SPC – Response Variable

This procedure creates control charts for data in the form of continuous variables. Such charts are widely used to monitor manufacturing processes, where the data often represent measurements of important features of the items being produced. Measurements may be taken either one at a time (individuals data) or in subgroups of *n* items each. It also performs a capability analysis and calculates capability indices.

The data for this analysis consist of *m* samples from a population, each consisting of one or more measurements. Let

x_{ij} = i-th measurement in sample j

n_j = size of sample j

Depending on the type of data provided, five types of charts can be created:

- **X Chart** – for individuals data, plots the observation in each sample.
- **X-Bar Chart** – for subgroup data, plots the average of the measurements in each sample.
- **MR(2) Chart** – for individuals data, plots moving ranges calculated from the absolute difference between successive measurements.
- **R Chart** – for subgroup data, plots the range of the measurements in each sample.
- **S Chart** – for subgroup data, plots the standard deviation of the measurements in each sample.

The capability indices that may be calculated include Cp, Cpk, Pp, Ppk, K, DPM, and SQL.

Access

Highlight: one or more *Response* columns. A *Time index* column may also be selected to locate the points along the X axis.

Select: *SPC* from the main menu.

Output Page 1: An X chart if one *Response* column is highlighted. An X-bar chart if more than one column is highlighted.

Output Page 2: An MR(2) chart if one *Response* column is highlighted. An R chart or S chart if more than one column is highlighted, depending on the properties of the leftmost *Response* column.

Output Page 3: Capability histogram and indices.

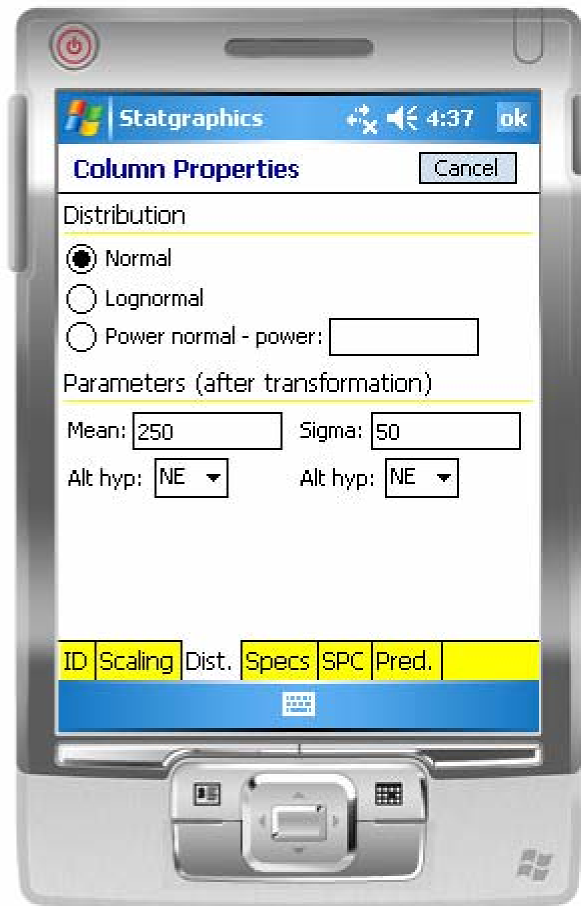
© 2006 by StatPoint, Inc.

Options

The type of analysis performed depends on the properties of the leftmost *Response* column selected. One important property is the assumed distribution for the data. To specify the distribution:

