

Chapter 38

PCHART Statement

Chapter Table of Contents

OVERVIEW	1303
GETTING STARTED	1304
Creating p Charts from Count Data	1304
Creating p Charts from Summary Data	1306
Saving Proportions of Nonconforming Items	1308
Saving Control Limits	1309
Reading Preestablished Control Limits	1312
SYNTAX	1314
Summary of Options	1316
DETAILS	1324
Constructing Charts for Proportion Nonconforming (p Charts)	1324
Output Data Sets	1326
ODS Tables	1329
Input Data Sets	1329
Axis Labels	1332
Missing Values	1333
EXAMPLES	1334
Example 38.1 Applying Tests for Special Causes	1334
Example 38.2 Specifying Standard Average Proportion	1336
Example 38.3 Working with Unequal Subgroup Sample Sizes	1337
Example 38.4 Creating a Chart with Revised Control Limits	1340
Example 38.5 OC Curve for Chart	1342

Chapter 38

PCHART Statement

Overview

The PCHART statement creates p charts for the proportions of nonconforming (defective) items in subgroup samples.

You can use options in the PCHART statement to

- compute control limits from the data based on a multiple of the standard error of the proportions or as probability limits
- tabulate subgroup sample sizes, proportions of nonconforming items, control limits, and other information
- save control limits in an output data set
- save subgroup sample sizes and proportions of nonconforming items in an output data set
- read preestablished control limits from a data set
- apply tests for special causes (also known as runs tests and Western Electric rules)
- specify a known (standard) proportion of nonconforming items for computing control limits
- specify the data as counts, proportions, or percentages of nonconforming items
- display distinct sets of control limits for data from successive time phases
- add block legends and symbol markers to reveal stratification in process data
- superimpose stars at points to represent related multivariate factors
- clip extreme points to make the chart more readable
- display vertical and horizontal reference lines
- control axis values and labels
- control layout and appearance of the chart

Getting Started

This section introduces the PCHART statement with simple examples that illustrate commonly used options. Complete syntax for the PCHART statement is presented in the “Syntax” section on page 1314, and advanced examples are given in the “Examples” section on page 1334.

Creating p Charts from Count Data

See SHWPCHR
in the SAS/QC
Sample Library

An electronics company manufactures circuits in batches of 500 and uses a p chart to monitor the proportion of failing circuits. Thirty batches are examined, and the failures in each batch are counted. The following statements create a SAS data set named CIRCUIITS,* which contains the failure counts:

```
data circuits;
  input batch fail @@;
datalines;
  1 5 2 6 3 11 4 6 5 4
  6 9 7 17 8 10 9 12 10 9
 11 8 12 7 13 7 14 15 15 8
 16 18 17 12 18 16 19 4 20 7
 21 17 22 12 23 8 24 7 25 15
 26 6 27 8 28 12 29 7 30 9
;
```

A partial listing of CIRCUIITS is shown in Figure 38.1.

Number of Failing Circuits	
batch	fail
1	5
2	6
3	11
4	6
5	4
.	.
.	.
.	.

Figure 38.1. The Data Set CIRCUIITS

There is a single observation for each batch. The variable BATCH identifies the subgroup sample and is referred to as the *subgroup-variable*. The variable FAIL contains the number of nonconforming items in each subgroup sample and is referred to as the *process variable* (or *process* for short).

The following statements create the p chart shown in Figure 38.2:

*This data set is also used in the “Getting Started” section of Chapter 37, “NPCHART Statement.”

```

title 'p Chart for the Proportion of Failing Circuits';
symbol v=dot;
proc shewhart data=circuits;
  pchart fail*batch / subgroupn=500;
run;

```

This example illustrates the basic form of the PCHART statement. After the keyword PCHART, you specify the *process* to analyze (in this case, FAIL), followed by an asterisk and the *subgroup-variable* (BATCH).

The input data set is specified with the DATA= option in the PROC SHEWHART statement. The SUBGROUPN= option specifies the number of items in each subgroup sample and is required with a DATA= input data set. The SUBGROUPN= option specifies one of the following:

- a constant subgroup sample size (as in this case)
- a variable in the input data set whose values provide the subgroup sample sizes (see the next example)

Options such as SUBGROUPN= are specified after the slash (/) in the PCHART statement. A complete list of options is presented in the “Syntax” section on page 1314.

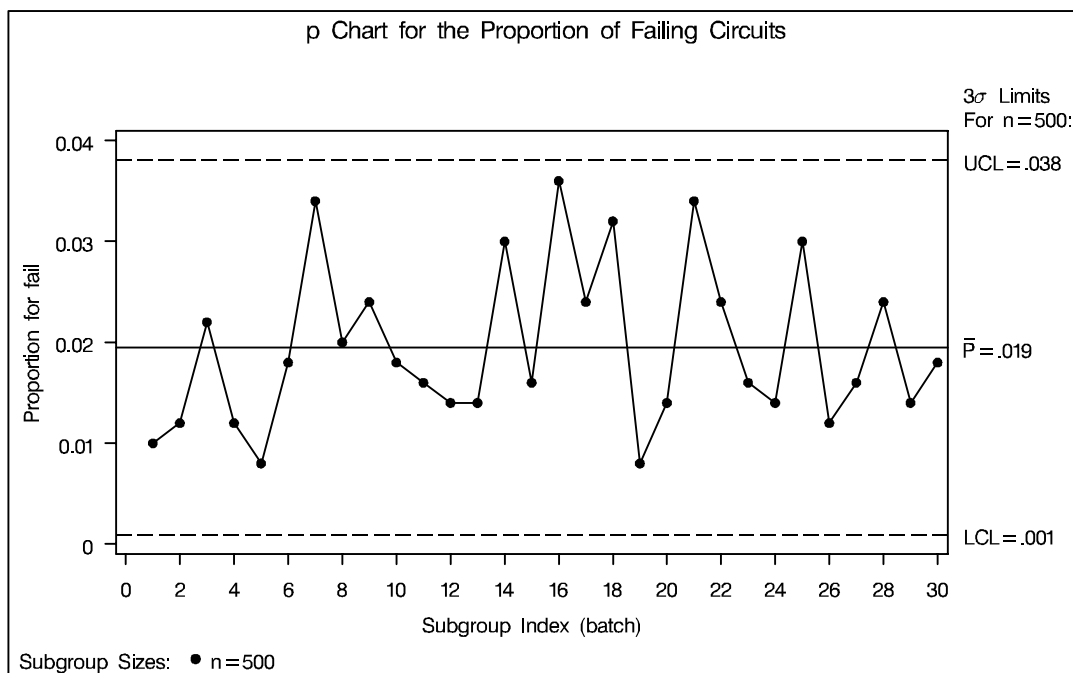


Figure 38.2. A p Chart for Circuit Failures

Each point on the p chart represents the proportion of nonconforming items for a particular subgroup. For instance, the value plotted for the first batch is $5/500 = 0.01$, as illustrated in Figure 38.3.

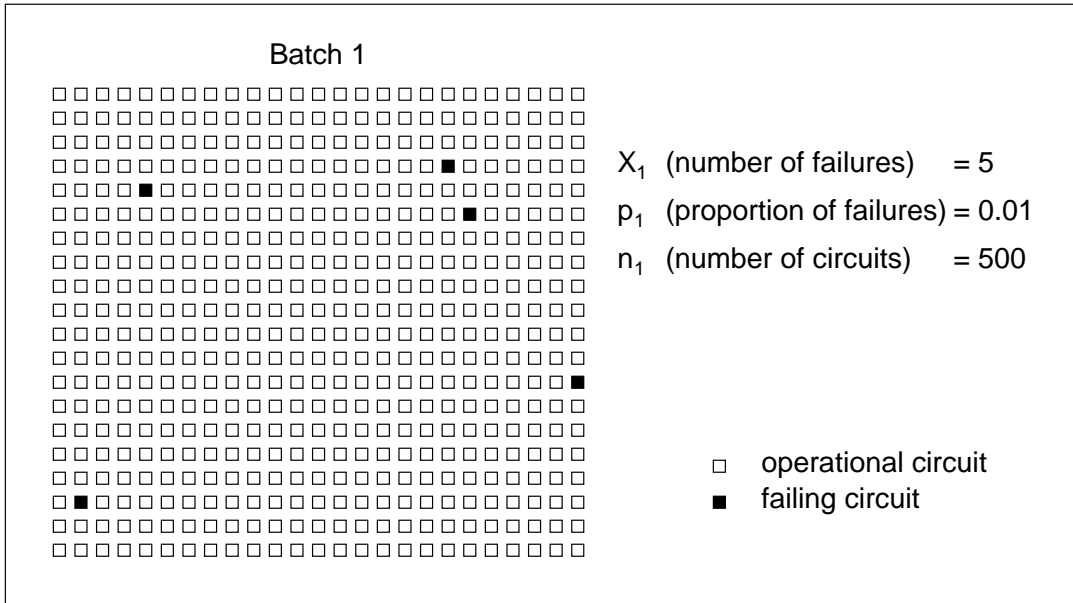


Figure 38.3. Proportions Versus Counts

Since all the points fall within the control limits, it can be concluded that the process is in statistical control.

By default, the control limits shown are 3σ limits estimated from the data; the formulas for the limits are given in “Control Limits” on page 1325. You can also read control limits from an input data set; see “Reading Preestablished Control Limits” on page 1312. For computational details, see “Constructing Charts for Proportion Nonconforming (p Charts)” on page 1324. For more details on reading counts of nonconforming items, see “DATA= Data Set” on page 1329.

Creating p Charts from Summary Data

See SHWPCHR
in the SAS/QC
Sample Library

The previous example illustrates how you can create p charts using raw data (counts of nonconforming items). However, in many applications, the data are provided in summarized form as proportions or percentages of nonconforming items. This example illustrates how you can use the PCHART statement with data of this type.

The following data set provides the data from the preceding example in summarized form:

```

data cirprop;
  input batch pfailed @@;
  sample=500;
datalines;
  1  0.010  2  0.012  3  0.022  4  0.012  5  0.008
  6  0.018  7  0.034  8  0.020  9  0.024 10  0.018
 11  0.016 12  0.014 13  0.014 14  0.030 15  0.016
 16  0.036 17  0.024 18  0.032 19  0.008 20  0.014
 21  0.034 22  0.024 23  0.016 24  0.014 25  0.030
 26  0.012 27  0.016 28  0.024 29  0.014 30  0.018
;
    
```

A partial listing of CIRPROP is shown in Figure 38.4. The subgroups are still indexed by BATCH. The variable PFAILED contains the proportions of nonconforming items, and the variable SAMPSIZE contains the subgroup sample sizes.

Subgroup Proportions of Nonconforming Items		
batch	pfailed	sampsize
1	0.010	500
2	0.012	500
3	0.022	500
4	0.012	500
5	0.008	500
.	.	.
.	.	.
.	.	.

Figure 38.4. The Data Set CIRPROP

The following statements create a *p* chart identical to the one in Figure 38.2:

```

title 'p Chart for the Proportion of Failing Circuits';
symbol v=dot;
proc shewhart data=cirprop;
    pchart pfailed*batch / subgroupn=sampsize
                dataunit =proportion;
label pfailed = 'Proportion for FAIL';
run;

```

The DATAUNIT= option specifies that the values of the *process* (PFAILED) are proportions of nonconforming items. By default, the values of the *process* are assumed to be counts of nonconforming items (see the previous example).

Alternatively, you can read the data set CIRPROP by specifying it as a HISTORY= data set in the PROC SHEWHART statement. A HISTORY= data set used with the PCHART statement must contain the following variables:

- subgroup variable
- subgroup proportion of nonconforming items variable
- subgroup sample size variable

Furthermore, the names of the subgroup proportion and sample size variables must begin with the *process* name specified in the PCHART statement and end with the special suffix characters *P* and *N*, respectively.

To specify CIRPROP as a HISTORY= data set and FAIL as the *process*, you must rename the variables PFAILED and SAMPSIZE to FAILP and FAILN, respectively. The following statements temporarily rename PFAILED and SAMPSIZE for the duration of the procedure step:

```

title 'p Chart for the Proportion of Failing Circuits';
proc shewhart history=cirprop lineprinter (rename=(pfailed =failp
                                                    sampsize=failn ));
    pchart fail*batch='*';
run;

```

The resulting p chart is shown in Figure 38.5. Since the LINEPRINTER option is specified in the PROC SHEWHART statement, line printer output is produced. * The asterisk specified in single quotes after the *subgroup-variable* indicates the character used to plot points. This character must follow an equal sign.

