# Chapter 39
# RCHART Statement

## Chapter Table of Contents

# Chapter 39
# RCHART Statement

## Overview

The RCHART statement creates an $R$ chart for subgroup ranges, which is used to analyze the variability of a process.*

You can use options in the RCHART statement to

- compute control limits from the data based on a multiple of the standard error of the plotted ranges or as probability limits
- tabulate subgroup sample sizes, subgroup ranges, control limits, and other information
- save control limits in an output data set
- save subgroup sample sizes, subgroup means, and subgroup ranges in an output data set
- read preestablished control limits from a data set
- specify the 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 $s$ charts for this purpose; see Chapter 40, "SCHART Statement." In general, $s$ charts are recommended with large subgroup sample sizes ($n_i \geq 10$).

# Getting Started

This section introduces the RCHART statement with simple examples that illustrate the most commonly used options. Complete syntax for the RCHART statement is presented in the "Syntax" section on page 1358, and advanced examples are given in the "Examples" section on page 1377.

## Creating Range Charts from Raw Data

See SHWRCHR in the SAS/QC Sample Library

A disk drive manufacturer performs a battery of tests to evaluate its drives. The following statements create a data set named DISKS, which contains the time (in milliseconds) required to complete one of these tests for six drives in each of 25 lots:

```
data disks;
   input lot @;
   do i=1 to 6;
      input time @;
      output;
      end;
   drop i;
 datalines;
 1 8.05 7.90 8.04 8.06 8.01 7.99
 2 8.03 8.06 8.02 8.02 7.97 8.03
 3 8.00 7.94 7.97 7.95 8.00 8.01
 4 8.00 8.06 8.06 7.99 7.97 7.96
 5 7.93 8.01 8.00 8.09 8.06 8.02
 6 7.98 7.99 8.01 8.09 8.00 7.97
 7 8.00 7.94 7.93 8.03 7.93 8.08
 8 8.01 7.98 7.98 8.07 8.05 8.09
 9 7.97 7.96 8.01 8.11 8.06 8.07
10 7.93 8.03 8.03 8.00 7.93 8.03
11 8.00 8.00 8.02 7.92 7.98 8.01
12 7.98 7.93 8.01 7.97 8.02 8.00
13 8.06 7.93 7.98 7.98 8.02 7.96
14 8.05 7.98 8.05 7.99 7.95 7.99
15 7.94 8.01 7.97 8.04 7.91 8.03
16 8.03 8.03 8.02 8.06 8.00 7.97
17 8.03 7.94 8.05 8.05 8.04 7.94
18 7.99 7.99 7.86 7.99 8.06 8.03
19 7.95 7.96 7.99 7.96 7.94 8.12
20 8.03 8.07 7.98 7.97 8.00 8.04
21 8.04 7.90 8.03 8.02 7.98 7.97
22 7.95 8.05 7.98 8.01 7.97 8.15
23 8.06 8.00 8.03 8.02 7.99 7.95
24 7.97 8.02 8.00 7.96 7.96 8.00
25 8.12 7.97 7.99 8.09 8.05 8.00
;
```

A partial listing of DISKS is shown in Figure 39.1.

**The Data Set DISKS**

| lot | time |
|---|---|
| 1 | 8.05 |
| 1 | 7.90 |
| 1 | 8.04 |
| 1 | 8.06 |
| 1 | 8.01 |
| 1 | 7.99 |
| 2 | 8.03 |
| 2 | 8.06 |
| 2 | 8.02 |
| 2 | 8.02 |
| 2 | 7.97 |
| 2 | 8.03 |
| 3 | 8.00 |
| 3 | 7.94 |
| 3 | 7.97 |
| 3 | 7.95 |
| 3 | 8.00 |
| 3 | 8.01 |
| . | . |
| . | . |
| . | . |

**Figure 39.1.** Partial Listing of the Data Set DISKS

The data set DISKS is said to be in "strung-out" form since each observation contains the lot number and test time for a single disk drive. The first five observations contain the times for the first lot, the second five observations contain the times 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 TIME contains the time measurements and is referred to as the *process variable* (or *process* for short).

You can use an $R$ chart to determine whether the variability in the performance of the disk drives is in control. The following statements create the $R$ chart shown in Figure 39.2:

```
title 'Range Chart for Disk Drive Test Times';
symbol v=dot;
proc shewhart data=disks;
   rchart time*lot;
run;
```

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

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

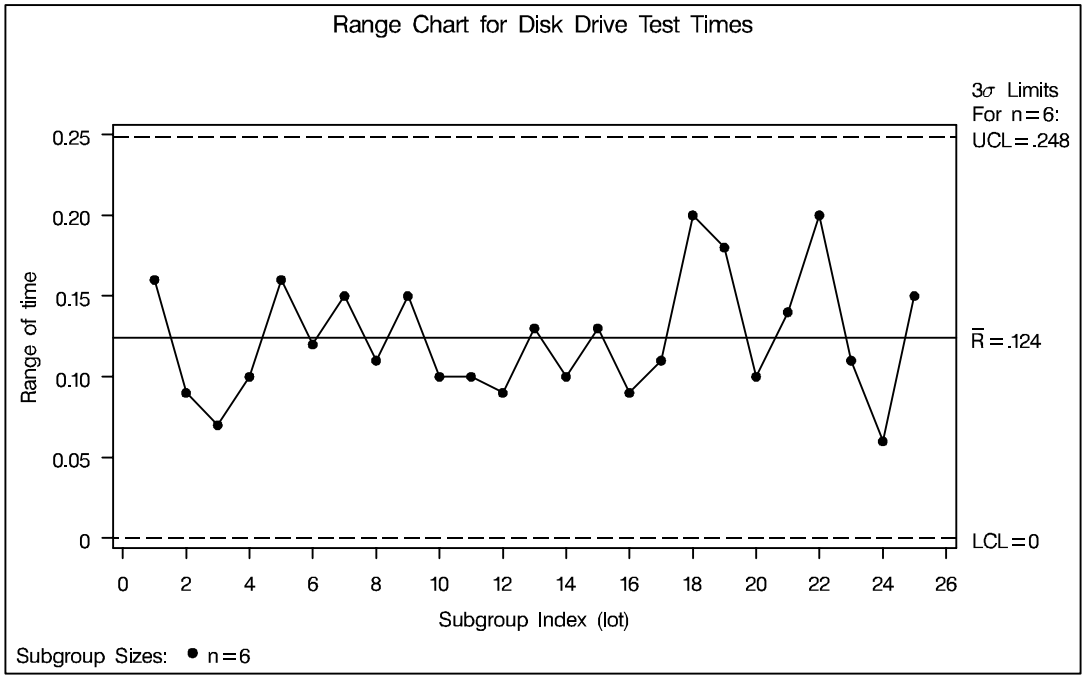



**Figure 39.2.** $R$ Chart for the Data Set DISKS

Each point on the $R$ chart represents the range of the measurements for a particular lot. For instance, the range plotted for the first lot is $8.06 - 7.90 = 0.16$. Since all of the subgroup ranges lie within the control limits, you can conclude that the variability in the performance of the disk drives is in statistical control.

