Chapter 42

XCHART Statement

Chapter Table of Contents

Chapter 42
XCHART Statement

Overview

The XCHART statement creates an $\bar{X}$ chart for subgroup means, which is used to analyze the central tendency of a process.

You can use options in the XCHART statement to

- compute control limits from the data based on a multiple of the standard error of the plotted means or as probability limits
- tabulate subgroup sample sizes, subgroup means, control limits, and other information
- save control limits in an output data set
- save subgroup sample sizes and subgroup means 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 one of several methods for estimating the process standard deviation
- specify whether subgroup standard deviations or subgroup ranges are used to estimate the process standard deviation
- specify a known (standard) process mean and standard deviation for computing control limits
- create a secondary chart that displays a time trend removed from the data (see "Displaying Trends in Process Data" on page 1711)
- 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

Note: When working with variables data, you should analyze the variability of the process as well as its central tendency. You can use the XRCHART statement or the XSCHART statement in the SHEWHART procedure for this purpose.

Getting Started

This section introduces the XCHART statement with simple examples that illustrate the most commonly used options. Complete syntax for the XCHART statement is presented in the "Syntax" section on page 1469, and advanced examples are given in the "Examples" section on page 1492.

Creating Charts for Means from Raw Data

See SHWXCHR in the SAS/QC Sample Library

Subgroup samples of five parts are taken from the manufacturing process at regular intervals, and the width of a critical gap in each part is measured in millimeters. The following statements create a SAS data set named PARTGAPS, which contains the gap width measurements for 21 samples:

```
data partgaps;
   input sample @;
   do i=1 to 5;
      input partgap @;
      output;
      end;
   drop i;
   label partgap='Gap Width'
         sample ='Sample Index';
   datalines;
 1 255 270 268 290 267
 2 260 240 265 262 263
 3 238 236 260 250 256
 4 260 242 281 254 263
 5 268 260 279 289 269
 6 270 249 265 253 263
 7 280 260 256 256 243
 8 229 266 250 243 252
 9 250 270 245 273 262
10 248 258 247 266 256
11 280 251 252 270 287
12 245 253 243 279 245
13 268 260 289 275 273
14 264 286 275 271 279
15 271 257 263 247 247
16 291 250 273 265 266
17 228 253 240 260 264
18 270 260 269 245 276
19 259 257 246 271 257
20 252 244 230 266 248
21 254 251 239 233 263
;
```

A partial listing of PARTGAPS is shown in Figure 42.1.

```
The Data Set PARTGAPS

  sample    partgap

     1        255
     1        270
     1        268
     1        290
     1        267
     2        260
     2        240
     2        265
     2        262
     2        263
     .         .
     .         .
     .         .
```

Figure 42.1. Partial Listing of the Data Set PARTGAPS

The data set PARTGAPS is said to be in "strung-out" form, since each observation contains the sample number and gap width measurement for a single part. The first five observations contain the gap widths for the first sample, the second five observations contain the gap widths for the second sample, and so on. Because the variable SAMPLE classifies the observations into rational subgroups, it is referred to as the *subgroup-variable*. The variable PARTGAP contains the gap width measurements and is referred to as the *process variable* (or *process* for short).

The within-subgroup variability of the gap widths is known to be stable. You can use an $\bar{X}$ chart to determine whether their mean level is in control. The following statements create the $\bar{X}$ chart shown in Figure 42.2:

```
title 'Mean Chart for Gap Widths';
symbol v=dot;
proc shewhart data=partgaps;
   xchart partgap*sample;
run;
```

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

The input data set is specified with the DATA= option in the PROC SHEWHART statement.

For more information on the SYMBOL statement, refer to *SAS/GRAPH Software: Reference.*

Each point on the $\bar{X}$ chart represents the average (mean) of the measurements for a particular sample. For instance, the mean plotted for the first sample is

$$\frac{255 + 270 + 268 + 290 + 267}{5} = 270$$

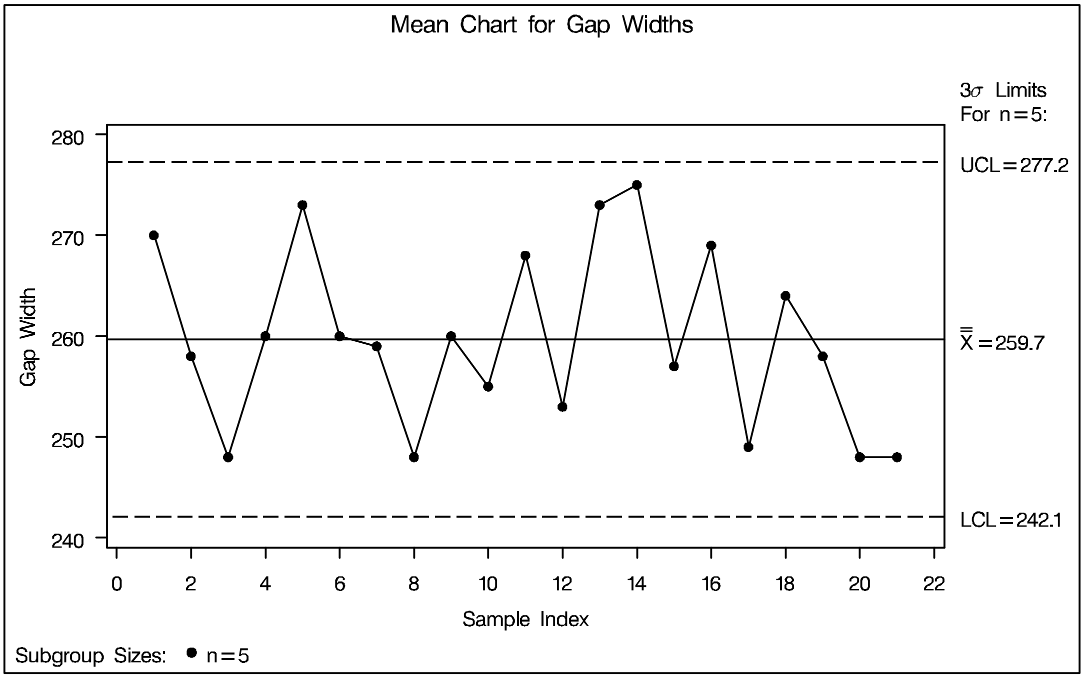


Figure 42.2. $\bar{X}$ Chart for Gap Width Data

Since all of the subgroup means lie within the control limits, it can be concluded that the mean level of 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 Table 42.22 on page 1481. You can also read control limits from an input data set; see "Reading Preestablished Control Limits" on page 1467.

For computational details, see "Constructing Charts for Means" on page 1480. For details on reading raw measurements, see "DATA= Data Set" on page 1485.

Creating Charts for Means from Subgroup Summary Data

See SHWXCHR in the SAS/QC Sample Library

The previous example illustrates how you can create $\bar{X}$ charts using raw data (process measurements). However, in many applications, the data are provided as subgroup summary statistics. This example illustrates how you can use the XCHART statement with data of this type.

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

```
data parts;
   input sample partgapx partgapr;
   partgapn=5;
   label partgapx='Mean of Gap Width'
         sample  ='Sample Index';
   datalines;
```

```
 1  270  35
 2  258  25
 3  248  24
 4  260  39
 5  273  29
 6  260  21
 7  259  37
 8  248  37
 9  260  28
10  255  19
11  268  36
12  253  36
13  273  29
14  275  22
15  257  24
16  269  41
17  249  36
18  264  31
19  258  25
20  248  36
21  248  30
;
```

A partial listing of PARTS is shown in Figure 42.3. There is exactly one observation for each subgroup (note that the subgroups are still indexed by SAMPLE). The variable PARTGAPX contains the subgroup means, the variable PARTGAPR contains the subgroup ranges, and the variable PARTGAPN contains the subgroup sample sizes (these are all five).

The Data Set PARTS

sample	partgapx	partgapr	partgapn
1	270	35	5
2	258	25	5
3	248	24	5
4	260	39	5
5	273	29	5
.	.	.	.
.	.	.	.
.	.	.	.

Figure 42.3. The Summary Data Set PARTS

You can read this data set by specifying it as a HISTORY= data set in the PROC SHEWHART statement, as follows:

```
title 'Mean Chart for Gap Width';
proc shewhart history=parts lineprinter;
   xchart partgap*sample='*';
run;
```

The resulting $\bar{X}$ chart is shown in Figure 42.4. 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.

