Chapter 35
MCHART Statement

Chapter Table of Contents

Chapter 35
MCHART Statement

Overview

The MCHART statement creates a chart for subgroup medians, which is used to monitor the central tendency of a process.

You can use options in the MCHART statement to

- compute control limits from the data based on a multiple of the standard error of the plotted medians or as probability limits
- tabulate subgroup sample sizes, subgroup medians, control limits, and other information
- save control limits in an output data set
- save subgroup sample sizes and subgroup medians 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 charts more readable
- display vertical and horizontal reference lines
- control axis values and labels
- control layout and appearance of the chart

Note: When analyzing variables data, you should examine the variability of the process as well as the mean level. You can use the MRCHART statement in the SHEWHART procedure to monitor both the mean level and variability.

Getting Started

This section introduces the MCHART statement with simple examples that illustrate commonly used options. Complete syntax for the MCHART statement is presented in the "Syntax" section on page 1195.

Creating Charts for Medians from Raw Data

See SHWMCHR in the SAS/QC Sample Library

A consumer products company weighs detergent boxes (in pounds) to determine whether the fill process is in control. The following statements create a SAS data set named DETERGNT, which contains the weights for five boxes in each of 28 lots. A lot is considered a rational subgroup.

```
data detergnt;
   input lot @;
   do i=1 to 5;
      input weight @;
      output;
   end;
drop i;
datalines;
 1 17.39 26.93 19.34 22.56 24.49
 2 23.63 23.57 23.54 20.56 22.17
 3 24.35 24.58 23.79 26.20 21.55
 4 25.52 28.02 28.44 25.07 23.39
 5 23.25 21.76 29.80 23.09 23.70
 6 23.01 22.67 24.70 20.02 26.35
 7 23.86 24.19 24.61 26.05 24.18
 8 26.00 26.82 28.03 26.27 25.85
 9 21.58 22.31 25.03 20.86 26.94
10 22.64 21.05 22.66 29.26 25.02
11 26.38 27.50 23.91 26.80 22.53
12 23.01 23.71 25.26 20.21 22.38
13 23.15 23.53 22.98 21.62 26.99
14 26.83 23.14 24.73 24.57 28.09
15 26.15 26.13 20.57 25.86 24.70
16 25.81 23.22 23.99 23.91 27.57
17 25.53 22.87 25.22 24.30 20.29
18 24.88 24.15 25.29 29.02 24.46
19 22.32 25.96 29.54 25.92 23.44
20 25.63 26.83 20.95 24.80 27.25
21 21.68 21.11 26.07 25.17 27.63
22 26.72 27.05 24.90 30.08 25.22
23 31.58 22.41 23.67 23.47 24.90
24 28.06 23.44 24.92 24.64 27.42
25 21.10 22.34 24.96 26.50 24.51
26 23.80 24.03 24.75 24.82 27.21
27 25.10 26.09 27.21 24.28 22.45
28 25.53 22.79 26.26 25.85 25.64
;
```

A partial listing of DETERGNT is shown in Figure 35.1.

The Data Set DETERGNT

lot	weight
1	17.39
1	26.93
1	19.34
1	22.56
1	24.49
2	23.63
2	23.57
2	23.54
2	20.56
2	22.17
3	24.35
3	24.58
3	23.79
3	26.20
3	21.55
.	.
.	.
.	.

Figure 35.1. Partial Listing of the Data Set DETERGNT

The data set DETERGNT is said to be in "strung-out" form, since each observation contains the lot number and weight of a single box. The first five observations contain the weights for the first lot, the second five observations contain the weights for the second lot, and so on. Because the variable LOT classifies the observations into rational subgroups, it is referred to as the *subgroup-variable*. The variable WEIGHT contains the weights and is referred to as the *process variable* (or *process* for short).

The within-subgroup variability of the weights is known to be stable. You can use a median chart to determine whether the mean level of the weights is in control. The following statements create the median chart shown in Figure 35.2:

```
title 'Median Chart for Detergent Box Weight';
symbol v=dot;
proc shewhart data=detergnt;
   mchart weight*lot;
run;
```