1. Access the *Properties* dialog box for the leftmost *Response* column by double-clicking on the column header.

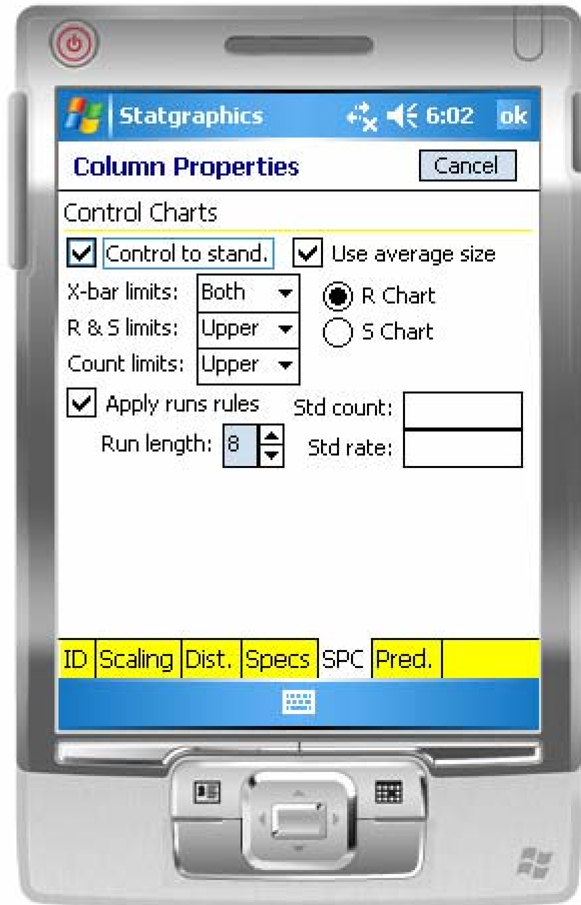
2. On the *Dist.* tab, select the assumed distribution. The default selection assumes that the data follow a normal distribution. If you select *Lognormal*, the logarithms of the data will be assumed to follow a normal distribution. If you select *Power normal*, the data will be assumed to follow a normal distribution after raising them to the indicated *power*.



Both *Phase I* (Initial Studies) and *Phase II* (Control to Standard) control charts can be created. In a Phase I control chart, the centerline and control limits are determined from the current data. In a Phase II control chart, the centerline and limits are determined from a user-specified standard. Phase I is the default. To create a Phase II control chart:

1. Specify standard values for the *Mean* and *Sigma* on the *Dist*. tab of the *Column Properties* dialog box for the leftmost *Response* column. These values will be used to position the control limits.

2. On the *SPC* tab, check the box labeled *Control to stand.*



The *X-bar limits* and *R & S limits* pulldown lists indicate whether lower and/or upper control limits should be plotted. The *Use average size* field indicates whether the average sample size should be used to generate horizontal control limits, or whether individual sample sizes should be used (resulting in jagged control limits if the sample sizes are not all equal).

The choice between an *R Chart* and an *S Chart* affects the output for subgroup data. In addition to defining which chart will be displayed, it determines the method used to estimate the short-term process sigma.

If *Apply runs rules* is checked, any unusual sequences of points will be indicated, defined as follows:

1. Runs up or down of length greater than or equal to the number indicated in the *Run length* field.

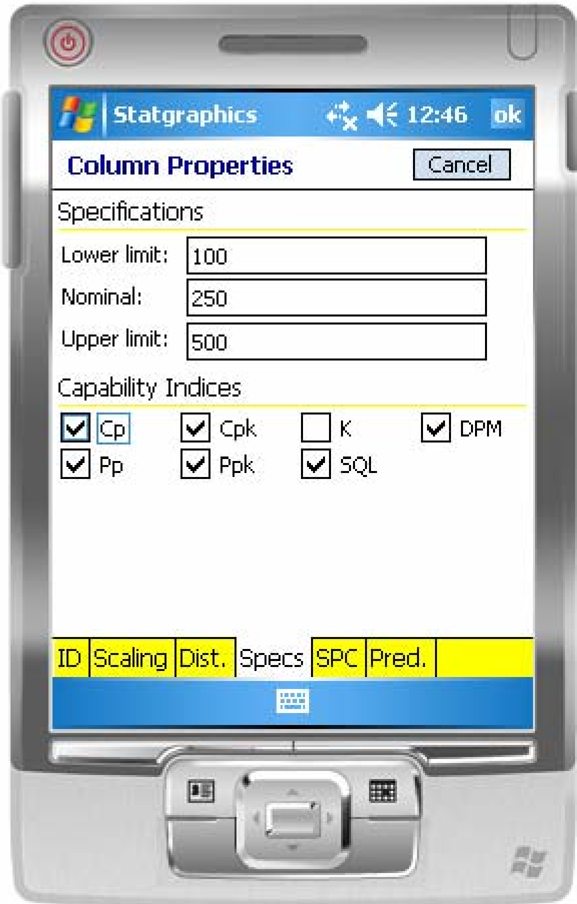
© 2006 by StatPoint, Inc.

2. Runs above or below the centerline of length greater than or equal to the number indicated in the *Run length* field.

3. 2 out of 3 points in a row beyond 2-sigma, on the same side of the centerline.

4. 4 out of 5 points in a row beyond 1-sigma, on the same side of the centerline.