Note that PARTGAP is *not* the name of a SAS variable in the data set but is, instead, the common prefix for the names of the three SAS variables PARTGAPX, PARTGAPR, and PARTGAPN. The suffix characters *X*, *R*, and *N* indicate *mean*, *range*, and *sample size*, respectively. Thus, you can specify three subgroup summary variables in a HISTORY= data set with a single name (PARTGAP), which is referred to as the *process*. The name SAMPLE specified after the asterisk is the name of the *subgroup-variable*.

```
                           Mean Chart for Gap Width

                                                                 3 Sigma Limits
                                                                 For n=5:
       --------------------------------------------------------
M 280 +                                                        |
e     |========================================================| UCL = 277.2
a     |                                  +*                    |
n     |             *                   * +                    |
  270 +   *        ++                   +  +   *               |
o     |   +        + +             *   +   +   +               |
f     |    +      +  +            ++   +    + + +   *          |
      |    +      +   +           + +  +    + + +   ++         | =
G 260 +-----+----*----*++*----*--+--+-+------+---+-+--+--------| X = 259.7
a     |     *    +       +   + + +  + +      *   + +   *       |
p     |      +  +         +  +  *    +           + +    +      |
      |       + +         + +        *            +     +      |
W 250 +       ++           ++                     *      +     |
i     |        *           *                             *++*  |
d     |                                                        |
t     |========================================================| LCL = 242.1
h 240 +                                                        |
       +----+----+----+----+----+----+----+----+----+----+----+
       0    2    4    6    8   10   12   14   16   18   20   22

                             Sample Index

Subgroup Sizes:  * n=5
```

Figure 42.4. $\bar{X}$ Chart from the Summary Data Set PARTS

In general, a HISTORY= input data set used with the XCHART statement must contain the following variables:

- subgroup variable
- subgroup mean variable
- either a subgroup range variable or a subgroup standard deviation variable
- subgroup sample size variable

Furthermore, the names of the subgroup mean, range (or standard deviation), and sample size variables must begin with the *process* name specified in the XCHART statement and end with the special suffix characters *X*, *R* (or *S*), and *N*, respectively. If the names do not follow this convention, you can use the RENAME option in the PROC SHEWHART statement to rename the variables for the duration of the SHEWHART procedure step (see page 1507).

If you specify the STDDEVIATIONS option in the XCHART statement, the HISTORY= data set must contain a subgroup standard deviation variable; otherwise, the HISTORY= data set must contain a subgroup range variable. The STDDEVIATIONS option specifies that the estimate of the process standard deviation σ is to be calculated from subgroup standard deviations rather than subgroup ranges. For example, in the following statements, the data set PARTS2 must contain a subgroup standard deviation variable named PARTGAPS:

```
title 'Mean Chart for Gap Width';
proc shewhart history=parts2;
   xchart partgap*sample='*' / stddeviations;
run;
```

Options such as STDDEVIATIONS are specified after the slash (/) in the XCHART statement. A complete list of options is presented in the "Syntax" section on page 1469.

In summary, the interpretation of *process* depends on the input data set.

- If raw data are read using the DATA= option (as in the previous example), *process* is the name of the SAS variable containing the process measurements.
- If summary data are read using the HISTORY= option (as in this example), *process* is the common prefix for the names of the variables containing the summary statistics.

For more information, see "HISTORY= Data Set" on page 1486.

Saving Summary Statistics

See SHWXCHR in the SAS/QC Sample Library

In this example, the XCHART statement is used to create a summary data set that can be read later by the SHEWHART procedure (as in the preceding example). The following statements read measurements from the data set PARTGAPS and create a summary data set named GAPHIST:

```
title 'Summary Data Set for Gap Widths';
proc shewhart data=partgaps;
   xchart partgap*sample / outhistory = gaphist
                           nochart;
run;
```

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 42.2.

Figure 42.5 contains a partial listing of GAPHIST.

```
                    Summary Data Set for Gap Widths

                 sample    partgapX   partgapR   partgapN

                    1        270         35         5
                    2        258         25         5
                    3        248         24         5
                    4        260         39         5
                    5        273         29         5
                    .         .           .         .
                    .         .           .         .
                    .         .           .         .
```

Figure 42.5. The Summary Data Set GAPHIST

There are four variables in the data set GAPHIST.

- SAMPLE contains the subgroup index.
- PARTGAPX contains the subgroup means.
- PARTGAPR contains the subgroup ranges.
- PARTGAPN contains the subgroup sample sizes.

Note that the summary statistic variables are named by adding the suffix characters *X*, *R*, and *N* to the *process* PARTGAP specified in the XCHART statement. In other words, the variable naming convention for OUTHISTORY= data sets is the same as that for HISTORY= data sets.

If you specify the STDDEVIATIONS option, the OUTHISTORY= data set includes a subgroup standard deviation variable rather than a subgroup range variable, as demonstrated by the following statements:

```
title 'Summary Data Set with Subgroup Standard Deviations';
proc shewhart data=partgaps;
   xchart partgap*sample / outhistory = gaphist2
                           stddeviations
                           nochart;
run;
```

Figure 42.6 contains a partial listing of GAPHIST2.

```
           Summary Data Set with Subgroup Standard Deviations

                 sample    partgapX   partgapS   partgapN

                    1        270      12.6293       5
                    2        258      10.2225       5
                    3        248      10.6771       5
                    4        260      14.2302       5
                    5        273      11.2027       5
                    .         .          .          .
                    .         .          .          .
                    .         .          .          .
```

Figure 42.6. The Summary Data Set GAPHIST2

The variable PARTGAPS, which contains the subgroup standard deviations, is named by adding the suffix character *S* to the *process* PARTGAP.

For more information, see "OUTHISTORY= Data Set" on page 1483.

Saving Control Limits

See SHWXCHR in the SAS/QC Sample Library

You can save the control limits for an $\bar{X}$ chart in a SAS data set; this enables you to apply the control limits to future data (see "Reading Preestablished Control Limits" on page 1467) or modify the limits with a DATA step program.

The following statements read measurements from the data set PARTGAPS (see page 1458) and save the control limits displayed in Figure 42.2 in a data set named GAPLIM:

```
title 'Control Limits for Gap Width Measurements';
proc shewhart data=partgaps;
   xchart partgap*sample / outlimits = gaplim
                           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 GAPLIM is listed in Figure 42.7.

Control Limits for Gap Width Measurements

VAR	_SUBGRP_	_TYPE_	_LIMITN_	_ALPHA_	_SIGMAS_
partgap	sample	ESTIMATE	5	.002699796	3

LCLX	_MEAN_	_UCLX_	_LCLR_	_R_	_UCLR_	_STDDEV_
242.087	259.667	277.246	0	30.4762	64.4419	13.1028

Figure 42.7. The Data Set GAPLIM Containing Control Limit Information

The data set GAPLIM contains one observation with the limits for *process* PARTGAP. The variables _LCLX_ and _UCLX_ contain the lower and upper control limits for the means, and the variable _MEAN_ contains the central line. The value of _MEAN_ is an estimate of the process mean, and the value of _STDDEV_ is an estimate of the process standard deviation σ. 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 values of _MEAN_ and _STDDEV_ are estimates or standard values.

The variables _LCLR_, _R_, and _UCLR_ are not used to create $\bar{X}$ charts, but they are included so the data set GAPLIM can be used to create an R chart; see Chapter 43, "XRCHART Statement." If you specify the STDDEVIATIONS option in the XCHART statement, the variables _LCLS_, _S_, and _UCLS_ are included in the OUTLIMITS= data set. These variables can be used to create an s chart; see Chapter 44, "XSCHART Statement." For more information, see "OUTLIMITS= Data Set" on page 1481.

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 'Summary Statistics and Control Limit Information';
proc shewhart data=partgaps;
   xchart partgap*sample / outtable=gtable
                           nochart;
run;
```

The data set GTABLE is listed in Figure 42.8.

