P Chart

Summary

The **P Chart** procedure creates a control chart for data that describes the proportion of times an event occurs in *m* samples taken from a product or process. The data might represent the proportion of defective items in a manufacturing process, the proportion of customers that return a product, or any other attribute that can be classified as acceptable or unacceptable. Out-of-control signals are highlighted, including both points beyond the control limits and any unusual runs in the data. The chart 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.

The data for this analysis consist of *m* samples from a population detailing:

n_j = number of items in sample *j*

d_j = number of defective or otherwise unacceptable items in sample *j*

Sample StatFolio: *pchart.sgp*

Sample Data:

The file *cans.sgd* contains inspection results from *m* = 94 samples taken from a manufacturing process, each corresponding *n* = 50 cardboard orange juice cans. The data is presented by Montgomery (2005). Each can was inspected and the number of defective cans was tabulated. The table below shows a partial list of the data in that file:

Sample	n	Defects
1	50	12
2	50	15
3	50	8
4	50	10
5	50	4
6	50	7
7	50	16
8	50	9
9	50	14
10	50	10
11	50	5
12	50	6
13	50	17
14	50	12
15	50	22

The first *m* = 30 samples were taken prior to a machine adjustment. The next 24 samples were then used to establish control limits for the process. The final 40 samples were plotted using the established control limits.

Data Input

The data required by this procedure is a numeric column containing the proportions of unacceptable items in the *m* samples.


- **Proportions:** a column containing the proportion of unacceptable items in each sample.

- **Sample Sizes**: the number of items in each sample. If all sample sizes are equal, you can enter a single value such as "50". Otherwise, enter a column with the sizes corresponding to each proportion.

- **Sample Labels:** optional labels for each sample.

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

- **Select:** subset selection.

Note the use of the FIRST operator to select only the first *m* = 30 rows for the initial Phase 1 analysis. Note also that the proportions of defective items are calculated "on-the-fly" by dividing the numbers of defective items by the sample sizes:

$$p_j = \frac{d_j}{n_j} \qquad (1)$$

P Chart

The *P Chart* plots the sample proportions p_j.

p Chart for defects/n
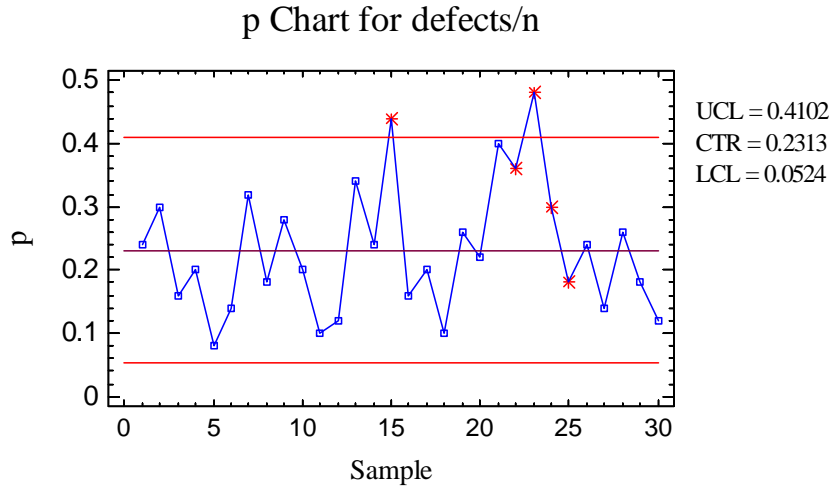
In Phase 1 (*Initial Studies*) mode, the centerline and control limits are determined from the data. The centerline is located at the weighted average of the sample proportions:

$$\overline{p} = \frac{\sum_{j=1}^{m} n_j p_j}{\sum_{j=1}^{m} n_j} \tag{2}$$

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

$$\overline{p} \pm k \sqrt{\frac{\overline{p}(1-\overline{p})}{n}} \tag{3}$$

where *k* is the sigma multiple specified on the *Control Charts* tab of the *Preferences* dialog box (*k* = 3 except in rare cases) and *n* is the sample size. If the sample sizes are not equal, then depending on *Analysis Options*, *n* is replaced by either:

(1) $\overline{n}$, the average sample size. In this case, the control limits are the same for all samples.

(2) n_j, the individual sample sizes. In this case, the control limits are step functions.

