Chapter 40
SCHART Statement

Chapter Table of Contents

Chapter 40
SCHART Statement

Overview

The SCHART statement creates an s chart for subgroup standard deviations, which is used to analyze the variability of a process.[*]

You can use options in the SCHART statement to

- compute control limits from the data based on a multiple of the standard error of the plotted standard deviations or as probability limits

- tabulate subgroup sample sizes, subgroup standard deviations, control limits, and other information

- save control limits in an output data set

- save subgroup sample sizes, subgroup means, and subgroup standard deviations in an output data set

- read preestablished control limits from a data set

- specify a method for estimating the process standard deviation

- specify a known (standard) process standard deviation for computing control limits

- 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

[*]You can also use R charts for this purpose; see Chapter 39, "RCHART Statement." In general, s charts are recommended with large subgroup sample sizes ($n_i \geq 10$).

Getting Started

This section introduces the SCHART statement with simple examples that illustrate commonly used options. Complete syntax for the SCHART statement is presented in the "Syntax" section on page 1393, and advanced examples are given in the "Examples" section on page 1414.

Creating Standard Deviation Charts from Raw Data

See SHWSCHR
in the SAS/QC
Sample Library

A petroleum company uses a turbine to heat water into steam, which is then pumped into the ground to make oil more viscous and easier to extract. This heating process occurs 20 times daily, and the amount of power (in kilowatts) used to heat the water to the desired temperature is recorded. The following statements create a SAS data set named TURBINE, which contains the power output measurements for 20 days:

```
data turbine;
   informat day date7.;
   format day date5.;
   input day @;
   do i=1 to 10;
      input kwatts @;
      output;
      end;
   drop i;
   datalines;
04JUL94 3196 3507 4050 3215 3583 3617 3789 3180 3505 3454
04JUL94 3417 3199 3613 3384 3475 3316 3556 3607 3364 3721
05JUL94 3390 3562 3413 3193 3635 3179 3348 3199 3413 3562
05JUL94 3428 3320 3745 3426 3849 3256 3841 3575 3752 3347
 .
 .
 .
23JUL94 3421 3787 3454 3699 3307 3917 3292 3310 3283 3536
23JUL94 3756 3145 3571 3331 3725 3605 3547 3421 3257 3574
;
```

A partial listing of TURBINE is shown in Figure 40.1.

```
            Kilowatt Power Output Data

          Obs      day     kwatts

           1     04JUL      3196
           2     04JUL      3507
           3     04JUL      4050
           4     04JUL      3215
           .       .          .
           .       .          .
           .       .          .
          400    23JUL      3574
```

Figure 40.1.   Partial Listing of the Data Set TURBINE

1384

The data set TURBINE is said to be in "strung-out" form, since each observation contains the day and power output for a single heating. The first 20 observations contain the power outputs for the first day, the second 20 observations contain the power outputs for the second day, and so on. Because the variable DAY classifies the observations into rational subgroups, it is referred to as the *subgroup-variable*. The variable KWATTS contains the power output measurements and is referred to as the *process variable* (or *process* for short).

You can use an s chart to determine whether the variability in the heating process is in control. The following statements create the s chart shown in Figure 40.2:

```
title 'Chart for Standard Deviations of Power Output';
symbol v=dot;
proc shewhart data=turbine;
   schart kwatts*day;
run;
```

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

The input data set is specified with the DATA= option in the PROC SHEWHART statement.
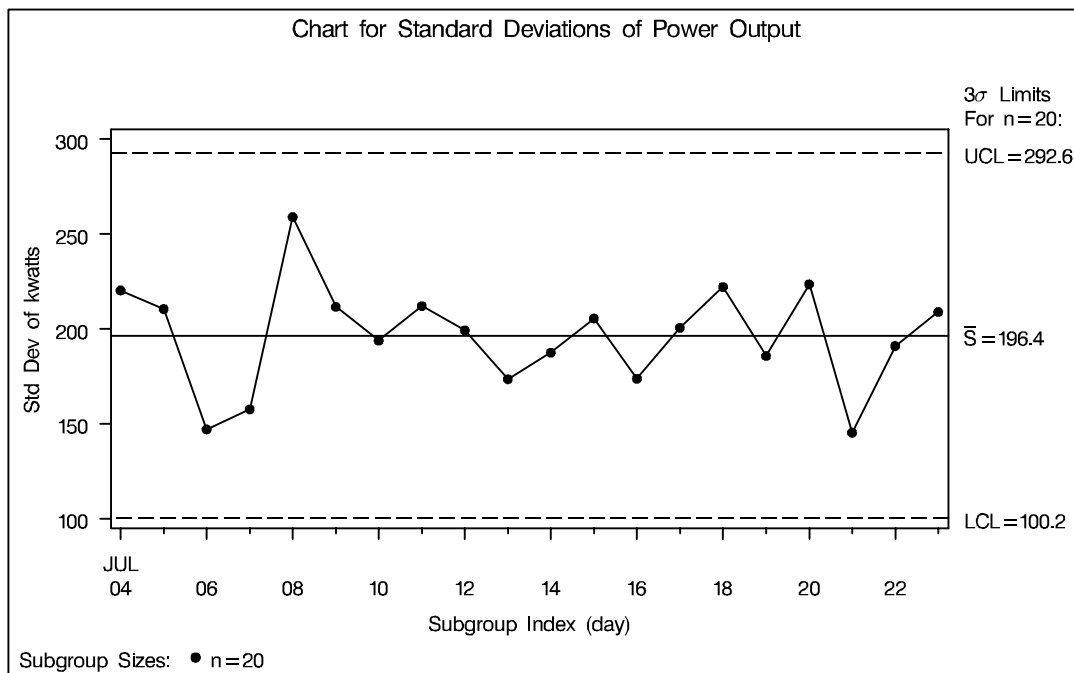
Figure 40.2. s Chart for Power Output Data