Summary Statistics and Control Limit Information

VAR	sample	_SIGMAS_	_LIMITN_	_SUBN_	_LCLX_	_SUBX_	_MEAN_	_UCLX_	_EXLIM_
partgap	1	3	5	5	242.087	270	259.667	277.246	
partgap	2	3	5	5	242.087	258	259.667	277.246	
partgap	3	3	5	5	242.087	248	259.667	277.246	
partgap	4	3	5	5	242.087	260	259.667	277.246	
partgap	5	3	5	5	242.087	273	259.667	277.246	
partgap	6	3	5	5	242.087	260	259.667	277.246	
partgap	7	3	5	5	242.087	259	259.667	277.246	
partgap	8	3	5	5	242.087	248	259.667	277.246	
partgap	9	3	5	5	242.087	260	259.667	277.246	
partgap	10	3	5	5	242.087	255	259.667	277.246	
partgap	11	3	5	5	242.087	268	259.667	277.246	
partgap	12	3	5	5	242.087	253	259.667	277.246	
partgap	13	3	5	5	242.087	273	259.667	277.246	
partgap	14	3	5	5	242.087	275	259.667	277.246	
partgap	15	3	5	5	242.087	257	259.667	277.246	
partgap	16	3	5	5	242.087	269	259.667	277.246	
partgap	17	3	5	5	242.087	249	259.667	277.246	
partgap	18	3	5	5	242.087	264	259.667	277.246	
partgap	19	3	5	5	242.087	258	259.667	277.246	
partgap	20	3	5	5	242.087	248	259.667	277.246	
partgap	21	3	5	5	242.087	248	259.667	277.246	

Figure 42.8. The Data Set GTABLE

This data set contains one observation for each subgroup sample. The variables _SUBX_ and _SUBN_ contain the subgroup means and sample sizes. The variables _LCLX_ and _UCLX_ contain the lower and upper control limits, and the variable _MEAN_ contains the central line. The variables _VAR_ and SAMPLE contain the *process* name and values of the *subgroup-variable*, respectively. For more information, see "OUTTABLE= Data Set" on page 1484.

An OUTTABLE= data set can be read later as a TABLE= data set. For example, the following statements read GTABLE and display an $\bar{X}$ chart (not shown here) identical to the chart in Figure 42.2:

```
title 'Mean Chart for Gap Widths';
proc shewhart table=gtable;
   xchart partgap*sample;
label _SUBX_ = 'Gap Width';
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 1488.

Reading Preestablished Control Limits

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

See SHWXCHR in the SAS/QC Sample Library

```
data gaps2;
   input sample @;
   do i=1 to 5;
      input partgap @;
      output;
      end;
   drop i;
   datalines;
22 287 265 248 263 271
23 267 253 285 251 271
24 249 252 277 269 241
25 243 248 263 282 261
26 287 266 256 278 242
27 251 262 243 274 245
28 256 245 244 243 272
29 262 247 252 277 266
30 244 269 263 278 261
31 245 264 246 242 273
32 272 257 277 265 241
33 251 249 240 260 261
34 289 277 275 273 261
35 267 286 275 261 272
36 266 256 247 255 241
37 291 267 267 252 262
38 258 245 264 245 281
39 277 267 241 272 244
40 252 267 272 245 252
41 243 241 245 263 248
;
```

The following statements create an $\bar{X}$ chart for the data in GAPS2 using the control limits in GAPLIM:

```
title 'Mean Chart for Gap Widths';
proc shewhart data=gaps2 limits=gaplim;
   xchart partgap*sample;
run;
```

The chart is shown in Figure 42.7.

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 PARTGAP
- the value of _SUBGRP_ matches the *subgroup-variable* name SAMPLE

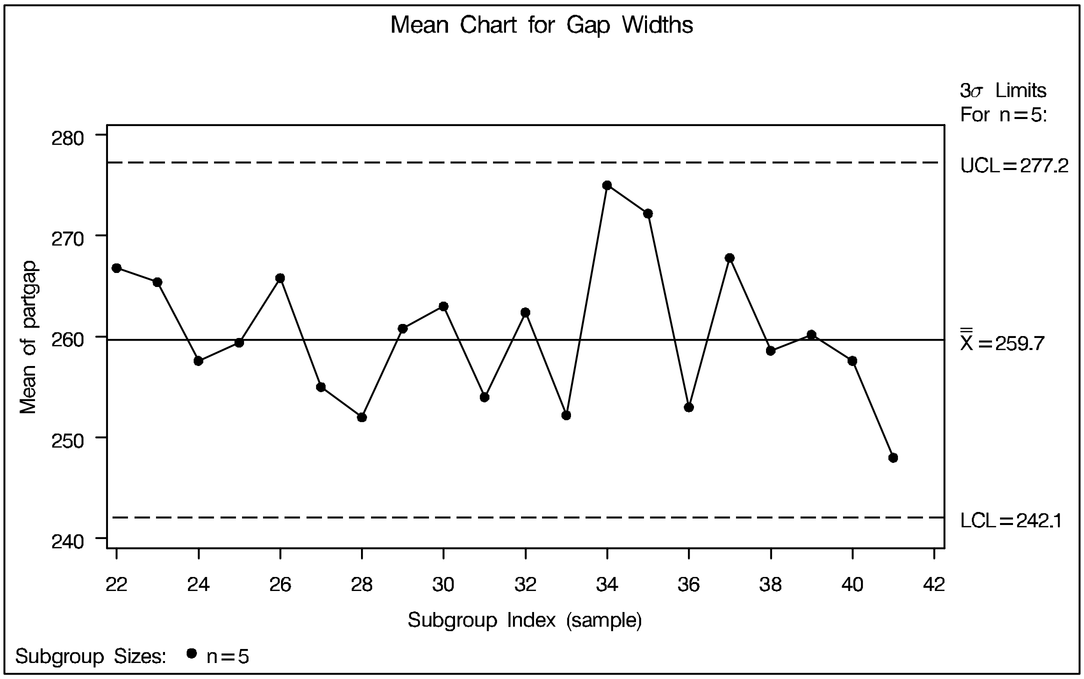


Figure 42.9. $\bar{X}$ Chart for Second Set of Gap Width Data

The chart indicates that the process is in control, since all the means lie within the control limits.

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 1486 for details concerning the variables that you must provide.

*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.

Syntax

The basic syntax for the XCHART statement is as follows:

XCHART *process*******subgroup-variable* **;**

The general form of this syntax is as follows:

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

You can use any number of XCHART statements in the SHEWHART procedure. The components of the XCHART 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 raw data are read from a DATA= data set, *process* must be the name of the variable containing the raw measurements. For an example, see "Creating Charts for Means from Raw Data" on page 1458.
- If summary data 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 Charts for Means from Subgroup Summary Data" on page 1460.
- If summary data 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 1465.

A *process* is required. If you specify more than one process, enclose the list in parentheses. For example, the following statements request distinct $\bar{X}$ charts for WEIGHT, LENGTH, and WIDTH:

```
proc shewhart data=measures;
   xchart (weight length width)*day;
run;
```

subgroup-variable
is the variable that identifies subgroups in the data. The *subgroup-variable* is required. In the preceding XCHART statement, DAY 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 the means.

- 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 SYMBOLn 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 an $\bar{X}$ chart using an asterisk (*) to plot the points:

```
proc shewhart data=values;
   xchart weight*day='*';
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 XCHART statement options by function. For complete descriptions, see Chapter 46, "Dictionary of Options."

Table 42.1. Tabulation Options

TABLE	creates a basic table of subgroup means, subgroup sample sizes, 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 42.2. Options for Specifying Tests for Special Causes

NO3SIGMACHECK	allows tests to be applied with control limits other than 3σ limits
TESTS=*value-list*| *customized-pattern-list*	specifies tests for special causes
TEST2RUN=*n*	specifies length of pattern for Test 2
TEST3RUN=*n*	specifies length of pattern for Test 3
TESTACROSS	applies tests across *phase* boundaries
TESTLABEL=’*label*’| (*variable*)|*keyword*	provides labels for points where test is positive
TESTLABEL*n*=’*label*’	specifies label for n^{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=*n*	specifies position of ZONEVALUES labels
ZONEVALUES	labels zone lines with their values

Table 42.3. Graphical Options for Displaying Tests for Special Causes

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

Table 42.4. Line Printer Options for Displaying Tests for Special Causes

TESTCHAR=’*character*’	specifies character for line segments that connect any sequence of points for which a test for special causes is positive
ZONECHAR=’*character*’	specifies character for lines that delineate zones for tests for special causes