By default, the control limits shown are $3\sigma$ limits estimated from the data; the formulas for the limits are given in Table 39.21 on page 1368. You can also read control limits from an input data set; see "Reading Preestablished Control Limits" on page 1356.

For computational details, see "Constructing Range Charts" on page 1368. For more details on reading raw data, see "DATA= Data Set" on page 1372.

## Creating Range Charts from Summary Data

See SHWRCHR in the SAS/QC Sample Library

The previous example illustrates how you can create $R$ 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 RCHART statement with data of this type.

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

```
data disksum;
   input lot timex timer;
   timen=6;
datalines;
 1  8.00833  0.16
 2  8.02167  0.09
 3  7.97833  0.07
 4  8.00667  0.10
 5  8.01833  0.16
 6  8.00667  0.12
 7  7.98500  0.15
 8  8.03000  0.11
 9  8.03000  0.15
10  7.99167  0.10
11  7.98833  0.10
12  7.98500  0.09
13  7.98833  0.13
14  8.00167  0.10
15  7.98333  0.13
16  8.01833  0.09
17  8.00833  0.11
18  7.98667  0.20
19  7.98667  0.18
20  8.01500  0.10
21  7.99000  0.14
22  8.01833  0.20
23  8.00833  0.11
24  7.98500  0.06
25  8.03667  0.15
;
```

A partial listing of DISKSUM is shown in Figure 39.3. There is exactly one observation for each subgroup (note that the subgroups are still indexed by LOT). The variable TIMEX contains the subgroup means, the variable TIMER contains the subgroup ranges, and the variable TIMEN contains the subgroup sample sizes (these are all six).

**The Summary Data Set of Disk Drive Test Times**

| lot | timex | timer | timen |
|---|---|---|---|
| 1 | 8.00833 | 0.16 | 6 |
| 2 | 8.02167 | 0.09 | 6 |
| 3 | 7.97833 | 0.07 | 6 |
| 4 | 8.00667 | 0.10 | 6 |
| 5 | 8.01833 | 0.16 | 6 |
| . | . | . | . |
| . | . | . | . |
| . | . | . | . |

**Figure 39.3.** The Summary Data Set DISKSUM

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

```
title 'Range Chart for Disk Drive Test Times';
proc shewhart history=disksum lineprinter;
   rchart time*lot='X';
run;
```

The resulting $R$ chart is shown in Figure 39.4. Since the LINEPRINTER option is specified in the PROC SHEWHART statement, line printer output is produced. The character (X) specified in single quotes after the *subgroup-variable* specifies the *character* used to plot points. This character must follow an equal sign.

Note that TIME is *not* the name of a SAS variable in the data set DISKSUM but is, instead, the common prefix for the names of the SAS variables TIMER and TIMEN. The suffix characters $R$ and $N$ indicate *range* and *sample size*, respectively. Thus, you can specify two subgroup summary variables in the HISTORY= data set with a single name (TIME), which is referred to as the *process*. The name LOT specified after the asterisk is the name of the *subgroup-variable*.

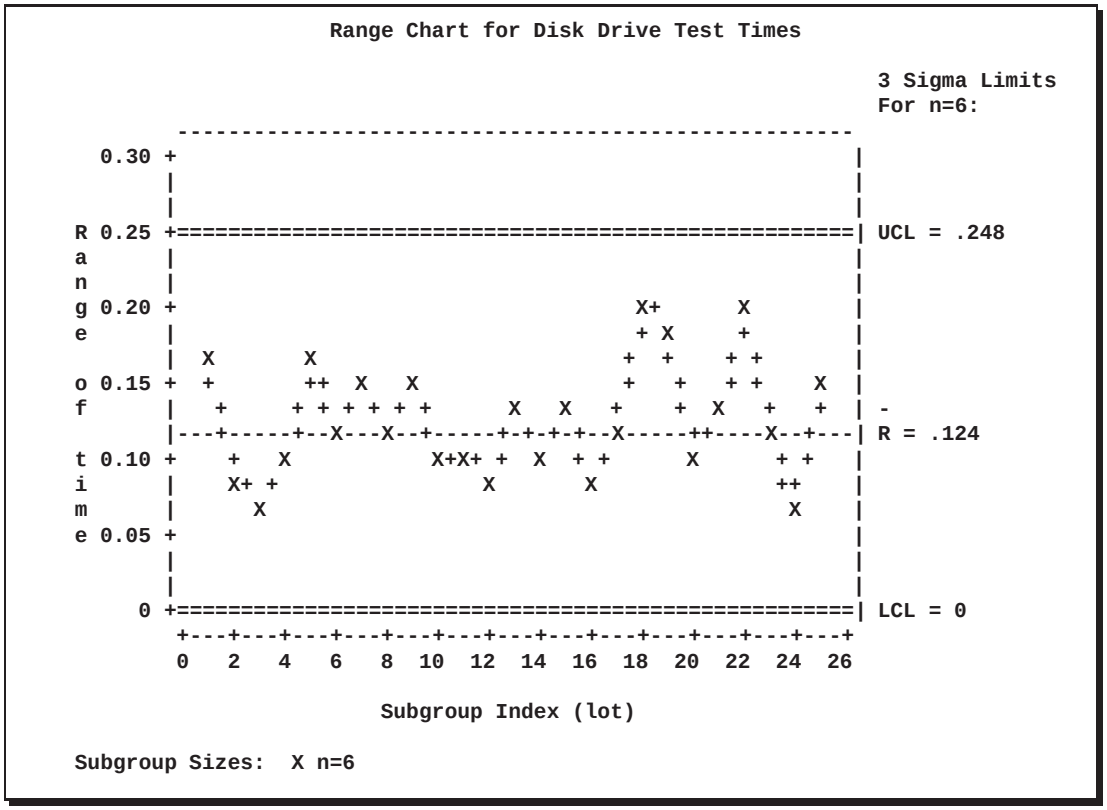



**Figure 39.4.** $R$ Chart from the Summary Data Set DISKSUM

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

- subgroup variable
- subgroup range variable
- subgroup sample size variable

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

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 1374.

