.57187

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

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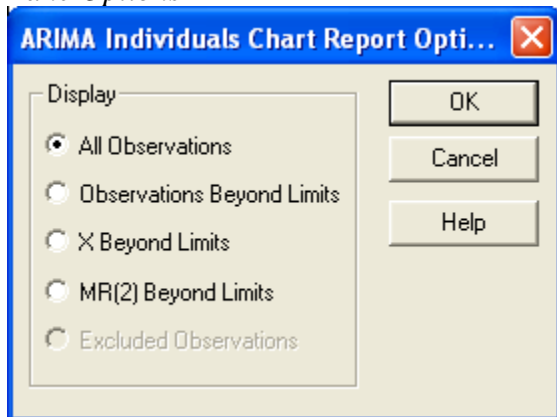
- **Display:** specify the observations or subgroups to display in the report.

Runs Tests

If the type of ARIMA chart selected plots the residuals, then the *Runs Tests* pane displays the results of standard tests designed to look for unusual sequence of points.

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	Residual Chart	MR(2) Chart
65		D
87		A
88		A
89		A
90		A
91		A
92		A
93		A
94		A

For a detailed description of these tests, see the *X-Bar and R Charts* documentation.
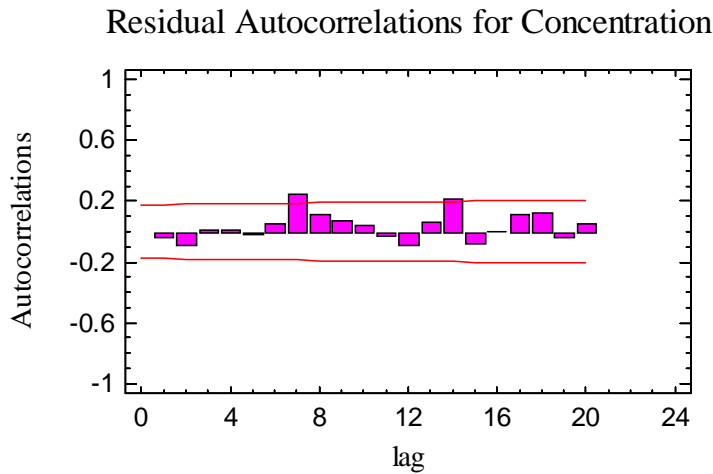
Pane Options
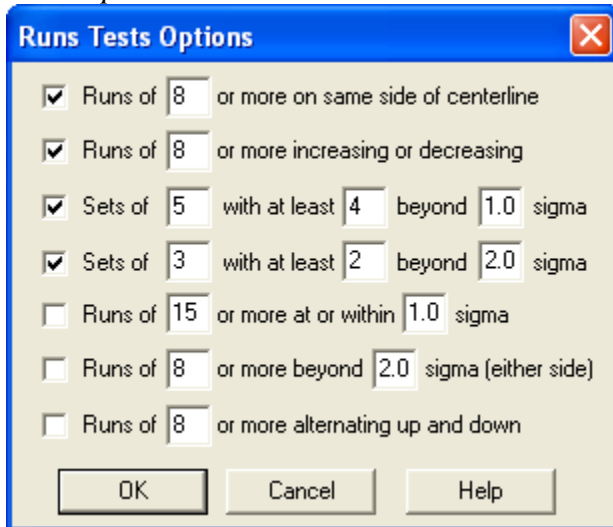
Select the runs tests to be applied and the parameters that define those tests. For example, some practitioners prefer to test for runs of length 7 rather than 8.
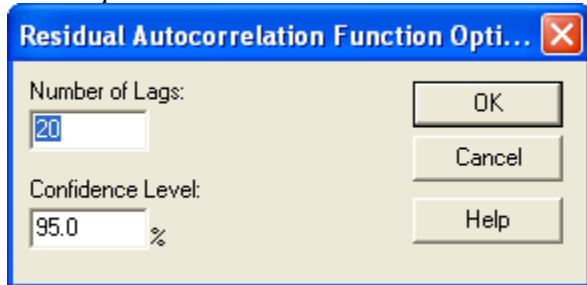
Residual Autocorrelation Function

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

Residual Autocorrelations for Concentration

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.