If you intend to perform a capability analysis, you should also set the options on the *Specs* tab:



Enter the desired lower specification limit, the nominal or target value, and the upper specification limit. Also, select the capability indices that you would like to calculate.

Sample Data

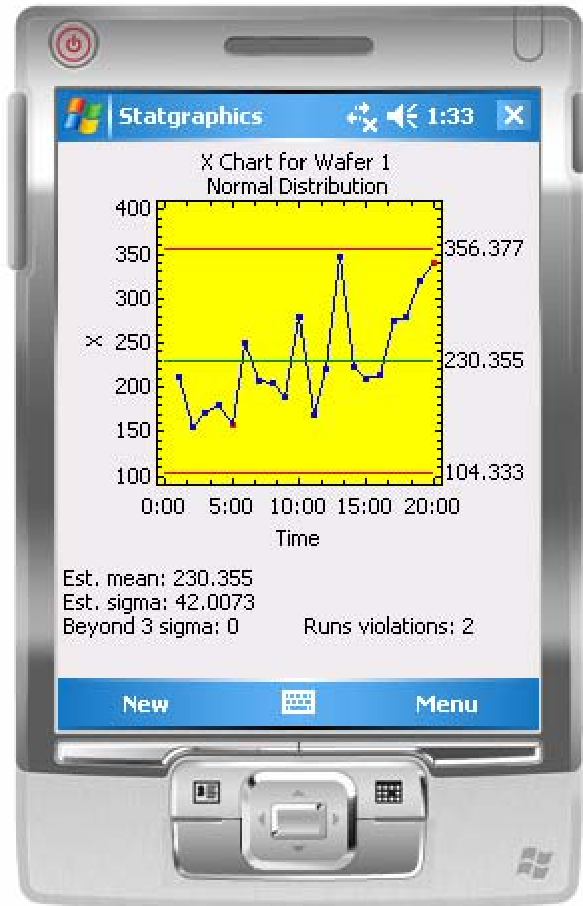
The file *wafers.sgm* contains data on m = 20 samples, each containing the measured resistivity of n = 5 silicon wafers. The first five samples are shown below:

Sample	Time	Wafer 1	Wafer 2	Wafer 3	Wafer 4	Wafer 5
1	1:00	211.2	128.4	154.9	186.8	156.7
2	2:00	155.4	255.7	165.8	227.9	178.8
3	3:00	171.0	205.6	285.7	284.9	207.7
4	4:00	181.1	169.0	393.1	246.2	228.2
5	5:00	158.0	239.6	200.2	250.8	189.1

© 2006 by StatPoint, Inc.

X Chart

The *X Chart* plots the measurements in a single *Response* column. Selecting only the data in the columns named *Time* and *Wafer 1* yields the following chart:



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:

$$\bar{x} = \frac{\sum_{j=1}^{m} x_j}{m} \tag{1}$$

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

$$\bar{x} \pm 3\hat{\sigma} \tag{2}$$

$\hat{\sigma}$ is the estimate of the short-term process sigma, calculated from

© 2006 by StatPoint, Inc.

$$\hat{\sigma} = \frac{\sum_{j=2}^{m} MR_j}{(m-1)d_2(2)} \tag{3}$$

where MR_j is the moving range at time period j, defined as

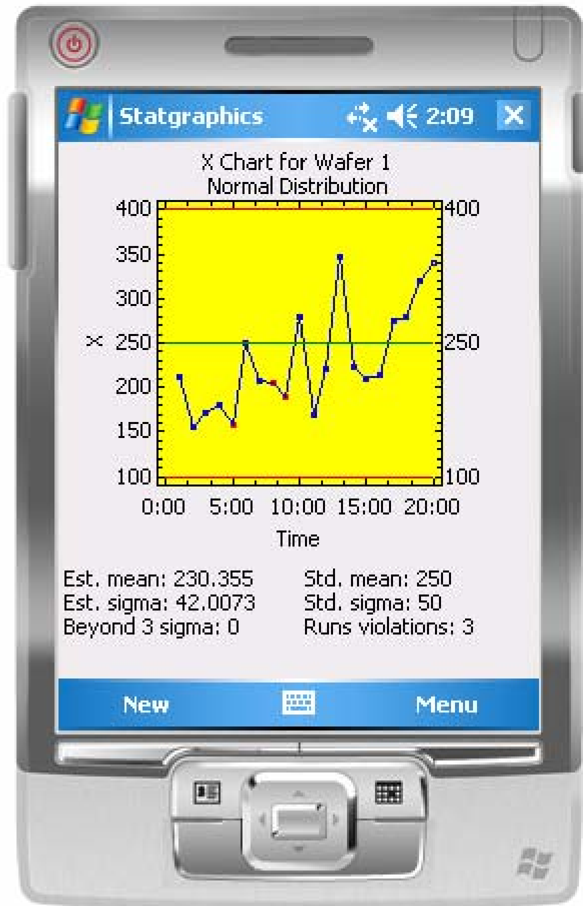
$$MR_j = \left| X_j - X_{j-1} \right| \tag{4}$$