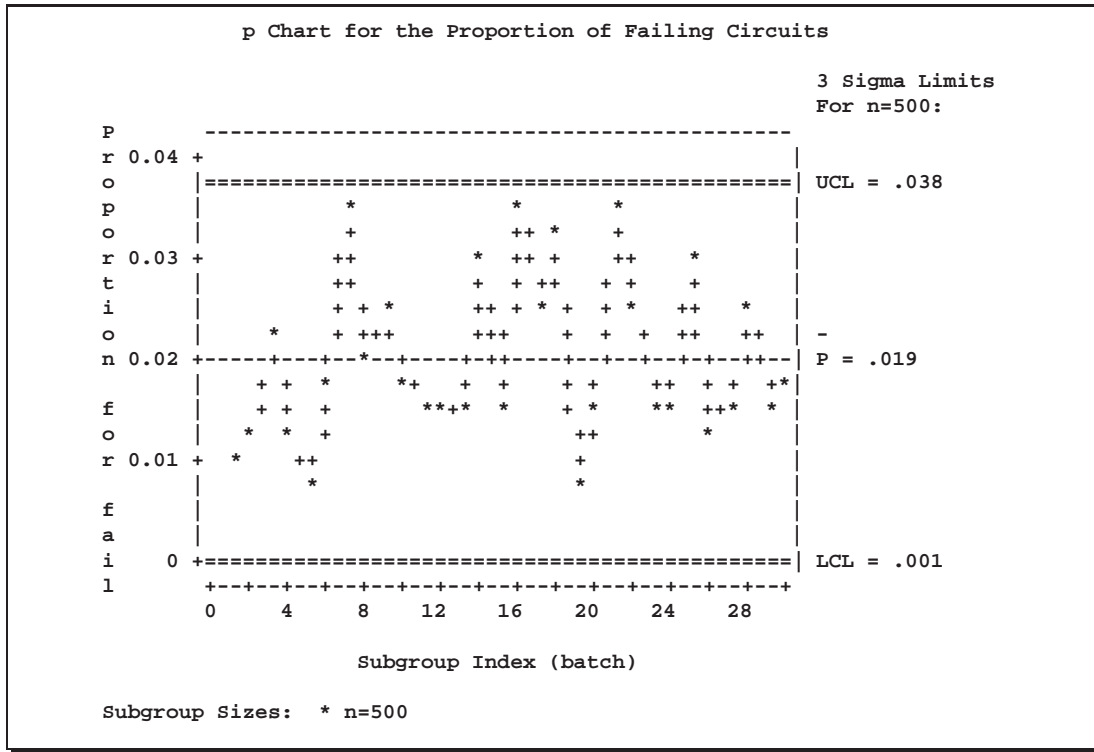


Figure 38.5. A p Chart from Subgroup Proportions
In this example, it is more convenient to use CIRPROP as a DATA= data set than as a HISTORY= data set. In general, it is more convenient to use the HISTORY= option for input data sets that have been previously created by the SHEWHART procedure as OUTHISTORY= data sets, as illustrated in the next example. For more information, see “HISTORY= Data Set” on page 1330.

Saving Proportions of Nonconforming Items

See SHWPCHR
in the SAS/QC
Sample Library

In this example, the PCHART statement is used to create a data set that can later be read by the SHEWHART procedure (as in the preceding example). The following statements read the number of nonconforming items from the data set CIRCUITS (see page 1304) and create a summary data set named CIRHIST:

```

title 'Subgroup Proportions of Failing Circuits';
proc shewhart data=circuits;
    pchart fail*batch / subgroupn =500
                    outhistory=cirhist
                    nochart;
run;
    
```

*In Release 6.12 and previous releases of SAS/QC software, the keyword GRAPHICS was required in the PROC SHEWHART statement to specify that the chart be created with a graphics device. In Version 7, you can specify the LINEPRINTER option to request line printer plots.

The OUTHISTORY= option names the output data set, and the NOCHART option suppresses the display of the chart, which would be identical to the chart in Figure 38.2. Figure 38.6 contains a partial listing of CIRHIST.

Subgroup Proportions and Control Limit Information		
batch	failP	failN
1	0.010	500
2	0.012	500
3	0.022	500
4	0.012	500
.	.	.
.	.	.
.	.	.

Figure 38.6. The Data Set CIRHIST

There are three variables in the data set CIRHIST.

- BATCH contains the subgroup index.
- FAILP contains the subgroup proportion of nonconforming items.
- FAILN contains the subgroup sample size.

Note that the variables containing the subgroup proportions of nonconforming items and subgroup sample sizes are named by adding the suffix characters *P* and *N* to the *process* FAIL specified in the PCHART statement. In other words, the variable naming convention for OUTHISTORY= data sets is the same as that for HISTORY= data sets. For more information, see “OUTHISTORY= Data Set” on page 1327.

Saving Control Limits

You can save the control limits for a *p* chart in a SAS data set; this enables you to apply the control limits to future data (see “Reading Preestablished Control Limits” on page 1312) or modify the limits with a DATA step program.

See SHWPCHR
in the SAS/QC
Sample Library

The following statements read the number of nonconforming items per subgroup from the data set CIRCUITS (see page 1304) and save the control limits displayed in Figure 38.2 in a data set named CIRLIM:

```

title 'Control Limits for the Proportion of Failing Circuits';
proc shewhart data=circuits;
  pchart fail*batch / subgroupn=500
                    outlimits=cirlim
                    nochart;
run;

```

The OUTLIMITS= option names the data set containing the control limits, and the NOCHART option suppresses the display of the chart. The data set CIRLIM is listed in Figure 38.7.

Control Limits for the Proportion of Failing Circuits									
	—		—		—				—
	S		L		S				
	U		I		A I				
—	B	T	M		L G		L		U
V	G	Y	I		P M		C		C
A	R	P	T		H A		L		L
R	P	E	N		A S		P	P	P
—	—	—	—		—		—	—	—
fail	batch	ESTIMATE	500	.005040334	3	.000930786	0.019467	0.038003	

Figure 38.7. The Data Set CIRLIM Containing Control Limit Information

The data set CIRLIM contains one observation with the limits for *process* FAIL. The variables `_LCLP_` and `_UCLP_` contain the lower and upper control limits, and the variable `_P_` contains the central line. The value of `_LIMITN_` is the nominal sample size associated with the control limits, and the value of `_SIGMAS_` is the multiple of σ associated with the control limits. The variables `_VAR_` and `_SUBGRP_` are bookkeeping variables that save the *process* and *subgroup-variable*. The variable `_TYPE_` is a bookkeeping variable that indicates whether the value of `_P_` is an estimate or standard value.

For more information, see “OUTLIMITS= Data Set” on page 1326.

You can create an output data set containing both control limits and summary statistics with the `OUTTABLE=` option, as illustrated by the following statements:

```

title 'Proportion Nonconforming and Control Limit Information';
proc shewhart data=circuits;
  pchart fail*batch / subgroupn=500
                    outtable=cirtable
                    nochart;
run;

```

The data set CIRTABLE is listed in Figure 38.8.

Subgroup Proportions and Control Limit Information									
VAR	batch	_SIGMAS_	_LIMITN_	_SUBN_	_LCLP_	_SUBP_	_P_	_UCLP_	_EXLIM_
fail	1	3	500	500	.000930786	0.010	0.019467	0.038003	
fail	2	3	500	500	.000930786	0.012	0.019467	0.038003	
fail	3	3	500	500	.000930786	0.022	0.019467	0.038003	
fail	4	3	500	500	.000930786	0.012	0.019467	0.038003	
fail	5	3	500	500	.000930786	0.008	0.019467	0.038003	
fail	6	3	500	500	.000930786	0.018	0.019467	0.038003	
fail	7	3	500	500	.000930786	0.034	0.019467	0.038003	
fail	8	3	500	500	.000930786	0.020	0.019467	0.038003	
fail	9	3	500	500	.000930786	0.024	0.019467	0.038003	
fail	10	3	500	500	.000930786	0.018	0.019467	0.038003	
fail	11	3	500	500	.000930786	0.016	0.019467	0.038003	
fail	12	3	500	500	.000930786	0.014	0.019467	0.038003	
fail	13	3	500	500	.000930786	0.014	0.019467	0.038003	
fail	14	3	500	500	.000930786	0.030	0.019467	0.038003	
fail	15	3	500	500	.000930786	0.016	0.019467	0.038003	
fail	16	3	500	500	.000930786	0.036	0.019467	0.038003	
fail	17	3	500	500	.000930786	0.024	0.019467	0.038003	
fail	18	3	500	500	.000930786	0.032	0.019467	0.038003	
fail	19	3	500	500	.000930786	0.008	0.019467	0.038003	
fail	20	3	500	500	.000930786	0.014	0.019467	0.038003	
fail	21	3	500	500	.000930786	0.034	0.019467	0.038003	
fail	22	3	500	500	.000930786	0.024	0.019467	0.038003	
fail	23	3	500	500	.000930786	0.016	0.019467	0.038003	
fail	24	3	500	500	.000930786	0.014	0.019467	0.038003	
fail	25	3	500	500	.000930786	0.030	0.019467	0.038003	
fail	26	3	500	500	.000930786	0.012	0.019467	0.038003	
fail	27	3	500	500	.000930786	0.016	0.019467	0.038003	
fail	28	3	500	500	.000930786	0.024	0.019467	0.038003	
fail	29	3	500	500	.000930786	0.014	0.019467	0.038003	
fail	30	3	500	500	.000930786	0.018	0.019467	0.038003	

Figure 38.8. The Data Set CIRTABLE

This data set contains one observation for each subgroup sample. The variables `_SUBP_` and `_SUBN_` contain the subgroup proportions of nonconforming items and subgroup sample sizes. The variables `_LCLP_` and `_UCLP_` contain the lower and upper control limits, and the variable `_P_` contains the central line. The variables `_VAR_` and `BATCH` contain the *process* name and values of the *subgroup-variable*, respectively. For more information, see “OUTTABLE= Data Set” on page 1328.

An OUTTABLE= data set can be read later as a TABLE= data set. For example, the following statements read the information in CIRTABLE and display a *p* chart (not shown here) identical to the chart in Figure 38.2:

```

title 'p Chart for the Proportion of Failing Circuits';
proc shewhart table=cirtable;
    pchart fail*batch;
run;

```

Because the SHEWHART procedure simply displays the information in a TABLE= data set, you can use TABLE= data sets to create specialized control charts (see Chapter 49, “Specialized Control Charts”). For more information, see “TABLE= Data Set” on page 1331.

Reading Preestablished Control Limits

See SHWPCHR in the SAS/QC Sample Library
--

In the previous example, the OUTLIMITS= data set CIRLIM saved control limits computed from the data in CIRCUITS. This example shows how these limits can be applied to new data provided in the following data set:

```
data circuit2;
  input batch fail @@;
datalines;
31 12 32 9 33 16 34 9
35 3 36 8 37 20 38 4
39 8 40 6 41 12 42 16
43 9 44 2 45 10 46 8
47 14 48 10 49 11 50 9
;
```

The following statements create a p chart for the data in CIRCUIT2 using the control limits in CIRLIM:

```
title 'p Chart for the Proportion of Failing Circuits';
symbol v=dot;
proc shewhart data=circuit2 limits=cirlim;
  pchart fail*batch / subgroupn=500;
run;
```

The LIMITS= option in the PROC SHEWHART statement specifies the data set containing the control limits. By default,* this information is read from the first observation in the LIMITS= data set for which

- the value of `_VAR_` matches the *process* name FAIL
- the value of `_SUBGRP_` matches the *subgroup-variable* name BATCH

The resulting p chart is shown in Figure 38.9.

*In Release 6.09 and in earlier releases, it is also necessary to specify the READLIMITS option to read control limits from a LIMITS= data set.