The above plot does show two small but significant autocorrelations, at separations of 7 and 14 periods. Since measurements were taken once every two hours, there could be a cycle in the data with a period of 14 hours. Note however that the strong correlations at very low lags in the original measurements have been accounted for by the AR(2) model.
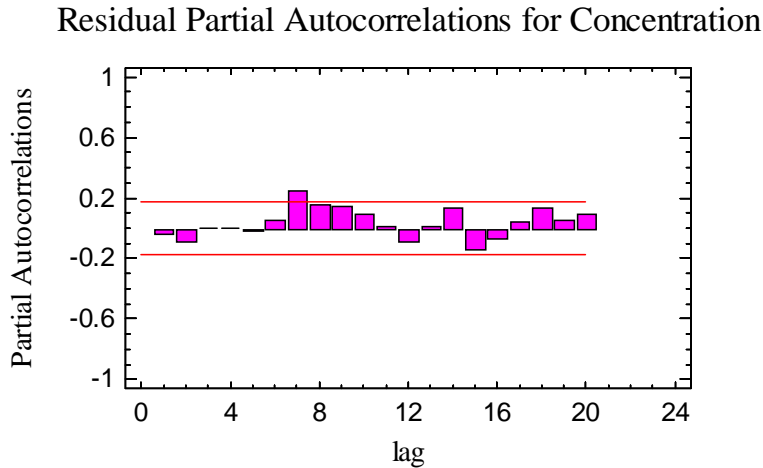
Pane Options

- **Number of Lags**: maximum lag for estimating the autocorrelation.

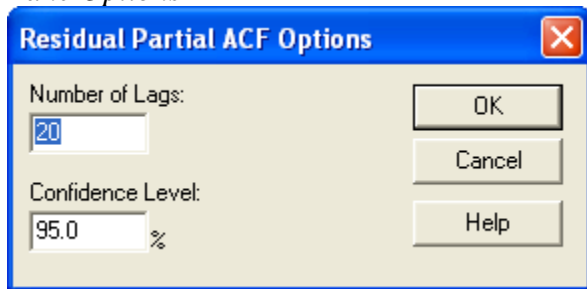- **Confidence Level**: level used to calculate the probability limits.

Residual Partial Autocorrelation Function

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

Residual Partial Autocorrelations for Concentration

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 AR(2) model appears to be sufficient to describe most of the correlation in the data. Again, a possible correlation at lag 7 is indicated, which would be worth investigating.

Pane Options

- **Number of Lags**: maximum lag for estimating the partial autocorrelation.

- **Confidence Level**: level used to calculate the probability limits.

Capability Indices

The *Capability Indices* pane displays the values of selected indices that measure how well the data conform to the specification limits.

Capability Indices for Concentration
Specifications
    USL = 18.5
    Nom = 17.0
    LSL = 15.5

	Short-Term *Capability*	*Long-Term* *Performance*
Sigma	0.415913	0.422027
Cp/Pp	1.20217	1.18476
CR/PR	83.1826	84.4053
Cpk/Ppk	1.20161	1.1842
K		0.000470617

Based on 6 sigma limits. Short-term sigma estimated from average moving range.

The indices displayed by default depend on the settings of the *Capability* tab on the *Preferences* dialog box. A detailed discussion of these indices may be found in the documentation for *Process Capability (Variables)*.

Pane Options
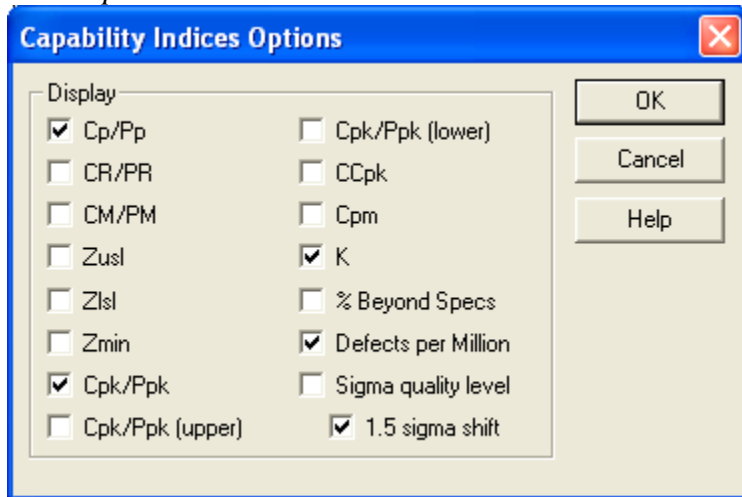
- **Display**: select the indices to be displayed.

Save Results

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

1. *Observations* – the original observations or subgroup means.
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.