Table 42.5. Reference Line Options

CHREF=*color*	specifies color for lines requested by HREF= and HREF2= options
CVREF=*color*	specifies color for lines requested by VREF= and VREF2= options
HREF=*values*| *SAS-data-set*	specifies position of reference lines perpendicular to horizontal axis on $\bar{X}$ chart
HREF2=*values*| *SAS-data-set*	specifies position of reference lines perpendicular to horizontal axis on trend chart
HREFCHAR='*character*'	specifies line character for HREF= and HREF2= lines
HREFDATA= *SAS-data-set*	specifies position of reference lines perpendicular to horizontal axis on $\bar{X}$ chart
HREF2DATA= *SAS-data-set*	specifies position of reference lines perpendicular to horizontal axis on trend chart
HREFLABELS= *'label1'...'labeln'*	specifies labels for HREF= lines
HREF2LABELS= *'label1'...'labeln'*	specifies labels for HREF2= lines
HREFLABPOS=*n*	specifies position of HREFLABELS= and HREF2LABELS= labels
LHREF=*linetype*	specifies line type for HREF= and HREF2= lines
LVREF=*linetype*	specifies line type for VREF= and VREF2= lines
NOBYREF	specifies that reference line information in a data set applies uniformly to charts created for all BY groups
VREF=*values*| *SAS-data-set*	specifies position of reference lines perpendicular to vertical axis on $\bar{X}$ chart
VREF2=*values*| *SAS-data-set*	specifies position of reference lines perpendicular to vertical axis on trend chart
VREFCHAR='*character*'	specifies line character for VREF= and VREF2= lines
VREFLABELS= *'label1'...'labeln'*	specifies labels for VREF= lines
VREF2LABELS= *'label1'...'labeln'*	specifies labels for VREF2= lines
VREFLABPOS=*n*	position of VREFLABELS= and VREF2LABELS= labels

Table 42.6. Block Variable Legend Options

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

Table 42.7. Axis and Axis Label Options

CAXIS=*color*	specifies color for axis lines and tick marks
CFRAME=*color*| (*color-list*)	specifies fill colors for frame for plot area
CTEXT=*color*	specifies color for tick mark values and axis labels
HAXIS=*values*|AXISn	specifies major tick mark values for horizontal axis
HEIGHT=*value*	specifies height of axis label and axis legend text
HMINOR=*n*	specifies number of minor tick marks between major tick marks on horizontal axis
HOFFSET=*value*	specifies length of offset at both ends of horizontal axis
INTSTART=*value*	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 subgroup values is to be labeled on horizontal axis
NOVANGLE	requests vertical axis labels that are strung out vertically
SKIPHLABELS=*n*	specifies thinning factor for tick mark labels on horizontal axis
SPLIT='*character*'	specifies splitting character for axis labels
TURNHLABELS	requests horizontal axis labels that are strung out vertically
VAXIS=*values*|AXISn	specifies major tick mark values for vertical axis of $\bar{X}$ chart
VAXIS2=*values*|AXISn	specifies major tick mark values for vertical axis of trend chart
VMINOR=*n*	specifies number of minor tick marks between major tick marks on vertical axis
VOFFSET=*value*	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=*n*	specifies width of axis lines

Table 42.8. Options for Specifying Control Limits

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

Table 42.9. Options for Displaying Control Limits

CINFILL=*color*	specifies color for area inside control limits
CLIMITS=*color*	specifies color of control limits, central line, and related labels
LCLLABEL=’*label*’	specifies label for lower control limit
LIMLABSUBCHAR= ’*character*’	specifies a substitution character for labels provided as quoted strings; the character is replaced with the value of the control limit
LLIMITS=*linetype*	specifies line type for control limits
NDECIMAL=*n*	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
NOUCL	suppresses display of upper control limit
UCLLABEL=’*string*’	specifies label for upper control limit
WLIMITS=*n*	specifies width for control limits and central line
XSYMBOL=’*string*’| *keyword*	specifies label for central line

Table 42.10. Specification Limit Options

CIINDICES=(ALPHA=*value* TYPE=*keyword*)	specifies α value and type for computing capability index confidence limits
LSL=*value-list*	specifies list of lower specification limits
TARGET=*value-list*	specifies list of target values
USL=*value-list*	specifies list of upper specification limits

Table 42.11. Options for Plotting and Labeling Points

ALLLABEL=VALUE|(*variable*)	labels every point on $\bar{X}$ chart
ALLLABEL2=VALUE|(*variable*)	labels every point on trend chart
CCONNECT=*color*	specifies color for line segments that connect points on chart
CFRAMELAB=*color*	specifies fill color for frame around labeled points
CNEEDLES=*color*	specifies color for needles that connect points to central line
CONNECTCHAR=’*character*’	specifies character used to form line segments that connect points on chart
COUT=*color*	specifies color for portions of line segments that connect points outside control limits
COUTFILL=*color*	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
NOTRENDCONNECT	suppresses line segments that connect points on trend chart
OUTLABEL=VALUE|(*variable*)	labels points outside control limits
SYMBOLCHARS=’*characters*’	specifies characters indicating *symbol-variable*
SYMBOLLEGEND=NONE|*name*	specifies LEGEND statement for levels of *symbol-variable*
SYMBOLORDER=*keyword*	specifies order in which symbols are assigned for levels of *symbol-variable*
TURNALL|TURNOUT	turns point labels so that they are strung out vertically

Table 42.12. Clipping Options

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

Table 42.13. Phase Options

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

Table 42.14. Process Mean and Standard Deviation Options

MU0=*value*	specifies known value of μ_0 for process mean μ
SIGMA0=*value*	specifies known value σ_0 for process standard deviation σ
SMETHOD=*keyword*	specifies method for estimating process standard deviation σ
STDDEVIATIONS	specifies that estimate of process standard deviation σ is to be calculated from subgroup standard deviations
TYPE=*keyword*	identifies parameters as estimates or standard values and specifies value of _TYPE_ in the OUTLIMITS= data set

Table 42.15. Options for Interactive Control Charts

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

Table 42.16. Input Data Set Options

MISSBREAK	specifies that observations with missing values are not to be processed

Table 42.17. Output Data Set Options

OUTHISTORY=*SAS-data-set*	creates output data set containing subgroup summary statistics
OUTINDEX='*string*'	specifies value of _INDEX_ in the OUTLIMITS= data set
OUTLIMITS=*SAS-data-set*	creates output data set containing control limits
OUTTABLE=*SAS-data-set*	creates output data set containing subgroup summary statistics and control limits

Table 42.18. Grid Options

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

Table 42.19. Plot Layout Options

ALLN	plots means for all subgroups
BILEVEL	creates control charts using half-screens and half-pages
EXCHART	creates control charts for a process only when exceptions occur
INTERVAL=*keyword*	natural time interval between consecutive subgroup positions when time, date, or datetime format is associated with a numeric subgroup variable
MAXPANELS=*n*	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=*n*	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=*n*	specifies number of pages or screens to be used to display chart
TRENDVAR=*variable*| (*variable-list*)	specifies list of trend variables
YPCT1=*value*	specifies length of vertical axis on $\bar{X}$ chart as a percentage of sum of lengths of vertical axes for $\bar{X}$ and trend charts
ZEROSTD	displays $\bar{X}$ chart regardless of whether $\hat{\sigma} = 0$

Table 42.20. Graphical Enhancement Options

ANNOTATE= *SAS-data-set*	specifies annotate data set that adds features to $\bar{X}$ chart
ANNOTATE2= *SAS-data-set*	specifies annotate data set that adds features to trend chart
DESCRIPTION='*string*'	specifies string that appears in the description field of the PROC GREPLAY master menu for $\bar{X}$ chart
FONT=*font*	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 for $\bar{X}$ chart
PAGENUM='*string*'	specifies the form of the label used in pagination
PAGENUMPOS= *keyword*	specifies the position of the page number requested with the PAGENUM= option
WTREND=*n*	specifies width of line segments connecting points on trend chart

Table 42.21. Star Options

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

Details

Constructing Charts for Means

The following notation is used in this section:

μ	process mean (expected value of the population of measurements)
σ	process standard deviation (standard deviation of the population of measurements)
$\bar{X}_i$	mean of measurements in i^{th} subgroup
R_i	range of measurements in i^{th} subgroup
n_i	sample size of i^{th} subgroup
N	number of subgroups
$\overline{\overline{X}}$	weighted average of subgroup means
z_p	$100p^{\text{th}}$ percentile of the standard normal distribution

Plotted Points

Each point on an $\bar{X}$ chart indicates the value of a subgroup mean ($\bar{X}_i$). For example, if the tenth subgroup contains the values 12, 15, 19, 16, and 14, the value plotted for this subgroup is

$$\bar{X}_{10} = \frac{12 + 15 + 19 + 16 + 14}{5} = 15.2$$

Central Line

By default, the central line on an $\bar{X}$ chart indicates an estimate for μ, which is computed as

$$\hat{\mu} = \overline{\overline{X}} = \frac{n_1\bar{X}_1 + \cdots + n_N\bar{X}_N}{n_1 + \cdots + n_N}$$

If you specify a known value (μ_0) for μ, the central line indicates the value of μ_0.