Each point on the chart represents the standard deviation of the measurements for a particular day. For instance, the standard deviation plotted for the first day is

$$\sqrt{\frac{(3196 - 3487.4)^2 + (3507 - 3487.4)^2 + \cdots + (3721 - 3487.4)^2}{19}} = 220.26$$

1385

Since all of the subgroup standard deviations lie within the control limits, you can conclude that the variability 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 40.21 on page 1404. You can also read control limits from an input data set; see "Reading Preestablished Control Limits" on page 1391.

For computational details, see "Constructing Charts for Standard Deviations" on page 1403. For more details on reading raw data, see "DATA= Data Set" on page 1408.

Creating Standard Deviation Charts from Subgroup Summary Data

See SHWSCHR in the SAS/QC Sample Library

The previous example illustrates how you can create s 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 SCHART statement with data of this type.

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

```
data oilsum;
   input day kwattsx kwattss kwattsn;
   informat day date7. ;
   format day date5. ;
   label day   ='Date of Measurement';
   datalines;
04JUL94 3487.40 220.260 20
05JUL94 3471.65 210.427 20
06JUL94 3488.30 147.025 20
07JUL94 3434.20 157.637 20
08JUL94 3475.80 258.949 20
09JUL94 3518.10 211.566 20
10JUL94 3492.65 193.779 20
11JUL94 3496.40 212.024 20
12JUL94 3398.50 199.201 20
13JUL94 3456.05 173.455 20
14JUL94 3493.60 187.465 20
15JUL94 3563.30 205.472 20
16JUL94 3519.05 173.676 20
17JUL94 3474.20 200.576 20
18JUL94 3443.60 222.084 20
19JUL94 3586.35 185.724 20
20JUL94 3486.45 223.474 20
21JUL94 3492.90 145.267 20
22JUL94 3432.80 190.994 20
23JUL94 3496.90 208.858 20
;
```

A partial listing of OILSUM is shown in Figure 40.3. There is exactly one observation for each subgroup (note that the subgroups are still indexed by DAY). The variable KWATTSX contains the subgroup means, the variable KWATTSS contains the subgroup standard deviations, and the variable KWATTSN contains the subgroup sample sizes (these are all 20).

```
           Summary Data Set for Power Outputs

         day     kwattsx     kwattss     kwattsn

        04JUL     3487.40     220.260        20
        05JUL     3471.65     210.427        20
        06JUL     3488.30     147.025        20
          .          .           .           .
          .          .           .           .
          .          .           .           .
```

Figure 40.3. The Summary Data Set OILSUM

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

```
title 'Chart for Standard Deviations of Power Output';
proc shewhart history=oilsum lineprinter;
   schart kwatts*day='*';
run;
```
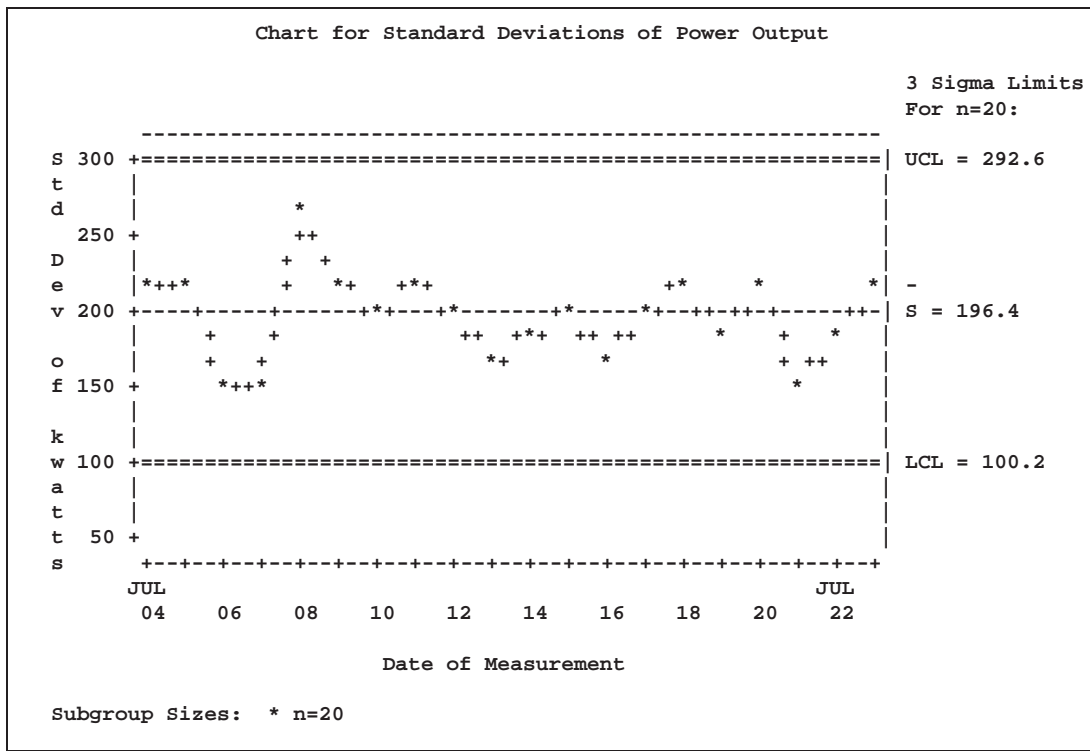
The resulting s chart is shown in Figure 40.4. Since the LINEPRINTER option is specified in the PROC SHEWHART statement, line printer output is produced. The asterisk (*) specified in single quotes after the *subgroup-variable* indicates the character used to plot points. This character must follow an equal sign.