Any points beyond the control limits are flagged using the *Highlighting* point symbol. If runs rules are being applied, any point at which a violation occurs is also highlighted. In the sample data, the last two points are highlighted, corresponding to the following runs rules violations:

(1) At 5:00, 4 of the last 5 points were more than 1 sigma below the centerline.

(2) At 20:00, 2 of the last 3 points were more than 2 sigma above the centerline.

In control to standard mode, the centerline $\bar{x}$ is replaced in the above equations by the specified *Mean* on the *Dist*. tab of the Column Properties dialog box, and $\hat{\sigma}$ is replaced by the specified *Sigma*. A typical example is shown below:



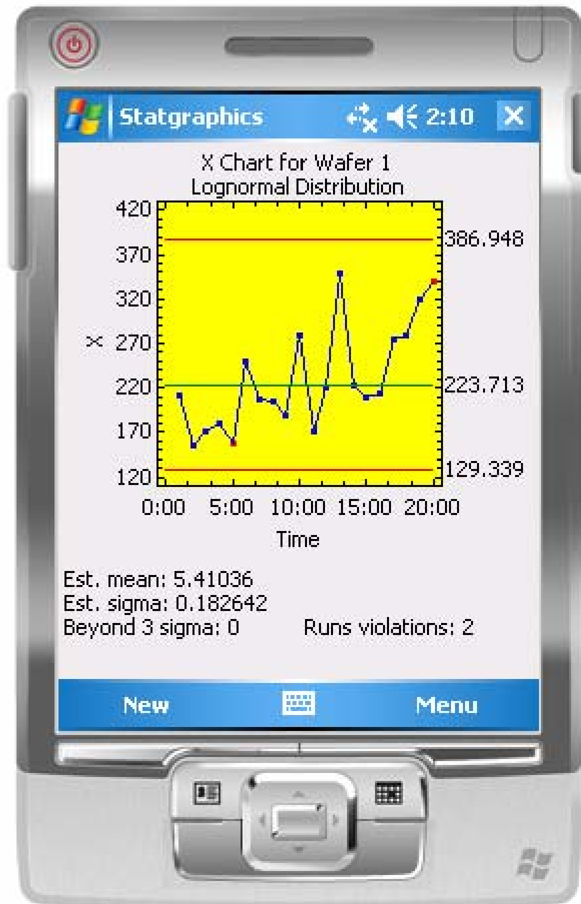
If the data are assumed to come from a distribution other than the normal distribution, then the following approach is used:

1. The data are transformed by taking logarithms if a lognormal distribution is specified on the *Dist*. tab of the *Column Properties* dialog box, or by raising the data to the indicated power if a power normal distribution is specified.

2. Control limits are calculated in the transformed metric.

3. The inverse transformation is applied to the control limits before displaying the graph.

For the resistivity data, selecting a lognormal distribution gives the following result:



Note that the control limits are not symmetrically spaced around the centerline after the inverse transformation is applied. This reflects the positive skewness of a lognormal distribution. Note also that the estimated mean and sigma are displayed for the transformed data, i.e., for the natural logarithm of resistivity.

