

Individuals Control Charts

Summary

The **Individuals Control Charts** procedure creates control charts for a single numeric variable where the data have been collected one-at-a-time rather than in subgroups. It creates both an X chart to monitor the process mean and a moving range (MR) chart to monitor the process variability. Out-of-control signals are highlighted, including both points beyond the control limits and any unusual runs in the data. 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.

Many of the tables and graphs created by this procedure are the same as those created by the *X-bar and R Charts*. For information about options not covered in this document, such as runs tests and ARL curves, consult the documentation for that procedure.

Because of the sensitivity of individuals charts to the assumption that the data follow a normal distribution, this document also describes several options for dealing with non-normal data. In particular, the use of transformations and alternative distributions such as the loglogistic distribution are illustrated.

Sample StatFolio: *individuals.sgp*

Sample Data:

The file *resistivity2.sgd* contains measurements made on the resistivity of 25 silicon wafers, presented by Montgomery (2005). The first 12 samples are shown below:

<i>sample</i>	<i>resistivity</i>	<i>date</i>
1	216	1/1/07
2	290	1/2/07
3	236	1/3/07
4	228	1/4/07
5	244	1/5/07
6	210	1/6/07
7	139	1/7/07
8	310	1/8/07
9	240	1/9/07
10	211	1/10/07
11	175	1/11/07
12	447	1/12/07

It was desired to perform a Phase 1 analysis of this data to determine whether or not the process was in a state of statistical control and to establish control limits for monitoring the process in the future.

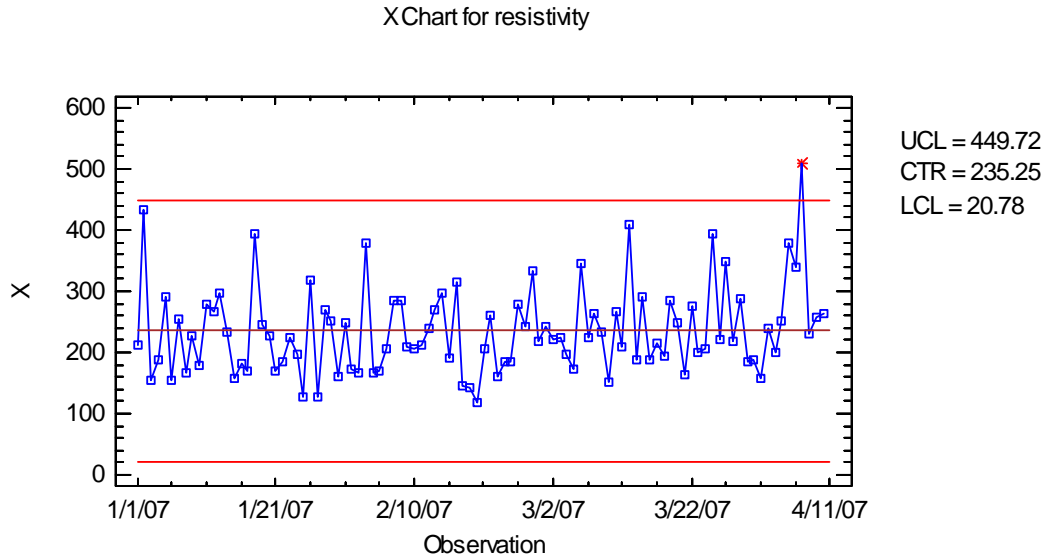
Data Input

The data required for this procedure should be placed in a single numeric column.

- **Observations:** a numeric column containing the measurements.
- **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.

X Chart

This chart plots the individual observations x_i .



In Phase 1 (*Initial Studies*) mode, the centerline and control limits are determined from the data. The centerline is located at the average of the observations:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \tag{1}$$

The control limits are placed above and below the centerline at:

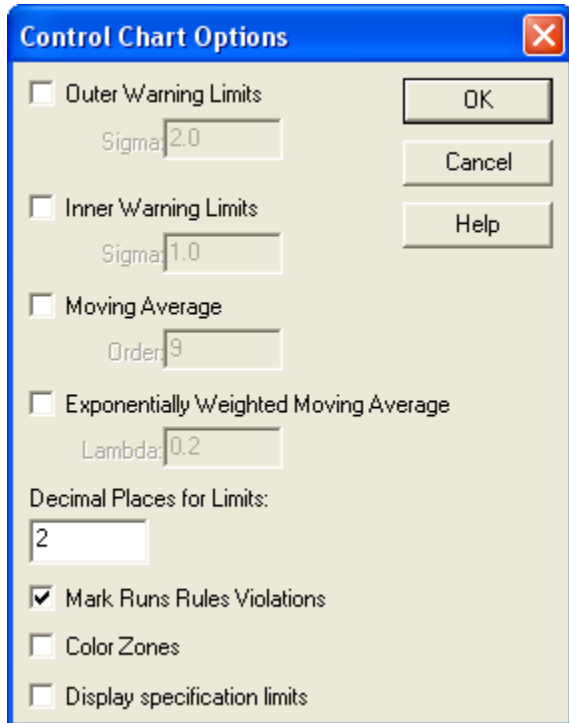
$$\bar{x} \pm k\hat{\sigma} \tag{2}$$

where k is the sigma multiple specified on the *Control Charts* tab of the *Preferences* dialog box ($k = 3$ except in rare cases), and $\hat{\sigma}$ is the estimate of the process sigma. The method for estimating the process sigma also depends on the settings on the *Control Charts* tab of the *Preferences* dialog box, as discussed in the *Analysis Summary* section below.