Note that KWATTS is *not* the name of a SAS variable in the data set OILSUM but is, instead, the common prefix for the names of the SAS variables KWATTSS and KWATTSN. The suffix characters S and N indicate *standard deviation* and *sample size*, respectively. Thus, you can specify two subgroup summary variables in the HISTORY= data set with a single name (KWATTS), which is referred to as the *process*. The name DAY, specified after the asterisk, is the name of the *subgroup-variable*.

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

- subgroup variable
- subgroup standard deviation variable
- subgroup sample size variable

Furthermore, the names of the subgroup standard deviation and sample size variables must begin with the *process* name specified in the SCHART statement and end with the special suffix characters 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).

```
                Chart for Standard Deviations of Power Output

                                                              3 Sigma Limits
                                                              For n=20:
            -----------------------------------------------------------
  S 300  +==========================================================|  UCL = 292.6
  t      |                                                          |
  d      |                  *                                       |
     250 +                 ++                                       |
  D      |                +   +                                     |
  e      |*++*           +    *+    +*+                 +*      *         *| -
  v 200  +----+-----+-------+*+---+*-------+*-----*+--++-++-+-----++-|  S = 196.4
         |       +     +               ++   +*+   ++ ++        *      +     *  |
  o      |     +   +                    *+        *             + ++  |
  f 150  +       *++*                                            *    |
         |                                                          |
  k      |                                                          |
  w 100  +==========================================================|  LCL = 100.2
  a      |                                                          |
  t      |                                                          |
  t  50  +                                                          |
  s      +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
         JUL                                         JUL
          04    06    08    10    12    14    16    18    20    22

                          Date of Measurement

  Subgroup Sizes:   * n=20
```

Figure 40.4. s Chart for Power Output Data

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 1409.

Saving Summary Statistics

See SHWSCHR in the SAS/QC Sample Library

In this example, the SCHART 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 TURBINE and create a summary data set named TURBHIST:

```
title 'Summary Data Set for Power Output';
proc shewhart data=turbine;
   schart kwatts*day / outhistory = turbhist
                       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

1388

40.2. Options such as OUTHISTORY= and NOCHART are specified after the slash
(/) in the SCHART statement. A complete list of options is presented in the "Syntax"
section on page 1393.

Figure 40.5 contains a partial listing of TURBHIST.

```
            Summary Data Set for Power Output

         day    kwattsX    kwattsS    kwattsN

        04JUL    3487.40    220.260       20
        05JUL    3471.65    210.427       20
        06JUL    3488.30    147.025       20
        07JUL    3434.20    157.637       20
        08JUL    3475.80    258.949       20
          .         .          .          .
          .         .          .          .
          .         .          .          .
```

Figure 40.5.   The Summary Data Set TURBHIST

There are four variables in the data set TURBHIST.

- DAY contains the subgroup index.
- KWATTSX contains the subgroup means.
- KWATTSS contains the subgroup standard deviations.
- KWATTSN contains the subgroup sample sizes.

The subgroup mean variable is included even though it is not required by the
SCHART statement. This allows the data set to be used as a HISTORY= data set
with the BOXCHART, XCHART, and XSCHART statements, as well as with the
SCHART statement. Note that the summary statistic variables are named by adding
the suffix characters *X*, *S*, and *N* to the *process* KWATTS specified in the SCHART
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 1406.

Saving Control Limits

You can save the control limits for an *s* chart in a SAS data set; this enables you to
apply the control limits to future data (see "Reading Preestablished Control Limits"
on page 1391) or modify the limits with a DATA step program.

> See SHWSCHR
> in the SAS/QC
> Sample Library

The following statements read measurements from the data set TURBINE (see
page 1384) and save the control limits displayed in Figure 40.2 in a data set named
TURBLIM:

```
   title 'Control Limits for Power Output Data';
   proc shewhart data=turbine;
      schart kwatts*day / outlimits=turblim
                          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 TURBLIM is listed in Figure 40.6.

```
                    Control Limits for Power Output Data

_VAR_      _SUBGRP_     _TYPE_      _LIMITN_       _ALPHA_     _SIGMAS_      _LCLX_

kwatts      day       ESTIMATE       20       .002792725        3         3351.92


 _MEAN_      _UCLX_      _LCLS_        _S_        _UCLS_      _STDDEV_

3485.41     3618.90     100.207      196.396     292.584     198.996
```

Figure 40.6. The Data Set TURBLIM Containing Control Limit Information

The data set TURBLIM contains one observation with the limits for *process* KWATTS. The variables _LCLS_ and _UCLS_ contain the lower and upper control limits, and the variable _S_ 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 _LCLX_ and _UCLX_, which contain the lower and upper control limits for subgroup means, are included so that the data set TURBLIM can be used to create an $\bar{X}$ chart (see Chapter 44, "XSCHART Statement"). For more information, see "OUTLIMITS= Data Set" on page 1404.