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, 2 points are beyond

the upper control limit. In addition, 3 signals were generated by runs rules violations, the first at j = 22 when the second of the last 3 samples fell more than 2 sigma above the centerline.

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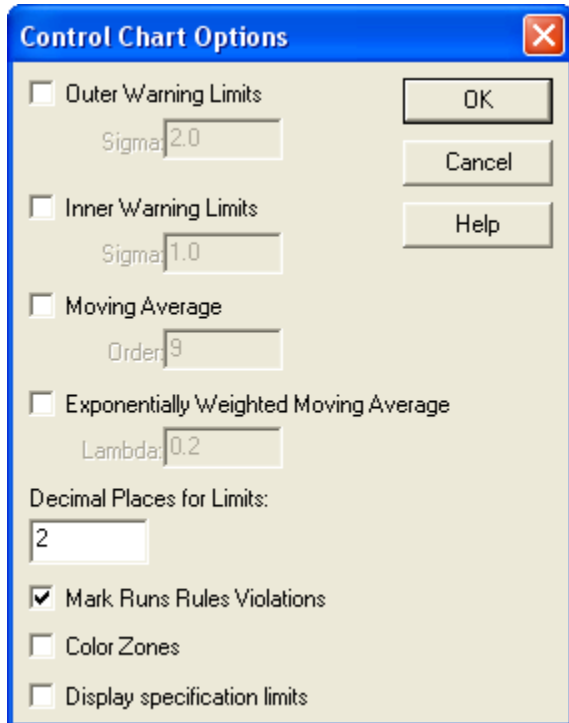
- **Outer Warning Limits**: check this box to add warning limits at the specified multiple of sigma, usually at 2 sigma.

- **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 sample proportions, 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 proportions 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 sample proportions, an exponentially weighted moving average of the sample proportions 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 proportions 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).

Example: Excluding Points

Montgomery (2005) reports that follow-up analyses of the 2 points beyond the control limits on the p-chart shown above revealed the following assignable causes for those large proportions:

       Sample #15: new material.
       Sample #23: new operator.

To determine how the control chart would change without those 2 samples, click on each with your mouse and press the *Exclude/Include* button on the analysis toolbar. The centerline and control limits will be recalculated, resulting in the following chart:

p Chart for defects/n

$UCL = 0.3893$
$CTR = 0.2150$
$LCL = 0.0407$
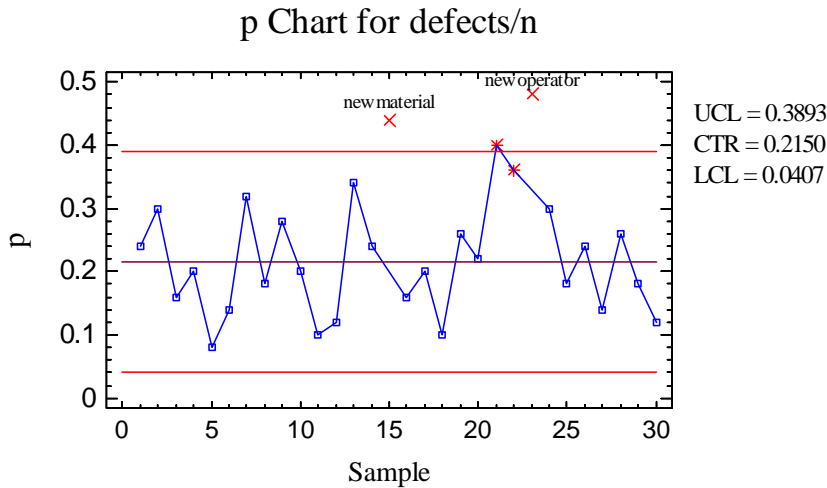
Each of the excluded points has been replaced by an X and a label added to the chart (using the *Add Text* button on the analysis toolbar).
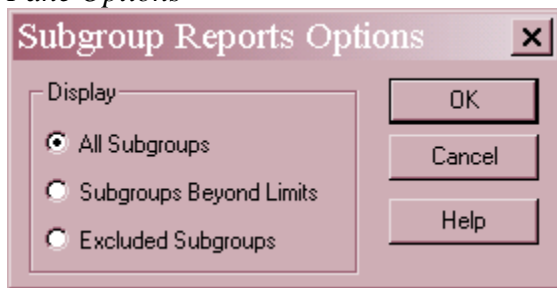
P Chart Report

This pane tabulates the values plotted on the control chart:

p Chart Report
Samples Beyond Limits
X = Excluded   * = Beyond Limits

Sample	Size	p
15	50	* 0.44
23	50	* 0.48

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

Pane Options

- **Display:** specify the samples (subgroups) to display in the report.