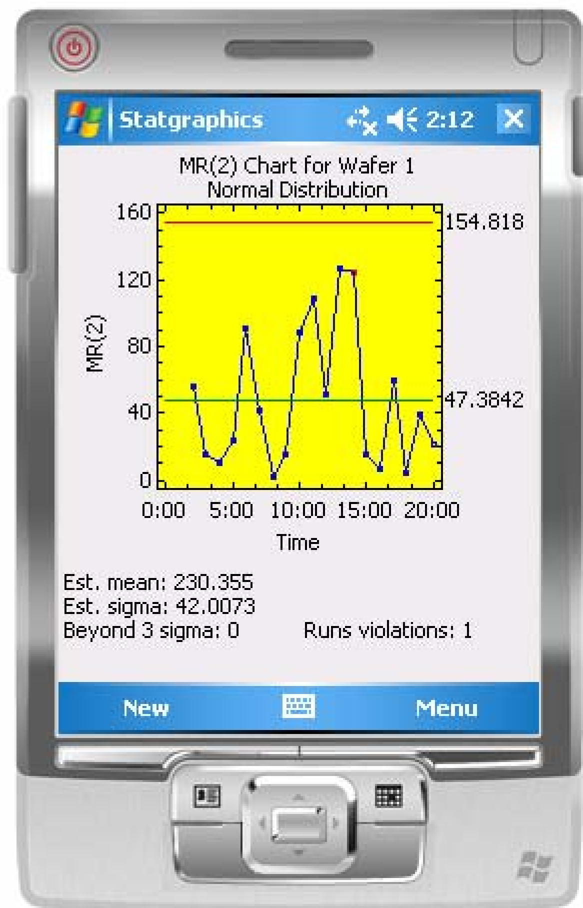
MR(2) Chart

The *MR(2) Chart* plots the moving ranges for individuals data, as defined in equation (4). The moving ranges provide an estimate of the short-term variability amongst the observations. In Phase 1 (*Initial Studies*) mode, the centerline and control limits are determined from the data. The centerline is located at:

$$CL = d_2(2)\hat{\sigma} \quad (5)$$

which is equivalent to the average of the moving ranges. The control limits are placed above and below the centerline at:

$$CL \pm 3d_3(2)\hat{\sigma} \quad (6)$$



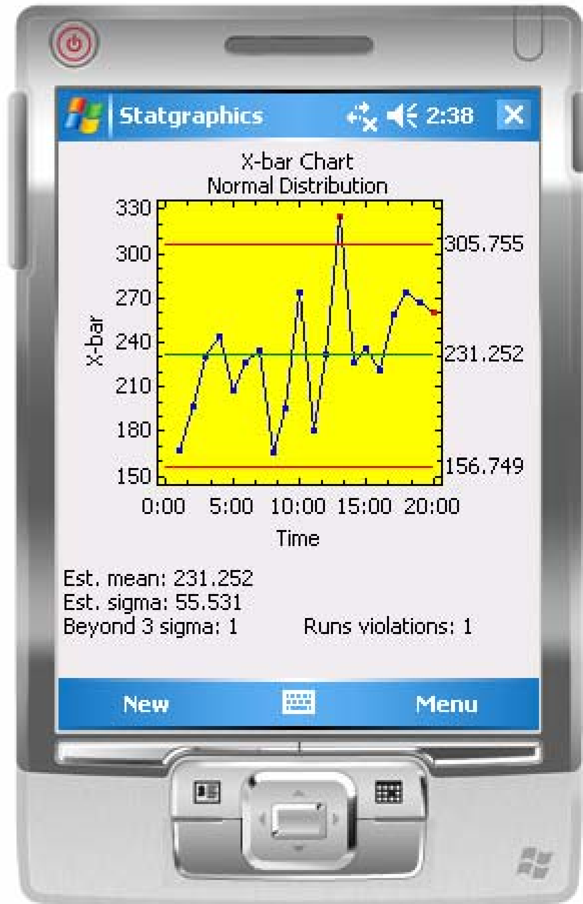
Note a runs rules violation at period 20, when the last 2 out of 3 observations were more than 2-sigma above the centerline.

In control to standard mode, $\hat{\sigma}$ is replaced by the specified *Sigma* on the *Dist*. tab of the *Column Properties* dialog box. If a non-normal distribution has been specified, the MR(2) chart is displayed in the transformed metric.

© 2006 by StatPoint, Inc.

X-Bar Chart

The *X-bar Chart* is displayed when one or more samples consists of more than one measurement. It plots the subgroup means $\bar{x}_j$.