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=turbine;
   schart kwatts*day / outtable=turbtab
                       nochart;
run;
```

The data set TURBTAB is listed in Figure 40.7.

This data set contains one observation for each subgroup sample. The variables _SUBS_ and _SUBN_ contain the subgroup standard deviations and subgroup sample sizes. The variables _LCLS_ and _UCLS_ contain the lower and upper control limits, and the variable _S_ 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 1406.

```
              Summary Statistics and Control Limit Information

                    _     _
                    S     L
                    I     I                                             _
                    G     M     _     _     _                     _     E
      _             M     I     S     L     S                     U     X
      V             A     T     U     C     U                     C     L
      A        d    S     N     B     L     B           _         L     I
      R        a    S     N     N     S     S           S         S     M
      _        y    _     _     _     _     _           _         _     _

   kwatts    04JUL   3    20    20   100.207  220.260  196.396  292.584
   kwatts    05JUL   3    20    20   100.207  210.427  196.396  292.584
   kwatts    06JUL   3    20    20   100.207  147.025  196.396  292.584
   kwatts    07JUL   3    20    20   100.207  157.637  196.396  292.584
   kwatts    08JUL   3    20    20   100.207  258.949  196.396  292.584
   kwatts    09JUL   3    20    20   100.207  211.566  196.396  292.584
   kwatts    10JUL   3    20    20   100.207  193.779  196.396  292.584
   kwatts    11JUL   3    20    20   100.207  212.024  196.396  292.584
   kwatts    12JUL   3    20    20   100.207  199.201  196.396  292.584
   kwatts    13JUL   3    20    20   100.207  173.455  196.396  292.584
   kwatts    14JUL   3    20    20   100.207  187.465  196.396  292.584
   kwatts    15JUL   3    20    20   100.207  205.472  196.396  292.584
   kwatts    16JUL   3    20    20   100.207  173.676  196.396  292.584
   kwatts    17JUL   3    20    20   100.207  200.576  196.396  292.584
   kwatts    18JUL   3    20    20   100.207  222.084  196.396  292.584
   kwatts    19JUL   3    20    20   100.207  185.724  196.396  292.584
   kwatts    20JUL   3    20    20   100.207  223.474  196.396  292.584
   kwatts    21JUL   3    20    20   100.207  145.267  196.396  292.584
   kwatts    22JUL   3    20    20   100.207  190.994  196.396  292.584
   kwatts    23JUL   3    20    20   100.207  208.858  196.396  292.584
```

Figure 40.7. The OUTTABLE= Data Set TURBTAB

An OUTTABLE= data set can be read later as a TABLE= data set. For example, the following statements read TURBTAB and display an s chart (not shown here) identical to the chart in Figure 40.2:

```
title 'Chart for Standard Deviations of Power Output';
symbol v=dot;
proc shewhart table=turbtab;
   schart kwatts*day;
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 1410.

Reading Preestablished Control Limits

In the previous example, the OUTLIMITS= data set TURBLIM saved control limits computed from the measurements in TURBINE. This example shows how these limits can be applied to new data.

See SHWSCHR
in the SAS/QC
Sample Library

The following statements create an s chart for new measurements in the data set TURBINE2 (not listed here) using the control limits in TURBLIM:

1391

```
title 'Chart for Standard Deviations of Power Output';
symbol v=dot;
proc shewhart data=turbine2 limits=turblim;
    schart kwatts*day;
run;
```

The chart is shown in Figure 40.8. 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 KWATTS
- the value of _SUBGRP_ matches the *subgroup-variable* name DAY
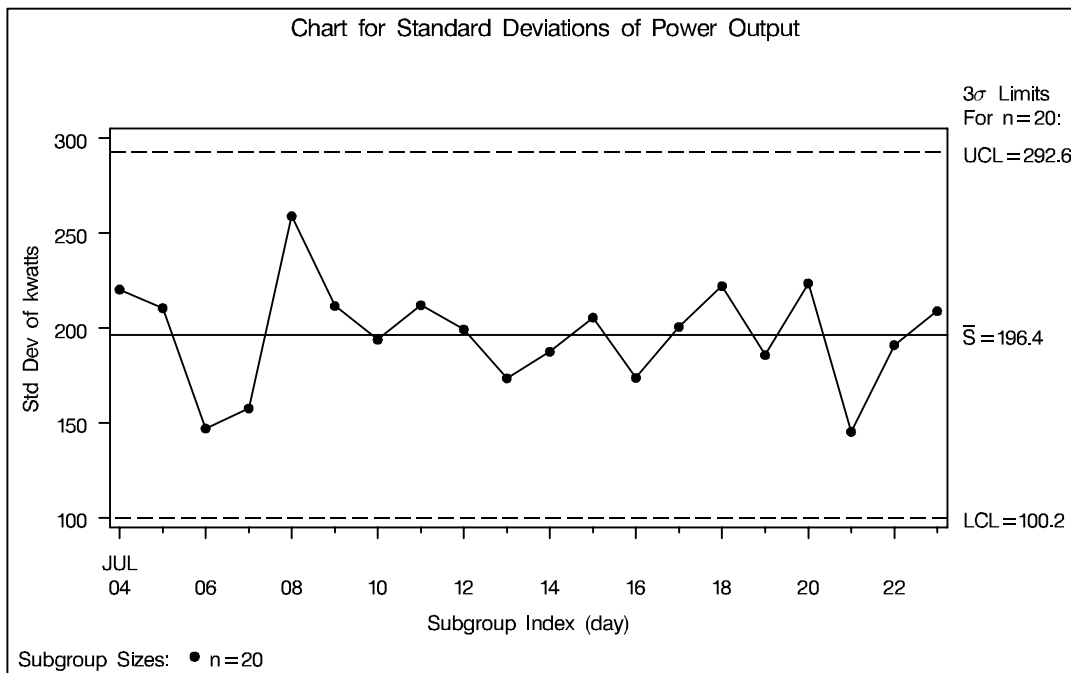
Figure 40.8. s Chart for Second Set of Power Output Data