Analysis Summary

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

p Chart - defects/n (FIRST(30))
Selection variable: FIRST(30)
Number of samples = 30
Sample size = 50.0
0 samples excluded

p Chart

Period	#1-30
UCL: +3.0 sigma	0.410239
Centerline	0.231333
LCL: -3.0 sigma	0.0524275

2 beyond limits

Estimates

Period	#1-30
Mean p	0.231333
Sigma	0.0596353

Included in the table are:

- **Sample Information**: the number of samples *m* and the average sample size

$$\bar{n} = \frac{\sum_{j=1}^{m} n_j}{m} \tag{4}$$

If any samples have been excluded from the calculations, that number is also displayed.

- **p Chart:** a summary of the centerline and control limits for the p chart. Using *Analysis Options*, separate limits may be calculated for different periods (sets of samples).

- **Estimates:** estimates of the process mean proportion of defective items:

$$\hat{\theta} = \bar{p} \tag{5}$$

and the process sigma

$$\hat{\sigma} = \sqrt{\frac{\bar{p}(1-\bar{p})}{\bar{n}}} \tag{6}$$

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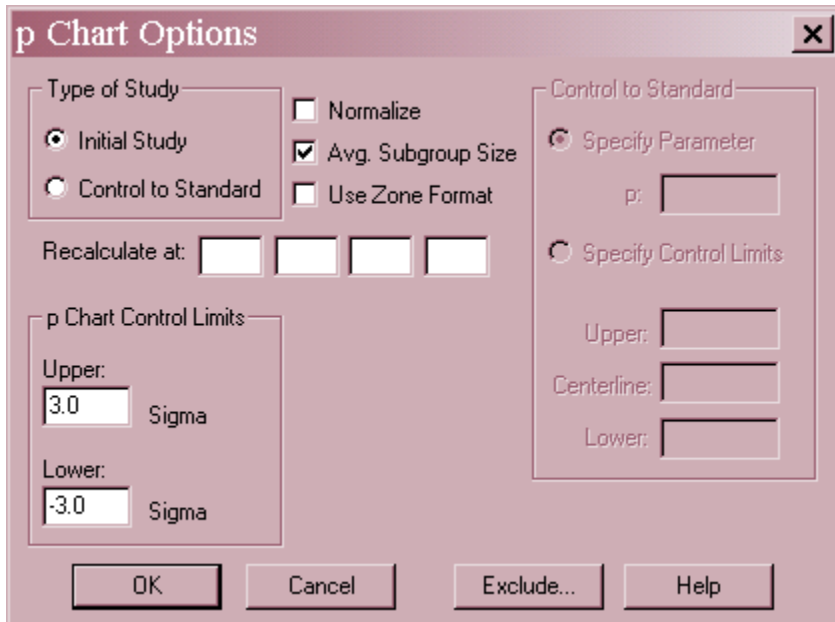
- **Type of Study:** determines how the control limits are set. For an *Initial Study* (Phase 1) chart, the limits are estimated from the current data. For a *Control to Standard* (Phase 2) chart, the control limits are determined from the information in the *Control to Standard* section of the dialog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

- **Avg. Subgroup Size:** if checked, the control limits will be horizontal lines based on the average sample size. If not checked, the individual sample sizes will be used, resulting in step function control limits if the sample sizes are not all equal.

- **Use Zone Format:** if checked, the statistics will be plotted using a zone chart rather than the usual format. See the example below.

- **Recalculate at:** the control limits of the chart may be recalculated at up to 4 locations along the X axis by specifying the sample numbers at which new estimates are to be initiated. Separate estimates of the mean proportion 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.

- **P Chart Control Limits:** specify the multiple k to use in determining the upper and lower control limits on the p 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 mean proportion of defective items.

  Specify Control Limits: specify the location of the centerline and control limits exactly where you wish them to be placed.