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

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

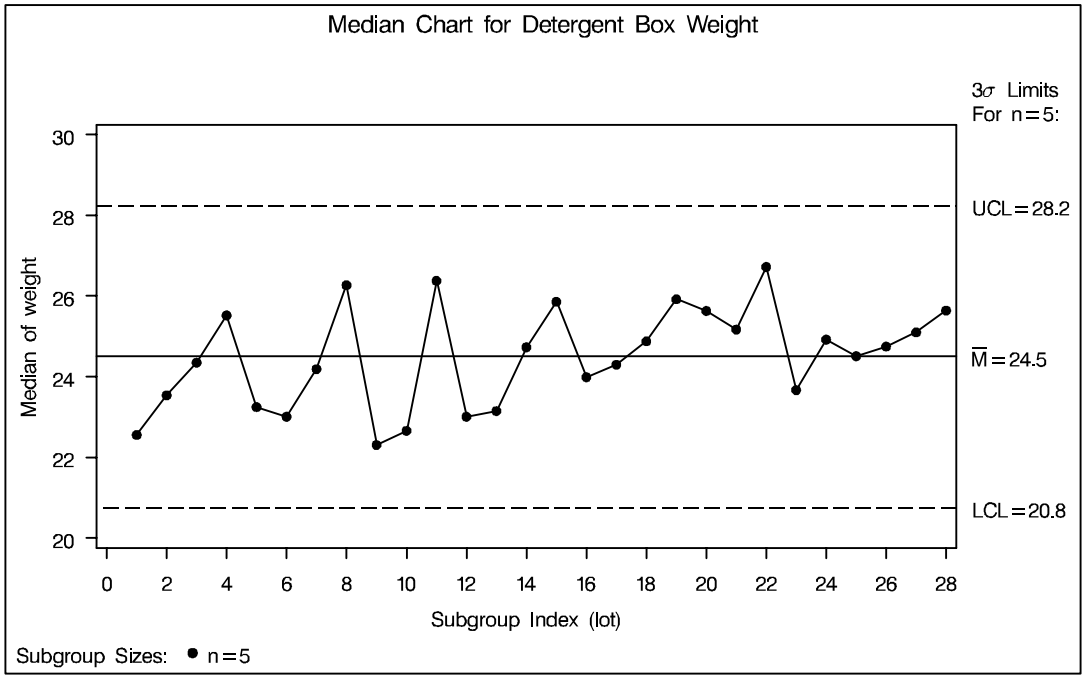

Figure 35.2. Median Chart for Detergent Box Weight Data

Each point on the chart represents the median of the weights for a particular lot. For instance, the weights for the first lot are 17.39, 19.34, 22.56, 24.49, and 26.93, and consequently, the median plotted for this lot is 22.56.

Since all of the subgroup medians lie within the control limits, you can conclude that the process is in statistical control. By default, the control limits shown are 3σ limits estimated from the data; the formulas for the limits are given in Table 35.22 on page 1207. You can also read control limits from an input data set; see "Reading Preestablished Control Limits" on page 1193.

For computational details, see "Constructing Median Charts" on page 1206. For more details on reading raw measurements, see "DATA= Data Set" on page 1211.

Creating Charts for Medians from Subgroup Summary Data

See SHWMCHR in the SAS/QC Sample Library

The previous example illustrates how you can create median 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 MCHART statement with data of this type.

The following data set (DETSUM) provides the data from the preceding example in summarized form. There is exactly one observation for each subgroup (note that the subgroups are still indexed by LOT). The variable WEIGHTM contains the subgroup medians, the variable WEIGHTR contains the subgroup ranges, and the variable WEIGHTN contains the subgroup sample sizes (these are all five).

```
data detsum;
   input lot weightm weightr;
   weightn = 5;
datalines;
 1  22.56  9.54
 2  23.54  3.07
 3  24.35  4.65
 4  25.52  5.05
 5  23.25  8.04
 6  23.01  6.33
 7  24.19  2.19
 8  26.27  2.18
 9  22.31  6.08
10  22.66  8.21
11  26.38  4.97
12  23.01  5.05
13  23.15  5.37
14  24.73  4.95
15  25.86  5.58
16  23.99  4.35
17  24.30  5.24
18  24.88  4.87
19  25.92  7.22
20  25.63  6.30
21  25.17  6.52
22  26.72  5.18
23  23.67  9.17
24  24.92  4.62
25  24.51  5.40
26  24.75  3.41
27  25.10  4.76
28  25.64  3.47
;
```

A partial listing of DETSUM is shown in Figure 35.3.

Summary Data Set for Detergent Box Weights

lot	weightm	weightr	weightn
1	22.56	9.54	5
2	23.54	3.07	5
3	24.35	4.65	5
4	25.52	5.05	5
5	23.25	8.04	5
.	.	.	.
.	.	.	.
.	.	.	.

Figure 35.3. The Summary Data Set DETSUM

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

```
title 'Median Chart for Weights';
proc shewhart history=detsum lineprinter;
   mchart weight*lot='*';
run;
```

The resulting median chart is shown in Figure 35.4. Since the LINEPRINTER option is included 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 WEIGHT is *not* the name of a SAS variable in the data set DETSUM but is, instead, the common prefix for the names of the three SAS variables WEIGHTM, WEIGHTR, and WEIGHTN. The suffix characters *M*, *R*, and *N* indicate *median*, *range*, and *sample size*, respectively. Thus, you can specify three subgroup summary variables in the HISTORY= data set with a single name (WEIGHT), which is referred to as the *process*. The name LOT specified after the asterisk is the name of the *subgroup-variable*.