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

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{m} n_j \bar{x}_j}{\sum_{j=1}^{m} n_j} \tag{7}$$

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

$$\bar{\bar{x}} \pm 3\frac{\hat{\sigma}}{\sqrt{n}} \tag{8}$$

where $\hat{\sigma}$ is the estimate of the short-term process sigma, and n is the subgroup size. If the subgroup sizes are not equal, then depending on the settings of *Use average size* on the *SPC* tab of the *Column Properties* dialog box for the leftmost *Response* column, n is replaced by either:

> (1) $\bar{n}$, the average subgroup size. In this case, the control limits are the same for all subgroups.
>
> (2) n_j, the individual subgroup sizes. In this case, the control limits are step functions.

The method for estimating the short-term process sigma also depends on the settings on the *SPC* tab of the leftmost *Response* column. If an accompanying *R chart* has been requested, then σ is estimated from the subgroup ranges R_j according to:

$$\hat{\sigma} = \frac{\sum_{j=1}^{m}\left(\frac{f_j R_j}{d_2(n_j)}\right)}{\sum_{j=1}^{m} f_j} \tag{9}$$

where

$$f_i = \frac{d_2^2(n_j)}{d_3^2(n_j)} \tag{10}$$

If an accompanying *S chart* has been requested, then σ is estimated from the subgroup standard deviations s_j according to:

$$\hat{\sigma} = \frac{\sum_{j=1}^{m}\left(\frac{h_j s_j}{c_4(n_j)}\right)}{\sum_{j=1}^{m} h_j} \tag{11}$$

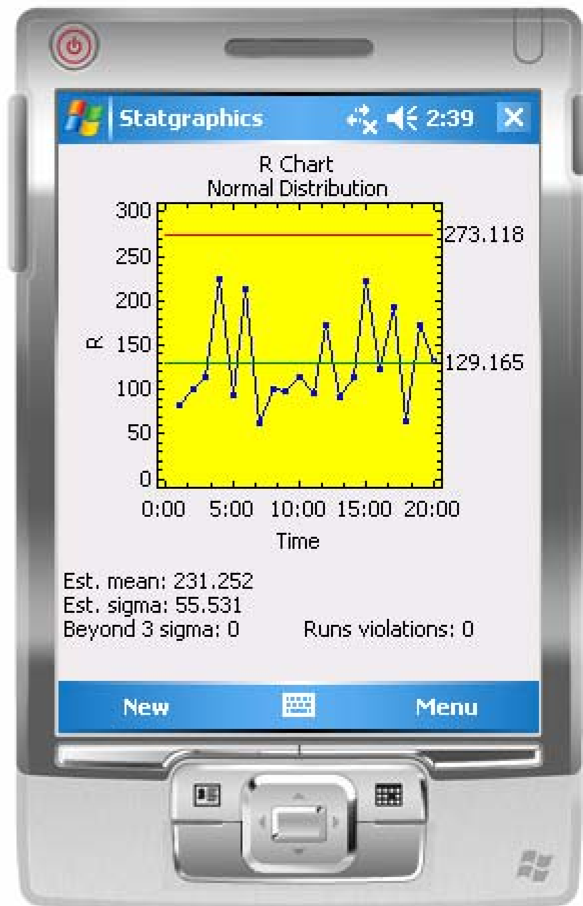
where

$$h_i = \frac{c_4^2(n_j)}{1 - c_4^2(n_j)} \tag{12}$$

For Phase II control charts or for data that is not assumed to follow a normal distribution, the same modifications are made as for the *X Chart* discussed earlier.

R Chart

The *R Chart* plots the subgroup ranges R_j.



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

$$CL = d_2(\bar{n})\hat{\sigma} \tag{13}$$

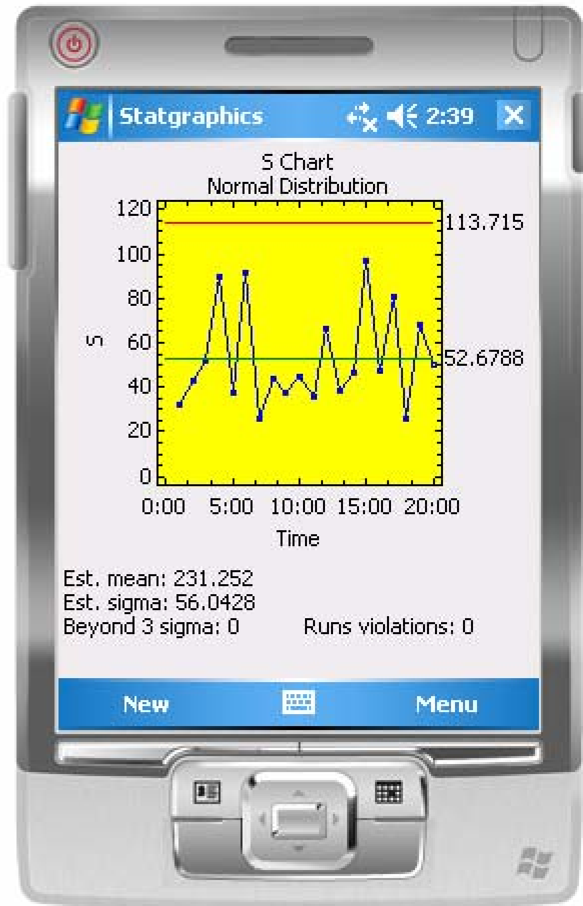
which equals $\bar{R}$. The control limits are placed above and below the centerline at the following locations:

$$CL \pm 3d_3(n)\hat{\sigma} \tag{14}$$

In control to standard mode, $\hat{\sigma}$ is replaced by the specified *Sigma* on the *Dist.* tab of the *Column Properties* dialog box.

S Chart

The *S Chart* plots the subgroup standard deviations s_j.



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

$$CL = \hat{\sigma} \tag{15}$$

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

$$CL \pm \frac{k\hat{\sigma}}{c_4(n)}\sqrt{1 - c_4^2(n)} \tag{16}$$

In control to standard mode, $\hat{\sigma}$ is replaced by the specified *Sigma* on the *Dist*. tab of the *Column Properties* dialog box.

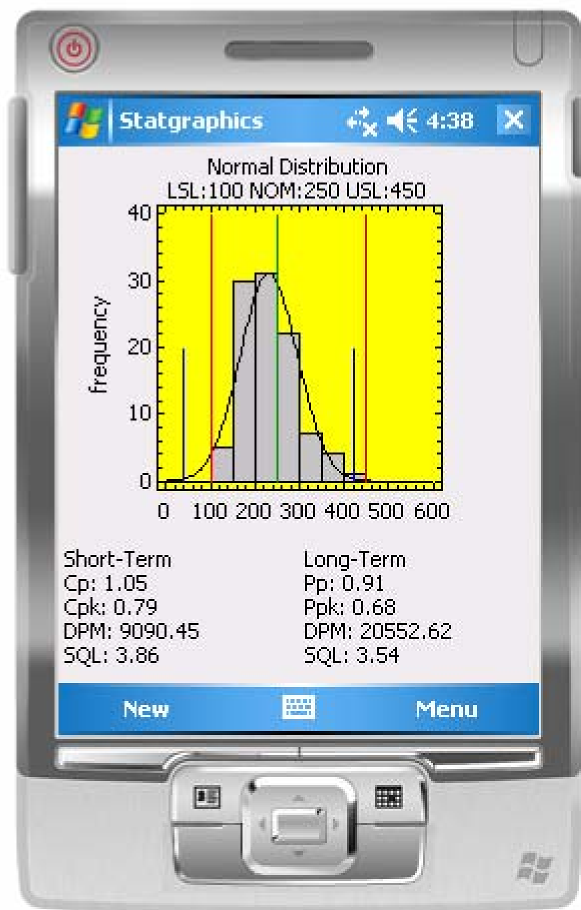
© 2006 by StatPoint, Inc.

Capability

If specifications exist for the *Response* variable, a capability analysis may be performed to determine how often items are likely to fall outside of the specification limits. As described earlier, specification limits are entered on the *Specs* tab of the *Column Properties* dialog box for the leftmost *Response* column. You may enter:

1. A lower specification limit.
2. A nominal or target value.
3. An upper specification limit.

At least one specification limit is required to conduct the analysis.



The output of the Capability Analysis includes:

1. A histogram of the data.

2. A plot of the fitted distribution. The type of distribution is specified on the *Dist*. tab of the *Column Properties* dialog box for the leftmost *Response* column. By default, a normal distribution is assumed.

3. Tall vertical lines at the specification limits and at the nominal value.

© 2006 by StatPoint, Inc.

4. Short vertical lines at percentiles of the fitted distribution that contain 99.73% of the distribution. If a normal distribution is assumed, these lines are located

$$\bar{x} \pm 3s \tag{17}$$

where $\bar{x}$ is the overall sample mean and s is the overall standard deviation.