- **Exclude button:** Use this button to exclude specific samples from the calculations. It will display the following dialog box:
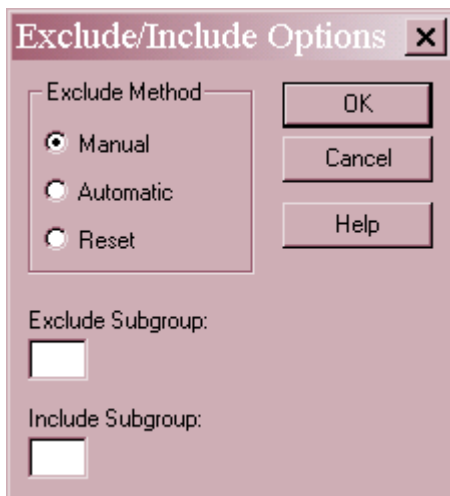
To exclude one or more values from the calculations:

  1. Select *Manual*, enter the sample (subgroup) number to exclude, and press OK. The control limits will be recalculated without that subgroup 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 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.

Example: Zone Chart

If the *Zone Format* option is selected, then the X-bar chart will be displayed in the following manner:

p Chart for defects/n
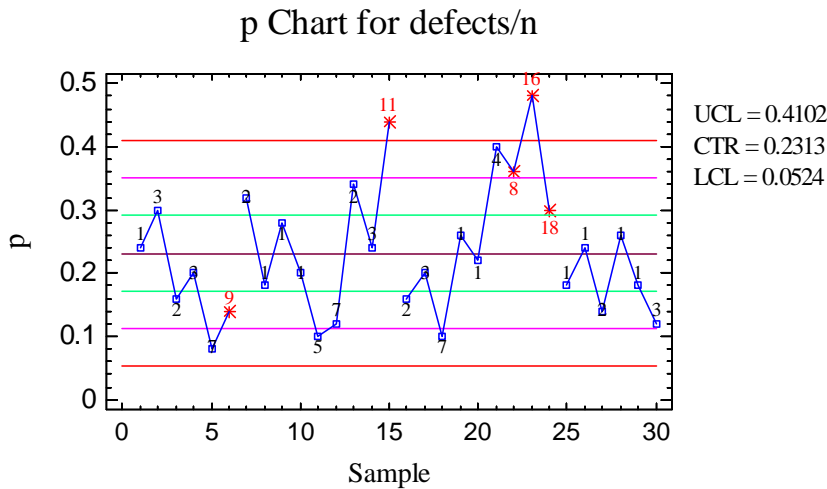
Above each observation is a running count based on the latest run of observations either above the centerline or below the centerline. Beginning at 0, a cumulative score is attached to each point on the chart in the following manner:

- o If the point is within the inner warning limits, the score is incremented by 1.
- o If the point is beyond the inner warning limits but within the outer warning limits, the score is incremented by 2.
- o If the point is beyond the outer warning limits but within the control limits, the score is incremented by 4.
- o If the point is beyond the control limits, the score is incremented by 8.

Whenever the process crosses the centerline, the score is reset to 0. Values exactly equal to the centerline do not change the score. When the score reaches a critical value (8 is the default), an out-of-control signal is generated.

The above rule is the default but may be overridden on the *Runs Tests* tab of the *Preferences* dialog box, accessible through the *Edit* menu.
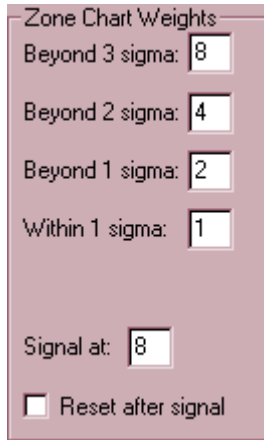
In addition, the score may be reset to 0 immediately after each out-of-control signal.

In the chart above, the maximum score reached 8 or more during 3 runs.