## Saving Summary Statistics

See SHWRCHR in the SAS/QC Sample Library

In this example, the RCHART statement procedure 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 DISKS and create a summary data set named DISKHIST:

```
title 'Summary Data Set for Disk Times';
proc shewhart data=disks;
   rchart time*lot / outhistory = diskhist
                     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 39.2. Options such as OUTHISTORY= and NOCHART are specified after the slash (/) in the RCHART statement. A complete list of options is presented in the "Syntax" section on page 1358.

Figure 39.5 contains a partial listing of DISKHIST.

**Summary Data Set for Disk Times**

| lot | timeX | timeR | timeN |
|---|---|---|---|
| 1 | 8.00833 | 0.16 | 6 |
| 2 | 8.02167 | 0.09 | 6 |
| 3 | 7.97833 | 0.07 | 6 |
| 4 | 8.00667 | 0.10 | 6 |
| 5 | 8.01833 | 0.16 | 6 |
| . | . | . | . |
| . | . | . | . |
| . | . | . | . |

**Figure 39.5.** The Summary Data Set DISKHIST

There are four variables in the data set DISKHIST.

- LOT contains the subgroup index.
- TIMEX contains the subgroup means.
- TIMER contains the subgroup ranges.
- TIMEN contains the subgroup sample sizes.

The subgroup mean variable is included in the OUTHISTORY= data set even though it is not required by the RCHART statement. This allows the data set to be used as a HISTORY= data set with the BOXCHART, XCHART, and XRCHART statements, as well as with the RCHART statement. Note that the summary statistic variables are named by adding the suffix characters *X*, *R*, and *N* to the *process* TIME specified in the RCHART 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 1370.

## Saving Control Limits

See SHWRCHR in the SAS/QC Sample Library

You can save the control limits for an $R$ chart in a SAS data set; this enables you to apply the control limits to future data (see "Reading Preestablished Control Limits" on page 1356) or modify the limits with a DATA step program.

The following statements read measurements from the data set DISKS (see page 1348) and save the control limits displayed in Figure 39.2 in a data set named DISKLIM:

```
title 'Control Limits for Disk Times';
proc shewhart data=disks;
   rchart time*lot / outlimits = disklim
                     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 DISKLIM is listed in Figure 39.6.

**Control Limits for Disk Times**

| _VAR_ | _SUBGRP_ | _TYPE_ | _LIMITN_ | _ALPHA_ | _SIGMAS_ |
|---|---|---|---|---|---|
| time | lot | ESTIMATE | 6 | .004447667 | 3 |

| _LCLX_ | _MEAN_ | _UCLX_ | _LCLR_ | _R_ | _UCLR_ | _STDDEV_ |
|---|---|---|---|---|---|---|
| 7.94314 | 8.00307 | 8.06299 | 0 | 0.124 | 0.24847 | 0.048927 |

**Figure 39.6.** The Data Set DISKLIM Containing Control Limit Information

The data set DISKLIM contains one observation with the limits for *process* TIME. The variables _LCLR_ and _UCLR_ contain the lower and upper control limits, and the variable _R_ 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 $\sigma$. The value of _LIMITN_ is the nominal sample size associated with the

control limits, and the value of _SIGMAS_ is the multiple of $\sigma$ 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 DISKLIM can be used to create an $\bar{X}$ chart (see Chapter 43, "XRCHART Statement"). For more information, see "OUTLIMITS= Data Set" on page 1369.

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=disks;
  rchart time*lot / outtable=disktab
                    nochart;
run;
```

The data set DISKTAB is listed in Figure 39.7.

**Summary Statistics and Control Limit Information**

| _VAR_ | lot | _SIGMAS_ | _LIMITN_ | _SUBN_ | _LCLR_ | _SUBR_ | _R_ | _UCLR_ | _EXLIM_ |
|---|---|---|---|---|---|---|---|---|---|
| time | 1 | 3 | 6 | 6 | 0 | 0.16 | 0.124 | 0.24847 | |
| time | 2 | 3 | 6 | 6 | 0 | 0.09 | 0.124 | 0.24847 | |
| time | 3 | 3 | 6 | 6 | 0 | 0.07 | 0.124 | 0.24847 | |
| time | 4 | 3 | 6 | 6 | 0 | 0.10 | 0.124 | 0.24847 | |
| time | 5 | 3 | 6 | 6 | 0 | 0.16 | 0.124 | 0.24847 | |
| time | 6 | 3 | 6 | 6 | 0 | 0.12 | 0.124 | 0.24847 | |
| time | 7 | 3 | 6 | 6 | 0 | 0.15 | 0.124 | 0.24847 | |
| time | 8 | 3 | 6 | 6 | 0 | 0.11 | 0.124 | 0.24847 | |
| time | 9 | 3 | 6 | 6 | 0 | 0.15 | 0.124 | 0.24847 | |
| time | 10 | 3 | 6 | 6 | 0 | 0.10 | 0.124 | 0.24847 | |
| time | 11 | 3 | 6 | 6 | 0 | 0.10 | 0.124 | 0.24847 | |
| time | 12 | 3 | 6 | 6 | 0 | 0.09 | 0.124 | 0.24847 | |
| time | 13 | 3 | 6 | 6 | 0 | 0.13 | 0.124 | 0.24847 | |
| time | 14 | 3 | 6 | 6 | 0 | 0.10 | 0.124 | 0.24847 | |
| time | 15 | 3 | 6 | 6 | 0 | 0.13 | 0.124 | 0.24847 | |
| time | 16 | 3 | 6 | 6 | 0 | 0.09 | 0.124 | 0.24847 | |
| time | 17 | 3 | 6 | 6 | 0 | 0.11 | 0.124 | 0.24847 | |
| time | 18 | 3 | 6 | 6 | 0 | 0.20 | 0.124 | 0.24847 | |
| time | 19 | 3 | 6 | 6 | 0 | 0.18 | 0.124 | 0.24847 | |
| time | 20 | 3 | 6 | 6 | 0 | 0.10 | 0.124 | 0.24847 | |
| time | 21 | 3 | 6 | 6 | 0 | 0.14 | 0.124 | 0.24847 | |
| time | 22 | 3 | 6 | 6 | 0 | 0.20 | 0.124 | 0.24847 | |
| time | 23 | 3 | 6 | 6 | 0 | 0.11 | 0.124 | 0.24847 | |
| time | 24 | 3 | 6 | 6 | 0 | 0.06 | 0.124 | 0.24847 | |
| time | 25 | 3 | 6 | 6 | 0 | 0.15 | 0.124 | 0.24847 | |

**Figure 39.7.** The Data Set DISKTAB

This data set contains one observation for each subgroup sample. The variables _SUBR_ and _SUBN_ contain the subgroup ranges and subgroup sample sizes. The variables _LCLR_ and _UCLR_ contain the lower and upper control limits, and the variable _R_ 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 1371. An OUTTABLE= data set can be read later as a TABLE= data set. For example, the following statements read

DISKTAB and display an $R$ chart (not shown here) identical to the chart in Figure 39.2:

```
title 'Range Chart for Disk Drive Test Times';
proc shewhart table=disktab;
   rchart time*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 1374.