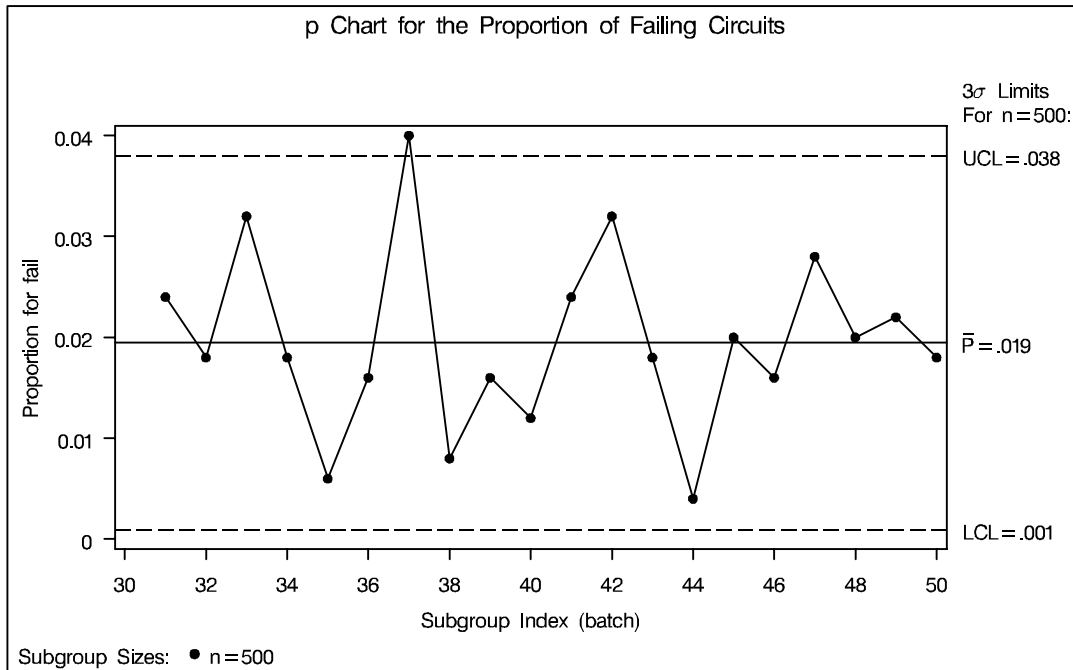


Figure 38.9. A p Chart for Second Set of Circuit Failures

The proportion of nonconforming items in the 37th batch exceeds the upper control limit, signaling that the process is out of control.

In this example, the LIMITS= data set was created in a previous run of the SHEWHART procedure. You can also create a LIMITS= data set with the DATA step. See “LIMITS= Data Set” on page 1330 for details concerning the variables that you must provide.

Syntax

The basic syntax for the PCHART statement is as follows:

```
PCHART process*subgroup-variable ;
```

The general form of this syntax is as follows:

```
PCHART (processes)*subgroup-variable <(block-variables ) >  
      < =symbol-variable | ='character' > < / options >;
```

You can use any number of PCHART statements in the SHEWHART procedure. The components of the PCHART statement are described as follows.

process

processes

identify one or more processes to be analyzed. The specification of *process* depends on the input data set specified in the PROC SHEWHART statement.

- If numbers of nonconforming items are read from a DATA= data set, *process* must be the name of the variable containing the numbers. For an example, see “Creating p Charts from Summary Data” on page 1306.
- If proportions of nonconforming items are read from a HISTORY= data set, *process* must be the common prefix of the summary variables in the HISTORY= data set. For an example, see “Creating p Charts from Summary Data” on page 1306.
- If proportions of nonconforming items and control limits are read from a TABLE= data set, *process* must be the value of the variable _VAR_ in the TABLE= data set. For an example, see “Saving Control Limits” on page 1309.

A *process* is required. If you specify more than one process, enclose the list in parentheses. For example, the following statements request distinct *p* charts for REJECTS and REWORKS:

```
proc shewhart data=measures;  
  pchart (rejects reworks)*sample / subgroupn=100;  
run;
```

Note that when data are read from a DATA= data set, the SUBGROUPN= option, which specifies subgroup sample sizes, is required.

subgroup-variable

is the variable that identifies subgroups in the data. The *subgroup-variable* is required. In the preceding PCHART statement, SAMPLE is the subgroup variable. For details, see “Subgroup Variables” on page 1534.

block-variables

are optional variables that group the data into blocks of consecutive subgroups. The blocks are labeled in a legend, and each *block-variable* provides one level of labels in the legend. See “Displaying Stratification in Blocks of Observations” on page 1684 for an example.

symbol-variable

is an optional variable whose levels (unique values) determine the symbol marker or character used to plot proportions of nonconforming items.

- If you produce a chart on a line printer, an ‘A’ is displayed for the points corresponding to the first level of the *symbol-variable*, a ‘B’ is displayed for the points corresponding to the second level, and so on.
- If you produce a chart on a graphics device, distinct symbol markers are displayed for points corresponding to the various levels of the *symbol-variable*. You can specify the symbol markers with SYMBOL n statements. See “Displaying Stratification in Levels of a Classification Variable” on page 1683 for an example.

character

specifies a plotting character for charts produced on line printers. For example, the following statements create a p chart using an asterisk (*) to plot the points:

```
proc shewhart data=values;
  pchart rejects*day='*' / subgroupn=100;
run;
```

options

enhance the appearance of the chart, request additional analyses, save results in data sets, and so on. The “Summary of Options” section, which follows, lists all options by function. Chapter 46, “Dictionary of Options,” describes each option in detail.

Summary of Options

The following tables list the PCHART statement options by function. For complete descriptions, see Chapter 46, “Dictionary of Options.”

Table 38.1. Tabulation Options

TABLE	creates a basic table of subgroup sample sizes, subgroup proportions of nonconforming items, and control limits
TABLEALL	is equivalent to the options TABLE, TABLECENTRAL, TABLEID, TABLELEGEND, TABLEOUTLIM, and TABLETESTS
TABLECENTRAL	augments basic table with values of central lines
TABLEID	augments basic table with columns for ID variables
TABLELEGEND	augments basic table with legend for tests for special causes
TABLEOUTLIM	augments basic table with columns indicating control limits exceeded
TABLETESTS	augments basic table with a column indicating which tests for special causes are positive

Note that specifying (EXCEPTIONS) after a tabulation option creates a table for exceptional points only.

Table 38.2. Reference Line Options

CHREF= <i>color</i>	specifies color for lines requested by the HREF= option
CVREF= <i>color</i>	specifies color for lines requested by the VREF= option
HREF= <i>values</i> <i>SAS-data-set</i>	specifies position of reference lines perpendicular to horizontal axis
HREFCHAR= <i>'character'</i>	specifies line character for HREF= lines
HREFDATA= <i>SAS-data-set</i>	specifies position of reference lines perpendicular to horizontal
HREFLABELS= <i>'label1'...'labeln'</i>	specifies labels for HREF= lines
HREFLABPOS= <i>n</i>	specifies position of HREFLABELS= labels
LHREF= <i>linetype</i>	specifies line type for HREF= lines
LVREF= <i>linetype</i>	specifies line type for VREF= lines
NOBYREF	specifies that reference line information in a data set applies uniformly to charts created for all BY groups
VREF= <i>values</i> <i>SAS-data-set</i>	specifies position of reference lines perpendicular to vertical axis
VREFCHAR= <i>'character'</i>	specifies line character for VREF= lines
VREFLABELS= <i>'label1'...'labeln'</i>	specifies labels for VREF= lines
VREFLABPOS= <i>n</i>	specifies position of VREFLABELS= labels

Table 38.3. Options for Specifying Tests for Special Causes

NO3SIGMACHECK	allows tests to be applied with control limits other than 3σ limits
TESTS= <i>value-list</i> <i>customized-pattern-list</i>	specifies tests for special causes
TEST2RUN= <i>n</i>	specifies length of pattern for Test 2
TEST3RUN= <i>n</i>	specifies length of pattern for Test 3
TESTACROSS	applies tests across <i>phase</i> boundaries
TESTLABEL= <i>'label'</i> <i>(variable)</i> <i>keyword</i>	provides labels for points where test is positive
TESTLABEL <i>n</i> = <i>'label'</i>	specifies label for <i>n</i> th test for special causes
TESTNMETHOD= STANDARDIZE	applies tests to standardized chart statistics
TESTOVERLAP	performs tests on overlapping patterns of points
ZONELABELS	adds labels A, B, and C to zone lines
ZONES	adds lines delineating zones A, B, and C
ZONEVALPOS= <i>n</i>	specifies position of ZONEVALUES labels
ZONEVALUES	labels zone lines with their values

Table 38.4. Graphical Options for Displaying Tests for Special Causes

CTESTS= <i>color</i> <i>test-color-list</i>	specifies color for labels indicating points where test is positive
CZONES= <i>color</i>	specifies color for lines and labels delineating zones A, B, and C
LABELFONT= <i>font</i>	specifies software font for labels at points where test is positive (alias for the TESTFONT= option)
LABELHEIGHT= <i>value</i>	specifies height of labels at points where test is positive (alias for the TESTHEIGHT= option)
LTESTS= <i>linetype</i>	specifies type of line connecting points where test is positive
LZONES= <i>linetype</i>	specifies line type for lines delineating zones A, B, and C
TESTFONT= <i>font</i>	specifies software font for labels at points where test is positive
TESTHEIGHT= <i>value</i>	specifies height of labels at points where test is positive

Table 38.5. Line Printer Options for Displaying Tests for Special Causes

TESTCHAR= <i>'character'</i>	specifies character for line segments that connect any sequence of points for which a test for special causes is positive
ZONECHAR= <i>'character'</i>	specifies character for lines that delineate zones for tests for special causes

Table 38.6. Block Variable Legend Options

BLOCKLABELPOS= <i>keyword</i>	specifies position of label for <i>block-variable</i> legend
BLOCKLABTYPE= <i>n keyword</i>	specifies text size of <i>block-variable</i> legend
BLOCKPOS= <i>n</i>	specifies vertical position of <i>block-variable</i> legend
BLOCKREP	repeats identical consecutive labels in <i>block-variable</i> legend
CBLOCKLAB= <i>color</i>	specifies color for filling background in <i>block-variable</i> legend
CBLOCKVAR= <i>variable </i> <i>(variables)</i>	specifies one or more variables whose values are colors for filling background of <i>block-variable</i> legend

Table 38.7. Axis and Axis Label Options

CAXIS= <i>color</i>	specifies color for axis lines and tick marks
CFRAME= <i>color </i> <i>(color-list)</i>	specifies fill colors for frame for plot area
CTEXT= <i>color</i>	specifies color for tick mark values and axis labels
HAXIS= <i>values </i> AXIS <i>n</i>	specifies major tick mark values for horizontal axis
HEIGHT= <i>value</i>	specifies height of axis label and axis legend text
HMINOR= <i>n</i>	specifies number of minor tick marks between major tick marks on horizontal axis
HOFFSET= <i>value</i>	specifies length of offset at both ends of horizontal axis
INTSTART= <i>value</i>	specifies first major tick mark value for numeric horizontal axis
NOHLABEL	suppresses label for horizontal axis
NOTICKREP	specifies that only the first occurrence of repeated, adjacent sub-group values is to be labeled on horizontal axis
NOTRUNC	suppresses vertical axis truncation at zero applied by default
NOVANGLE	requests vertical axis labels that are strung out vertically
SKIPHLABELS= <i>n</i>	specifies thinning factor for tick mark labels on horizontal axis
TURNHLABELS	requests horizontal axis labels that are strung out vertically
VAXIS= <i>values </i> AXIS <i>n</i>	specifies major tick mark values for vertical axis
VMINOR= <i>n</i>	specifies number of minor tick marks between major tick marks on vertical axis
VOFFSET= <i>value</i>	specifies length of offset at both ends of vertical axis
VZERO	forces origin to be included in vertical axis for primary chart
VZERO2	forces origin to be included in vertical axis for secondary chart
WAXIS= <i>n</i>	specifies width of axis lines
YSCALE=PERCENT	scales vertical axis in percent units (rather than proportions)