Example: Calculating Different Limits for Different Periods
In Montgomery's (2005) discussion of this data, he states that a machine adjustment was made after reviewing the above results and an additional 24 samples were then taken. To examine the combined 54 samples, modify the dialog box as shown below:
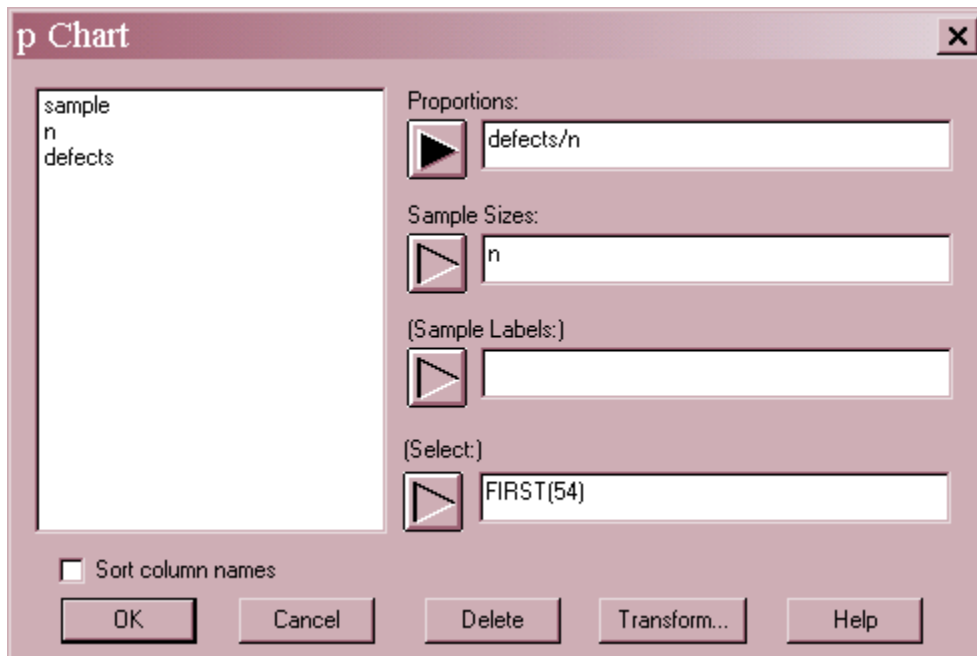
Then access *Analysis Options* and ask for new limits to be calculated beginning at sample 31:
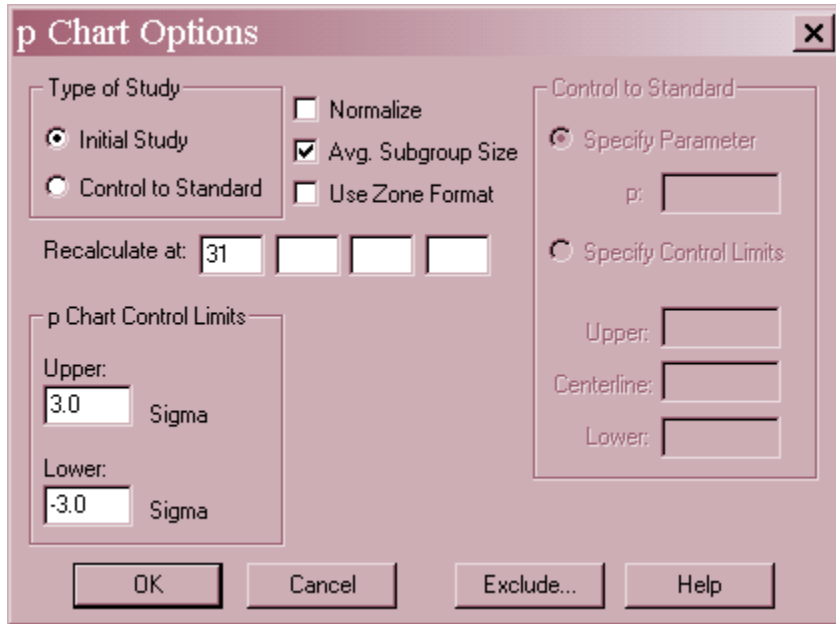
Separate control limits will then be computed for samples 1-30 and 31-54. The limits for the last section are displayed on the chart:
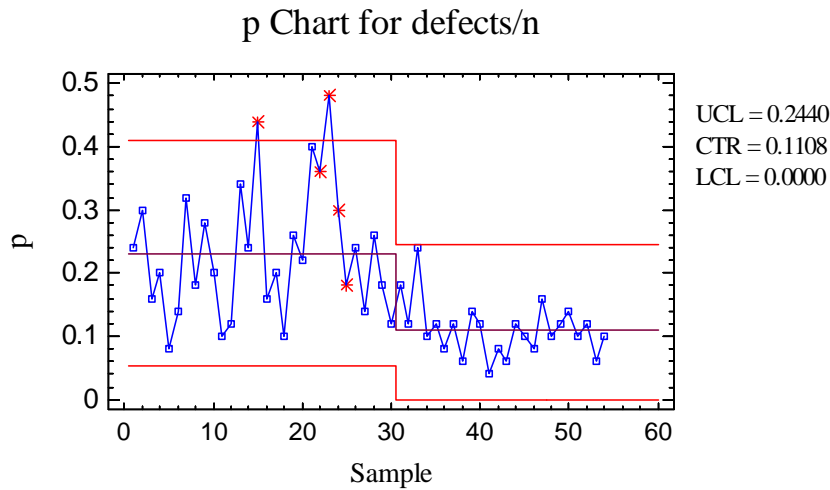
Obviously, the machine adjustment made a dramatic improvement.