Control Limits

You can compute the limits in the following ways:

- as a specified multiple (k) of the standard error of $\bar{X}_i$ 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 $\bar{X}_i$ exceeds the limits

The following table provides the formulas for the limits:

Table 42.22. Limits for $\bar{X}$ Charts

Control Limits
LCL = lower limit = $\overline{\overline{X}} - k\hat{\sigma}/\sqrt{n_i}$ UCL = upper limit = $\overline{\overline{X}} + k\hat{\sigma}/\sqrt{n_i}$

Probability Limits
LCL = lower limit = $\overline{\overline{X}} - z_{\alpha/2}(\hat{\sigma}/\sqrt{n_i})$ UCL = upper limit = $\overline{\overline{X}} + z_{\alpha/2}(\hat{\sigma}/\sqrt{n_i})$

Note that the limits vary with n_i. If standard values μ_0 and σ_0 are available for μ and σ, respectively, replace $\overline{\overline{X}}$ with μ_0 and $\hat{\sigma}$ with σ_0 in Table 42.22.

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 μ_0 with the MU0= option or with the variable _MEAN_ in a LIMITS= data set.
- Specify σ_0 with the SIGMA0= option or with the variable _STDDEV_ 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 42.23. OUTLIMITS= Data Set

Variable	Description
ALPHA	probability (α) of exceeding limits
_CP_	capability index C_p
CPK	capability index C_{pk}
CPL	capability index CPL
CPM	capability index C_{pm}
CPU	capability index CPU
INDEX	optional identifier for the control limits specified with the OUTINDEX= option
LCLR	lower control limit for subgroup range
LCLS	lower control limit for subgroup standard deviation
LCLX	lower control limit for subgroup mean

Table 42.23. (continued)

Variable	Description
LIMITN	sample size associated with the control limits
LSL	lower specification limit
MEAN	process mean ($\overline{\overline{X}}$ or μ_0)
_R_	value of central line on R chart
_S_	value of central line on s chart
SIGMAS	multiple (k) of standard error of $\bar{X}_i$
STDDEV	process standard deviation ($\hat{\sigma}$ or σ_0)
SUBGRP	*subgroup-variable* specified in the XCHART statement
TARGET	target value
TYPE	type (estimate or standard value) of _MEAN_ and _STDDEV_
UCLR	upper control limit for subgroup range
UCLS	upper control limit for subgroup standard deviation
UCLX	upper control limit for subgroup mean
USL	upper specification limit
VAR	*process* specified in the XCHART statement

Notes:

1. The variables _LCLS_, _S_, and _UCLS_ are included if you specify the STDDEVIATIONS option; otherwise, the variables _LCLR_, _R_, and _UCLR_ are included. These variables are not used to create $\bar{X}$ charts, but they allow the OUTLIMITS= data set to be used as a LIMITS= data set with the BOXCHART, MRCHART, RCHART, SCHART, XRCHART, and XSCHART statements.

2. If the control limits vary with subgroup sample size, the special missing value *V* is assigned to the variables _LIMITN_, _LCLX_, _UCLX_, _LCLR_, _R_, _UCLR_, _LCLS_, _S_, and _UCLS_.

3. If the limits are defined in terms of a multiple k of the standard error of $\bar{X}_i$, the value of _ALPHA_ is computed as $\alpha = 2(1 - \Phi(k))$, where $\Phi(\cdot)$ is the standard normal distribution function.

4. If the limits are probability limits, the value of _SIGMAS_ is computed as $k = \Phi^{-1}(1 - \alpha/2)$, where Φ^{-1} is the inverse standard normal distribution function.

5. The variables _CP_, _CPK_, _CPL_, _CPU_, _LSL_, and _USL_ are included only if you provide specification limits with the LSL= and USL= options. The variables _CPM_ and _TARGET_ are included if, in addition, you provide a target value with the TARGET= option. See "Capability Indices" on page 1537 for computational details.

6. Optional BY variables are saved in the OUTLIMITS= data set.

The OUTLIMITS= data set contains one observation for each *process* specified in the XCHART statement. For an example, see "Saving Control Limits" on page 1465.

OUTHISTORY= Data Set

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

- the *subgroup-variable*
- a subgroup mean variable named by *process* suffixed with *X*
- a subgroup sample size variable named by *process* suffixed with *N*
- a subgroup range variable named by *process* suffixed with *R*
- a subgroup standard deviation variable named by *process* suffixed with *S*

A subgroup standard deviation variable is included if you specify the STDDEVIATIONS option; otherwise, a subgroup range variable is included.

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* DIAMETER to DIAMTER before adding the suffix.

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

```
proc shewhart data=steel;
   xchart (width diameter)*lot / outhistory=summary;
run;
```

The data set SUMMARY contains variables named LOT, WIDTHX, WIDTHR, WIDTHN, DIAMTERX, DIAMTERR, and DIAMTERN. The variables WIDTHR and DIAMTERR are included, since the STDDEVIATIONS option is not specified. If you specified the STDDEVIATIONS option, the data set SUMMARY would contain the variables WIDTHS and DIAMTERS rather than WIDTHR and DIAMTERR.

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 Summary Statistics" on page 1463.

OUTTABLE= Data Set

The OUTTABLE= data set saves subgroup summary statistics, control limits, and related information. The following variables can be saved:

Variable	Description
ALPHA	probability (α) of exceeding control limits
EXLIM	control limit exceeded on $\bar{X}$ chart
LCLX	lower control limit for mean
LIMITN	nominal sample size associated with the control limits
MEAN	process mean
SIGMAS	multiple (*k*) of the standard error associated with control limits
subgroup	values of the subgroup variable
SUBN	subgroup sample size
SUBX	subgroup mean
TESTS	tests for special causes signaled on $\bar{X}$ chart
UCLX	upper control limit for mean
VAR	*process* specified in the XCHART 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)
- _TREND_ (if the TRENDVAR= 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= option, 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 all eight tests and Tests 2 and 8 are positive for a given subgroup, the value of _TESTS_ has a 2 for the second character, an 8 for the eighth 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 1465.

ODS Tables

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

Table 42.24. ODS Tables Produced with the XCHART Statement

Table Name	Description	Options
XCHART	$\bar{X}$ 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 (process measurements) from a DATA= data set specified in the PROC SHEWHART statement. Each *process* specified in the XCHART statement must be a SAS variable in the DATA= data set. This variable provides measurements that must be grouped into subgroup samples indexed by the *subgroup-variable*. The *subgroup-variable*, which is specified in the XCHART statement, must also be a SAS variable in the DATA= data set. Each observation in a DATA= data set must contain a value for each *process* and a value for the *subgroup-variable*. If the i^{th} subgroup contains n_i items, there should be n_i consecutive observations for which the value of the *subgroup-variable* is the index of the i^{th} subgroup. For example, if each subgroup contains five items and there are 30 subgroup samples, the DATA= data set should contain 150 observations.

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

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*) with the READPHASES= option (for an example, see "Displaying Stratification in Phases" on page 1689).

For an example of a DATA= data set, see "Creating Charts for Means from Raw Data" on page 1458.

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;
   xchart weight*batch;
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; see Table 42.23 on page 1481. 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 _LCLX_, _MEAN_, and _UCLX_, which specify the control limits directly
- the variables _MEAN_ and _STDDEV_, which are used to calculate the control limits according to the equations in Table 42.22 on page 1481

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`, `STANDARD`, `STDMU`, and `STDSIGMA`.
- BY variables are required if specified with a BY statement.

For an example, see "Reading Preestablished Control Limits" on page 1467.

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, CUSUM, or MACONTROL procedures or to read output data sets created with SAS summarization procedures, such as PROC MEANS.

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

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

- the *subgroup-variable*
- a subgroup mean variable for each *process*
- a subgroup sample size variable for each *process*
- either a subgroup range variable or subgroup standard deviation variable for each *process*

If you specify the STDDEVIATIONS option, the subgroup standard deviation variable must be included; otherwise, the subgroup range variable must be included.

The names of the subgroup mean, subgroup range or subgroup standard deviation, and subgroup sample size variables must be the *process* name concatenated with the suffix characters X, R or S, and N, respectively.

For example, consider the following statements:

```
proc shewhart history=summary;
   xchart (weight yldstren)*batch;
run;
```

The data set SUMMARY must include the variables BATCH, WEIGHTX, WEIGHTR, WEIGHTN, YLDSRENX, YLDSRENR, and YLDSRENN. If the STDDEVIATIONS option were specified in the preceding XCHART statement, it would be necessary for SUMMARY to include the variables BATCH, WEIGHTX, WEIGHTS, WEIGHTN, YLDSRENX, YLDSRENS, and YLDSRENN.