All the standard deviations lie within the control limits, indicating that the variability of the heating process is still in statistical 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 1409 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 SCHART statement is as follows:

SCHART *process*******subgroup-variable* ;

The general form of this syntax is as follows:

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

You can use any number of SCHART statements in the SHEWHART procedure. The components of the SCHART 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 Standard Deviation Charts from Raw Data" on page 1384.

- 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 Standard Deviation Charts from Subgroup Summary Data" on page 1386.

- 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 1389.

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

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

subgroup-variable
    is the variable that identifies subgroups in the data. The *subgroup-variable* is required. In the preceding SCHART 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.

1393

symbol-variable
> is an optional variable whose levels (unique values) determine the symbol marker or character used to plot the subgroup standard deviations.
>
> - 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 an *s* chart using an asterisk (*) to plot the points:
>
> ```
> proc shewhart data=values;
>     schart 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 SCHART statement options by function. For complete descriptions, see Chapter 46, "Dictionary of Options."

Table 40.1. Tabulation Options

TABLE	creates a basic table of subgroup standard deviations, 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 40.2. Options for Specifying Tests for Special Causes

TESTS2=*value-list* | *customized-pattern-list*	specifies tests for special causes for the *s* chart
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
ZONE2LABELS	adds labels A, B, and C to zone lines for *s* chart
ZONE2VALUES	labels zone lines with their values
ZONES2	adds lines to *s* chart delineating zones A, B, and C
ZONEVALPOS=*n*	specifies position of ZONE2VALUES labels

Table 40.3. 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 40.4. 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 40.5. 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 40.6. Reference Line Options

CHREF=*color*	specifies color for lines requested by HREF= option
CVREF=*color*	specifies color for lines requested by VREF= option
HREF=*values*│ *SAS-data-set*	specifies position of reference lines perpendicular to horizontal axis on s chart
HREFCHAR='*character*'	specifies line character for HREF= lines
HREFLABELS= '*label1*'...'*labeln*'	specifies labels for HREF= lines
HREFLABPOS=*n*	specifies position of HREFLABELS= labels
LHREF=*linetype*	specifies line type for HREF= lines
LVREF=*linetype*	specifies line type for VREF= lines
NOBYREF	specifies that reference line information in a data set is to be applied uniformly to charts created for all BY groups
VREF=*values*│ *SAS-data-set*	specifies position of reference lines perpendicular to vertical axis on s chart
VREFCHAR='*character*'	specifies line character for VREF= lines
VREFLABELS= '*label1*'...'*labeln*'	specifies labels for VREF= lines
VREFLABPOS=*n*	specifies position of VREFLABELS= labels

Table 40.7. 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 40.8. 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*|AXIS*n*	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
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
NOTRUNC	suppresses vertical axis truncation at zero applied by default to *s* chart
NOVANGLE	requests vertical axis labels that are strung out vertically
SKIPHLABELS=*n*	specifies thinning factor for tick mark labels on horizontal axis
TURNHLABELS	requests horizontal axis labels that are strung out vertically
VAXIS=*values*|AXIS*n*	specifies major tick mark values for vertical axis on *s* 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 40.9. Plot Layout Options

ALLN	plots summary statistics for all subgroups
BILEVEL	creates control charts using half-screens and half-pages
EXCHART	creates control charts for a process variable only when exceptions occur
INTERVAL=*keyword*	specifies natural time interval between consecutive subgroup positions when time, date, or datetime format is associated with a numeric subgroup variable
MAXPANELS=*n*	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 s 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
ZEROSTD	displays s chart regardless of whether $\hat{\sigma} = 0$

Table 40.10. Options for Specifying Control Limits

ALPHA=*value*	requests probability limits for control charts
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 from a LIMITS= data set (Release 6.10 and later releases)
READALPHA	reads the variable _ALPHA_ instead of the variable _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 statistic

Table 40.11. 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 on s chart
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 on s chart
NOCTL	suppresses display of central line on s chart
NOLCL	suppresses display of lower control limit on s chart
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 zero lower control limit on s chart
NOUCL	suppresses display of upper control limit on s chart
SSYMBOL='*string*' | *keyword*	specifies label for central line on s chart
UCLLABEL='*string*'	specifies label for upper control limit on s chart
WLIMITS=*n*	specifies width for control limits and central line

Table 40.12. 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 40.13. Options for Plotting and Labeling Points

ALLLABEL=VALUE| (*variable*)	labels every point on *s* 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
OUTLABEL=VALUE| (*variable*)	labels points outside control limits on *s* chart
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*

Table 40.14. Input Data Set Options

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

Table 40.15. 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
WEBOUT= *SAS-data-set*	creates OUTTABLE= data set with additional graph coordinate data

Table 40.16. 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
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 40.17. Specification Limit Options