Runs Tests

The *Runs Tests* pane displays the results of standard tests applied to the p chart to look for unusual sequences 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 subgroups with at least 4 beyond 1.0 sigma.
(D) sets of 3 subgroups with at least 2 beyond 2.0 sigma.
(E) sets of 15 subgroups at or within 1.0 sigma.
(F) sets of 8 subgroups beyond 2.0 sigma.
(G) sets of 8 observations alternating up and down

Violations

Observation	p Chart
20	G
21	G
22	DG
23	DG
24	CDG
25	C

Depending on the default settings on the *Runs Tests* tab of the *Preferences* dialog box, STATGRAPHICS will look for up to 7 different types of patterns:

A. A group of 8 or more points, all above or all below the centerline.

B. A group of 8 or more points, all increasing or all decreasing.

C. A group of 5 points in which at least 4 are more than 1-sigma away from the centerline, on the same side of the centerline.

D. A group of 3 points in which at least 2 are more than 2-sigma away from the centerline, on the same side of the centerline.

E. A group of 15 or more points, all within 1-sigma.

F. A group of 8 or more points, all beyond 2-sigma, but not necessarily on the same side of the centerline.

G. A group of 8 or more points, all following an alternating up and down pattern.

Any such runs will be indicated in the above table and also on the control charts (unless suppressed).

Runs tests are designed to make standard Shewhart charts more sensitive to small shifts in the process. In the current situation, rule G detects an up and down oscillation beginning at time period $j = 20$.

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.

Phase 2 Analyses

The discussion above deals with a Phase 1 study, in which the goal was to bring the process into a state of statistical control and estimate a standard value for the mean proportion of defective items. After the machine adjustment, the process appeared to stabilize around a proportion θ = 0.1108.