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 Charts for Means from Subgroup Summary Data” on page 1460.

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 in 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 XCHART statement:

Table 42.25. Variables Required in a TABLE= Data Set

Variable	Description
LCLX	lower control limit for mean
LIMITN	nominal sample size associated with the control limits
MEAN	process mean
subgroup-variable	values of the *subgroup-variable*
SUBN	subgroup sample size
SUBX	subgroup mean
UCLX	upper control limit for mean

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 1465.

Methods for Estimating the Standard Deviation

When control limits are computed from the input data, three methods (referred to as default, MVLUE, and RMSDF) are available for estimating the process standard deviation σ. The method depends on whether you specify the STDDEVIATIONS option. If you specify this option, σ is estimated using subgroup standard deviations; otherwise, σ is estimated using subgroup ranges.

For an illustration of the methods, see Example 42.2 on page 1494.

Default Method Based on Subgroup Ranges

If you do not specify the STDDEVIATIONS option, the default estimate for σ is

$$\hat{\sigma} = \frac{R_1/d_2(n_1) + \cdots + R_N/d_2(n_N)}{N}$$

where N is the number of subgroups for which $n_i \geq 2$, and R_i is the sample range of the observations $x_{i1}, \ldots, x_{in_i}$ in the i^{th} subgroup.

$$R_i = \max_{1 \leq j \leq n_i}(x_{ij}) - \min_{1 \leq j \leq n_i}(x_{ij})$$

A subgroup range R_i is included in the calculation only if $n_i \geq 2$. The unbiasing factor $d_2(n_i)$ is defined so that, if the observations are normally distributed, the expected value of R_i is $d_2(n_i)\sigma$. Thus, $\hat{\sigma}$ is the unweighted average of N unbiased estimates of σ. This method is described in the *ASTM Manual on Presentation of Data and Control Chart Analysis* (1976).

Default Method Based on Subgroup Standard Deviations

If you specify the STDDEVIATIONS option, the default estimate for σ is

$$\hat{\sigma} = \frac{s_1/c_4(n_1) + \cdots + s_N/c_4(n_N)}{N}$$

where N is the number of subgroups for which $n_i \geq 2$, s_i is the sample standard deviation of the i^{th} subgroup

$$s_i = \sqrt{\frac{1}{n_i - 1}\sum_{j=1}^{n_i}(x_{ij} - \bar{X}_i)^2}$$

and

$$c_4(n_i) = \frac{\Gamma(n_i/2)\sqrt{2/(n_i - 1)}}{\Gamma((n_i - 1)/2)}$$

Here $\Gamma(\cdot)$ denotes the gamma function, and $\bar{X}_i$ denotes the i^{th} subgroup mean. A subgroup standard deviation s_i is included in the calculation only if $n_i \geq 2$. If the observations are normally distributed, the expected value of s_i is $c_4(n_i)\sigma$. Thus, $\hat{\sigma}$ is the unweighted average of N unbiased estimates of σ. This method is described in the *ASTM Manual on Presentation of Data and Control Chart Analysis* (1976).

MVLUE Method Based on Subgroup Ranges

If you do not specify the STDDEVIATIONS option and you specify SMETHOD=MVLUE, a minimum variance linear unbiased estimate (MVLUE) is computed for σ. Refer to Burr (1969, 1976) and Nelson (1989, 1994). The MVLUE is a weighted average of N unbiased estimates of σ of the form $R_i/d_2(n_i)$, and it is computed as

$$\hat{\sigma} = \frac{f_1 R_1/d_2(n_1) + \cdots + f_N R_N/d_2(n_N)}{f_1 + \cdots + f_N}$$

where

$$f_i = \frac{[d_2(n_i)]^2}{[d_3(n_i)]^2}$$

A subgroup range R_i is included in the calculation only if $n_i \geq 2$, and N is the number of subgroups for which $n_i \geq 2$. The unbiasing factor $d_3(n_i)$ is defined so that, if the observations are normally distributed, the expected value of σ_{R_i} is $d_3(n_i)\sigma$. The MVLUE assigns greater weight to estimates of σ from subgroups with larger sample sizes, and it is intended for situations where the subgroup sample sizes vary. If the subgroup sample sizes are constant, the MVLUE reduces to the default estimate.

MVLUE Method Based on Subgroup Standard Deviations

If you specify the STDDEVIATIONS option and SMETHOD=MVLUE, a minimum variance linear unbiased estimate (MVLUE) is computed for σ. Refer to Burr (1969, 1976) and Nelson (1989, 1994). This estimate is a weighted average of N unbiased estimates of σ of the form $s_i/c_4(n_i)$, and it is computed as

$$\hat{\sigma} = \frac{h_1 s_1/c_4(n_1) + \cdots + h_N s_N/c_4(n_N)}{h_1 + \cdots + h_N}$$

where

$$h_i = \frac{[c_4(n_i)]^2}{1 - [c_4(n_i)]^2}$$

A subgroup standard deviation s_i is included in the calculation only if $n_i \geq 2$, and N is the number of subgroups for which $n_i \geq 2$. The MVLUE assigns greater weight to estimates of σ from subgroups with larger sample sizes, and it is intended for situations where the subgroup sample sizes vary. If the subgroup sample sizes are constant, the MVLUE reduces to the default estimate.

RMSDF Method Based on Subgroup Standard Deviations

If you specify the STDDEVIATIONS option and SMETHOD=RMSDF, a weighted root-mean-square estimate is computed for σ.

$$\hat{\sigma} = \frac{\sqrt{(n_1 - 1)s_1^2 + \cdots + (n_N - 1)s_N^2}}{c_4(n)\sqrt{n_1 + \cdots + n_N - N}}$$

The weights are the degrees of freedom $n_i - 1$. A subgroup standard deviation s_i is included in the calculation only if $n_i \geq 2$, and N is the number of subgroups for which $n_i \geq 2$.

If the unknown standard deviation σ is constant across subgroups, the root-mean-square estimate is more efficient than the minimum variance linear unbiased estimate. However, in process control applications, it is generally not assumed that σ is constant, and if σ varies across subgroups, the root-mean-square estimate tends to be more inflated than the MVLUE.

Default Method Based on Individual Measurements

When each subgroup sample contains a single observation ($n_i \equiv 1$), the process standard deviation σ is estimated as $\hat{\sigma} = \bar{R}/d_2(2)$, where $\bar{R}$ is the average of the moving ranges of consecutive measurements taken in pairs. This is the method used to estimate σ for individual measurements and moving range charts. See page 1172 in Chapter 34, “IRCHART Statement.”

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	*subgroup-variable*
Vertical	DATA=	*process*
Vertical	HISTORY=	subgroup mean variable
Vertical	TABLE=	_SUBX_

If you specify the TRENDVAR= option, you can provide distinct labels for the vertical axes of the $\bar{X}$ and trend charts by breaking the vertical axis into two parts with a split character. Specify the split character with the SPLIT= option. The first part labels the vertical axis of the $\bar{X}$ chart, and the second part labels the vertical axis of the trend chart.