Below the plot are estimates of selected capability indices. Capability indices summarize how well the process appears to meet the specification. Estimates are given for both short-term capability, using the estimated short-term process sigma $\hat{\sigma}$ from the control charts, and also for long-term capability using the sample standard deviation s of the entire set of data.

The indices that are displayed are selected on the *Specs* tab of the *Column Properties* dialog box for the leftmost *Response* column. The indices that may be calculated are:

- **Cp/Pp** – These two-sided capability indices compare the distance between the specification limits to 6 sigma:

$$C_P = \frac{USL - LSL}{6\hat{\sigma}} \tag{18}$$

$$P_P = \frac{USL - LSL}{6s} \tag{19}$$

  Many companies require that these indices be at least 1.33.

- **Cpk/Ppk** – These one-sided capability indices are based on the distance from the sample mean to the nearer specification limit:

$$C_{pk} = \min\left(\frac{\bar{x} - LSL}{3\hat{\sigma}}, \frac{USL - \hat{\mu}}{3\hat{\sigma}}\right) \tag{20}$$

$$P_{pk} = \min\left(\frac{\bar{x} - LSL}{3s}, \frac{USL - \hat{\mu}}{3s}\right) \tag{21}$$

  The one-sided indices may be considerably smaller than the two-sided indices if the sample mean is not close to the center of the specification limits.

- **K** – a measure of the distance from the nominal or target value to the estimated process mean, scaled by the distance between the specification limits:

$$K = \frac{\bar{x} - T}{USL - LSL} \tag{22}$$

- **DPM** – the estimated defects per million, based on the fitted distribution.

- **Sigma Quality Level** – an index of the level of quality for the process developed as part of the Six Sigma process. The short-term *SQL* equals

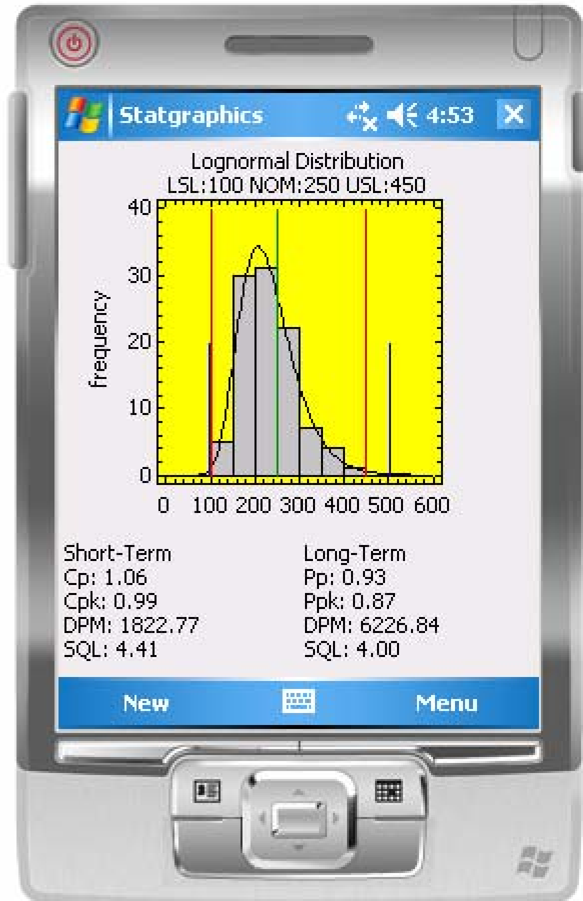
© 2006 by StatPoint, Inc.

$$SQL = 1.5 + \min\left(\frac{\bar{x} - LCL}{\hat{\sigma}}, \frac{UCL - \bar{x}}{\hat{\sigma}}\right) \quad (23)$$

The long-term SQL uses a similar equation with s replacing $\hat{\sigma}$. Achieving "Six Sigma" quality requires reducing process variation until *SQL* rises to 6.

Note: The *Six Sigma Calculator* on the *Tools* menu allows you to quickly convert between *Cpk*, *DPM*, and *SQL*.

In the case of a non-normal distribution, the fitted distribution will not be symmetric:



The vertical lines still cover 99.73% of the fitted distribution, but they are no longer located at plus and minus 3 sigma. When estimating the capability indices, both the data and the specification limits are first transformed, and the indices are then calculated. This insures that the same relationship exists between *Cpk*, *DPM*, and *SQL* as when the data follow a normal distribution.

© 2006 by StatPoint, Inc.