In a Phase 2 study, additional data is collected and plotted with control limits based on that standard. Rows 55-94 of the sample data file represent such a case.
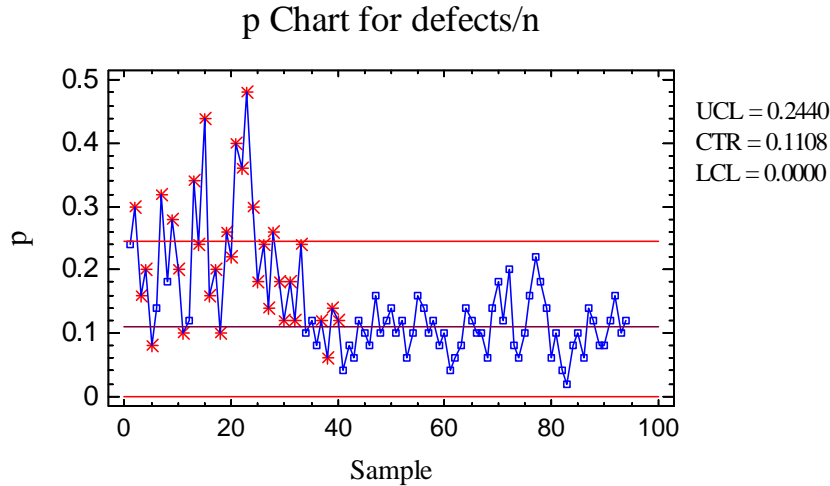
To perform a Phase 2 study on the sample data, the data input dialog box will first be modified to remove the earlier *Subset* selection:
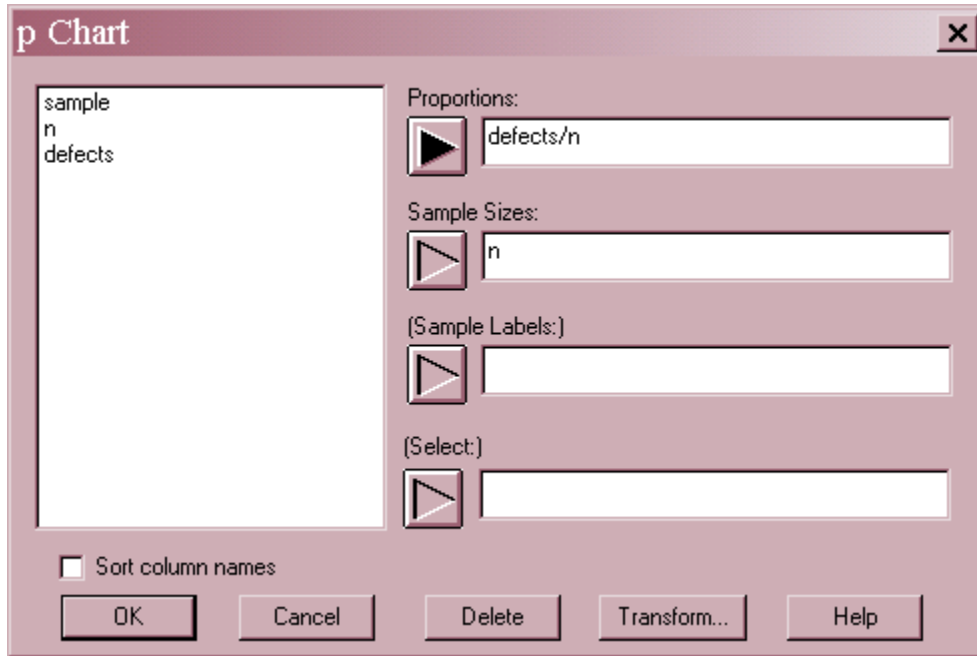
Next, the *Type of Study* will be switched to *Control to Standard* in the *Analysis Options* dialog box and the standard proportion entered:
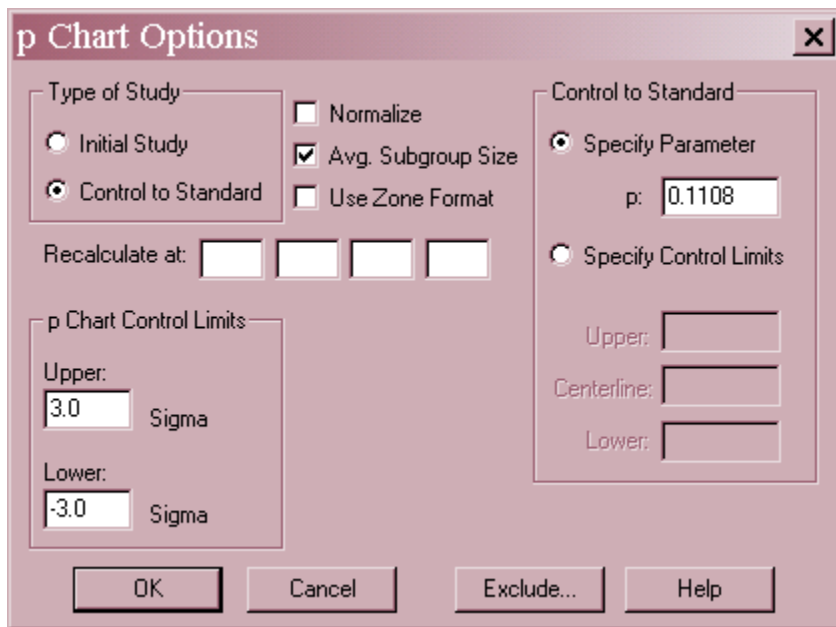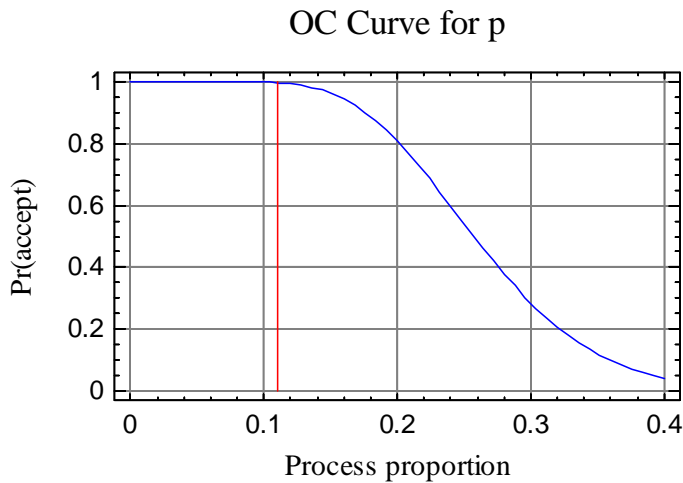