For example, the following sets of statements specify the label *Residual Mean* for the vertical axis of the $\bar{X}$ chart and the label *Fitted Mean* for the vertical axis of the trend chart:

```
proc shewhart data=toolwear;
   xchart diameter*hour / split    = '/'
                          trendvar = fitted ;
   label diameter = 'Residual Mean/Fitted Mean';
run;

proc shewhart history=regdata;
   xchart diameter*hour / split    = '/'
                          trendvar = fitted;
   label diamterx = 'Residual Mean/Fitted Mean';
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 XCHART statement.

Example 42.1. Applying Tests for Special Causes

See SHWTEST in the SAS/QC Sample Library

This example illustrates how you can apply tests for special causes to make $\bar{X}$ charts more sensitive to special causes of variation.

The following statements create an $\bar{X}$ chart for the gap width measurements in the data set PARTS on page 1460 and tabulate the results:

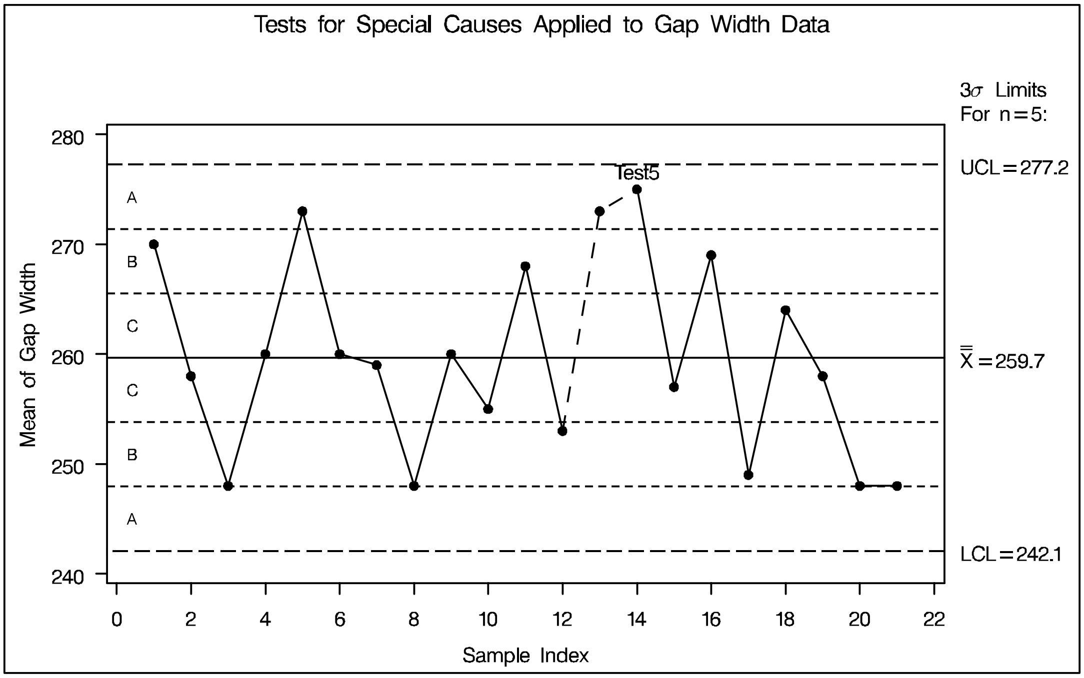
```
title 'Tests for Special Causes Applied to Gap Width Data';
symbol v=dot;
proc shewhart history=parts;
   xchart partgap*sample/ tests =1 to 5
                          ltests=20
                          tabletests
                          tablecentral
                          tablelegend
                          zonelabels
                          nolegend;
run;
```

The $\bar{X}$ chart is shown in Output 42.1.1 and the printed output is shown in Output 42.1.2. The TESTS= requests Tests 1, 2, 3, 4, and 5, which are described in Chapter 48, "Tests for Special Causes." The TABLECENTRAL option requests a table of the subgroup means, control limits, and central line. The TABLETESTS option adds a column indicating which subgroups tested positive for special causes, and the TABLELEGEND option adds a legend describing the tests that were signaled.

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 test patterns that were signaled. The NOLEGEND option suppresses the subgroup sample size legend that is displayed by default in the lower left corner of the chart.

Output 42.1.1 and Output 42.1.2 indicate that Test 5 was positive at sample 14, signaling a possible shift in the mean of the process.

Output 42.1.1. Tests for Special Causes Displayed on an $\bar{X}$ Chart



Output 42.1.2. Tabular Form of $\bar{X}$ Chart

Tests for Special Causes Applied to Gap Width Data

Means Chart Summary for partgap

sample	Subgroup Sample Size	3 Sigma Limits with n=5 for Mean: Lower Limit	Subgroup Mean	Average Mean	Upper Limit	Special Tests Signaled
1	5	242.08741	270.00000	259.66667	277.24592	
2	5	242.08741	258.00000	259.66667	277.24592	
3	5	242.08741	248.00000	259.66667	277.24592	
4	5	242.08741	260.00000	259.66667	277.24592	
5	5	242.08741	273.00000	259.66667	277.24592	
6	5	242.08741	260.00000	259.66667	277.24592	
7	5	242.08741	259.00000	259.66667	277.24592	
8	5	242.08741	248.00000	259.66667	277.24592	
9	5	242.08741	260.00000	259.66667	277.24592	
10	5	242.08741	255.00000	259.66667	277.24592	
11	5	242.08741	268.00000	259.66667	277.24592	
12	5	242.08741	253.00000	259.66667	277.24592	
13	5	242.08741	273.00000	259.66667	277.24592	
14	5	242.08741	275.00000	259.66667	277.24592	5
15	5	242.08741	257.00000	259.66667	277.24592	
16	5	242.08741	269.00000	259.66667	277.24592	
17	5	242.08741	249.00000	259.66667	277.24592	
18	5	242.08741	264.00000	259.66667	277.24592	
19	5	242.08741	258.00000	259.66667	277.24592	
20	5	242.08741	248.00000	259.66667	277.24592	
21	5	242.08741	248.00000	259.66667	277.24592	

Test Descriptions

Test 5   Two out of three points in a row in Zone A or beyond

Example 42.2. Estimating the Process Standard Deviation

See SHWXEX2 in the SAS/QC Sample Library

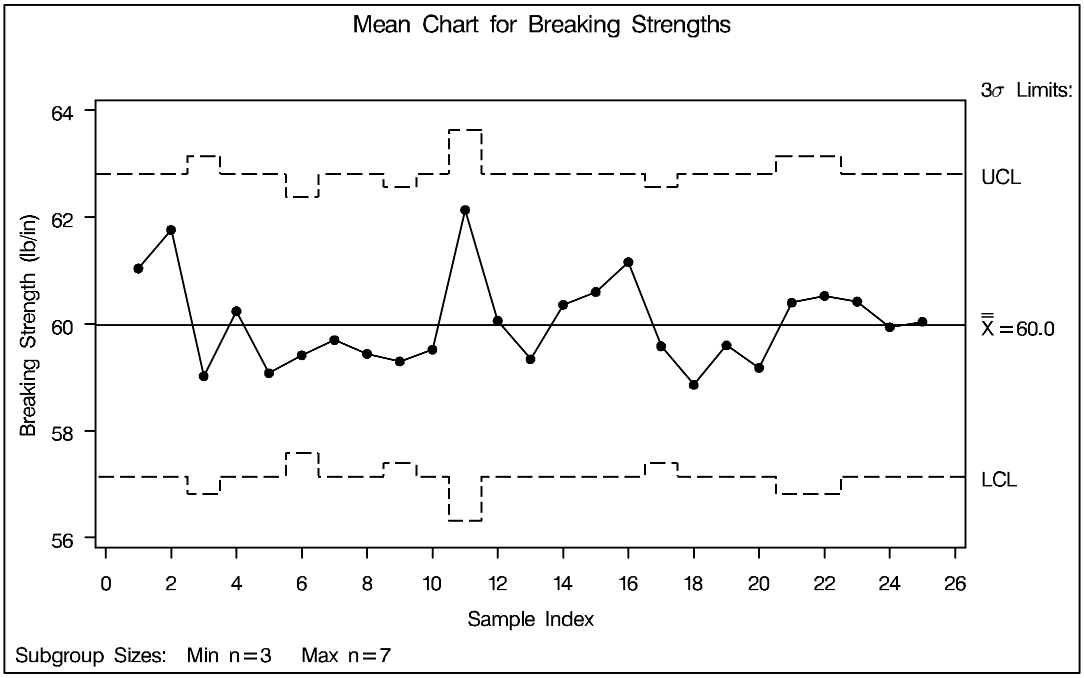
The following data set (WIRE) contains breaking strength measurements recorded in pounds per inch for 25 samples from a metal wire manufacturing process. The subgroup sample sizes vary between 3 and 7.