Table 38.8. Options for Specifying Control Limits

ALPHA= <i>value</i>	requests probability limits for control charts
LIMITN= <i>n</i> VARYING	specifies either nominal sample size for fixed control limits or varying limits
NOREADLIMITS	computes control limits for each <i>process</i> from the data rather than from a LIMITS= data set (Release 6.10 and later releases)
READALPHA	reads _ALPHA_ instead of _SIGMAS_ from a LIMITS= data set
READINDEXES=ALL ' <i>label1</i> '...' <i>labeln</i> '	reads multiple sets of control limits for each <i>process</i> from a LIMITS= data set
READLIMITS	reads single set of control limits for each <i>process</i> from a LIMITS= data set (Release 6.09 and earlier releases)
SIGMAS= <i>k</i>	specifies width of control limits in terms of multiple <i>k</i> of standard error of plotted proportion of nonconforming items

Table 38.9. Options for Displaying Control Limits

CINFILL= <i>color</i>	specifies color for area inside control limits
CLIMITS= <i>color</i>	specifies color of control limits, central line, and related labels
LCLLABEL=' <i>label</i> '	specifies label for lower control limit
LIMLABSUBCHAR= ' <i>character</i> '	specifies a substitution character for labels provided as quoted strings; the character is replaced with the value of the control limit
LLIMITS= <i>linetype</i>	specifies line type for control limits
NDECIMAL= <i>n</i>	specifies number of digits to right of decimal place in default labels for control limits and central line
NOCTL	suppresses display of central line
NOLCL	suppresses display of lower control limit
NOLIMITLABEL	suppresses labels for control limits and central line
NOLIMITS	suppresses display of control limits
NOLIMITSFRAME	suppresses default frame around control limit information when multiple sets of control limits are read from a LIMITS= data set
NOLIMITSLEGEND	suppresses legend for control limits
NOLIMIT0	suppresses display of lower control limit if it is 0
NOLIMIT1	suppresses display of upper control limit if it is 1 (100%)
NOUCL	suppresses display of upper control limit
PSYMBOL=' <i>string</i> ' <i>keyword</i>	specifies label for central line
UCLLABEL=' <i>string</i> '	specifies label for upper control limit
WLIMITS= <i>n</i>	specifies width for control limits and central line

Table 38.10. Grid Options

ENDGRID	adds grid after last plotted point
GRID	adds grid to control chart
LENDGRID= <i>linetype</i>	specifies line type for grid requested with the ENDGRID option
LGRID= <i>linetype</i>	specifies line type for grid requested with the GRID option
WGRID= <i>n</i>	specifies width of grid lines

Table 38.11. Options for Plotting and Labeling Points

ALLLABEL=VALUE (<i>variable</i>)	labels every point
CCONNECT= <i>color</i>	specifies color for line segments that connect points on chart
CFRAMELAB= <i>color</i>	specifies fill color for frame around labeled points
CNEEDLES= <i>color</i>	specifies color for needles that connect points to central line
CONNECTCHAR= ' <i>character</i> '	specifies character used to form line segments that connect points on chart
COUT= <i>color</i>	specifies color for portions of line segments that connect points outside control limits
COUTFILL= <i>color</i>	specifies color for shading areas between the connected points and control limits outside the limits
NEEDLES	connects points to central line with vertical needles
NOCONNECT	suppresses line segments that connect points on chart
OUTLABEL=VALUE (<i>variable</i>)	labels points outside control limits
SYMBOLCHARS= ' <i>characters</i> '	specifies characters indicating <i>symbol-variable</i>
SYMBOLLEGEND= NONE <i>name</i>	specifies LEGEND statement for levels of <i>symbol-variable</i>
SYMBOLORDER= <i>keyword</i>	specifies order in which symbols are assigned for levels of <i>symbol-variable</i>
TURNALL TURNOUT	turns point labels so that they are strung out vertically

Table 38.12. Options for Interactive Control Charts

HTML=(<i>variable</i>)	specifies a variable whose values are URLs to be associated with subgroups
HTML_LEGEND= (<i>variable</i>)	specifies a variable whose values are URLs to be associated with symbols in the symbol legend
TESTURLS= <i>SAS-data-set</i>	associates URLs with tests for special causes
WEBOUT= <i>SAS-data-set</i>	creates an OUTTABLE= data set with additional graphics coordinate data

Table 38.13. Clipping Options

CCLIP= <i>color</i>	specifies color for plot symbol for clipped points
CLIPCHAR= <i>'character'</i>	specifies plot character for clipped points
CLIPFACTOR= <i>value</i>	determines extent to which extreme points are clipped
CLIPLEGEND= <i>'string'</i>	specifies text for clipping legend
CLIPLEGPOS= <i>keyword</i>	specifies position of clipping legend
CLIPSUBCHAR= <i>'character'</i>	specifies substitution character for CLIPLEGEND= text
CLIPSYMBOL= <i>symbol</i>	specifies plot symbol for clipped points
CLIPSYMBOLHT= <i>value</i>	specifies symbol marker height for clipped points

Table 38.14. Phase Options

CPHASEBOX= <i>color</i>	specifies color for box enclosing all plotted points for a phase
CPHASEBOX- CONNECT= <i>color</i>	specifies color for line segments connecting adjacent enclosing boxes
CPHASEBOXFILL= <i>color</i>	specifies fill color for box enclosing all plotted points for a phase
CPHASELEG= <i>color</i>	specifies text color for <i>phase</i> legend
CPHASEMEAN- CONNECT= <i>color</i>	specifies color for line segments connecting average value points within a phase
NOPHASEFRAME	suppresses default frame for <i>phase</i> legend
OUTPHASE= <i>'string'</i>	specifies value of <code>_PHASE_</code> in the <code>OUTHISTORY=</code> data set
PHASEBREAK	disconnects last point in a <i>phase</i> from first point in next <i>phase</i>
PHASELABTYPE= <i>value</i> <i>keyword</i>	specifies text size of <i>phase</i> legend
PHASELEGEND	displays <i>phase</i> labels in a legend across top of chart
PHASELIMITS	labels control limits for each phase, provided they are constant within that phase
PHASEMEANSYMBOL= <i>symbol</i>	specifies symbol marker for average of values within a phase
PHASEREF	delineates <i>phases</i> with vertical reference lines
READPHASES= ALL <i>'label1' ... 'labeln'</i>	specifies <i>phases</i> to be read from an input data set

Table 38.15. Standard Value Options

P0= <i>value</i>	specifies known (standard) value p_0 for proportion of nonconforming items p
TYPE= <i>keyword</i>	identifies whether parameters are estimates or standard values and specifies value of <code>_TYPE_</code> in the <code>OUTLIMITS=</code> data set

Table 38.16. Input Data Set Options

DATAUNIT= <i>keyword</i>	specifies that input values are proportions or percentages (rather than counts) of nonconforming items
MISSBREAK	specifies that observations with missing values are not to be processed
SUBGROUPN= <i>n</i> <i>variable</i>	specifies subgroup sample sizes as constant number <i>n</i> or as values of <i>variable</i> in a DATA= data set

Table 38.17. Output Data Set Options

OUTHISTORY= <i>SAS-data-set</i>	creates output data set containing subgroup proportions of nonconforming items and subgroup sample sizes
OUTINDEX='string'	specifies value of _INDEX_ in the OUTLIMITS= data set
OUTLIMITS= <i>SAS-data-set</i>	creates output data set containing control limits
OUTTABLE= <i>SAS-data-set</i>	creates output data set containing subgroup proportions of nonconforming items, subgroup sample sizes, and control limits

Table 38.18. Plot Layout Options

ALLN	plots proportion of nonconforming items for all subgroups
BILEVEL	creates control charts using half-screens and half-pages
EXCHART	creates control charts for a <i>process</i> only when exceptions occur
INTERVAL= <i>keyword</i>	specifies natural time interval between consecutive subgroup positions when time, date, or datetime format is associated with a numeric subgroup variable
MAXPANELS= <i>n</i>	specifies maximum number of pages or screens for chart
NMARKERS	requests special markers for points corresponding to sample sizes not equal to nominal sample size for fixed control limits
NOCHART	suppresses creation of chart
NOFRAME	suppresses frame for plot area
NOLEGEND	suppresses legend for subgroup sample sizes
NPANELPOS= <i>n</i>	specifies number of subgroup positions per panel on each chart
REPEAT	repeats last subgroup position on panel as first subgroup position of next panel
TOTPANELS= <i>n</i>	specifies number of pages or screens to be used to display chart
ZEROSTD	displays <i>p</i> chart regardless of whether $\hat{\sigma} = 0$

Table 38.19. Graphical Enhancement Options

ANNOTATE= <i>SAS-data-set</i>	specifies annotate data set that adds features to chart
DESCRIPTION='string'	specifies string that appears in the description field of the PROC GREPLAY master menu
FONT= <i>font</i>	specifies software font for labels and legends on charts
NAME='string'	specifies name that appears in the name field of the PROC GREPLAY master menu
PAGENUM='string'	specifies the form of the label used in pagination
PAGENUMPOS= <i>keyword</i>	specifies the position of the page number requested with the PAGENUM= option

Table 38.20. Star Options

CSTARCIRCLES= <i>color</i>	specifies color for circles specified by the STARCIRCLES= option
CSTARFILL= <i>color</i> <i>(variable)</i>	specifies color for filling stars
CSTAROUT= <i>color</i>	specifies outline color for stars exceeding inner or outer circles
CSTARS= <i>color</i> <i>(variable)</i>	specifies color for outlines of stars
LSTARCIRCLES= <i>linetypes</i>	specifies line types for STARCIRCLES= circles
LSTARS= <i>linetype</i> <i>(variable)</i>	specifies line types for outlines of stars requested with the STARVERTICES= option
STARBDRADIUS= <i>value</i>	specifies radius of outer bound circle for vertices of stars
STARCIRCLES= <i>value-list</i>	specifies reference circles for stars
STARINRADIUS= <i>value</i>	specifies inner radius of stars
STARLABEL= <i>keyword</i>	specifies vertices to be labeled
STARLEGEND= <i>keyword</i>	specifies style of legend for star vertices
STARLEGENDLAB= <i>'label'</i>	specifies label for STARLEGEND= legend
STAROUTRADIUS= <i>value</i>	specifies outer radius of stars
STARSPEC= <i>value</i> <i>SAS-data-set</i>	specifies method used to standardize vertex variables
STARSTART= <i>value</i>	specifies angle for first vertex
STARTYPE= <i>keyword</i>	specifies graphical style of star
STARVERTICES= <i>variable</i> <i>(variables)</i>	superimposes star at each point on chart
WSTARCIRCLES= <i>n</i>	specifies width of circles requested by the STARCIRCLES= option
WSTARS= <i>n</i>	specifies width of stars requested by the STARVERTICES= option

Details

Constructing Charts for Proportion Nonconforming (p Charts)

The following notation is used in this section:

p	expected proportion of nonconforming items produced by the process
p_i	proportion of nonconforming items in the i^{th} subgroup
X_i	number of nonconforming items in the i^{th} subgroup
n_i	number of items in the i^{th} subgroup
\bar{p}	average proportion of nonconforming items taken across subgroups: $\bar{p} = \frac{n_1 p_1 + \cdots + n_N p_N}{n_1 + \cdots + n_N} = \frac{X_1 + \cdots + X_N}{n_1 + \cdots + n_N}$
N	number of subgroups
$I_T(\alpha, \beta)$	incomplete beta function: $I_T(\alpha, \beta) = (\Gamma(\alpha + \beta) / \Gamma(\alpha)\Gamma(\beta)) \int_0^T t^{\alpha-1} (1-t)^{\beta-1} dt$ for $0 < T < 1$, $\alpha > 0$, and $\beta > 0$, where $\Gamma(\cdot)$ is the gamma function

Plotted Points

Each point on a p chart represents the observed proportion ($p_i = X_i/n_i$) of nonconforming items in a subgroup. For example, suppose the second subgroup (see Figure 38.10) contains 16 items, of which two are nonconforming. The point plotted for the second subgroup is $p_2 = 2/16 = 0.125$.