Any points beyond the control limits will be flagged using a special point symbol. Any point excluded from the analysis, usually by clicking on a point on the chart and pressing the *Exclude/Include* button, will be indicated by an X. If so indicated on the *Pane Options* dialog box, unusual sequences of points may also be flagged.

In the current chart, no unusual points or out-of-control signals are indicated. However, the points do tend to vary further above the centerline than they do below the centerline. This evidence of possible skewness in the data is discussed in some detail below.

Pane Options



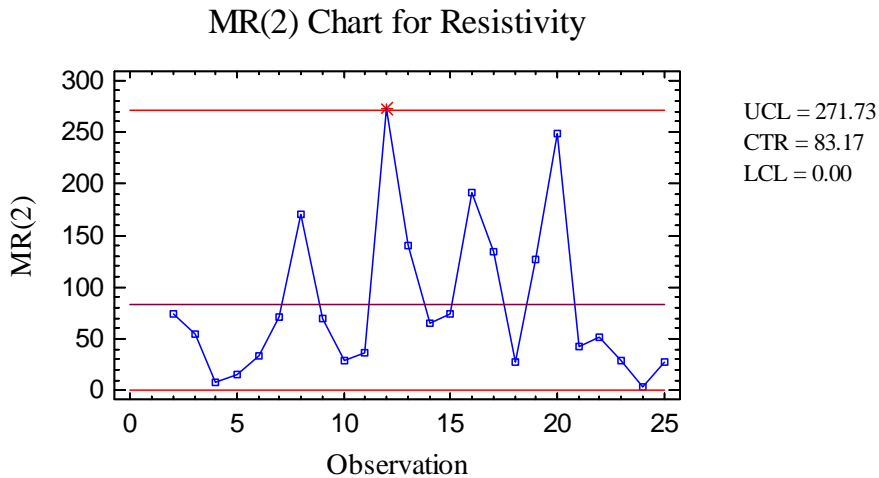
- **Outer Warning Limits:** check this box to add warning limits at the specified multiple of sigma, usually at 2 sigma.
- **Inner Warning Limits:** check this box to add warning limits at the specified multiple of sigma, usually at 1 sigma.
- **Moving Average:** check this box to add a moving average smoother to the chart. In addition to the observations, 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 observations 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 observations, an exponentially weighted moving average of the observations 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 observations 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).

MR(2) Chart

To monitor process variability, STATGRAPHICS calculates the moving range defined by

$$MR_i = |X_i - X_{i-1}| \tag{3}$$

The moving range takes each pair of consecutive observations and treats them as if they were a subgroup of size 2. The *MR(2) Chart* plots these ranges:



In Phase 1 (*Initial Studies*) mode, the centerline and control limits are determined from the data. The centerline is located at:

$$CL = d_2(2)\hat{\sigma} \tag{4}$$

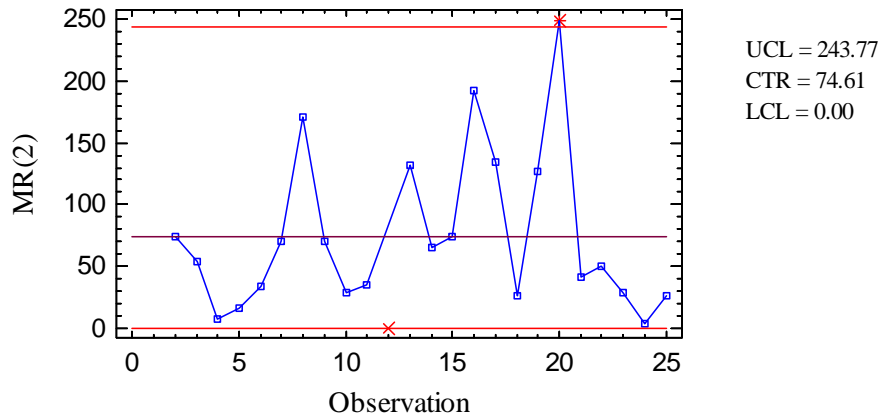
If sigma is estimated from the average range, this equals \overline{MR} . The control limits are placed above and below the centerline at the following locations:

$$CL \pm d_3(2)\hat{\sigma} \tag{5}$$

where k is the sigma multiple specified on the *Control Charts* tab of the *Preferences* dialog box ($k = 3$ except in rare cases) and $\hat{\sigma}$ is the estimate of the process sigma.

The MR(2) chart for the sample data shows one point beyond the upper control limit, which would normally be treated as an out-of-control signal. To see the impact of removing that value, you may click on it with your mouse and then press the *Exclude/Include* button on the analysis toolbar. The process mean and process sigma will be instantly recalculated and the X chart and MR(2) chart will be redrawn. The modified MR(2) chart is shown below:

MR(2) Chart for Resistivity



Having removed the point at $i = 12$, a new point at $i = 20$ now exceeds the upper control limit. Evidently, the moving ranges tend to be larger than expected for data from a stable, normally distributed population. Two explanations are possible:

1. The process is not in a state of statistical control.
2. The measurements do not come from a normal distribution.

Pane Options

The same options exist as for the X chart.