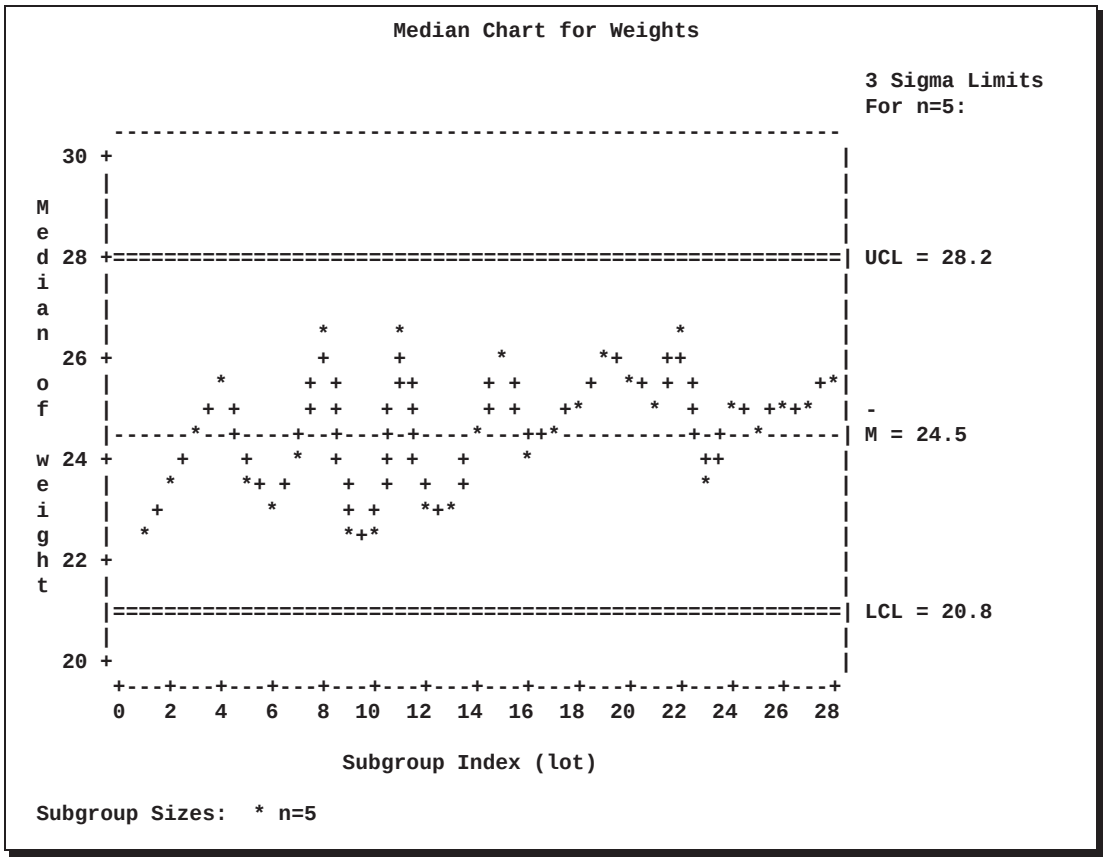

Figure 35.4. Median Chart from the Summary Data Set DETSUM

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

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

Furthermore, the names of the subgroup median, range (or standard deviation), and sample size variables must begin with the *process* name specified in the MCHART statement and end with the special suffix characters *M*, *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 1225).

If you specify the STDDEVIATIONS option in the MCHART 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 DETSUM2 must contain a subgroup standard deviation variable named WEIGHTS:

```
title 'Median Chart for Weights';
symbol v=dot;
proc shewhart history=detsum2;
   mchart weight*lot / stddeviations;
run;
```

Options such as STDDEVIATIONS are specified after the slash (/) in the MCHART statement. A complete list of options is presented in the "Syntax" section on page 1195.

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 1213.