## Reading Preestablished Control Limits

See SHWRCHR in the SAS/QC Sample Library

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

```
data disks2;
   input lot @;
   do i=1 to 6;
      input time @;
      output;
      end;
   drop i;
datalines;
26 7.93 7.97 7.89 7.81 7.88 7.92
27 7.86 7.91 7.87 7.89 7.83 7.87
28 7.93 7.95 7.90 7.89 7.88 7.90
29 7.97 8.00 7.86 7.89 7.84 7.78
30 7.91 7.93 7.98 7.93 7.83 7.88
31 7.85 7.94 7.88 7.98 7.96 7.84
32 7.86 8.01 7.88 7.95 7.90 7.89
33 7.87 7.93 7.96 7.89 7.81 8.00
34 7.87 7.97 7.95 7.89 7.92 7.84
35 7.92 7.97 7.90 7.88 7.89 7.86
36 7.96 7.90 7.90 7.84 7.90 8.00
37 7.92 7.90 7.98 7.92 7.94 7.94
38 7.88 7.99 8.02 7.98 7.88 7.92
39 7.89 7.91 7.92 7.90 7.94 7.94
40 7.84 7.88 7.91 7.98 7.87 7.93
41 7.91 7.87 7.96 7.91 7.89 7.92
42 7.96 7.93 7.86 7.93 7.86 7.94
43 7.84 7.82 7.87 7.91 7.91 8.01
44 7.93 7.91 7.92 7.88 7.91 7.86
45 7.95 7.92 7.93 7.90 7.86 8.00
;
```

The following statements create an $R$ chart using the control limits in DISKLIM:

```
symbol v=dot;
title 'Range Chart for Disk Drive Test Times';
proc shewhart data=disks2 limits=disklim;
   rchart time*lot;
run;
```

The chart is shown in Figure 39.8. 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 TIME
- the value of _SUBGRP_ matches the *subgroup-variable* name LOT

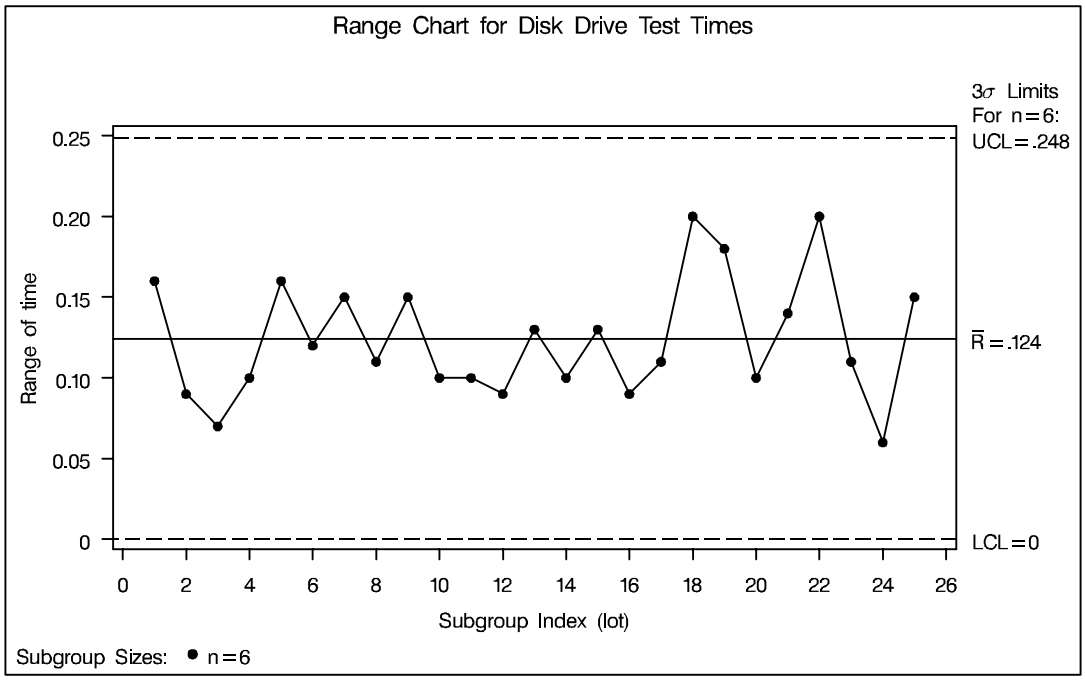



**Figure 39.8.** $R$ Chart for Second Set of Disk Drive Test Times

All the ranges lie within the control limits, indicating that the variability in disk drive performance 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 Example 39.2 on page 1379 and "LIMITS= Data Set" on page 1373 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 RCHART statement is as follows:

**RCHART** *process******subgroup-variable* **;**

The general form of this syntax is as follows:

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

You can use any number of RCHART statements in the SHEWHART procedure. The components of the RCHART 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 Range Charts from Raw Data" on page 1348.
- 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 Range Charts from Summary Data" on page 1350.
- 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 1354.