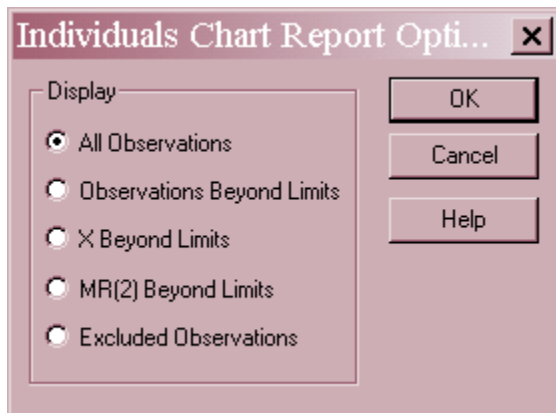
Individuals Charts Reports

This pane tabulates the values plotted on the control charts:

Individuals Chart Report		
All Observations		
X = Excluded * = Beyond Limits		
<i>Observation</i>	<i>X</i>	<i>MR(2)</i>
1/1/07	216.0	
1/2/07	290.0	74.0
1/3/07	236.0	54.0
1/4/07	228.0	8.0
1/5/07	244.0	16.0
1/6/07	210.0	34.0
1/7/07	139.0	71.0
1/8/07	310.0	171.0
1/9/07	240.0	70.0
1/10/07	211.0	29.0
1/11/07	175.0	36.0
1/12/07	X 447.0	X
1/13/07	307.0	132.0
1/14/07	242.0	65.0
1/15/07	168.0	74.0
1/16/07	360.0	192.0
1/17/07	226.0	134.0
1/18/07	253.0	27.0
1/19/07	380.0	127.0
1/20/07	131.0	* 249.0
1/21/07	173.0	42.0
1/22/07	224.0	51.0
1/23/07	195.0	29.0
1/24/07	199.0	4.0
1/25/07	226.0	27.0

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 to display in the report.

Analysis Summary

The *Analysis Summary* summarizes the data and the control charts.

<u>Individuals Charts - Resistivity</u>	
Number of observations = 25	
0 observations excluded	
Distribution: Normal	
Transformation: none	
X Chart	
<i>Period</i>	#1-25
UCL: +3.0 sigma	462.388
Centerline	241.2
LCL: -3.0 sigma	20.0121
0 beyond limits	
MR(2) Chart	
<i>Period</i>	#1-25
UCL: +3.0 sigma	271.729
Centerline	83.1667
LCL: -3.0 sigma	0.0
1 beyond limits	
Estimates	
<i>Period</i>	#1-25
Process mean	241.2
Process sigma	73.7293
Mean MR(2)	83.1667
Sigma estimated from average moving range	

Included in the table are:

- **Number of Observations:** the number of observations n . If any observations have been excluded from the calculations, that number is also displayed.
- **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.
- **X Chart:** a summary of the centerline and control limits for the X chart. Using *Analysis Options*, separate limits may be calculated for different periods (sets of observations).
- **MR(2) Chart:** a summary of the centerline and control limits for the moving range chart.
- **Estimates:** estimates of the process mean μ and the process standard deviation σ . The process mean is estimated from the average of the observations:

$$\hat{\mu} = \bar{x} \tag{6}$$

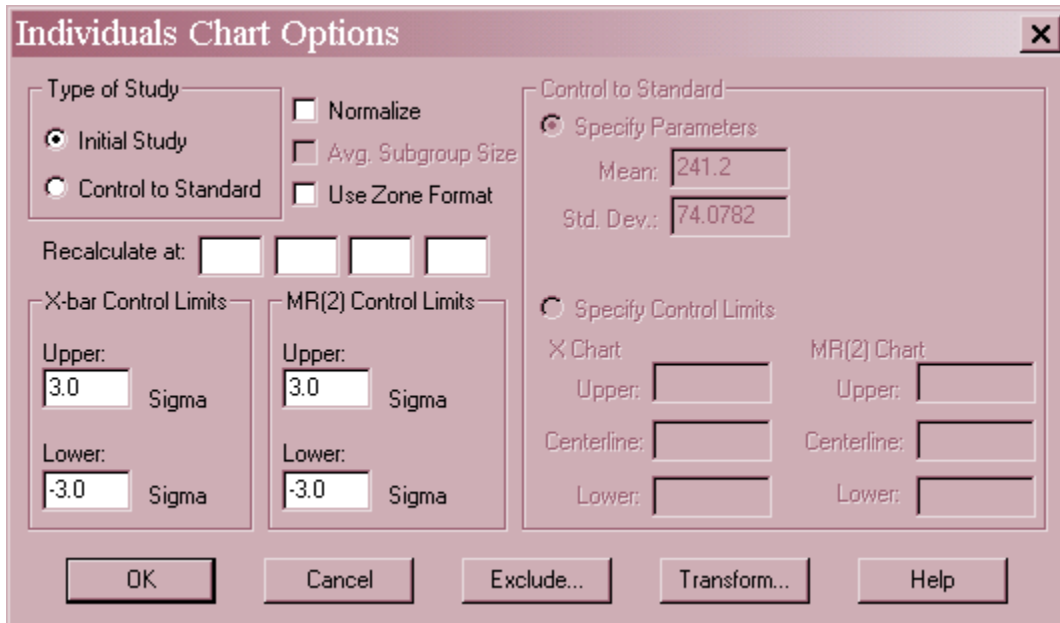
The process sigma may be estimated in any of various ways, depending upon the settings on the *Control Charts* tab of the *Preferences* dialog box, accessible through the *Edit* menu. There are 4 options:

- (1) *From average MR*: the process sigma is estimated from the average of the moving ranges. This is the estimator used in most SPC textbooks.
- (2) *From median MR*: the process sigma is estimated from the median of the moving ranges.
- (3) *From mean SSD with no bias correction*: the process sigma is estimated from the average squared successive difference, i.e., the squared differences between successive observations. An unbiased estimate of σ^2 is obtained.
- (4) *From mean SSD with bias correction*: the process sigma is estimated from the average squared successive difference, i.e., the squared differences between successive observations. The result is multiplied by a constant to create an unbiased estimate of σ .