Saving Summary Statistics

See SHWMCHR in the SAS/QC Sample Library

In this example, the MCHART 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 DETERGNT and create a summary data set named DETHIST:

```
title 'Summary Data Set DETHIST for Detergent Box Weights';
proc shewhart data=detergnt;
   mchart weight*lot / outhistory = dethist
                       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 35.2. Figure 35.5 contains a partial listing of DETHIST.

Summary Data Set DETHIST for Detergent Box Weights

lot	weightM	weightR	weightN
1	22.56	9.54	5
2	23.54	3.07	5
3	24.35	4.65	5
4	25.52	5.05	5
5	23.25	8.04	5
.	.	.	.
.	.	.	.
.	.	.	.

Figure 35.5. The Summary Data Set DETHIST

There are four variables in the data set DETHIST.

- LOT contains the subgroup index.
- WEIGHTM contains the subgroup medians.
- WEIGHTR contains the subgroup ranges.
- WEIGHTN contains the subgroup sample sizes.

Note that the summary statistic variables are named by adding the suffix characters *M*, *R*, and *N* to the *process* WEIGHT specified in the MCHART 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 instead of a subgroup range variable, as demonstrated by the following statements:

```
title 'Summary Data Set with Subgroup Standard Deviations';
proc shewhart data=detergnt;
   mchart weight*lot / outhistory = dethist2
                       stddeviations
                       nochart;
run;
```

Figure 35.6 contains a partial listing of DETHIST2.

Summary Data Set with Subgroup Standard Deviations

lot	weightM	weightS	weightN
1	22.56	3.84205	5
2	23.54	1.34050	5
3	24.35	1.68087	5
4	25.52	2.11558	5
5	23.25	3.14747	5
.	.	.	.
.	.	.	.
.	.	.	.

Figure 35.6. The Summary Data Set DETHIST2

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

For more information, see "OUTHISTORY= Data Set" on page 1209.

Saving Control Limits

See SHWMCHR in the SAS/QC Sample Library

You can save the control limits for a median chart in a SAS data set; this enables you to apply the control limits to future data (see "Reading Preestablished Control Limits" on page 1193) or modify the limits with a DATA step program.

The following statements read measurements from the data set DETERGNT (see page 1184) and save the control limits displayed in Figure 35.2 in a data set named DETLIM:

```
title 'Control Limits for Detergent Box Weights';
proc shewhart data=detergnt;
   mchart weight*lot / outlimits=detlim
                       nochart;
run;
```

The OUTLIMITS= option names the data set containing the control limits, and the NOCHART option suppresses the display of the charts. The data set DETLIM is listed in Figure 35.7.

Control Limits for Detergent Box Weights

VAR	_SUBGRP_	_TYPE_	_LIMITN_	_ALPHA_	_SIGMAS_	_LCLM_
weight	lot	ESTIMATE	5	.002909021	3	20.7554

MEAN	_UCLM_	_LCLR_	_R_	_UCLR_	_STDDEV_
24.4996	28.2439	0	5.42036	11.4613	2.33041

Figure 35.7. The Data Set DETLIM Containing Control Limit Information

The data set DETLIM contains one observation with the limits for the *process* WEIGHT. The variables _LCLM_ and _UCLM_ contain the lower and upper control limits for the medians, 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 median charts, but they are included so the data set DETLIM can be used to create an R chart; see Chapter 36, "MRCHART Statement," and Chapter 39, "RCHART Statement." If you specify the STDDEVIATIONS option in the MCHART 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 40, "SCHART Statement." For more information, see "OUTLIMITS= Data Set" on page 1208.

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=detergnt;
   mchart weight*lot / outtable=dtable
                       nochart;
run;
```

The data set DTABLE is listed in Figure 35.8.

Summary Statistics and Control Limit Information

VAR	lot	_SIGMAS_	_LIMITN_	_SUBN_	_LCLM_	_SUBMED_	_MEAN_	_UCLM_	_EXLIM_
weight	1	3	5	5	20.7554	22.56	24.4996	28.2439	
weight	2	3	5	5	20.7554	23.54	24.4996	28.2439	
weight	3	3	5	5	20.7554	24.35	24.4996	28.2439	
weight	4	3	5	5	20.7554	25.52	24.4996	28.2439	
weight	5	3	5	5	20.7554	23.25	24.4996	28.2439	
weight	6	3	5	5	20.7554	23.01	24.4996	28.2439	
weight	7	3	5	5	20.7554	24.19	24.4996	28.2439	
weight	8	3	5	5	20.7554	26.27	24.4996	28.2439	
weight	9	3	5	5	20.7554	22.31	24.4996	28.2439	
weight	10	3	5	5	20.7554	22.66	24.4996	28.2439	
weight	11	3	5	5	20.7554	26.38	24.4996	28.2439	
weight	12	3	5	5	20.7554	23.01	24.4996	28.2439	
weight	13	3	5	5	20.7554	23.15	24.4996	28.2439	
weight	14	3	5	5	20.7554	24.73	24.4996	28.2439	
weight	15	3	5	5	20.7554	25.86	24.4996	28.2439	
weight	16	3	5	5	20.7554	23.99	24.4996	28.2439	
weight	17	3	5	5	20.7554	24.30	24.4996	28.2439	
weight	18	3	5	5	20.7554	24.88	24.4996	28.2439	
weight	19	3	5	5	20.7554	25.92	24.4996	28.2439	
weight	20	3	5	5	20.7554	25.63	24.4996	28.2439	
weight	21	3	5	5	20.7554	25.17	24.4996	28.2439	
weight	22	3	5	5	20.7554	26.72	24.4996	28.2439	
weight	23	3	5	5	20.7554	23.67	24.4996	28.2439	
weight	24	3	5	5	20.7554	24.92	24.4996	28.2439	
weight	25	3	5	5	20.7554	24.51	24.4996	28.2439	
weight	26	3	5	5	20.7554	24.75	24.4996	28.2439	
weight	27	3	5	5	20.7554	25.10	24.4996	28.2439	
weight	28	3	5	5	20.7554	25.64	24.4996	28.2439	