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

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

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

- 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 $R$ chart using an asterisk (*) to plot the points:

```
proc shewhart data=values;
   rchart 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 RCHART statement options by function. For complete descriptions, see Chapter 46, "Dictionary of Options."

**Table 39.1.** Tabulation Options

| | |
|---|---|
| TABLE | creates a basic table of subgroup values, subgroup sample sizes, subgroup ranges, and control limits |
| TABLEALL | is equivalent to the options TABLE, TABLECENTRAL, TABLEID, TABLELEGEND, TABLEOUTLIM, and TABLETESTS |
| TABLECENTRAL | augments basic table with value of central line |
| 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 39.2.** Options for Specifying Tests for Special Causes

| | |
|---|---|
| TESTS2=*value-list*\|*customized-pattern-list* | specifies tests for special causes for the $R$ 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 $n^{\text{th}}$ 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 |
| ZONE2VALUES | labels zone lines with their values |
| ZONES2 | adds lines delineating zones A, B, and C |
| ZONEVALPOS=*n* | specifies position of ZONE2VALUES labels |

**Table 39.3.** 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=*line-type* | 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 39.4.** 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 39.5.** 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 39.6.** Reference Line Options

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

**Table 39.7.** Axis and Axis Label Options

| | |
|---|---|
| CAXIS=*color* | specifies color for axis lines and tick marks |
| CFRAME=*color*\|(*color-list*) | specifies fill colors for frame for plot area |
| CTEXT=*color* | specifies color for tick mark values and axis labels |
| HAXIS=*values*\|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 |
| 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 |
| 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 39.8.** 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 ranges |

**Table 39.9.** Options for Displaying Control Limits

| | |
|---|---|
| CINFILL=*color* | specifies color for area inside control limits |
| CLIMITS=*color* | specifies color of control limits, central line, and related labels |
| LCLLABEL='*label*' | specifies label for lower control limit |
| LIMLABSUBCHAR='*character*' | specifies a substitution character for labels provided as quoted strings; the character is replaced with the value of the control limit |
| LLIMITS=*linetype* | specifies line type for control limits |
| NDECIMAL=*n* | specifies number of digits to right of decimal place in default labels for control limits and central line |
| NOCTL | suppresses display of central line |
| NOLCL | suppresses display of lower control limit |
| NOLIMITLABEL | suppresses labels for control limits and central line |
| NOLIMITS | suppresses display of control limits |
| NOLIMITSFRAME | suppresses default frame around control limit information when multiple sets of control limits are read from a LIMITS= data set |
| NOLIMITSLEGEND | suppresses legend for control limits |
| NOLIMIT0 | suppresses display of zero lower control limit |
| NOUCL | suppresses display of upper control limit |
| RSYMBOL='*string*'\|*keyword* | specifies label for central line |
| UCLLABEL='*string*' | specifies label for upper control limit |
| WLIMITS=*n* | specifies width for control limits and central line |

**Table 39.10.** Process Mean and Standard Deviation Options

| | |
|---|---|
| SIGMA0=*value* | specifies known value $\sigma_0$ for process standard deviation $\sigma$ |
| SMETHOD=*keyword* | specifies method for estimating process standard deviation $\sigma$ |
| TYPE=*keyword* | identifies whether parameters are estimates or standard values and specifies value of _TYPE_ in the OUTLIMITS= data set |

**Table 39.11.** Specification Limit Options

| | |
|---|---|
| CIALPHA=*value* | specifies $\alpha$ 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 39.12.** 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 39.13.** 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 requested |

**Table 39.14.** Options for Plotting and Labeling Points

| | |
|---|---|
| ALLLABEL=VALUE\|(*variable*) | labels every point on 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 |
| 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 39.15.** 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 |
| 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 39.16.** Input Data Set Options

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

**Table 39.17.** 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 |
| WEBOUT=*SAS-data-set* | creates OUTTABLE= data set with additional graph coordinate data |

**Table 39.18.** Plot Layout Options

| | |
|---|---|
| ALLN | plots ranges 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 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 $R$ chart regardless of whether $\hat{\sigma} = 0$ |

**Table 39.19.** Graphical Enhancement Options

| | |
|---|---|
| ANNOTATE=*SAS-data-set* | specifies annotate data set that adds features to chart |
| DESCRIPTION='*string*' | specifies string that appears in the description field of the PROC GREPLAY master menu |
| FONT=*font* | specifies software font for labels and legends on chart |
| 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 |
| 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 39.20.** 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 chart |
| WSTARCIRCLES=*n* | specifies width of STARCIRCLES= circles |
| WSTARS=*n* | specifies width of STARVERTICES= stars |

# Details

## Constructing Range Charts

The following notation is used in this section:

| | |
|---|---|
| $\sigma$ | process standard deviation (standard deviation of the population of measurements) |
| $R_i$ | range of measurements in $i^{\text{th}}$ subgroup |
| $n_i$ | sample size of $i^{\text{th}}$ subgroup |
| $d_2(n)$ | expected value of the range of $n$ independent normally distributed variables with unit standard deviation |
| $d_3(n)$ | standard error of the range of $n$ independent observations from a normal population with unit standard deviation |
| $D_p(n)$ | $100p^{\text{th}}$ percentile of the distribution of the range of $n$ independent observations from a normal population with unit standard deviation |

### Plotted Points

Each point on an $R$ chart indicates the value of a subgroup range ($R_i$). For example, if the tenth subgroup contains the values 12, 15, 19, 16, and 14, the value plotted for this subgroup is $R_{10} = 19 - 12 = 7$.

### Central Line

By default, the central line for the $i^{\text{th}}$ subgroup indicates an estimate of the expected value of $R_i$, which is computed as $d_2(n_i)\hat{\sigma}$, where $\hat{\sigma}$ is an estimate of $\sigma$. If you specify a known value ($\sigma_0$) for $\sigma$, the central line indicates the value of $d_2(n_i)\sigma_0$. 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 $R_i$ above and below the central line. The default limits are computed with $k = 3$ (these are referred to as $3\sigma$ *limits*).
- as probability limits defined in terms of $\alpha$, a specified probability that $R_i$ exceeds the limits

The following table provides the formulas for the limits:

**Table 39.21.** Limits for $R$ Charts

| Control Limits |
|---|
| LCL = lower limit = $\max(d_2(n_i)\hat{\sigma} - kd_3(n_i)\hat{\sigma}, 0)$ |
| UCL = upper limit = $d_2(n_i)\hat{\sigma} + kd_3(n_i)\hat{\sigma}$ |

| Probability Limits |
|---|
| LCL = lower limit = $D_{\alpha/2}\hat{\sigma}$ |
| UCL = upper limit = $D_{1-\alpha/2}\hat{\sigma}$ |

The formulas assume that the data are normally distributed. Note that the control limits vary with $n_i$ and that the probability limits for $R_i$ are asymmetric around the central line. If a standard value $\sigma_0$ is available for $\sigma$, replace $\hat{\sigma}$ with $\sigma_0$ in Table 39.21.