- **Mean MR(2)**: the average of the moving ranges:

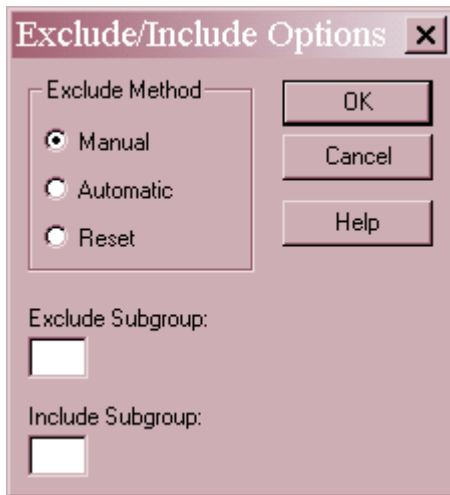
$$\overline{MR} = \frac{\sum_{i=2}^n (x_i - x_{i-1})}{n - 1} \tag{7}$$

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.

- **Normalize:** if selected, all statistics will be normalized by calculating Z-scores and the Z-scores plotted on the charts. The centerline on such a chart is always located at 0, and the control limits are always located at $\pm k$.
- **Use Zone Format:** if checked, the statistics will be plotted using a zone chart rather than the usual format. See the example in the *X-Bar and R Charts* documentation.
- **Recalculate at:** the control limits of the chart may be recalculated at up to 4 locations along the X axis by specifying the observation numbers at which new estimates are to be initiated. Separate estimates of the process mean and sigma will be obtained using the data in the different sections. In such cases, the control limits will be adjusted at the start of each new period.
- **X Control Limits:** specify the multiple k to use in determining the upper and lower control limits on the X-bar 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) 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 either of the following:
 - Specify Parameters:* specify the established standard process mean and sigma (or other parameters if not assuming a normal distribution). These values will then be used to position the centerlines and control limits on the charts.
 - Specify Control Limits:* specify the location of the centerlines and control limits exactly where you wish them to be placed.
- **Exclude button:** Use this button to exclude specific observations from the calculations. It will display the following dialog box:

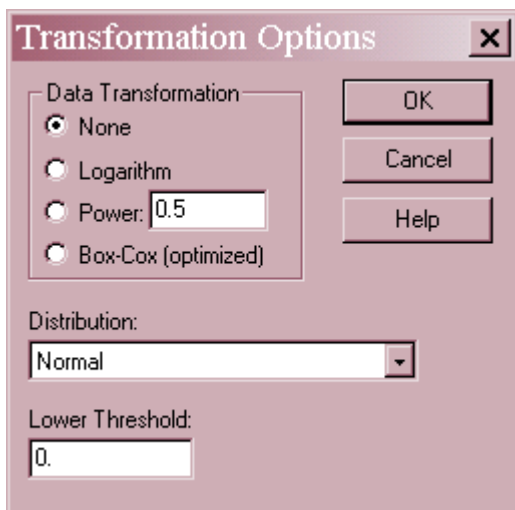


To exclude one or more values from the calculations:

1. Select *Manual*, enter the observation number to exclude, and press OK. The control limits will be recalculated without that observation and the chart replotted.
2. Select *Automatic* and press OK. The program will then remove points one at a time from the control charts, recalculating the limits after each point is removed. It will stop removing points once all remaining points are within the current control limits. This operation is done first to the MR(2) chart and then to the X chart. This option should be used with care and is designed primarily for “what if” types of analyses.

Points may also be excluded by clicking on them while viewing a control chart and then pressing the *Exclude/Include* button on the analysis toolbar.

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



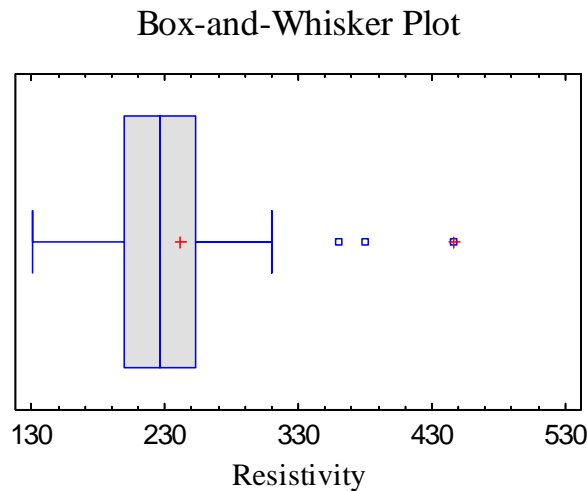
If a *Data Transformation* option is selected, the observations will first be transformed according to the option selected before any calculations are performed. When displaying the X chart, the inverse transformation will be applied to the centerline and control limits so that

the chart is plotted in the original units. On such a chart, the control limits will *not* be equally spaced around the centerline.