Figure 35.8. The Data Set DTABLE

This data set contains one observation for each subgroup sample. The variables _SUBMED_ and _SUBN_ contain the subgroup medians and subgroup sample sizes. The variables _LCLM_ and _UCLM_ contain the lower and upper control limits, and the variable _MEAN_ contains the central line. The variables _VAR_ and LOT contain the *process* name and values of the *subgroup-variable*, respectively. For more information, see "OUTTABLE= Data Set" on page 1210.

An OUTTABLE= data set can be read later as a TABLE= data set. For example, the following statements read DTABLE and display a median chart (not shown here) identical to the chart in Figure 35.2 on page 1186:

```
title 'Median Chart for Detergent Box Weight';
proc shewhart table=dtable;
   mchart weight*lot;
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 1214.

Reading Preestablished Control Limits

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

See SHWMCHR in the SAS/QC Sample Library

```
data detergt2;
   input lot @;
   do i=1 to 5;
      input weight @;
      output;
      end;
   drop i;
   datalines;
29 16.66 27.49 18.87 22.53 24.72
30 23.74 23.67 23.64 20.26 22.09
31 24.56 24.82 23.92 26.67 21.38
32 25.89 28.73 29.21 25.38 23.47
33 23.32 21.61 30.75 23.13 23.82
34 23.04 22.65 24.96 19.64 26.84
35 24.01 24.38 24.86 26.50 24.37
36 26.43 27.36 28.74 26.74 26.27
37 21.41 22.24 25.34 20.59 27.51
38 22.62 20.81 22.64 30.15 25.32
39 26.86 28.14 24.06 27.35 22.49
40 23.03 23.83 25.59 19.85 22.33
41 23.19 23.63 23.00 21.46 27.57
42 27.38 23.18 24.99 24.81 28.82
43 26.60 26.58 20.26 26.27 24.96
44 26.22 23.28 24.15 24.06 28.23
45 25.90 22.88 25.55 24.50 19.95
46 16.66 27.49 18.87 22.53 24.72
47 23.74 23.67 23.64 20.26 22.09
48 24.56 24.82 23.92 26.67 21.38
49 25.89 28.73 29.21 25.38 23.47
50 23.32 21.61 30.75 23.13 23.82
;
```

The following statements create a median chart for the data in DETERGT2 using the control limits in DETLIM:

```
title 'Median Chart for Weights';
symbol v=dot;
proc shewhart data=detergt2 limits=detlim;
   mchart weight*lot;
run;
```

The chart is shown in Figure 35.9. 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 WEIGHT
- the value of _SUBGRP_ matches the *subgroup-variable* name LOT

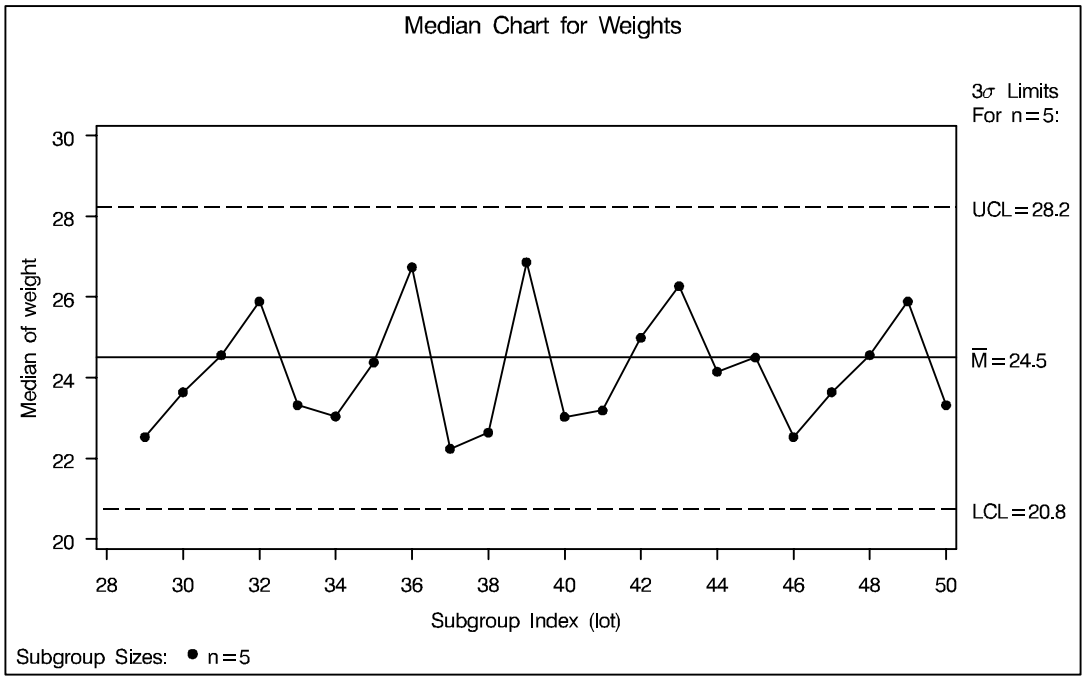

Figure 35.9. Median Chart for Second Set of Detergent Box Weight Data

The chart indicates that the process is in control, since all the medians 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 1212 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 MCHART statement is as follows:

MCHART *process******subgroup-variable* **;**

The general form of this syntax is as follows:

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

You can use any number of MCHART statements in the SHEWHART procedure. The components of the MCHART 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 Medians from Raw Data" on page 1184.
- 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 Medians from Subgroup Summary Data" on page 1186.
- 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 1191.

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

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

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

- 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 a median chart using an asterisk (*) to plot the points:

```
proc shewhart data=values;
   mchart weight*day='*';
run;
```

options
enhance the appearance of the charts, 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 MCHART statement options by function. For complete descriptions, see Chapter 46, "Dictionary of Options."

Table 35.1. Tabulation Options

TABLE	creates a basic table of subgroup medians, subgroup sample sizes, and control limits
TABLEALL	is equivalent to the options TABLE, TABLECENTRAL, TABLEID, TABLELEGEND, TABLEOUT, 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 35.2. 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 for the median 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
ZONELABELS	adds labels A, B, and C to zone lines for median chart
ZONES	adds lines to median chart delineating zones A, B, and C
ZONEVALPOS=*n*	specifies position of ZONEVALUES labels
ZONEVALUES	labels zone lines with their values

Table 35.3. Graphical Options for Displaying Tests for Special Causes

CTESTS=*color*| *test-color-list*	specifies color for labels used to identify 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 35.4. 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 35.5. 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 35.6. 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 35.7. 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 35.8. 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 median 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 median 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 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 median 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*	specifies position of VREFLABELS= and VREF2LABELS= labels

Table 35.9. 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 35.10. 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 median 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 35.11. 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 _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 statistic

Table 35.12. 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 a median chart
LIMLABSUBCHAR='*character*'	specifies a substitution character for labels provided as quoted strings; the character is replaced with the value of the control limit
on median chart	
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 in median chart
NOCTL	suppresses display of central line in median chart
NOLCL	suppresses display of lower control limit in median 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 LIMITS= data set
NOLIMITSLEGEND	suppresses legend for control limits
NOUCL	suppresses display of upper control limit in median chart
UCLLABEL='*string*'	specifies label for upper control limit in median chart
WLIMITS=*n*	specifies width for control limits and central line
XSYMBOL='*string*'| *keyword*	specifies label for central line in median chart

Table 35.13. Options for Plotting and Labeling Points

ALLLABEL=VALUE|(*variable*)	labels every point on median 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 on median 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*
TURNALL|TURNOUT	turns point labels so that they are strung out vertically

Table 35.14. Process Mean and Standard Deviation Options

MEDCENTRAL=*keyword*	specifies method for estimating process mean μ
MU0=*value*	specifies known value 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 whether parameters are estimates or standard values and specifies value of _TYPE_ in the OUTLIMITS= data set

Table 35.15. 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 35.16. Input Data Set Options

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

Table 35.17. 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 35.18. 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 median 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 median chart as a percentage of sum of lengths of vertical axes for median and trend charts
ZEROSTD	displays median chart regardless of whether $\hat{\sigma} = 0$

Table 35.19. Graphical Enhancement Options

ANNOTATE=*SAS-data-set*	specifies annotate data set that adds features to median 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 PROC GREPLAY master menu for median 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 median 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 35.20. 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 median chart
WSTARCIRCLES=*n*	specifies width of STARCIRCLES= circles
WSTARS=*n*	specifies width of STARVERTICES= stars

Table 35.21. 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 co-ordinate data

Details

Constructing Median Charts

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
n_i	sample size of i^{th} subgroup
N	the number of subgroups
x_{ij}	j^{th} measurement in the i^{th} subgroup, $j = 1, 2, 3, \ldots, n_i$
$x_{i(j)}$	j^{th} largest measurement in the i^{th} subgroup. Then $$x_{i(1)} \leq x_{i(2)} \leq \ldots \leq x_{i(n_i)}$$
$\overline{\overline{X}}$	weighted average of subgroup means
M_i	median of the measurements in the i^{th} subgroup: $$M_i = \begin{cases} x_{i((n_i+1)/2)} & \text{if } n_i \text{ is odd} \\ (x_{i(n_i/2)} + x_{i((n_i/2)+1)})/2 & \text{if } n_i \text{ is even} \end{cases}$$
$\bar{M}$	average of the subgroup medians: $$\bar{M} = (n_1 M_1 + \ldots + n_N M_N)/(n_1 + \ldots + n_N)$$
$\tilde{M}$	median of the subgroup medians. Denote the j^{th} largest median by $M_{(j)}$ so that $M_{(1)} \leq M_{(2)} \leq \ldots \leq M_{(N)}$. Then $$\tilde{M} = \begin{cases} M_{((N+1)/2)} & \text{if } N \text{ is odd} \\ (M_{(N/2)} + M_{(N/2)+1})/2 & \text{if } N \text{ is even} \end{cases}$$
$e_M(n)$	standard error of the median of n independent, normally distributed variables with unit standard deviation (the value of $e_M(n)$ can be calculated with the STDMED function in a DATA step)
$Q_p(n)$	$100p^{\text{th}}$ percentile $(0 < p < 1)$ of the distribution of the median of n independent observations from a normal population with unit standard deviation
z_p	$100p^{\text{th}}$ percentile of the standard normal distribution
$D_p(n)$	$100p^{\text{th}}$ percentile of the distibution of the range of n independent observations from a normal population with unit standard deviation

Plotted Points

Each point on a median chart indicates the value of a subgroup median (M_i). For example, if the tenth subgroup contains the values 12, 15, 19, 16, and 14, the value plotted for this subgroup is $M_{10} = 15$.

Central Line

The value of the central line indicates an estimate for μ, which is computed as

- $\bar{M}$ by default
- $\overline{\overline{X}}$ when you specify MEDCENTRAL=AVGMEAN
- $\tilde{M}$ when you specify MEDCENTRAL=MEDMED
- μ_0 when you specify μ_0 with the MU0= option

Control Limits