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 $\alpha$ 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 $\sigma_0$ 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 39.22.** OUTLIMITS= Data Set

| Variable | Description |
|---|---|
| _ALPHA_ | probability ($\alpha$) 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 with the OUTINDEX= option |
| _LCLR_ | lower control limit for subgroup range |
| _LCLX_ | lower control limit for subgroup mean |
| _LIMITN_ | sample size associated with the control limits |
| _LSL_ | lower specification limit |
| _MEAN_ | process mean ($\overline{\overline{X}}$) |
| _R_ | value of central line on $R$ chart |
| _SIGMAS_ | multiple ($k$) of standard error of $R_i$ |
| _STDDEV_ | process standard deviation ($\hat{\sigma}$ or $\sigma_0$) |
| _SUBGRP_ | *subgroup-variable* specified in the RCHART statement |
| _TARGET_ | target value |
| _TYPE_ | type (estimate or standard value) of _MEAN_ and _STDDEV_ |
| _UCLR_ | upper control limit for subgroup range |
| _UCLX_ | upper control limit for subgroup mean |
| _USL_ | upper specification limit |
| _VAR_ | *process* specified in the RCHART 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 XRCHART statements.

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

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

$$F_R(\text{\_LCLR\_}/\text{\_STDDEV\_}) + 1 - F_R(\text{\_UCLR\_}/\text{\_STDDEV\_})$$

   where $F_R(\cdot)$ is the cumulative distribution function of the range 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 $(\text{\_UCLR\_} - \text{\_R\_})/e$, where $e$ is the standard error of the range 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 RCHART statement. For an example, see "Saving Control Limits" on page 1354.

### 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 range variable named by *process* suffixed with *R*
- a subgroup sample size variable named by *process* suffixed with *N*

The subgroup mean variable is saved so that the data set can be reused as a HISTORY= data set with the BOXCHART, XCHART, and XRCHART statements, as well as the RCHART 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 RCHART statement. For example, consider the following statements:

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

The data set SUMMARY contains variables named LOT, WIDTHX, WIDTHR, WIDTHN, DIAMTERX, DIAMTERR, 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 1353.

### 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 ($\alpha$) of exceeding control limits |
| _EXLIM_ | control limit exceeded on $R$ chart |
| _LCLR_ | lower control limit for range |
| _LIMITN_ | nominal sample size associated with the control limits |
| _R_ | average range |
| _SIGMAS_ | multiple ($k$) of the standard error associated with control limits |
| *subgroup* | values of the subgroup variable |
| _SUBN_ | subgroup sample sizes |
| _SUBR_ | subgroup range |
| _TESTS2_ | tests for special causes signaled on $R$ chart |
| _UCLR_ | upper control limit for range |
| _VAR_ | *process* specified in the RCHART 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 1354.

## ODS Tables

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

**Table 39.23.** ODS Tables Produced with the RCHART Statement

| Table Name | Description | Options |
|---|---|---|
| RCHART | $R$ 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 RCHART statement must be a SAS variable in the DATA= data set. This variable provides measurements that must be grouped in subgroup samples indexed by the *subgroup-variable*. The *subgroup-variable*, which is specified in the RCHART statement, must also be a SAS variable in the DATA= data set. Each observation in a DATA= data set must contain a raw measurement for each *process* and a value for the *subgroup-variable*. If the $i^{\text{th}}$ 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 $i^{\text{th}}$ 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 Range Charts from Raw Data" on page 1348.

### 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;
   rchart 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 39.22 on page 1369. 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 _LCLR_, _R_, and _UCLR_, which specify the control limits directly
- the variable _STDDEV_, which is used to calculate the control limits according to the equations in Table 39.21 on page 1368

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.

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

For an example, see “Reading Preestablished Control Limits” on page 1356.

### 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 MEANS.

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

- the *subgroup-variable*
- a subgroup range variable for each *process*
- a subgroup sample size variable for each *process*

The names of the subgroup range and subgroup sample size variables must be the prefix *process* concatenated with the special suffix characters $R$ and $N$, respectively.

For example, consider the following statements:

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

The data set SUMMARY must include the variables BATCH, WEIGHTR, WEIGHTN, YLDSRENR, and YLDSRENN.

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

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

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

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

For an example of a HISTORY= data set, see “Creating Range Charts from Summary Data” on page 1350.

### 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 RCHART statement:

**Table 39.24.** Variables Required in a TABLE= Data Set

| Variable | Description |
|---|---|
| _LCLR_ | lower control limit for range |
| _LIMITN_ | nominal sample size associated with the control limits |
| _R_ | average range |
| *subgroup-variable* | values of the *subgroup-variable* |
| _SUBN_ | subgroup sample size |
| _SUBR_ | subgroup range |
| _UCLR_ | upper control limit for range |

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 1354.

## Methods for Estimating the Standard Deviation

When control limits are determined from the input data, two methods (referred to as default and MVLUE) are available for estimating $\sigma$.

### *Default Method*

The default estimate for $\sigma$ is

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

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

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

A subgroup range $R_i$ is included in the calculation only if $n_i \geq 2$. The unbiasing factor $d_2(n_i)$ is defined so that, if the observations are normally distributed, the expected

value of $R_i$ is $d_2(n_i)\sigma$. Thus, $\hat{\sigma}$ is the unweighted average of $N$ unbiased estimates of $\sigma$. 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 $\sigma$. Refer to Burr (1969, 1976) and Nelson (1989, 1994). The MVLUE is a weighted average of $N$ unbiased estimates of $\sigma$ of the form $R_i/d_2(n_i)$, and it is computed as

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

where

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

A subgroup range $R_i$ is included in the calculation only if $n_i \geq 2$, and $N$ is the number of subgroups for which $n_i \geq 2$. The unbiasing factor $d_3(n_i)$ is defined so that, if the observations are normally distributed, the expected value of $\sigma_{R_i}$ is $d_3(n_i)\sigma$. The MVLUE assigns greater weight to estimates of $\sigma$ 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.

## 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 range variable |
| Vertical | TABLE= | _SUBR_ |

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 RCHART statement.

## Example 39.1. Computing Probability Limits

See SHWREX1 in the SAS/QC Sample Library

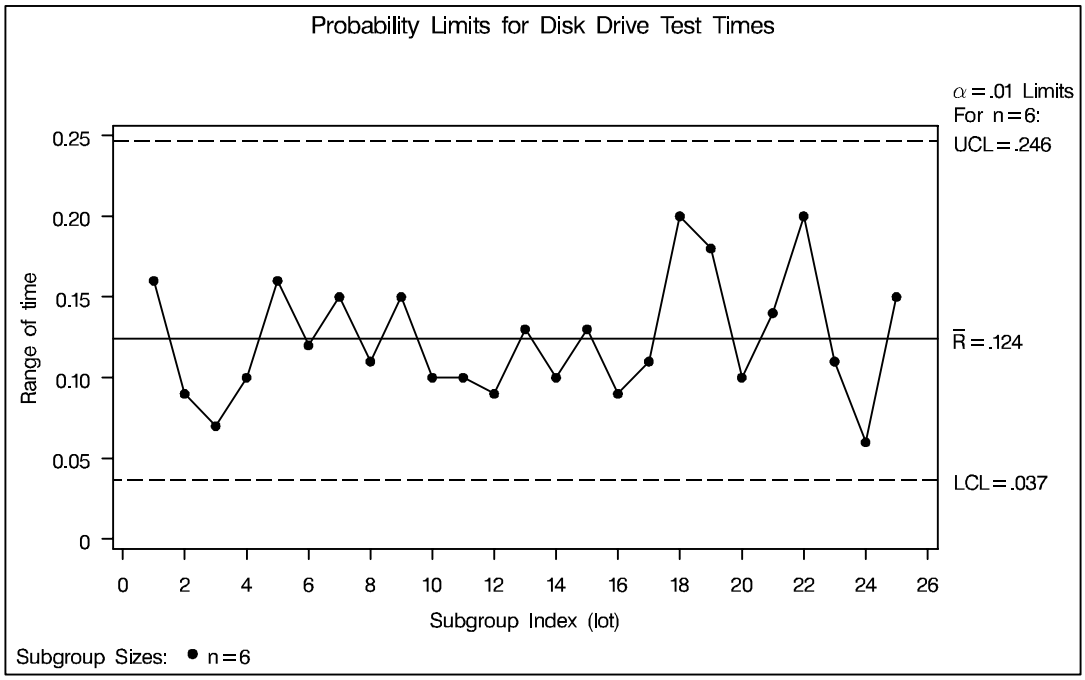
This example demonstrates how to create $R$ charts with probability limits. The following statements read the disk drive test times from the data set DISKS (see page 1348) and create the $R$ chart shown in Output 39.1.1:

```
title 'Probability Limits for Disk Drive Test Times';
symbol v=dot;
proc shewhart data=disks;
   rchart time*lot / alpha    =.01
                     outlimits=dlimits;
run;
```

The ALPHA= option specifies the probability ($\alpha$) that a subgroup range exceeds its limits. Here, the limits are computed so that the probability that a range is less than the lower limit is $\alpha/2 = 0.005$, and the probability that a range is greater than the upper limit is $\alpha/2 = 0.005$. This assumes that the measurements are normally distributed. The OUTLIMITS= option names an output data set that saves the probability limits. A listing of DLIMITS is shown in Output 39.1.2.

The variable _ALPHA_ saves the value of $\alpha$. Note that, in this case, the upper probability limit is equivalent to an upper $2.95\sigma$ limit.

**Output 39.1.1.** $R$ Chart with Probability Limits



**Output 39.1.2.** Probability Limits Data Set

Probablity Limits for Disk Drive Test Times

| _VAR_ | _SUBGRP_ | _TYPE_ | _LIMITN_ | _ALPHA_ | _SIGMAS_ | _LCLX_ |
|---|---|---|---|---|---|---|
| time | lot | ESTIMATE | 6 | 0.01 | 2.94715 | 7.95162 |

| _MEAN_ | _UCLX_ | _LCLR_ | _R_ | _UCLR_ | _STDDEV_ |
|---|---|---|---|---|---|
| 8.00307 | 8.05452 | 0.036645 | 0.124 | 0.24628 | 0.048927 |

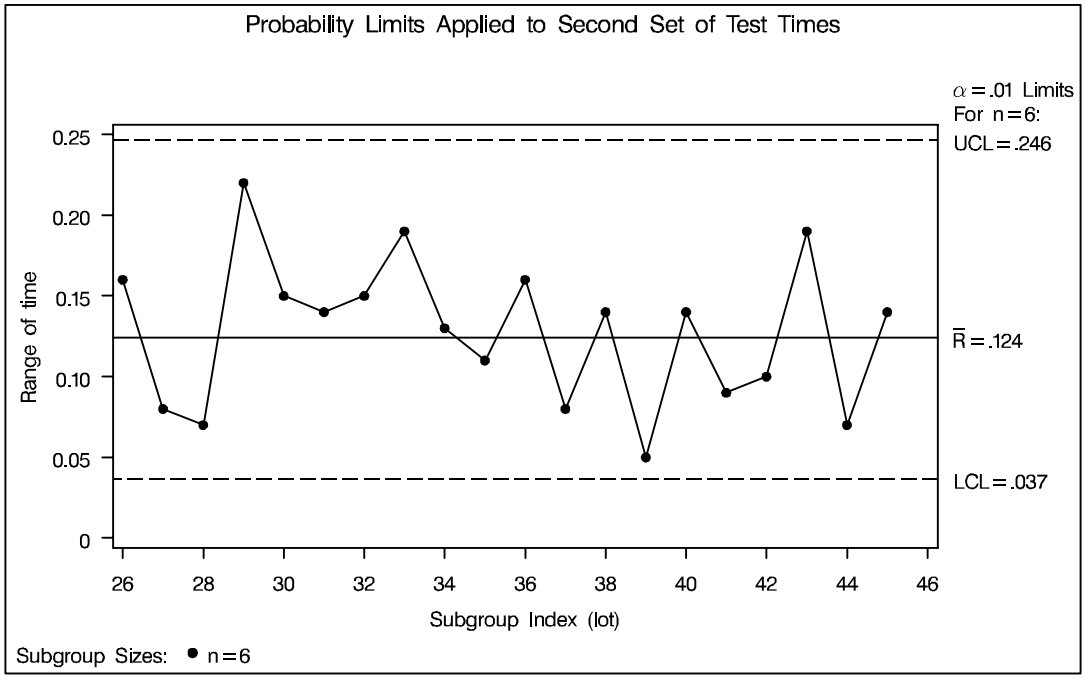
Since all the points fall within the probability limits, it can be concluded that the variability in the disk drive performance is in statistical control.

The following statements apply the limits in DLIMITS to the times in the data set DISKS2 (see page 1356):