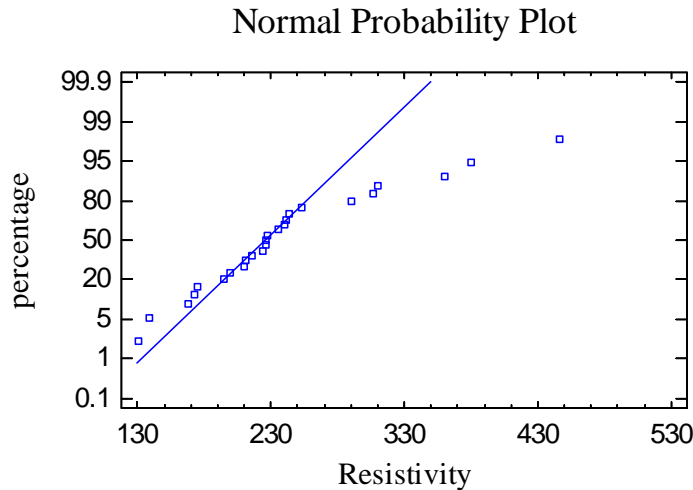
If a *Distribution* other than the Normal is selected, then the data will be normalized (transformed to equivalent Z-scores) before the control charts are calculated. When displaying the X chart, the normalization will be reversed so that the plot may be done in the original metric. As with a data transformation, the control limits will *not* be equally spaced around the centerline. For distribution such as the 3-parameter Weibull and 3-parameter lognormal distributions, you must also specify the value of the lower threshold parameter.

Handling Non-Normal Data

It was commented during the discussion of the X chart above that the data tend to be somewhat skewed in the positive direction. This was indicated by a general tendency to fall farther above the centerline than below it. A box-and-whisker plot for the data, created using the *One Variable Analysis* procedure, shows the classic pattern of positive skewness:



The median, shown by the vertical line in the middle of the box, is below the mean (shown as a + sign). There are also 3 “outside” points on the high side, including 1 far outside point. Further examination of the data using a normal probability plot (with reference line fit using quartiles rather than least squares) shows significant departure from normality, particularly at large values of resistivity:



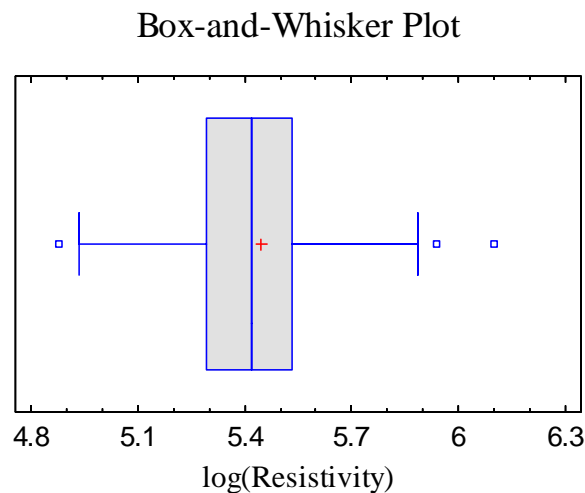
The heavy skewness in the data can distort the interpretation of the X chart and the MR(2) chart, leading to wrong conclusions and to control limits which are likely to be impossible to meet in the future.

There are 2 basic choices in dealing with such positive skewness:

1. Transform the data by taking a transformation such as a logarithm and create the X chart and R chart using $\log(\text{resistivity})$.
2. Base the control charts on a distribution other than the normal.

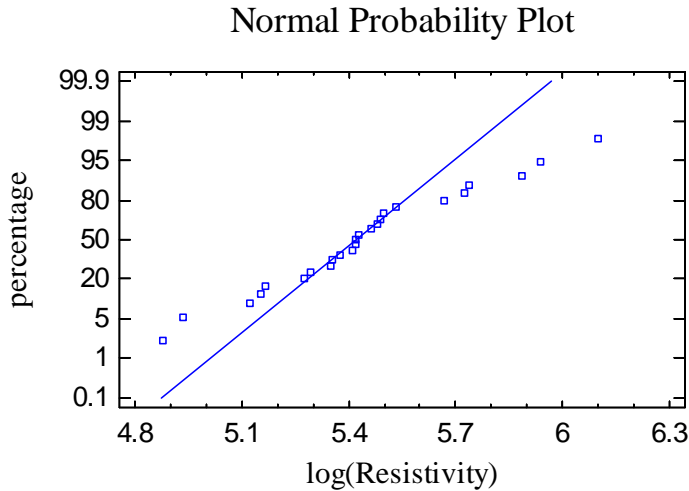
Example: Using a Logarithmic Transformation

Montgomery (2005) demonstrates that a logarithmic transformation, when applied to resistivity, makes the result much more normally distributed. For example, the chart below shows a box-and-whisker plot for the logs:

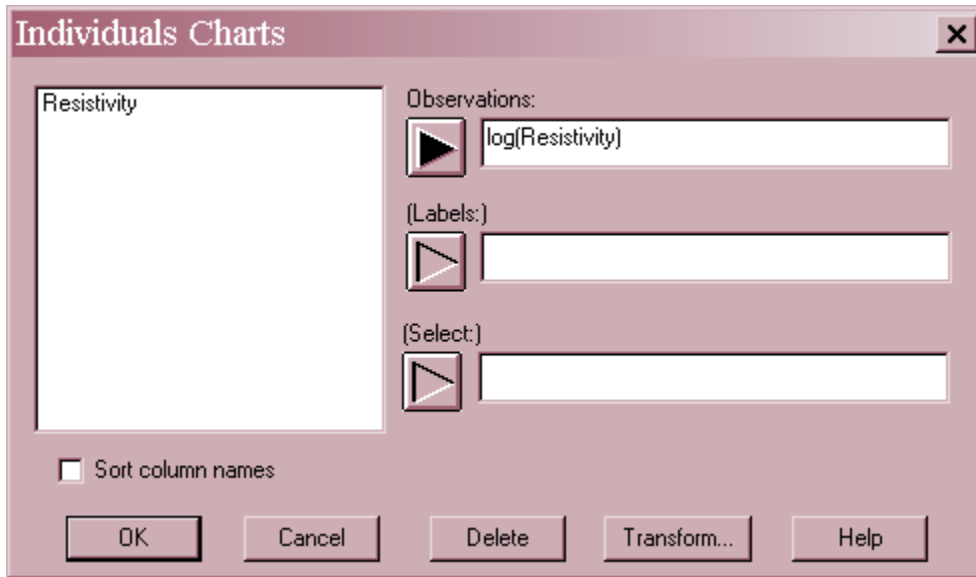


The transformation has made the data considerably more symmetric, as can be seen by the fact that median and mean are now much closer to each other. On the other hand, 3 out of 20 points are still classified as outside, suggesting that the tails of the distribution may be longer than those

of a normal distribution, a condition referred to as having positive *kurtosis*. The normal probability plot of the logs confirms the long tails:

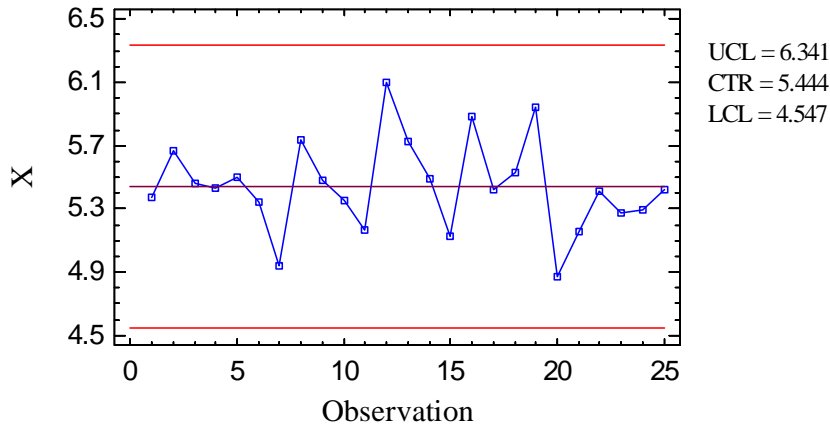


Removing the skewness in the data will improve the performance of the X and MR(2) charts. In STATGRAPHICS, the transformation can be applied in either of two ways. First, the data input dialog box can be modified to request an analysis of the logarithms:

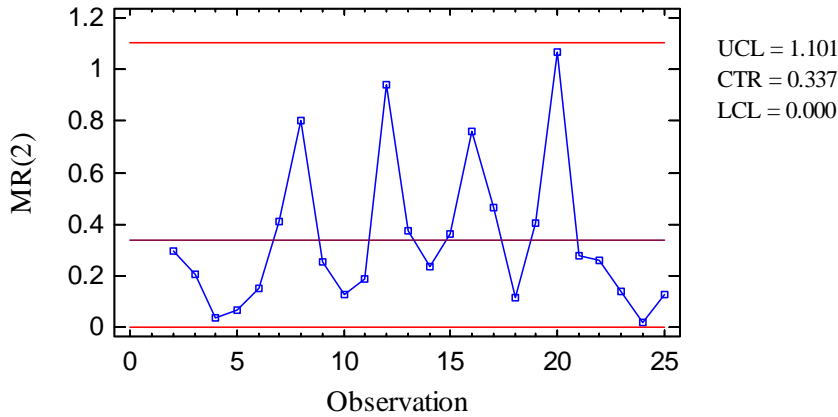


It will be recalled that the LOG operator in STATGRAPHICS calculates the natural logarithms of its argument. The resulting control charts then show the observations in the transformed metric:

X Chart for log(Resistivity)

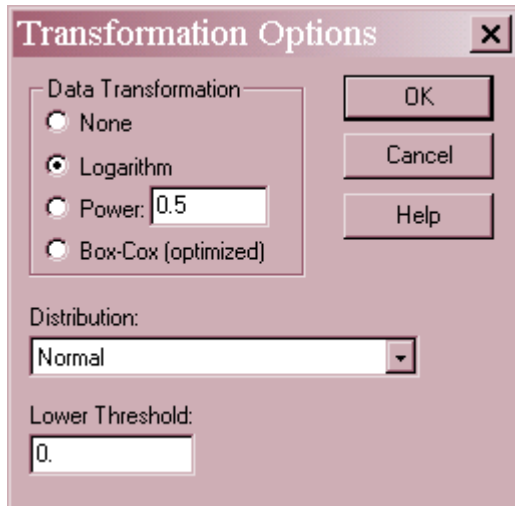


MR(2) Chart for log(Resistivity)