You can compute the limits

- as a specified multiple (k) of the standard error of M_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 M_i exceeds the limits

The following table provides the formulas for the limits:

Table 35.22. Limits for Median Charts

Control Limits
LCLM = lower limit $= \bar{M} - k\hat{\sigma}e_M(n_i)$ UCLM = upper limit $= \bar{M} + k\hat{\sigma}e_M(n_i)$

Probability Limits
LCLM = lower limit $= \bar{M} - Q_{\alpha/2}(n_i)\hat{\sigma}$ UCLM = upper limit $= \bar{M} + Q_{1-\alpha/2}(n_i)\hat{\sigma}$

Note that the limits vary with n_i. In Table 35.22, replace $\bar{M}$ with $\overline{\overline{X}}$ if you specify MEDCENTRAL=AVGMEAN, and replace $\bar{M}$ with $\tilde{M}$ if you specify MEDCENTRAL=MEDMED. Replace $\bar{M}$ with μ_0 if you specify μ_0 with the MU0= option, and replace $\hat{\sigma}$ with σ_0 if you specify σ_0 with the SIGMA0= option. The formulas assume that the data are normally distributed.

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 the LIMITS= data set.
- Specify σ_0 with the SIGMA0= option or with the variable _STDDEV_ in the 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 35.23. 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
LCLM	lower control limit for subgroup median
LCLR	lower control limit for subgroup range
LCLS	lower control limit for subgroup standard deviation
LIMITN	sample size associated with the control limits
LSL	lower specification limit
MEAN	value of central line on median chart ($\bar{M}$, $\tilde{M}$, $\overline{\overline{X}}$, 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 M_i
STDDEV	process standard deviation ($\hat{\sigma}$ or σ_0)
SUBGRP	*subgroup-variable* specified in the MCHART statement
TARGET	target value
TYPE	type (estimate or standard value) of _MEAN_ and _STDDEV_
UCLM	upper control limit for subgroup median
UCLR	upper control limit for subgroup range
UCLS	upper control limit for subgroup standard deviation
USL	upper specification limit
VAR	*process* specified in the MCHART 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 median charts, but they allow the OUTLIMITS= data set to be used as a LIMITS= data set with the BOXCHART, XRCHART, XSCHART, and MRCHART statements.

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

3. If the limits are defined in terms of a multiple k of the standard error of M_i, the value of _ALPHA_ is computed as $\alpha = 2(1 - F_{med}(k, n))$, where $F_{med}(\cdot, n)$ is the cumulative distribution function of the median of a random sample of n standard normally distributed observations, 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 $k = F^{-1}_{med}(1 - \alpha/2, n)$, where $F^{-1}_{med}(\cdot, n)$ is the inverse distribution function of the median of a random sample of n standard normally distributed observations, and n is the value of _LIMITN_. 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 MCHART statement. For an example, see "Saving Control Limits" on page 1191.

OUTHISTORY= Data Set

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

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

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.

Variables containing subgroup summary statistics are created for each *process* specified in the MCHART statement. For example, consider the following statements:

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

The data set SUMMARY contains variables named LOT, WIDTHM, WIDTHR, WIDTHN, DIAMTERM, 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 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 1189.

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 median chart
LCLM	lower control limit for median
LIMITN	nominal sample size associated with the control limits
MEAN	estimate of process mean ($\bar{M}$, $\tilde{M}$, $\overline{\overline{X}}$, or μ_0)
SIGMAS	multiple (*k*) of the standard error associated with control limits
subgroup	values of the subgroup variable
SUBMED	subgroup median
SUBN	subgroup sample size
TESTS	tests for special causes signaled on median chart
UCLM	upper control limit for median
VAR	*process* specified in the MCHART 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= options, respectively, or with the corresponding variables in a LIMITS= data set).

2. The variable _TESTS_ is saved if you specify the TESTS= option. The k^{th} character of a value of _TESTS_ is k if Test k is positive at that subgroup. For example, if you request 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 1191.

ODS Tables

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

Table 35.24. ODS Tables Produced with the MCHART Statement

Table Name	Description	Options
MCHART	median 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 MCHART 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 values of the *subgroup-variable*. The *subgroup-variable*, which is specified in the MCHART 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= data set includes the variable _PHASE_, you can read selected groups of observations (referred to as *phases*) by specifying the READPHASES= option (for an example, see "Displaying Stratification in Phases" on page 1689).

For an example of a DATA= data set, see "Creating Charts for Medians from Raw Data" on page 1184.

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;
   mchart 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. 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 _LCLM_, _MEAN_, and _UCLM_, 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 35.22 on page 1207

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 the "Reading Preestablished Control Limits" section on page 1193.

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

HISTORY= Data Set

You can read subgroup summary statistics from a HISTORY= data set specified in the PROC SHEWHART statement. This allows you to reuse OUTHISTORY= data sets that have been created in previous runs of the SHEWHART procedure or to read output data sets created with SAS summarization procedures, such as PROC UNIVARIATE.

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

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

The names of the subgroup summary statistics variables must be the *process* name concatenated with the following special suffix characters:

Subgroup Summary Statistic	Suffix Character
subgroup median	M
subgroup mean	X
subgroup sample size	N
subgroup range	R
subgroup standard deviation	S

You must provide the subgroup mean variable only if you specify the MEDCENTRAL=AVGMEAN option. If you specify the STDDEVIATIONS option, the subgroup standard deviation variable must be included; otherwise, the subgroup range variable must be included.

For example, consider the following statements:

```
proc shewhart history=summary;
   mchart (weight yldstren)*batch / medcentral=avgmean;
run;
```

The data set SUMMARY must include the variables BATCH, WEIGHTX, WEIGHTM, WEIGHTR, WEIGHTN, YLDSRENX, YLDSRENM, YLDSRENR, and YLDSRENN. If the STDDEVIATIONS option were specified in the preceding MCHART statement, it would be necessary for SUMMARY to include the variables BATCH, WEIGHTX, WEIGHTM, WEIGHTS, WEIGHTN, YLDSRENX, YLDSRENM, 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 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 Medians from Subgroup Summary Data" on page 1186.

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 MCHART statement:

Table 35.25. Variables Required in a TABLE= Data Set

Variable	Description
LCLM	lower control limit for median
LIMITN	nominal sample size associated with the control limits
MEAN	process mean
subgroup-variable	values of the *subgroup-variable*
SUBMED	subgroup median
SUBN	subgroup sample size
UCLM	upper control limit for median

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 1191.

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 with the MCHART statement for estimating the process standard deviation σ. The method used to calculate σ depends on whether you specify the STDDEVIATIONS option in the MCHART statement. If this option is specified, σ is estimated using subgroup standard deviations; otherwise, σ is estimated using subgroup ranges. For further details and formulas, see "Methods for Estimating the Standard Deviation" on page 1488.

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 median variable
Vertical	TABLE=	_SUBMED_

If you specify the TRENDVAR= option, you can provide distinct labels for the vertical axes of the median 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 median chart, and the second part labels the vertical axis of the trend chart.

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.

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