CIALPHA=*value*	specifies α value for computing capability index confidence limits
CITYPE=*keyword*	specifies capability index confidence limits type
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 40.18. Process Mean and Standard Deviation Options

SIGMA0=*value*	specifies known value σ_0 for process standard deviation σ
SMETHOD=*keyword*	specifies method for estimating process standard deviation σ
TYPE=*keyword*	identifies whether parameters are estimates or standard values and specifies value of _TYPE_ in the OUTLIMITS= data set

Table 40.19. Graphical Enhancement Options

ANNOTATE= *SAS-data-set*	specifies annotate data set that adds features to *s* chart
DESCRIPTION='*string*'	specifies string that appears in the description field of PROC GREPLAY master menu for *s* chart
FONT=*font*	specifies software font for labels and legends on charts
HTML=(*variable*)	specifies a variable whose values are URLs to be associated with subgroups
NAME='*string*'	specifies name that appears in the name field of the PROC GREPLAY master menu for *s* 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

Table 40.20. 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 *s* chart
WSTARCIRCLES=*n*	specifies width of STARCIRCLES= circles
WSTARS=*n*	specifies width of STARVERTICES= stars

Details

Constructing Charts for Standard Deviations

The following notation is used in this section:

σ	process standard deviation (standard deviation of the population of measurements)
s_i	standard deviation of measurements in i^{th} subgroup $$s_i = \sqrt{(1/(n_i - 1))((x_{i1} - \bar{X}_i)^2 + \cdots + (x_{in_i} - \bar{X}_i)^2)}$$
n_i	sample size of ith subgroup
$c_4(n)$	expected value of the standard deviation of n independent normally distributed variables with unit standard deviation
$c_5(n)$	standard error of the standard deviation of n independent observations from a normal population with unit standard deviation
$\chi^2_p(n)$	$100p^{\text{th}}$ percentile $(0 < p < 1)$ of the χ^2 distribution with n degrees of freedom

Plotted Points

Each point on an s chart indicates the value of a subgroup standard deviation (s_i). For example, if the tenth subgroup contains the values 12, 15, 19, 16, and 13, the value plotted for this subgroup is

$$s_{10} = \sqrt{((12 - 15)^2 + (15 - 15)^2 + (19 - 15)^2 + (16 - 15)^2 + (13 - 15)^2)/4} = 2.739$$

Central Line

By default, the central line for the ith subgroup indicates an estimate for the expected value of s_i, which is computed as $c_4(n_i)\hat{\sigma}$, where $\hat{\sigma}$ is an estimate of σ. If you specify a known value (σ_0) for σ, the central line indicates the value of $c_4(n_i)\sigma_0$. Note that the central line varies with n_i.

Control Limits

You can compute the limits in the following ways:

- as a specified multiple (k) of the standard error of s_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 s_i exceeds the limits

The following table provides the formulas for the limits:

Table 40.21. Limits for s Charts

Control Limits
LCL = lower limit = $\max(c_4(n_i)\hat{\sigma} - kc_5(n_i)\hat{\sigma}, 0)$ UCL = upper limit = $c_4(n_i)\hat{\sigma} + kc_5(n_i)\hat{\sigma}$

Probability Limits
LCL = lower limit = $\hat{\sigma}\sqrt{\chi^2_{\alpha/2}(n_i - 1)/(n_i - 1)}$
UCL = upper limit = $\hat{\sigma}\sqrt{\chi^2_{1-\alpha/2}(n_i - 1)/(n_i - 1)}$

The formulas assume that the data are normally distributed. If a standard value σ_0 is available for σ, replace $\hat{\sigma}$ with σ_0 in Table 40.21. Note that the upper and lower limits vary with n_i and that the probability limits are asymmetric around the central line.

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 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 are saved:

Table 40.22. 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
LCLS	lower control limit for subgroup standard deviation
LCLX	lower control limit for subgroup mean
LIMITN	sample size associated with the control limits
LSL	lower specification limit
MEAN	process mean ($\overline{\overline{X}}$ or μ_0)

1404

Table 40.22. (continued)

Variable	Description
_S_	value of central line on s chart
SIGMAS	multiple (k) of standard error of $\bar{X}_i$ or s_i
STDDEV	process standard deviation ($\hat{\sigma}$ or σ_0)
SUBGRP	*subgroup-variable* specified in the SCHART statement
TARGET	target value
TYPE	type (estimate or standard value) of _MEAN_ and _STDDEV_
UCLS	upper control limit for subgroup standard deviation
UCLX	upper control limit for subgroup mean
USL	upper specification limit
VAR	*process* specified in the SCHART statement

Notes:

1. The variables _LCLX_, _MEAN_, and _UCLX_ are saved to allow the OUTLIMITS= data set to be used as a LIMITS= data set with the BOXCHART, XCHART, 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_, _LCLS_, _S_, and _UCLS_.

3. If the limits are defined in terms of a multiple k of the standard error of s_i, the value of _ALPHA_ is computed as

$$F_S(\_LCLS\_/\_STDDEV\_) + 1 - F_S(\_UCLS\_/\_STDDEV\_)$$

   where $F_S(\cdot)$ is the cumulative distribution function of the standard deviation of a sample of n observations from a normal population with unit standard deviation, and n is the value of _LIMITN_. If _LIMITN_ has the special missing value V, this value is assigned to _ALPHA_.