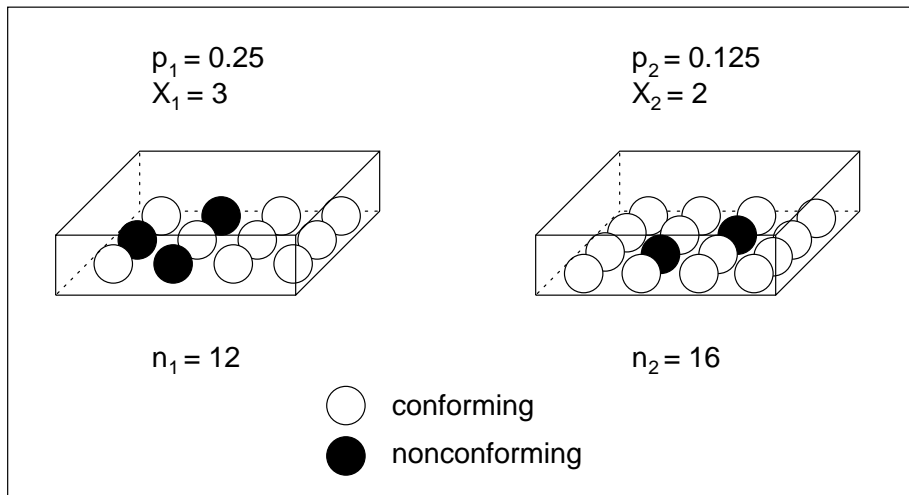


Figure 38.10. Proportions Versus Counts

Note that an np chart displays the number (count) of nonconforming items X_i . You can use the NPCHART statement to create np charts; see Chapter 37, “NPCHART Statement.”

Central Line

By default, the central line on a p chart indicates an estimate of p that is computed as \bar{p} . If you specify a known value (p_0) for p , the central line indicates the value of p_0 .

Control Limits

You can compute the limits in the following ways:

- as a specified multiple (k) of the standard error of p_i above and below the central line. The default limits are computed with $k = 3$ (these are referred to as 3σ limits).
- as probability limits defined in terms of α , a specified probability that p_i exceeds the limits

The lower and upper control limits, LCL and UCL, respectively, are computed as

$$\begin{aligned} \text{LCL} &= \max\left(\bar{p} - k\sqrt{\bar{p}(1-\bar{p})/n_i}, 0\right) \\ \text{UCL} &= \min\left(\bar{p} + k\sqrt{\bar{p}(1-\bar{p})/n_i}, 1\right) \end{aligned}$$

A lower probability limit for p_i can be determined using the fact that

$$\begin{aligned} P\{p_i < \text{LCL}\} &= 1 - P\{p_i \geq \text{LCL}\} \\ &= 1 - P\{X_i \geq n_i \text{LCL}\} \\ &= 1 - I_{\bar{p}}(n_i \text{LCL}, n_i + 1 - n_i \text{LCL}) \\ &= I_{1-\bar{p}}(n_i + 1 - n_i \text{LCL}, n_i \text{LCL}) \end{aligned}$$

Refer to Johnson, Kotz, and Kemp (1992). This assumes that the process is in statistical control and that X_i is binomially distributed. The lower probability limit LCL is then calculated by setting

$$I_{1-\bar{p}}(n_i + 1 - n_i \text{LCL}, n_i \text{LCL}) = \alpha/2$$

and solving for LCL. Similarly, the upper probability limit for p_i can be determined using the fact that

$$\begin{aligned} P\{p_i > \text{UCL}\} &= P\{p_i > \text{UCL}\} \\ &= P\{X_i > n_i \text{UCL}\} \\ &= I_{\bar{p}}(n_i \text{UCL}, n_i + 1 - n_i \text{UCL}) \end{aligned}$$

The upper probability limit UCL is then calculated by setting

$$I_{\bar{p}}(n_i \text{UCL}, n_i + 1 - n_i \text{UCL}) = \alpha/2$$

and solving for UCL. The probability limits are asymmetric around the central line. Note that both the control limits and probability limits vary with n_i .

You can specify parameters for the limits as follows:

- Specify k with the SIGMAS= option or with the variable `_SIGMAS_` in a LIMITS= data set.
- Specify α with the ALPHA= option or with the variable `_ALPHA_` in a LIMITS= data set.
- Specify a constant nominal sample size $n_i \equiv n$ for the control limits with the LIMITN= option or with the variable `_LIMITN_` in a LIMITS= data set.
- Specify p_0 with the P0= option or with the variable `_P_` in a LIMITS= data set.

Output Data Sets

OUTLIMITS= Data Set

The OUTLIMITS= data set saves control limits and control limit parameters. The following variables can be saved:

Table 38.21. OUTLIMITS= Data Set

Variable	Description
<code>_ALPHA_</code>	probability (α) of exceeding limits
<code>_INDEX_</code>	optional identifier for the control limits specified with the OUTINDEX= option
<code>_LCLP_</code>	lower control limit for proportion of nonconforming items
<code>_LIMITN_</code>	nominal sample size associated with the control limits
<code>_P_</code>	average proportion of nonconforming items (\bar{p} or p_0)
<code>_SIGMAS_</code>	multiple (k) of standard error of p_i
<code>_SUBGRP_</code>	<i>subgroup-variable</i> specified in the PCHART statement
<code>_TYPE_</code>	type (standard or estimate) of <code>_P_</code>
<code>_UCLP_</code>	upper control limit for proportion of nonconforming items
<code>_VAR_</code>	<i>process</i> specified in the PCHART statement

Notes:

1. If the control limits vary with subgroup sample size, the special missing value V is assigned to the variables `_LIMITN_`, `_LCLP_`, `_UCLP_`, and `_SIGMAS_`.
2. If the limits are defined in terms of a multiple k of the standard error of p_i , the value of `_ALPHA_` is computed as $\alpha = P\{p_i < _LCLP_\} + P\{p_i > _UCLP_\}$, using the incomplete beta function.
3. If the limits are probability limits, the value of `_SIGMAS_` is computed as $k = (_UCLP_ - _P_) / \sqrt{_P_(1 - _P_) / _LIMITN_}$. If `_LIMITN_` has the special missing value V , this value is assigned to `_SIGMAS_`.
4. Optional BY variables are saved in the OUTLIMITS= data set.

The OUTLIMITS= data set contains one observation for each *process* specified in the PCHART statement. For an example, see “Saving Control Limits” on page 1309.

OUTHISTORY= Data Set

The OUTHISTORY= data set saves subgroup summary statistics. The following variables are saved:

- the *subgroup-variable*
- a subgroup proportion of nonconforming items variable named by *process* suffixed with *P*
- a subgroup sample size variable named by *process* suffixed with *N*

Given a *process* name that contains eight characters, the procedure first shortens the name to its first four characters and its last three characters, and then it adds the suffix. For example, the procedure shortens the *process* REJECTED to REJETED before adding the suffix.

Subgroup summary variables are created for each *process* specified in the PCHART statement. For example, consider the following statements:

```
proc shewhart data=input;
  pchart (rework rejected)*batch / outhistory=summary
                                     subgroupn =30;
run;
```

The data set SUMMARY contains variables named BATCH, REWORKP, REWORKN, REJETEDP, and REJETEDN.

Additionally, the following variables, if specified, are included:

- BY variables
- *block-variables*
- *symbol-variable*
- ID variables
- `_PHASE_` (if the OUTPHASE= option is specified)

For an example of an OUTHISTORY= data set, see “Saving Proportions of Nonconforming Items” on page 1308.

Note that an OUTHISTORY= data set created with the PCHART statement can be reused as a HISTORY= data set by either the PCHART statement or the NPCHART statement.

OUTTABLE= Data Set

The OUTTABLE= data set saves subgroup summary statistics, control limits, and related information. The variables shown in the following table are saved:

Variable	Description
ALPHA	probability (α) of exceeding control limits
EXLIM	control limit exceeded on p chart
LCLP	lower control limit for proportion of nonconforming items
LIMITN	nominal sample size associated with the control limits
SIGMAS	multiple (k) of the standard error of p_i associated with the control limits
<i>subgroup</i>	values of the subgroup variable
SUBP	subgroup proportion of nonconforming items
SUBN	subgroup sample size
TESTS	tests for special causes signaled on p chart
UCLP	upper control limit for proportion of nonconforming items
VAR	<i>process</i> specified in the PCHART statement

In addition, the following variables, if specified, are included:

- BY variables
- *block-variables*
- *symbol-variable*
- ID variables
- _PHASE_ (if the READPHASES= option is specified)

Notes:

1. Either the variable _ALPHA_ or the variable _SIGMAS_ is saved depending on how the control limits are defined (with the ALPHA= or SIGMAS= options, respectively, or with the corresponding variables in a LIMITS= data set).
2. The variable _TESTS_ is saved if you specify the TESTS= option. The k^{th} character of a value of _TESTS_ is k if Test k is positive at that subgroup. For example, if you request the first four tests (the tests appropriate for p charts) and Tests 2 and 4 are positive for a given subgroup, the value of _TESTS_ has a 2 for the second character, a 4 for the fourth character, and blanks for the other six characters.
3. The variables _VAR_, _EXLIM_, and _TESTS_ are character variables of length 8. The variable _PHASE_ is a character variable of length 16. All other variables are numeric.

For an example, see “Saving Control Limits” on page 1309.

ODS Tables

The following table summarizes the ODS tables that you can request with the PCHART statement.

Table 38.22. ODS Tables Produced with the PCHART Statement

Table Name	Description	Options
PCHART	<i>p</i> chart summary statistics	TABLE, TABLEALL, TABLEC, TABLEID, TABLELEG, TABLEOUT, TABLETESTS
Tests	descriptions of tests for special causes requested with the TESTS= option for which at least one positive signal is found	TABLEALL, TABLELEG

Input Data Sets

DATA= Data Set

You can read raw data (counts of nonconforming items) from a DATA= data set specified in the PROC SHEWHART statement. Each *process* specified in the PCHART statement must be a SAS variable in the DATA= data set. This variable provides counts for subgroup samples indexed by the values of the *subgroup-variable*. The *subgroup-variable*, which is specified in the PCHART statement, must also be a SAS variable in the DATA= data set. Each observation in a DATA= data set must contain a count for each *process* and a value for the *subgroup-variable*. The data set must contain one observation for each subgroup. Note that you can specify the DATAUNIT= option in the PCHART statement to read proportions or percentages of nonconforming items instead of counts. Other variables that can be read from a DATA= data set include

- `_PHASE_` (if the READPHASES= option is specified)
- *block-variables*
- *symbol-variable*
- BY variables
- ID variables

When you use a DATA= data set with the PCHART statement, the SUBGROUPN= option (which specifies the subgroup sample size) is required. By default, the SHEWHART procedure reads all of the observations in a DATA= data set. However, if the data set includes the variable `_PHASE_`, you can read selected groups of observations (referred to as *phases*) by specifying the READPHASES= option (for an example, see “Displaying Stratification in Phases” on page 1689).

For an example of a DATA= data set, see “Creating p Charts from Count Data” on page 1304.

LIMITS= Data Set

You can read preestablished control limits (or parameters from which the control limits can be calculated) from a LIMITS= data set specified in the PROC SHEWHART statement. For example, the following statements read control limit information from the data set CONLIMS:*

```
proc shewhart data=info limits=conlims;  
  pchart rejects*batch / subgroupn= 100;  
run;
```

The LIMITS= data set can be an OUTLIMITS= data set that was created in a previous run of the SHEWHART procedure. Such data sets always contain the variables required for a LIMITS= data set. The LIMITS= data set can also be created directly using a DATA step. When you create a LIMITS= data set, you must provide one of the following:

- the variables `_LCLP_`, `_P_`, and `_UCLP_`, which specify the control limits directly
- the variable `_P_`, without providing `_LCLP_` and `_UCLP_`. The value of `_P_` is used to calculate the control limits according to the equations on page 1325.

In addition, note the following:

- The variables `_VAR_` and `_SUBGRP_` are required. These must be character variables of length 8.
- The variable `_INDEX_` is required if you specify the `READINDEX=` option; this must be a character variable of length 16.
- The variables `_LIMITN_`, `_SIGMAS_` (or `_ALPHA_`), and `_TYPE_` are optional, but they are recommended to maintain a complete set of control limit information. The variable `_TYPE_` must be a character variable of length 8; valid values are `ESTIMATE` and `STANDARD`.
- `BY` variables are required if specified with a `BY` statement.

For an example, see “Reading Preestablished Control Limits” on page 1312.

HISTORY= Data Set

You can read subgroup summary statistics from a HISTORY= data set specified in the PROC SHEWHART statement. This allows you to reuse OUTHISTORY= data sets that have been created in previous runs of the SHEWHART procedure or to create your own HISTORY= data set.