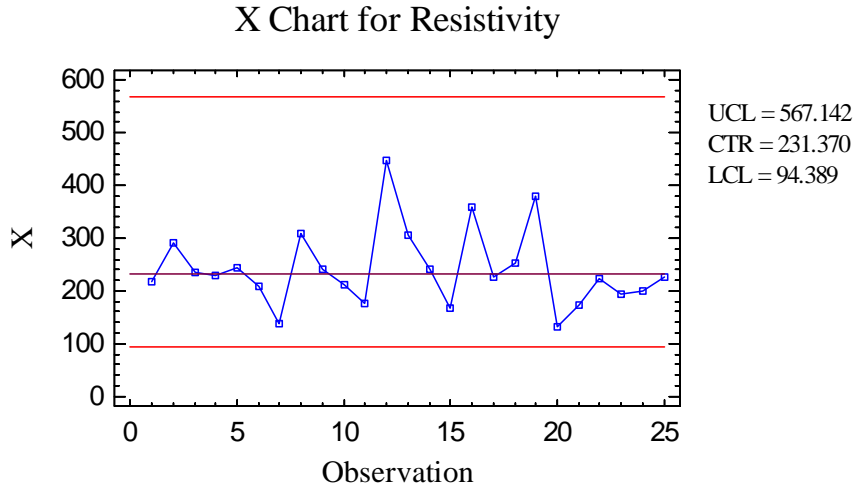


Notice that the effect of the log transformation has been to pull in the upper tails. In fact, the transformation was strong enough to remove the out-of-control signal on the MR(2) chart.

You can also specify a transformation using the *Transform* button on the *Analysis Options* dialog box:



In such a case, you would *not* modify the data input dialog box. The data will then be transformed, control limits calculated, and the transformation reversed before displaying the X chart:



Note that the effect of the transformation has been to space the limits asymmetrically around the centerline, mirroring the positive skewness in the data. These control limits still cover the usual 99.73% of the distribution of X, but now are based on the assumption that the natural logarithms of the data come from a normal distribution.

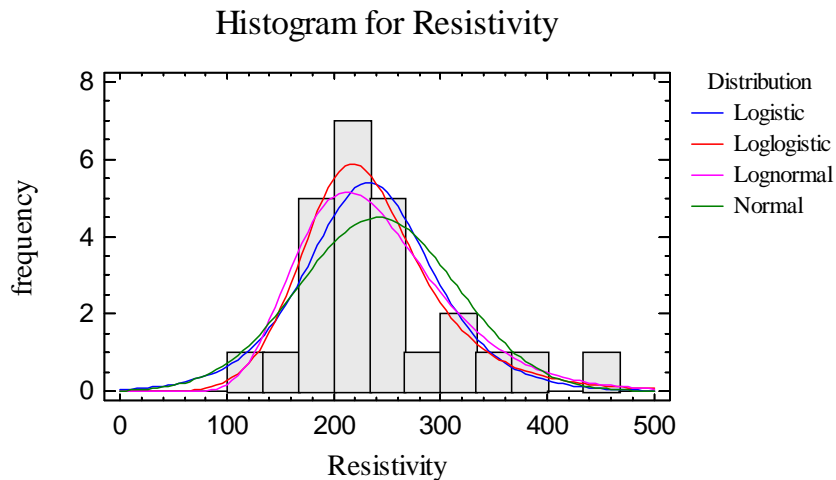
Note: the MR(2) chart is displayed in the transformed metric and looks identical to that created for log(resistivity).

Example: Using a Non-Normal Distribution

Rather than transforming the data, you may select a non-normal distribution instead. The *Distribution Fitting (Uncensored Data)* procedure is a good place to quickly try many alternative distributions. For the resistivities, that procedure creates the following table:

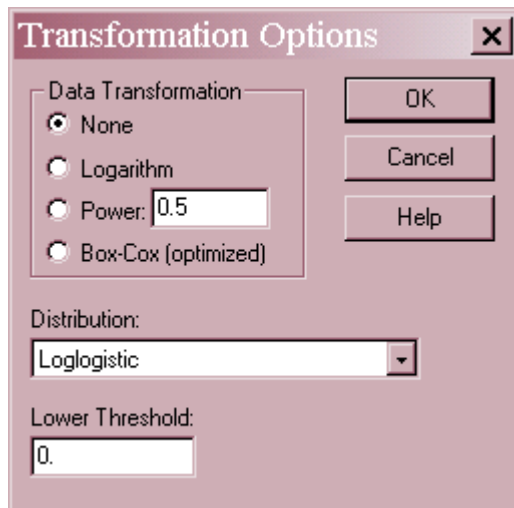
Comparison of Alternative Distributions			
<i>Distribution</i>	<i>Est. Parameters</i>	<i>KS D</i>	<i>A^2</i>
Loglogistic	2	0.122333	0.308004
Largest Extreme Value	2	0.131639	0.338341
Lognormal	2	0.147519	0.379814
Inverse Gaussian	2	0.147339	0.383268
Birnbaum-Saunders	2	0.148146	0.386353
Gamma	2	0.165927	0.479865
Laplace	2	0.0984938	0.501885
Exponential Power	3	0.0985228	0.501945
Rayleigh	2	0.172541	0.517638
Maxwell	2	0.18329	0.584741
Logistic	2	0.148088	0.596628
Normal	2	0.204922	0.846412
Weibull	2	0.202172	0.911631
Smallest Extreme Value	2	0.243874	1.75747

14 different distributions were fit and sorted according to the value of the Anderson-Darling goodness-of-fit statistic. The best fit was obtained by the loglogistic distribution. The logistic distribution is a symmetric distribution, similar to the normal, but with a different shaped peak and tails. The analysis above suggest that a good approach for resistivity is to fit the logistic distribution to $\log(\text{resistivity})$. A plot of several fitted distribution, generated by the *Distribution Fitting* procedure, is shown below:



Of the 4 distributions displayed, the loglogistic is clearly the best in modeling the peak of the histogram.