```
data wire;
   input sample size @;
   do i=1 to size;
      input brstr @@;
      output;
      end;
   drop i size;
   label brstr  ='Breaking Strength (lb/in)'
         sample ='Sample Index';
   datalines;
 1  5 60.6 62.3 62.0 60.4 59.9
 2  5 61.9 62.1 60.6 58.9 65.3
 3  4 57.8 60.5 60.1 57.7
 4  5 56.8 62.5 60.1 62.9 58.9
 5  5 63.0 60.7 57.2 61.0 53.5
 6  7 58.7 60.1 59.7 60.1 59.1 57.3 60.9
 7  5 59.3 61.7 59.1 58.1 60.3
 8  5 61.3 58.5 57.8 61.0 58.6
 9  6 59.5 58.3 57.5 59.4 61.5 59.6
10  5 61.7 60.7 57.2 56.5 61.5
11  3 63.9 61.6 60.9
12  5 58.7 61.4 62.4 57.3 60.5
13  5 56.8 58.5 55.7 63.0 62.7
14  5 62.1 60.6 62.1 58.7 58.3
15  5 59.1 60.4 60.4 59.0 64.1
16  5 59.9 58.8 59.2 63.0 64.9
17  6 58.8 62.4 59.4 57.1 61.2 58.6
18  5 60.3 58.7 60.5 58.6 56.2
19  5 59.2 59.8 59.7 59.3 60.0
20  5 62.3 56.0 57.0 61.8 58.8
21  4 60.5 62.0 61.4 57.7
22  4 59.3 62.4 60.4 60.0
23  5 62.4 61.3 60.5 57.7 60.2
24  5 61.2 55.5 60.2 60.4 62.4
25  5 59.0 66.1 57.7 58.5 58.9
;
```

The following statements request an $\bar{X}$ chart, shown in Output 42.2.1, for the breaking strength measurements:

```
title 'Mean Chart for Breaking Strengths';
symbol v=dot;
proc shewhart data=wire;
   xchart brstr*sample;
run;
```

Output 42.2.1. $\bar{X}$ Chart with Varying Subgroup Sample Sizes



Note that the control limits vary with the subgroup sample size. The sample size legend in the lower left corner displays the minimum and maximum subgroup sample sizes.

By default, the control limits are 3σ limits estimated from the data. You can use the STDDEVIATIONS option and the SMETHOD= option to specify how the estimate of the process standard deviation σ is to be computed, as illustrated by the following statements:

```
title 'Estimates of the Process Standard Deviation';
proc shewhart data=wire;
   xchart brstr*sample / outlimits=wirelim1
                         outindex ='Default-Ranges';
   xchart brstr*sample / outlimits=wirelim2
                         stddeviations
                         outindex ='Default-Stds';
   xchart brstr*sample / outlimits=wirelim3
                         smethod  =mvlue
                         outindex ='MVLUE  -Ranges';
   xchart brstr*sample / outlimits=wirelim4
                         stddeviations
                         smethod  =mvlue
                         outindex ='MVLUE  -Stds';
   xchart brstr*sample / outlimits=wirelim5
                         stddeviations
                         smethod  =rmsdf
                         outindex ='RMSDF  -Stds';
run;
```

The STDDEVIATIONS option specifies that the estimate is to be calculated from subgroup standard deviations rather than subgroup ranges, the default. The SMETHOD= option specifies the method for estimating σ. The default method estimates σ as an unweighted average of subgroup estimates of σ. Specifying SMETHOD=MVLUE requests a minimum variance linear unbiased estimate, and specifying SMETHOD=RMSDF requests a weighted root-mean-square estimate. For details, see "Methods for Estimating the Standard Deviation" on page 1488.

The variable _STDDEV_ in each OUTLIMITS= data set contains the estimate of σ. The OUTINDEX= option specifies the value of the variable _INDEX_ in the OUTLIMITS= data set and is used here to identify the estimation method.

The following statements merge the five OUTLIMITS= data sets into a single data set, which is listed in Output 42.2.2:

```
data wlimits;
   set wirelim1 wirelim2 wirelim3 wirelim4 wirelim5;
   keep _index_ _stddev_;
run;
```

Output 42.2.2. The Data Set WLIMITS

```
            Estimates of the Process Standard Deviation

                      _INDEX_        _STDDEV_

                   Default-Ranges     2.11146
                   Default-Stds       2.15453
                   MVLUE  -Ranges     2.11240
                   MVLUE  -Stds       2.14790
                   RMSDF  -Stds       2.17479
```

The $\bar{X}$ chart shown in Output 42.2.1 uses the default estimate listed first in Output 42.2.2 ($\sigma = 2.11146$). In this case, there is very little difference in the five estimates, since the sample sizes do not differ greatly. In general, the MVLUE's are recommended with large sample sizes ($n_i \geq 10$).

Example 42.3. Plotting OC Curves for Mean Charts

See SHWOC1 in the SAS/QC Sample Library

This example uses the GPLOT procedure and the DATA step function PROBNORM to plot operating characteristic (OC) curves for $\bar{X}$ charts with 3σ limits. An OC curve is plotted for each of the subgroup samples sizes 1, 2, 3, 4, and 16. Refer to page 226 in Montgomery (1996). Each curve plots the probability β of not detecting a shift of magnitude $\nu\sigma$ in the process mean as a function of ν. The value of β is computed using the following formula:

$$\begin{aligned} \beta &= P\{LCL \leq \bar{X}_i \leq UCL\} \\ &= \Phi(3 - \nu\sqrt{n}) - \Phi(-3 - \nu\sqrt{n}). \end{aligned}$$

The following statements compute β (the variable BETA) as a function of ν (the variable NU). The variable NSAMPLE contains the sample size.

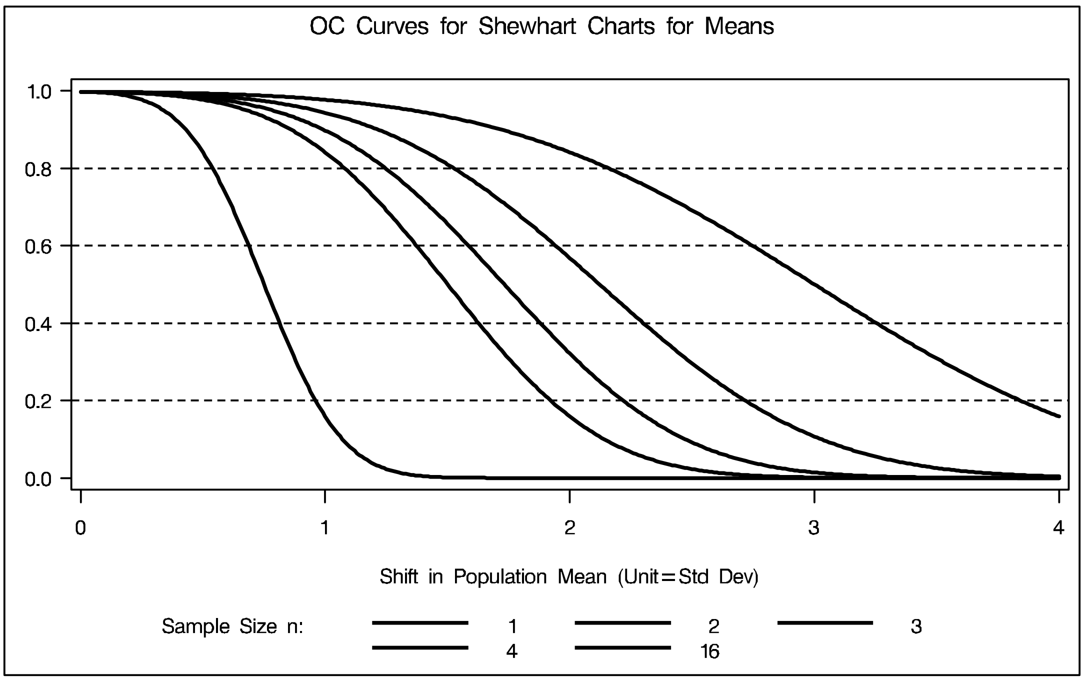
```
data oc;
   keep beta nsample nu;
   do nsample=1, 2, 3, 4, 16;
      do j=0 to 400;
         nu=j/100;
         beta=probnorm( 3-nu*sqrt(nsample)) -
              probnorm(-3-nu*sqrt(nsample));
         output;
         end;
      end;
   label nu  ='Shift in Population Mean (Unit=Std Dev)'
         beta='Probability of Not Detecting Shift';
run;
```

The following statements use the GPLOT procedure to display the OC curves shown in Output 42.3.1:

```
symbol1 v=none i=join l=1  w=2;
symbol2 v=none i=join l=2  w=2;
symbol3 v=none i=join l=8  w=2;
symbol4 v=none i=join l=15 w=2;
symbol5 v=none i=join l=20 w=2;
title1 'OC Curves for Shewhart Charts for Means';
proc gplot data=oc;
   plot beta*nu=nsample /
      frame
      legend=legend1
      vaxis=axis1
      haxis=axis2
      autovref
      autohref
      lvref = 2
      lhref = 2
      vzero
      hzero;
```

```
      axis1 label =(r=0 a=90)
            value =(t=1 ' ')
            order =(0.0 0.2 0.4 0.6 0.8 1.0)
            minor =none
            offset=(0,0);
      axis2 order =(0 1 2 3 4)
            offset=(0,0)
            minor =(n=3);
      legend1 label=('Sample Size n:');
   run;
```

Output 42.3.1. OC Curves for Different Subgroup Sample Sizes



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

ISBN 1–58025–493–4

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

1st printing, October 1999

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