```
title 'Probability Limits Applied to Second Set of Test Times';
symbol v=dot;
proc shewhart data=disks2 limits=dlimits;
   rchart time*lot / readalpha;
run;
```

The READALPHA option* specifies that the variable _ALPHA_, rather than the variable _SIGMAS_, is to be read from the LIMITS= data set. Thus the limits displayed in the chart, shown in Output 39.1.3, are probability limits.

**Output 39.1.3.** Reading Probability Limits from a LIMITS= Data Set




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

## Example 39.2. Specifying Control Limit Information

This example illustrates how you can use a DATA step program to create a LIMITS= data set. You can provide previously established values for the limits and central line with the variables _LCLR_, _R_, and _UCLR_, as in the following statements:

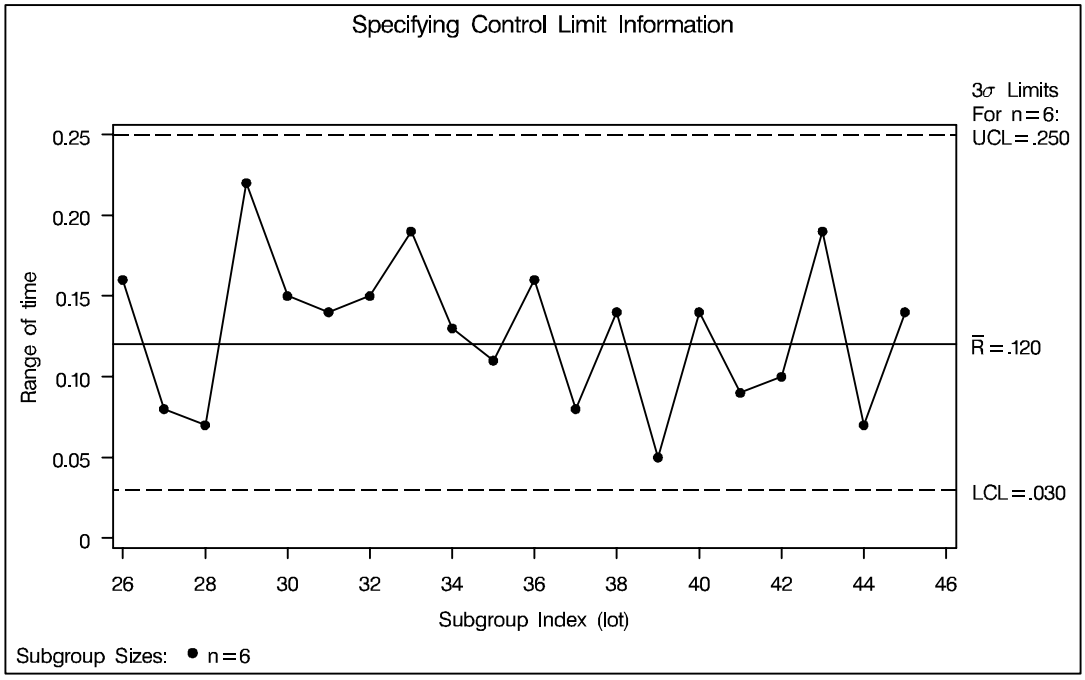
See SHWREX2 in the SAS/QC Sample Library

```
data dlimits2;
   length _var_ _subgrp_ _type_ $8;
   _var_    = 'time';
   _subgrp_ = 'lot';
   _type_   = 'STANDARD';
   _limitn_ = 6;
   _lclr_   = .03;
   _r_      = .12;
   _uclr_   = .25;
run;
```

The following statements* apply the control limits in DLIMITS2 to the measurements in DISKS2 (see page 1356) and create the $R$ chart shown in Output 39.2.1:

```
title 'Specifying Control Limit Information';
symbol v=dot;
proc shewhart data=disks2 limits=dlimits2;
   rchart time*lot;
run;
```

**Output 39.2.1.** Reading Control Limits from DLIMITS2




In some cases, a standard value ($\sigma_0$) may be available for the process standard deviation. The following DATA step creates a data set named DLIMITS3 that provides this value:

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

```
data dlimits3;
   length _var_ _subgrp_ _type_ $8;
   _var_    = 'time';
   _subgrp_ = 'lot';
   _stddev_ = .045;
   _limitn_ = 6;
   _type_   = 'STDSIGMA';
run;
```

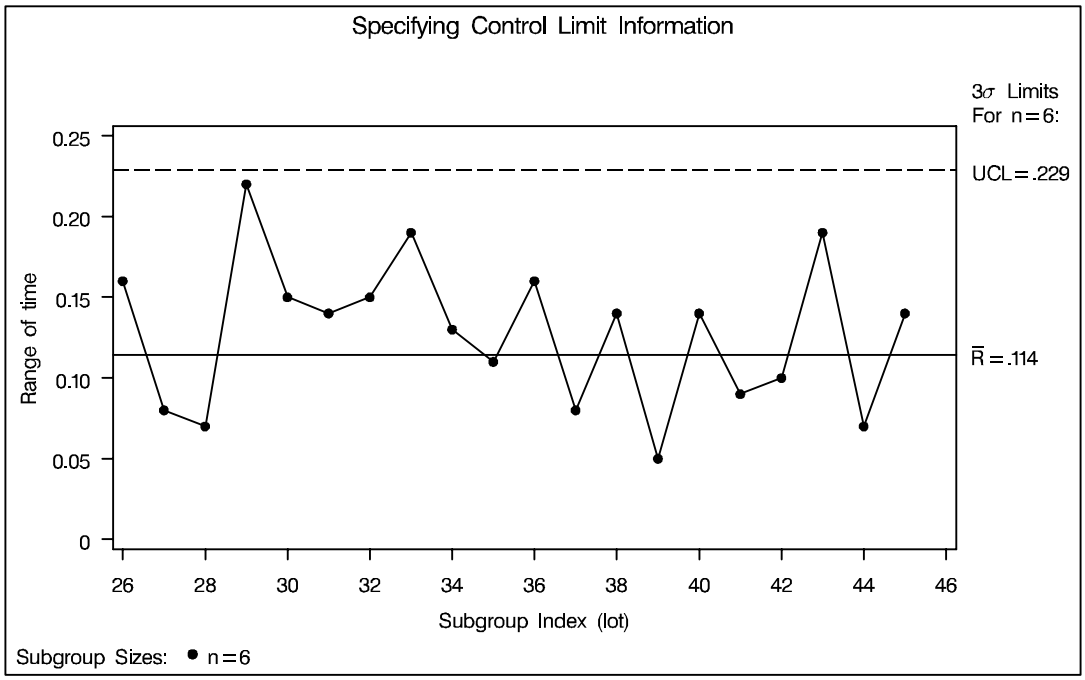
The variable _TYPE_ is a bookkeeping variable whose value indicates that the value of _STDDEV_ is a standard value rather than an estimate.

The following statements read the value of $\sigma_0$ from DLIMITS3 and create the $R$ chart shown in Output 39.2.2:

```
title 'Specifying Control Limit Information';
symbol v=dot;
proc shewhart data=disks2 limits=dlimits3;
   rchart time*lot / nolimit0;
run;
```

The NOLIMIT0 option suppresses the display of a fixed lower control limit if the value of the limit is zero (which is the case in this example).

**Output 39.2.2.** Reading in Standard Value for Process Standard Deviation




Instead of specifying $\sigma_0$ with the variable _STDDEV_ in a LIMITS= data set, you can use the SIGMA0= option in the RCHART statement. The following statements create an $R$ chart identical to the chart shown in Output 39.2.2:

```
proc shewhart data=disks;
   rchart time*lot / sigma0=.045 nolimit0;
run;
```

For more information, see "LIMITS= Data Set" on page 1373.

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