4. If the limits are probability limits, the value of _SIGMAS_ is computed as $(\_UCLS\_ - \_S\_)/e$, where e is the standard error of the standard deviation of n observations from a normal population with unit standard deviation. If _LIMITN_ has the special missing value V, this value is assigned to _SIGMAS_.

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 SCHART statement. For an example, see "Saving Control Limits" on page 1389.

OUTHISTORY= Data Set

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

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

The subgroup mean variable is included so that the data set can be reused as a HISTORY= data set with the BOXCHART, XCHART, and XSCHART statements, as well as the SCHART statement.

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 SCHART statement. For example, consider the following statements:

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

The data set SUMMARY contains variables named LOT, WIDTHX, WIDTHS, WIDTHN, DIAMTERX, DIAMTERS, and DIAMTERN.

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 1388.

OUTTABLE= Data Set

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

Variable	Description
ALPHA	probability (α) of exceeding control limits
EXLIM	control limit exceeded on s chart
LCLS	lower control limit for standard deviation
LIMITN	nominal sample size associated with the control limits
_S_	average standard deviation
SIGMAS	multiple (k) of the standard error associated with control limits
subgroup	values of the subgroup variable
SUBN	subgroup sample size
SUBS	subgroup standard deviation
TESTS2	tests for special causes signaled on s chart
UCLS	upper control limit for standard deviation
VAR	*process* specified in the SCHART 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= option, respectively, or with the corresponding variables in a LIMITS= data set).

2. The variable _TESTS2_ is saved if you specify the TESTS2= option.

3. The variables _VAR_, _EXLIM_, and _TESTS2_ 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 1389.

ODS Tables

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

Table 40.23. ODS Tables Produced with the SCHART Statement

Table Name	Description	Options
SCHART	s 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 SCHART statement must be a SAS variable in the DATA= data set. This variable provides measurements, which must be grouped into subgroup samples indexed by the values of the *subgroup-variable*. The *subgroup-variable*, which is specified in the SCHART 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 ith 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 ith 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= 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 Standard Deviation Charts from Raw Data" on page 1384.

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;
    schart 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 40.22 on page 1404. 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 _LCLS_, _S_, and _UCLS_, which specify the control limits directly

- the variable _STDDEV_, which is used to calculate the control limits according to the equations in Table 40.21 on page 1404

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, ESTIMATE, STDMU, and STDSIGMA.

- BY variables are required if specified with a BY statement.

For an example, see "Reading Preestablished Control Limits" on page 1391.

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 SCHART statement must contain the following:

- the *subgroup-variable*
- a subgroup standard deviation variable for each *process*
- a subgroup sample size variable for each *process*

The names of the subgroup standard deviation and subgroup sample size variables must be the *process* name concatenated with the special suffix characters S and N, respectively. For example, consider the following statements:

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

The data set SUMMARY must include the variables BATCH, WEIGHTS, WEIGHTN, 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 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 READ-PHASES= option (see "Displaying Stratification in Phases" on page 1689 for an example).

For an example of a HISTORY= data set, see "Creating Standard Deviation Charts from Subgroup Summary Data" on page 1386.

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 SHE-WHART 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 SCHART statement:

Table 40.24. Variables Required in a TABLE= Data Set

Variable	Description
LCLS	lower control limit for standard deviation
LIMITN	nominal sample size associated with the control limits
_S_	average standard deviation
subgroup-variable	values of the *subgroup-variable*
SUBN	subgroup sample size
SUBS	subgroup standard deviation
UCLS	upper control limit for standard deviation

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.

- _TESTS2_ (if the TESTS2= 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 1389.

1411

Methods for Estimating the Standard Deviation

When control limits are determined from the input data, three methods (referred to as default, MVLUE, and RMSDF) are available for estimating σ.

Default Method

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 ith 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 ith 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, then 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

If you specify 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

If you specify 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.

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 standard deviation variable
Vertical	TABLE=	_SUBS_

For an example, see "Labeling Axes" on page 1719.

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 SCHART statement.

Example 40.1. Specifying a Known Standard Deviation

See SHWSEX1
in the SAS/QC
Sample Library

In some applications, a standard value σ_0 may be available for the process standard deviation σ. This example shows how you can specify σ_0 to compute the control limits.

Suppose that the amount of power needed to heat water in the heating process described on page 1384 has a known standard deviation of 200. The following statements specify this known value and create an *s* chart, shown in Output 40.1.1, for the power output measurements in the data set TURBINE:

```
title 's Chart for Power Output';
title2 'Using Known Process Standard Deviation of 200';
symbol v=dot;
proc shewhart data=turbine;
   schart kwatts*day / sigma0 =200
                       ssymbol=s0;
run;
```

The SIGMA0= option specifies σ_0, and the SSYMBOL= option specifies a label for the central line indicating that the central line is computed from σ_0. Since all the points lie within the limits, you can conclude that the variability of the process is stable.