To construct a control chart assuming that resistivity follows a loglogistic distribution, select *Analysis Options* and press the *Transform* button.

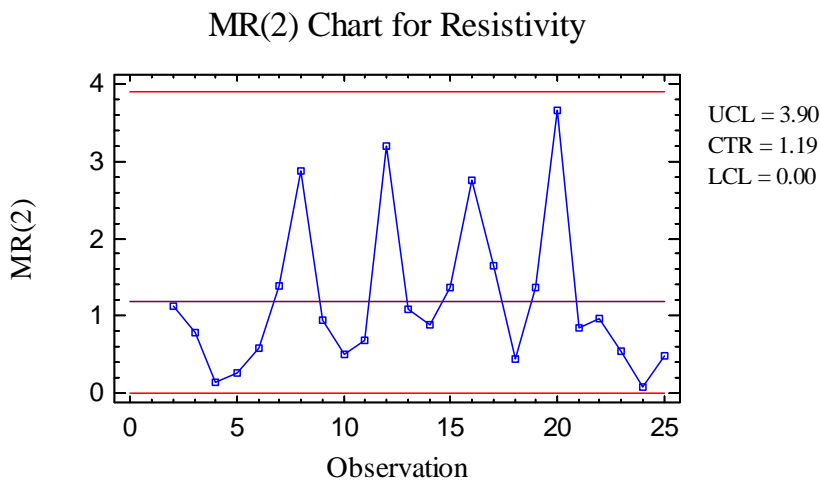
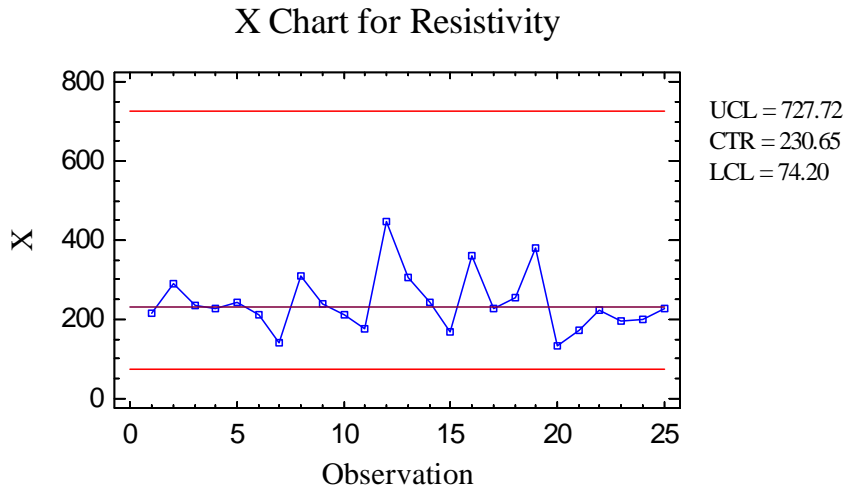


Set the *Data Transformation* field to *None* and select *Loglogistic* from the pulldown list under *Distribution*. All of the control charts will now be created by:

1. Fitting a loglogistic distribution to the observed values of resistivity.
2. Transforming each data value into an equivalent Z-score by evaluating the inverse cumulative distribution function at the observed value.
3. Creating normalized control charts for the Z-scores.

- Reversing the transformation so that the X chart may be displayed in the original units of resistivity.

The resulting charts are shown below:



The MR(2) chart is very similar to that obtained when using the logarithms of resistivity. The control limits of the X chart are farther apart, however, corresponding to the longer tails of the logistic distribution.

Note: the analysis based on the log transformation of resistivity could have been duplicated exactly by selecting the *lognormal* distribution instead of the *loglogistic*.

Save Results

The following results can be saved to the datasheet:

1. *Means* – the observations, transformed if so requested.
2. *Moving Ranges* – the moving ranges.
3. *Labels* – the observation labels.
4. *Process Mean* – the estimated process mean.
5. *Process Sigma* – the estimated process standard deviation.
6. *Included Observations* – a column of 0's and 1's for excluded and included observations, respectively. This column can then be used in the *Select* field on other data input dialog boxes.

Calculations**Estimate of Process Sigma**

(1) From average MR:

$$\hat{\sigma} = \frac{\sum_{i=2}^n MR_i}{(n-1)d_2(2)} \quad (8)$$

(2) From median MR:

$$\hat{\sigma} = \frac{1}{d_4(2)} \tilde{MR} \quad (9)$$

(3) From mean SSD with no bias correction:

$$\hat{\sigma} = \sqrt{\frac{1}{2(n-1)} \sum_{i=2}^n (x_i - x_{i-1})^2} \quad (10)$$

(4) From mean SSD with bias correction:

$$\hat{\sigma} = \frac{1}{c'_4(2)} \sqrt{\frac{1}{2(n-1)} \sum_{i=2}^n (x_i - x_{i-1})^2} \quad (11)$$