*In Release 6.09 and in earlier releases, it is necessary to specify the `READLIMITS` option.

A HISTORY= data set used with the PCHART statement must contain the following:

- the *subgroup-variable*
- a subgroup proportion of nonconforming items variable for each *process*
- a subgroup sample size variable for each *process*

The names of the proportion sample size variables must be the *process* name concatenated with the special suffix characters *P* and *N*, respectively.

For example, consider the following statements:

```
proc shewhart history=summary;
    pchart (rework rejected)*batch / subgroupn=50;
run;
```

The data set SUMMARY must include the variables BATCH, REWORKP, REWORKN, REJETEDP, and REJETEDN.

Note that if you specify a *process* name that contains eight characters, the names of the summary variables must be formed from the first four characters and the last three characters of the *process* name, suffixed with the appropriate character.

Other variables that can be read from a HISTORY= data set include

- `_PHASE_` (if the READPHASES= option is specified)
- *block-variables*
- *symbol-variable*
- BY variables
- ID variables

By default, the SHEWHART procedure reads all of the observations in a HISTORY= data set. However, if the data set includes the variable `_PHASE_`, you can read selected groups of observations (referred to as *phases*) by specifying the READPHASES= option (see “Displaying Stratification in Phases” on page 1689 for an example).

For an example of a HISTORY= data set, see “Creating p Charts from Summary Data” on page 1306.

TABLE= Data Set

You can read summary statistics and control limits from a TABLE= data set specified in the PROC SHEWHART statement. This enables you to reuse an OUTTABLE= data set created in a previous run of the SHEWHART procedure. Because the SHEWHART procedure simply displays the information read from a TABLE= data set, you can use TABLE= data sets to create specialized control charts. Examples are provided in Chapter 49, “Specialized Control Charts.”

The following table lists the variables required in a TABLE= data set used with the PCHART statement:

Table 38.23. Variables Required in a TABLE= Data Set

Variable	Description
LCLP	lower control limit for proportion of nonconforming items
LIMITN	nominal sample size associated with the control limits
P	average proportion of nonconforming items
<i>subgroup-variable</i>	values of the <i>subgroup-variable</i>
SUBN	subgroup sample size
SUBP	subgroup proportion of nonconforming items
UCLP	upper control limit for proportion of nonconforming items

Other variables that can be read from a TABLE= data set include

- *block-variables*
- *symbol-variable*
- BY variables
- ID variables
- _PHASE_ (if the READPHASES= option is specified). This variable must be a character variable of length 16.
- _TESTS_ (if the TESTS= option is specified). This variable is used to flag tests for special causes and must be a character variable of length 8.
- _VAR_. This variable is required if more than one *process* is specified or if the data set contains information for more than one *process*. This variable must be a character variable of length 8.

For an example of a TABLE= data set, see “Saving Control Limits” on page 1309.

Axis Labels

You can specify axis labels by assigning labels to particular variables in the input data set, as summarized in the following table:

Axis	Input Data Set	Variable
Horizontal	all	<i>subgroup-variable</i>
Vertical	DATA=	<i>process</i>
Vertical	HISTORY=	subgroup proportion nonconforming variable
Vertical	TABLE=	_SUBP_

For example, the following sets of statements specify the label *Proportion Nonconforming* for the vertical axis of the *p* chart:


```

proc shewhart data=circuits;
  pchart fail*batch / subgroupn=500;
  label fail = 'Proportion Nonconforming';
run;

proc shewhart history=cirhist;
  pchart fail*batch ;
  label failp = 'Proportion Nonconforming';
run;

proc shewhart table=cirtable;
  pchart fail*batch ;
  label _SUBP_ = 'Proportion Nonconforming';
run;

```

In this example, the label assignments are in effect only for the duration of the procedure step, and they temporarily override any permanent labels associated with the variables.

Missing Values

An observation read from a DATA=, HISTORY=, or TABLE= data set is not analyzed if the value of the subgroup variable is missing. For a particular process variable, an observation read from a DATA= data set is not analyzed if the value of the process variable is missing. Missing values of process variables generally lead to unequal subgroup sample sizes. For a particular process variable, an observation read from a HISTORY= or TABLE= data set is not analyzed if the values of any of the corresponding summary variables are missing.

Examples

This section provides advanced examples of the PCHART statement.

Example 38.1. Applying Tests for Special Causes

See SHWPEX1
 in the SAS/QC
 Sample Library

This example shows how you can apply tests for special causes to make p charts more sensitive to special causes of variation. The following statements create a SAS data set named CIRCUI3, which contains the number of failing circuits for 20 batches from the circuit manufacturing process introduced in “Creating p Charts from Count Data” on page 1304:

```

data circuit3;
  input batch fail @@;
datalines;
  1 12    2 21    3 16    4  9
  5  3    6  4    7  6    8  9
  9 11   10 13   11 12   12  7
 13  2   14 14   15  9   16  8
 17 14   18 10   19 11   20  9
;

```

The following statements create the p chart, apply several tests to the chart, and tabulate the results:

```

title1 'p Chart for the Proportion of Failing Circuits';
title2 'Tests=1 to 4';
symbol v=dot;
proc shewhart data=circuit3;
  pchart fail*batch / subgroupn=500
                    tests    =1 to 4
                    ltests   =20
                    zonelabels
                    tabletests
                    tablelegend;
run;

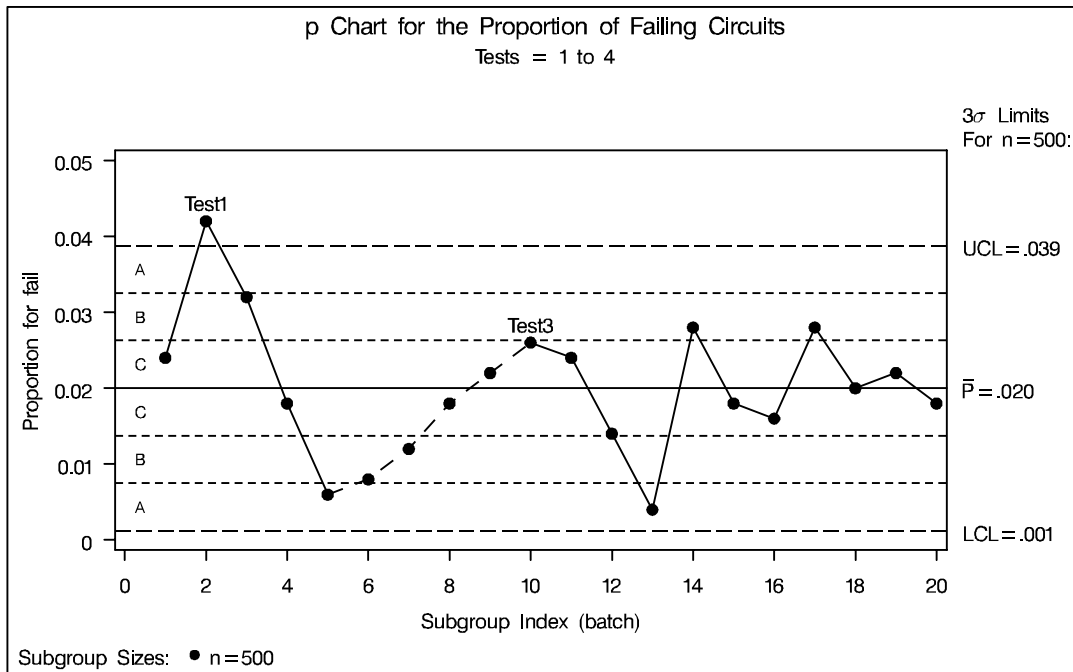
```

The chart is shown in Output 38.1.1, and the printed output is shown in Output 38.1.2. The TESTS= option requests Tests 1, 2, 3, and 4, which are described in Chapter 48, “Tests for Special Causes.” The TABLETESTS option requests a table of proportions of nonconforming items and control limits, with a column indicating which subgroups tested positive for special causes. The TABLELEGEND option adds a legend describing the tests that are positive.

The ZONELABELS option displays zone lines and zone labels on the chart. The zones are used to define the tests. The LTESTS= option specifies the line type used to connect the points in a pattern for a test that is signaled.

Output 38.1.1 and Output 38.1.2 indicate that Test 1 is positive at batch 2 and Test 3 is positive at batch 10.

Output 38.1.1. Tests for Special Causes Displayed on *p* Chart



Output 38.1.2. Tabular Form of *p* Chart

p Chart for the Proportion of Failing Circuits
Tests = 1 to 4

p Chart Summary for fail

batch	Subgroup Sample Size	-3 Sigma Limits with n=500 for Proportion-			Special Tests Signaled
		Lower Limit	Subgroup Proportion	Upper Limit	
1	500	0.00121703	0.02400000	0.03878297	
2	500	0.00121703	0.04200000	0.03878297	1
3	500	0.00121703	0.03200000	0.03878297	
4	500	0.00121703	0.01800000	0.03878297	
5	500	0.00121703	0.00600000	0.03878297	
6	500	0.00121703	0.00800000	0.03878297	
7	500	0.00121703	0.01200000	0.03878297	
8	500	0.00121703	0.01800000	0.03878297	
9	500	0.00121703	0.02200000	0.03878297	
10	500	0.00121703	0.02600000	0.03878297	3
11	500	0.00121703	0.02400000	0.03878297	
12	500	0.00121703	0.01400000	0.03878297	
13	500	0.00121703	0.00400000	0.03878297	
14	500	0.00121703	0.02800000	0.03878297	
15	500	0.00121703	0.01800000	0.03878297	
16	500	0.00121703	0.01600000	0.03878297	
17	500	0.00121703	0.02800000	0.03878297	
18	500	0.00121703	0.02000000	0.03878297	
19	500	0.00121703	0.02200000	0.03878297	
20	500	0.00121703	0.01800000	0.03878297	

Test Descriptions

Test 1 One point beyond Zone A (outside control limits)
 Test 3 Six points in a row steadily increasing or decreasing

Example 38.2. Specifying Standard Average Proportion

See SHWPEX2
in the SAS/QC
Sample Library

In some situations, a standard (known) value (p_0) is available for the expected proportion of nonconforming items, based on extensive testing or previous sampling. This example illustrates how you can specify p_0 to create a p chart.

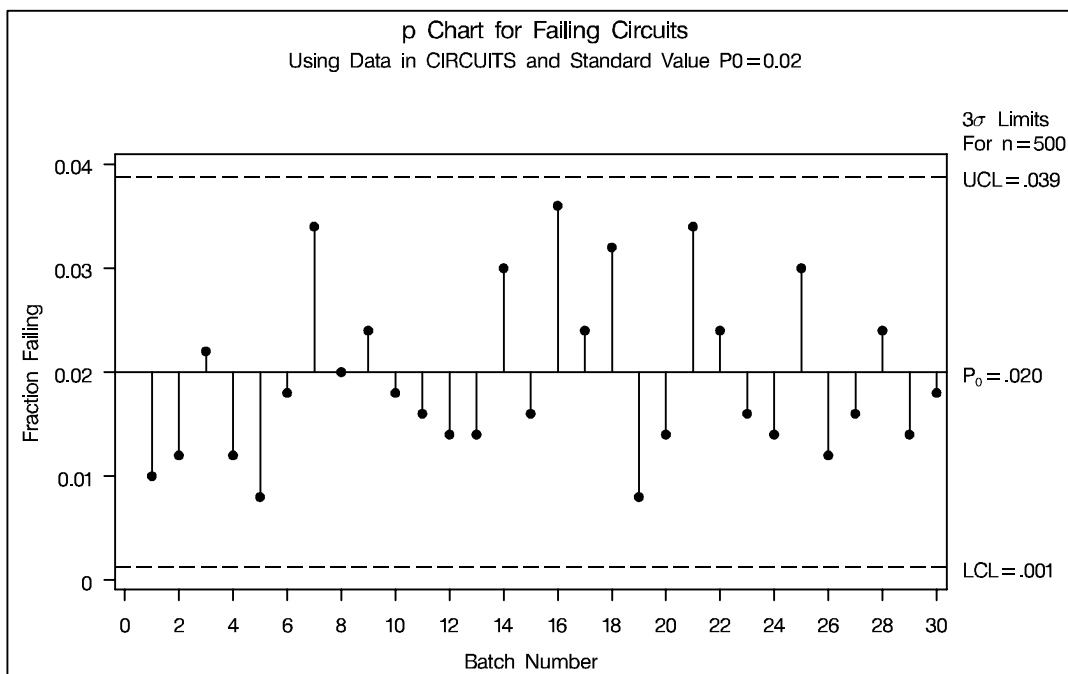
A p chart is used to monitor the proportion of failing circuits in the data set CIRCUITS, which is introduced on page 1304. The expected proportion is known to be $p_0 = 0.02$. The following statements create a p chart, shown in Output 38.2.1, using p_0 to compute the control limits:

```

title1 'p Chart for Failing Circuits';
title2 'Using Data in CIRCUITS and Standard Value P0=0.02';
symbol v=star;
proc shewhart data=circuits;
    pchart fail*batch / subgroupn = 500
                        p0         = 0.02
                        psymbol    = p0
                        nolegend
                        needles;
    label batch = 'Batch Number'
           fail  = 'Fraction Failing';
run;

```

Output 38.2.1. A p Chart with Standard Value p_0



The chart indicates that the process is in control. The P0= option specifies p_0 . The PSYMBOL= option specifies a label for the central line indicating that the line represents a standard value. The NEEDLES option connects points to the central line with vertical needles. The NOLEGEND option suppresses the default legend for subgroup sample sizes. Labels for the vertical and horizontal axes are provided with

the LABEL statement. For details concerning axis labeling, see “Axis Labels” on page 1332.