The resulting chart has the same centerline and control limits as before:

p Chart for defects/n

The process appears to have remained in control during the period in which the additional data was collected.

OC Curve

The *OC (Operating Characteristic) Curve* is designed to illustrate the properties of a Phase 2 control chart.

OC Curve for p

The chart displays the probability that a sample proportion will be within the control limits on the p chart, as a function of the true process proportion of unacceptable items. For example, if the process proportion were to shift to 0.2, the sample estimate would remain within the limits approximately 80% of the time and thus not generate an out-of-control signal. That probability drops to about 5% if the proportion of unacceptable items equals 0.4.

ARL Curve

The *ARL Curve* is another way to view the performance of a Phase 2 p chart.

ARL Curve for p
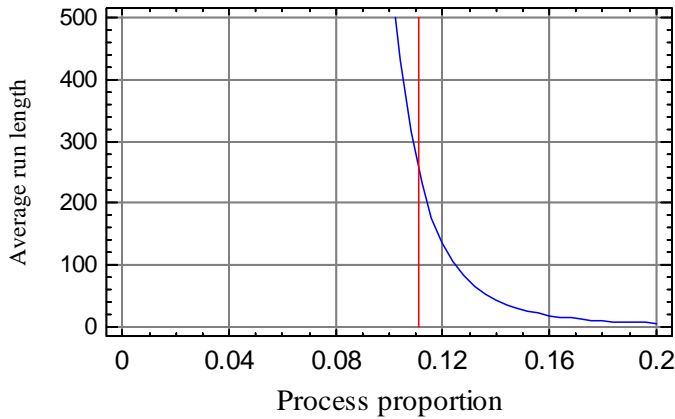
The ARL curve plots the average run length (average number of samples plotted up to and including the first point beyond the control limits) as a function of the true process proportion. Assuming that the proportion of unacceptable items suddenly shifts to a new value, the chart shows how long it takes on average until an out-of-control signal is generated (not counting any signals from run rules violations). For very small shifts, it can take in excess of 350 samples on average to detect the shift. Expanding the chart in the vicinity of 0.2 shows:

ARL Curve for p
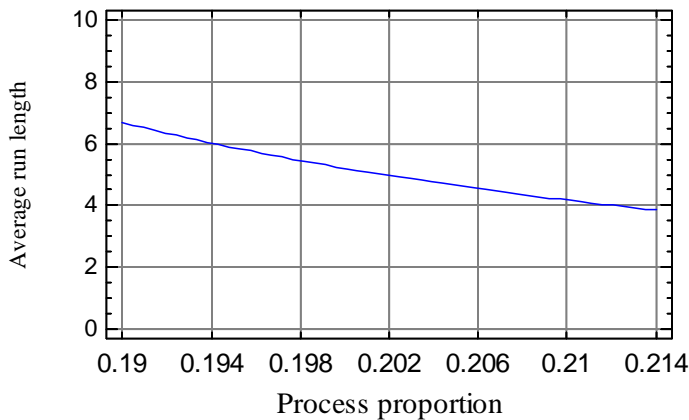
At a shift to $\theta = 0.2$, the ARL is approximately 5 samples.

The responsiveness of the p chart depends on the sample size *n*. To help determine a good sample size, see the documentation for the *Control Chart Design* procedure.

Save Results

The following results can be saved to the datasheet:

1. *Proportions* – the sample proportions.
2. *Sizes* – the sample sizes.
3. *Labels* – the sample labels.
4. *Mean p* – the estimated mean proportion of unacceptable items.
5. *Sigma* – the estimated standard error of the proportions.