Output 40.1.1. Reading in Standard Value for Process Standard Deviation
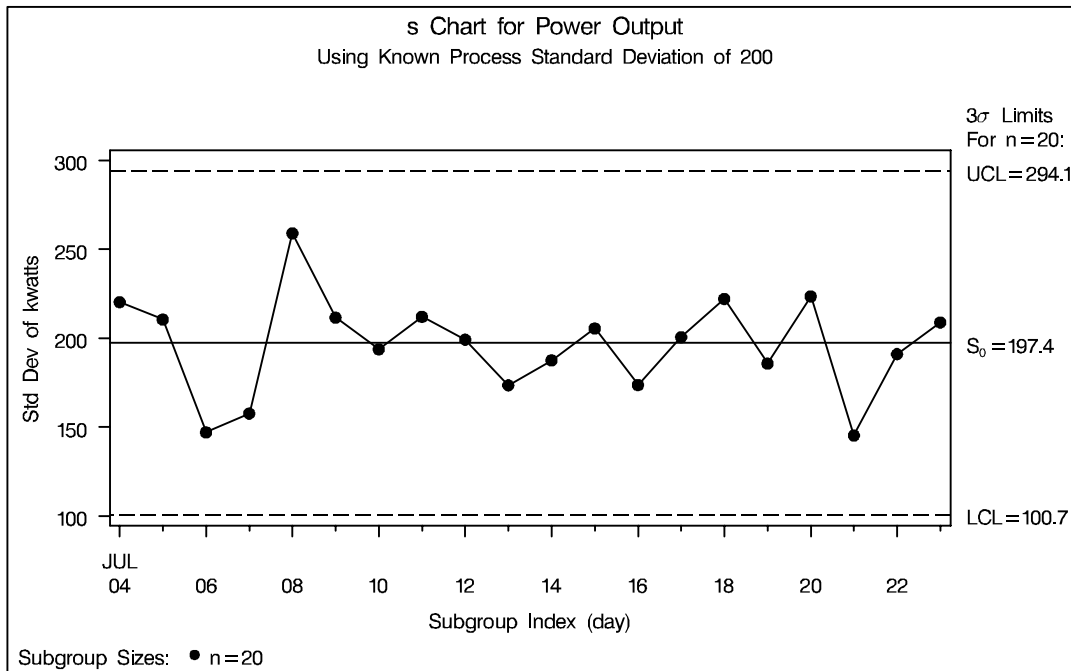
You can also specify σ_0 as the value of the variable _STDDEV_ in a LIMITS= data set, as illustrated by the following statements:*

```
data plimits;
   length _var_ _subgrp_ _type_ $8;
   _var_    = 'kwatts';
   _subgrp_ = 'day';
   _type_   = 'STDSIGMA';
   _limitn_ = 20;
   _stddev_ = 200;
run;

title 'Chart Using Known Process Standard Deviation';
symbol v=dot;
proc shewhart data=turbine limits=plimits;
   schart kwatts*day / ssymbol=s0;
run;
```

The resulting s chart (not shown here) is identical to the one shown in Output 40.1.1. For more information, see "LIMITS= Data Set" on page 1409.

Example 40.2. Computing Average Run Lengths for s Charts

This example illustrates how you can compute the average run length of an s chart. The data used here are the power measurements in the data set TURBINE, which is introduced on page 1384.

<div style="float:right; border:1px solid; padding:4px;">
See SHWSARL in the SAS/QC Sample Library
</div>

The in-control average run length of a Shewhart chart is $\text{ARL} = \frac{1}{p}$, where p is the probability that a single point exceeds its control limits. Since this probability is saved as the value of the variable _ALPHA_ in an OUTLIMITS= data set, you can compute ARL for an s chart as follows:

```
title 'Average In-Control Run Length';
proc shewhart data=turbine;
   schart kwatts*day / outlimits=turblim nochart;

data arlcomp;
   keep _var_ _sigmas_ _alpha_ arl;
   set turblim;
   arl = 1 / _alpha_;
run;
```

The data set ARLCOMP is listed in Output 40.2.1, which shows that the ARL is equal to 358.

Output 40.2.1.   The Data Set ARLCOMP

```
            Average In-Control Run Length

      _VAR_          _ALPHA_     _SIGMAS_       arl

      kwatts     .002792725         3        358.073
```

---

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

To compute out-of-control average run lengths, define f as the slippage factor for the process standard deviation σ, where $f > 1$. In other words, the "shifted" standard deviation to be detected by the chart is $f\sigma$. The following statements compute the ARL as a function of f:

```
title 'Average Run Length Analysis';
data arlshift;
   keep f f_std p arl_f;
   set turblim;
   df = _limitn_ - 1;
   do f = 1 to 1.5 by 0.05;
      f_std = f * _stddev_;
      low   = df * ( _lcls_ / f_std )**2;
      upp   = df * ( _ucls_ / f_std )**2;
      p     = probchi( low, df ) + 1 - probchi( upp, df );
      arl_f = 1 / p;
      output;
   end;
run;
```

The data set ARLSHIFT is listed in Output 40.2.2. For example, on average, 53 samples are required to detect a ten percent increase in σ (a shifted standard deviation of approximately 219). The computations use the fact that $(n_i - 1)s_i^2/\sigma^2$ has a χ^2 distribution with $n_i - 1$ degrees of freedom, assuming that the measurements are normally distributed.

Output 40.2.2. The Data Set ARLSHIFT

```
            Average Run Length Analysis

         f      f_std       p        arl_f

       1.00    198.996   0.00279    358.073
       1.05    208.945   0.00758    131.922
       1.10    218.895   0.01875     53.322
       1.15    228.845   0.03984     25.102
       1.20    238.795   0.07388     13.535
       1.25    248.745   0.12239      8.171
       1.30    258.694   0.18475      5.413
       1.35    268.644   0.25834      3.871
       1.40    278.594   0.33923      2.948
       1.45    288.544   0.42298      2.364
       1.50    298.494   0.50546      1.978
```

SAS/QC® User's Guide, Version 8

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