Alternatively, you can specify p_0 using the variable `_P_` in a LIMITS= data set,* as follows:

```
data climits;
  length _var_ _subgrp_ _type_ $8;
  _p_      = 0.02;
  _subgrp_ = 'batch';
  _var_    = 'fail';
  _type_   = 'STANDARD';
  _limitn_ = 500;
run;

proc shewhart data=circuits limits=climits;
  pchart fail*batch / subgroupn = 500
                    psymbol    = p0
                    nolegend
                    needles;
  label batch = 'Batch Number'
        fail  = 'Fraction Failing';
run;
```

The bookkeeping variable `_TYPE_` indicates that `_P_` has a standard value. The chart produced by these statements is identical to the chart in Output 38.2.1.

Example 38.3. Working with Unequal Subgroup Sample Sizes

The following statements create a SAS data set named BATTERY, which contains the number of alkaline batteries per lot failing an acceptance test. The number of batteries tested in each lot varies but is approximately 150.

See SHWP3EX3
in the SAS/QC
Sample Library

```
data battery;
  length lot $3;
  input lot nfailed sampsize @@;
  label nfailed = 'Number Failed'
        lot     = 'Lot Identifier'
        sampsize = 'Number Sampled';
  datalines;
AE3  6  151    AE4  5  142    AE9  6  145
BR3  9  149    BR7  3  150    BR8  0  156
BR9  4  150    DB1  9  158    DB2  4  152
DB3  0  162    DB5  9  140    DB6  7  161
DS4  6  154    DS6  1  144    DS8  5  154
JG1  3  151    MC3  8  148    MC4  2  143
MK6  4  150    MM1  4  147    MM2  0  150
RT5  2  154    RT9  8  149    SP1  3  160
SP3  9  153
;
```

*In Release 6.09 and in earlier releases, it is also necessary to specify the READLIMITS option

Part 9. The CAPABILITY Procedure

The variable NFAILED contains the number of battery failures, the variable LOT contains the lot number, and the variable SAMPSIZE contains the lot sample size. The following statements request a *p* chart for this data:

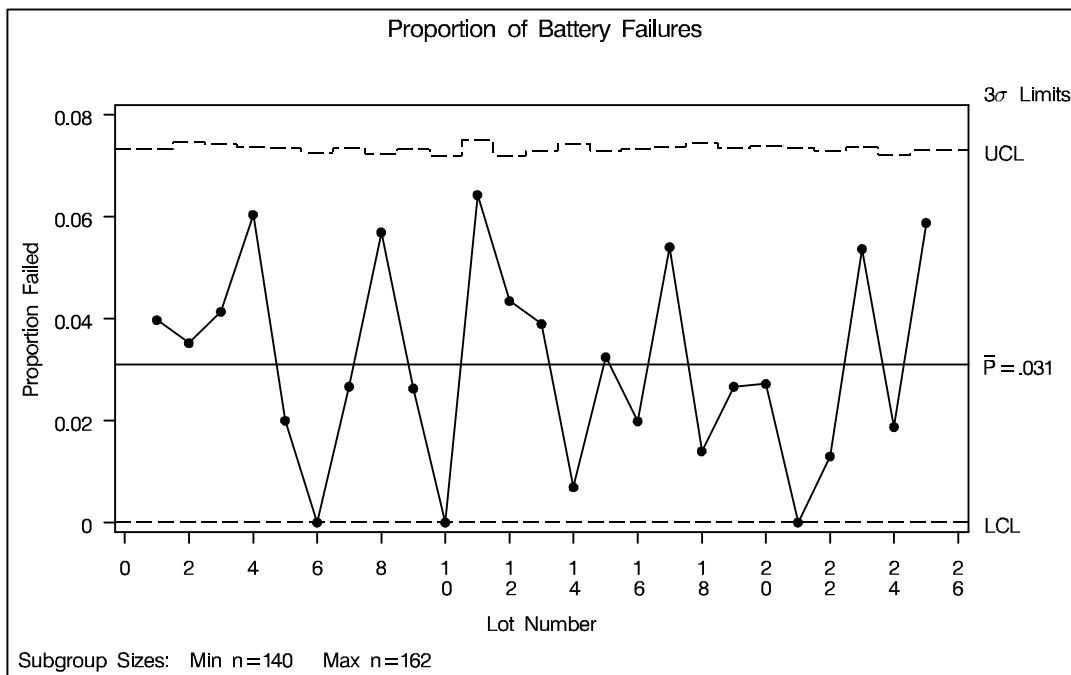
```

title 'Proportion of Battery Failures';
symbol v=dot;
proc shewhart data=battery;
  pchart nfailed*lot / subgroupn=sampsize
    turnhlabels
    outlimits=batlim;
  label nfailed='Proportion Failed';
run;

```

The chart is shown in Output 38.3.1 and the OUTLIMITS= data set BATLIM is listed in Output 38.3.2.

Output 38.3.1. A *p* Chart with Varying Subgroup Sample Sizes



Note that the upper control limit varies with the subgroup sample size. The lower control limit is truncated at zero. The sample size legend indicates the minimum and maximum subgroup sample sizes.

Output 38.3.2. Listing of the Control Limits Data Set BATLIM

Control Limits for Battery Failures								
VAR	_SUBGRP_	_TYPE_	_LIMITN_	_ALPHA_	_SIGMAS_	_LCLP_	_P_	_UCLP_
nfailed	lot	ESTIMATE	V	V	3	V	0.031010	V

The variables in BATLIM whose values vary with subgroup sample size are assigned the special missing value V.

The SHEWHART procedure provides various options for working with unequal subgroup sample sizes. For example, you can use the LIMITN= option to specify a fixed (nominal) sample size for computing the control limits, as illustrated by the following statements:

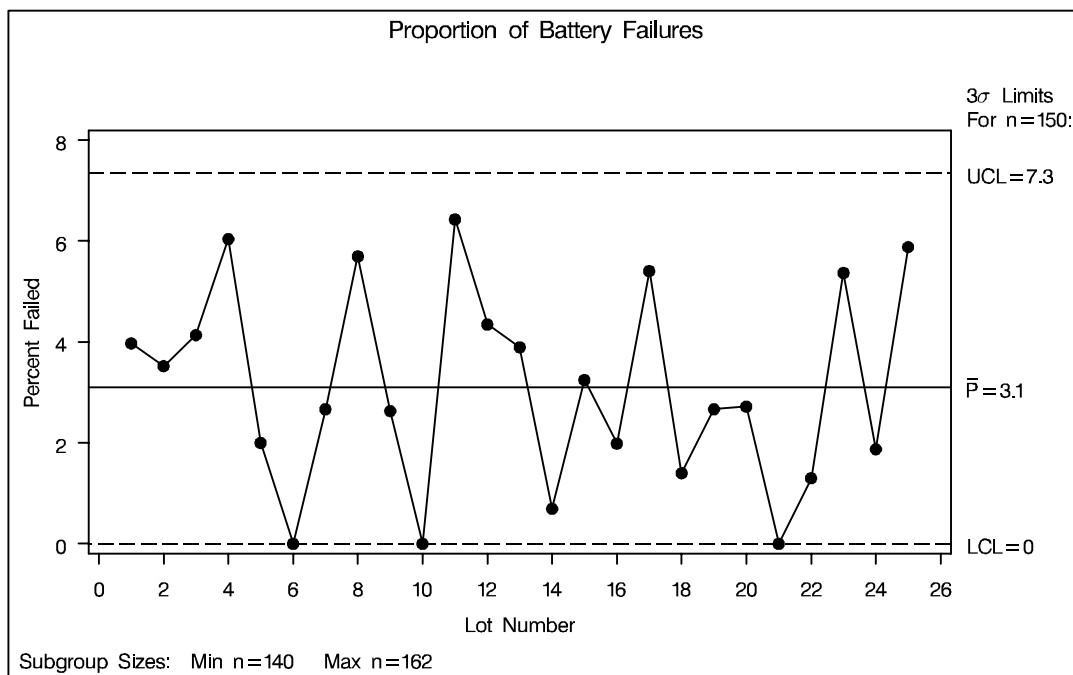
```

title 'Proportion of Battery Failures';
proc shewhart data=battery;
  pchart nfailed*lot / subgroupn=sampsize
                    limitn   =150
                    alln
                    yscale   =percent;
  label nfailed='Percent Failed';
run;

```

The ALLN option specifies that all points (regardless of subgroup sample size) are to be displayed. By default, only points for subgroups whose sample size matches the LIMITN= value are displayed. The YSCALE= option specifies that the vertical axis is to be scaled in percentages rather than proportions. The chart is shown in Output 38.3.3.

Output 38.3.3. Control Limits Based on Fixed Subgroup Sample Size



All the points are inside the control limits, indicating that the process is in statistical control. Since there is relatively little variation in the sample sizes, the control limits in Output 38.3.3 provide a close approximation to the exact control limits in Output 38.3.1, and the same conclusions can be drawn from both charts. In general, care should be taken when interpreting charts that use a nominal sample size to compute control limits, since these limits are only approximate when the sample sizes vary.

Example 38.4. Creating a Chart with Revised Control Limits

See SHWPEX4
in the SAS/QC
Sample Library

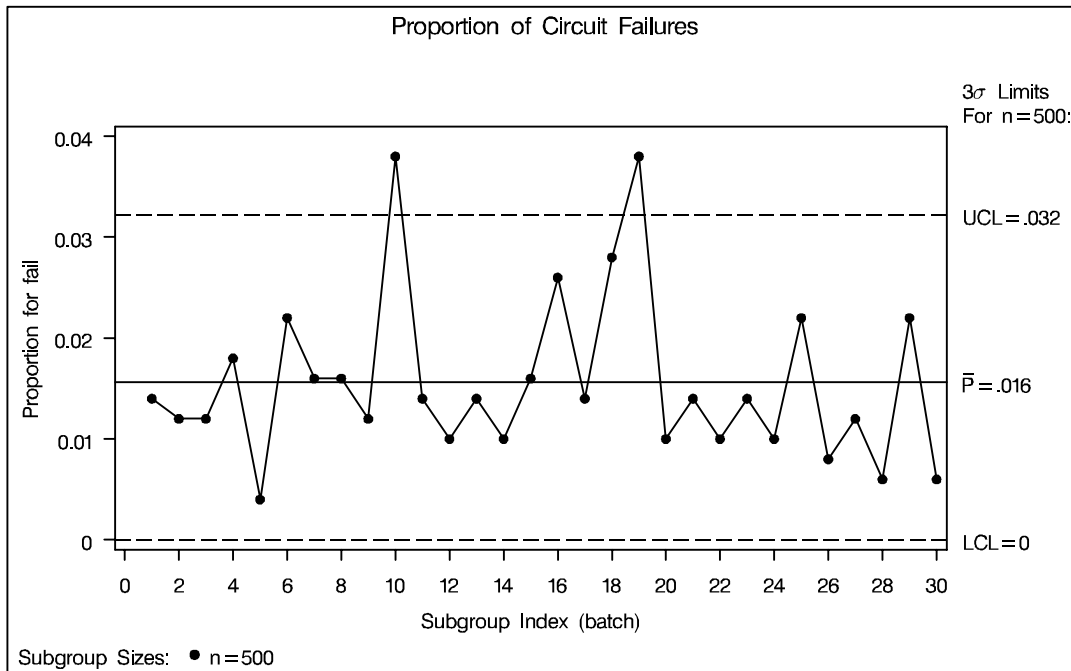
The following statements create a SAS data set named CIRC_ONE, which contains the number of failing circuits for 30 batches produced by the circuit manufacturing process introduced in the “Getting Started” section on page 1304:

```
data circ_one;
  input batch fail @@;
  datalines;
  1 7 2 6 3 6 4 9 5 2
  6 11 7 8 8 8 9 6 10 19
  11 7 12 5 13 7 14 5 15 8
  16 13 17 7 18 14 19 19 20 5
  21 7 22 5 23 7 24 5 25 11
  26 4 27 6 28 3 29 11 30 3
  ;
```

A *p* chart is used to monitor the proportion of failing circuits. The following statements create the chart shown in Output 38.4.1:

```
title 'Proportion of Circuit Failures';
symbol v=dot;
proc shewhart data=circ_one;
  pchart fail*batch / subgroupn=500
    outlimits=faillim1
    outindex ='Trial Limits';
run;
```

Output 38.4.1. A *p* Chart for Circuit Failures



Batches 10 and 19 have unusually high proportions of failing circuits. Subsequent investigation identifies special causes for both batches, and it is decided to eliminate these batches from the data set and recompute the control limits. The following statements create a data set named FAILLIM2 that contains the revised control limits:

```
proc shewhart data=circ_one;
  where batch^=10 and batch^=19;
  pchart fail*batch / subgroupn= 500
    nochart
    outindex ='Revised Limits'
    outlimits= faillim2;
run;

data faillims;
  set faillim1 faillim2;
run;
```

The data set FAILLIMS, which contains the true and revised control limits, is listed in Output 38.4.2.

Output 38.4.2. Listing of the Data Set FAILLIMS

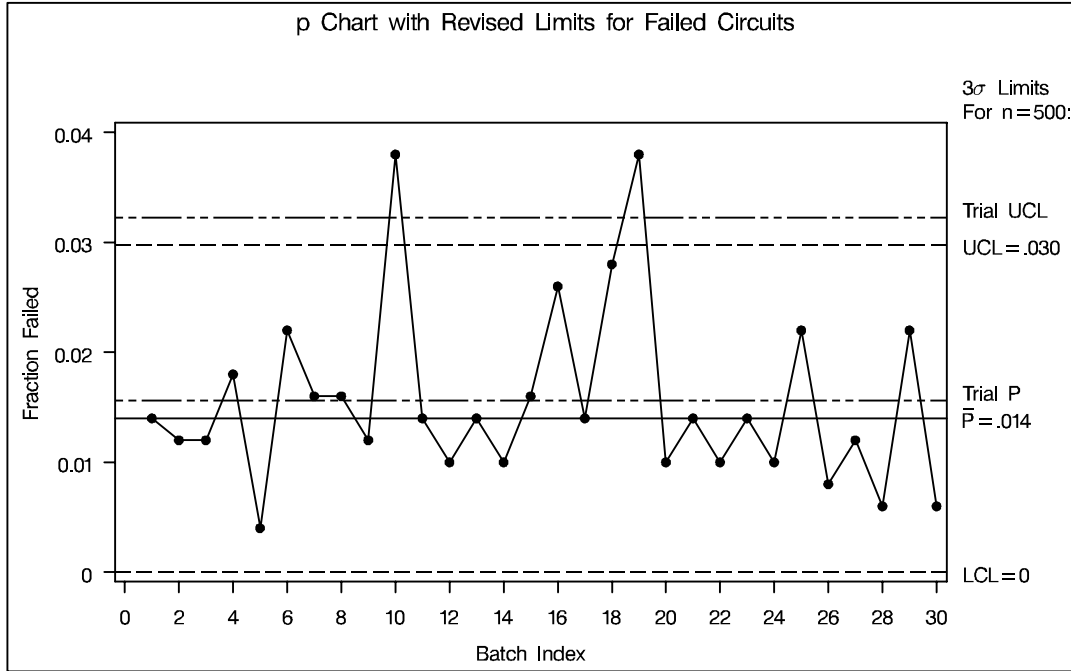
Proportion of Circuit Failures								
—	S	—	—	L	—	S	—	
—	U	I	—	I	—	A	I	
—	B	N	—	M	—	L	G	
V	G	D	Y	I	—	P	M	
A	R	E	P	T	—	H	A	
R	P	X	E	N	—	A	S	
—	—	—	—	—	—	—	—	
fail batch	Trial Limits	ESTIMATE	500	.005620297	3	0	0.0156	0.032226
fail batch	Revised Limits	ESTIMATE	500	.005942336	3	0	0.0140	0.029763

The following statements create a *p* chart displaying both sets of control limits:

```
title 'p Chart with Revised Limits for Failed Circuits';
symbol v=dot;
proc shewhart data=circ_one limits=faillims;
  pchart fail*batch / subgroupn = 500
    readindex = 'Revised Limits'
    vref      = 0.0156 0.032226
    vreflabels = ('Trial p'
                 'Trial UCL')
    vreflabpos = 3
    lvref      = 15
    nolegend;
  label fail = 'Fraction Failed'
    batch = 'Batch Index';
run;
```

The READINDEX= option is used to select the revised limits displayed on the p chart in Output 38.4.3. See “Displaying Multiple Sets of Control Limits” on page 1692. The VREF=, VREFLABELS=, and VREFLABPOS= options are used to display and label the trial limits. You can also pass in the values of the trial limits with macro variables. For an illustration of this technique, see Example 32.6 on page 1096.

Output 38.4.3. p Chart with Revised Limits



Example 38.5. OC Curve for Chart

See SHWPOC
in the SAS/QC
Sample Library

This example uses the GPLOT procedure and the OUTLIMITS= data set FAILLIM2 from the previous example to plot an OC curve for the p chart shown in Output 38.4.3.

The OC curve displays β (the probability that p_i lies within the control limits) as a function of p (the true proportion nonconforming). The computations are exact, assuming that the process is in control and that the number of nonconforming items (X_i) has a binomial distribution.

The value of β is computed as follows:

$$\begin{aligned}
 \beta &= P(p_i \leq \text{UCL}) - P(p_i < \text{LCL}) \\
 &= P(X_i \leq n\text{UCL}) - P(X_i < n\text{LCL}) \\
 &= P(X_i < n\text{UCL}) + P(X_i = n\text{UCL}) - P(X_i < n\text{LCL}) \\
 &= I_{1-p}(n + 1 - n\text{UCL}, n\text{UCL}) + P(X_i = n\text{UCL}) - I_{1-p}(n + 1 - n\text{LCL}, n\text{LCL}) \\
 &= I_p(n\text{LCL}, n + 1 - n\text{LCL}) + P(X_i = n\text{UCL}) - I_p(n\text{UCL}, n + 1 - n\text{UCL})
 \end{aligned}$$

Here, $I_p(\cdot, \cdot)$ denotes the incomplete beta function. The following DATA step computes β (the variable BETA) as a function of p (the variable P):

```

data ocpchart;
  set faillim2;
  keep beta fraction;
  nucl=_limitn*_uclp_;
  nlcl=_limitn*_lclp_;
  do p=0 to 500;
    fraction=p/1000;
    if nucl=floor(nucl) then
      adjust=probbnml(fraction,_limitn_,nucl) -
        probbnml(fraction,_limitn_,nucl-1);
    else adjust=0;
    if nlcl=0 then
      beta=1 - probbeta(fraction,nucl,_limitn_-nucl+1) + adjust;
    else beta=probbeta(fraction,nlcl,_limitn_-nlcl+1) -
      probbeta(fraction,nucl,_limitn_-nucl+1) +
      adjust;
    if beta >= 0.001 then output;
  end;
  call symput('lcl', put(_lclp_,5.3));
  call symput('mean',put(_p_, 5.3));
  call symput('ucl', put(_uclp_,5.3));
run;

```

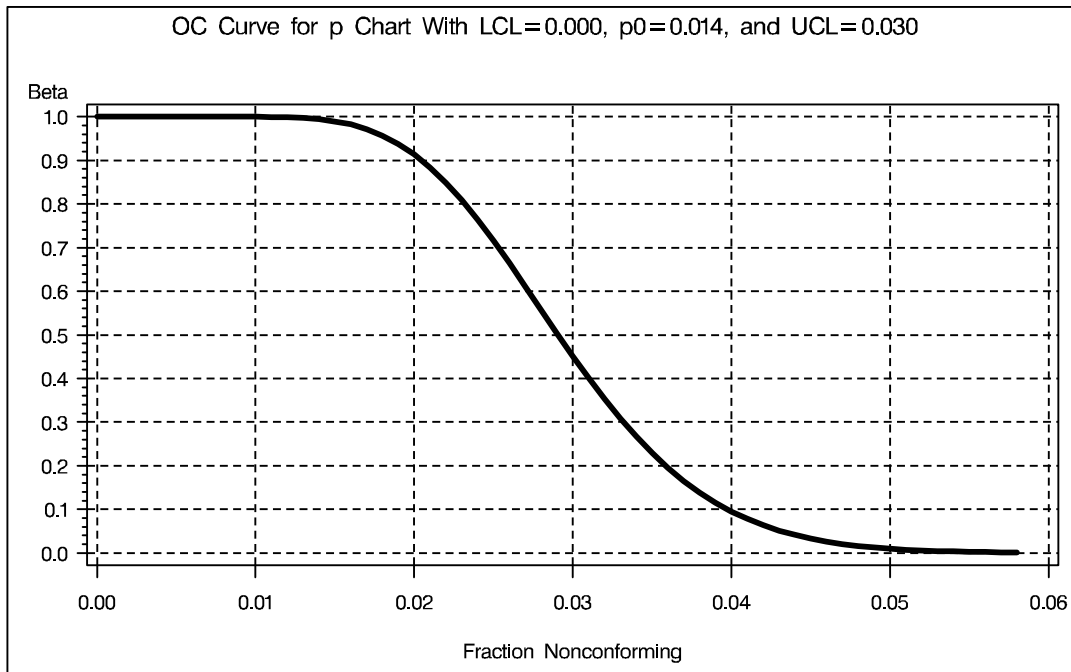
The following statements display the OC curve shown in Output 38.5.1:

```

title 'OC Curve for p Chart With LCL=&LCL, p0=&MEAN, and UCL=&UCL';
symbol i=j w=2 v=none;
proc gplot data=ocpchart;
  plot beta*fraction /
    vaxis=axis1
    haxis=axis2
    frame
    autovref
    autohref
    lvref = 2
    lhref = 2
    vzero
    hzero;
  label fraction='Fraction Nonconforming'
    beta = 'Beta';
axis1 offset=(0,.5) minor=none order=0 to 1.0 by 0.1;
axis2 offset=(0,0) minor=none order=0 to 0.06 by 0.005;
run;

```

Output 38.5.1. OC Curve for p Chart



The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *SAS/QC[®] User's Guide, Version 8*, Cary, NC: SAS Institute Inc., 1999. 1994 pp.

SAS/QC[®] User's Guide, Version 8

Copyright © 1999 SAS Institute Inc., Cary, NC, USA.

ISBN 1-58025-493-4

All rights reserved. Printed in the United States of America. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, by any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

U.S. Government Restricted Rights Notice. Use, duplication, or disclosure of the software by the government is subject to restrictions as set forth in FAR 52.227-19 Commercial Computer Software-Restricted Rights (June 1987).

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st printing, October 1999

SAS[®] and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute in the USA and other countries. [®] indicates USA registration.

IBM[®], ACF/VTAM[®], AIX[®], APPN[®], MVS/ESA[®], OS/2[®], OS/390[®], VM/ESA[®], and VTAM[®] are registered trademarks or trademarks of International Business Machines Corporation. [®] indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

The Institute is a private company devoted to the support and further development of its software and related services.