

Analyzing Quality Data with SAS®

Excerpt from

*Just Enough SAS®: A Quick-Start
Guide to SAS® for Engineers*

Robert A. Rutledge

The correct bibliographic citation for this manual is as follows: Rutledge, Robert A. 2009. *Analyzing Quality Data with SAS® (Excerpt from Just Enough SAS®: A Quick-Start Guide to SAS® for Engineers)*. Cary, NC: SAS Institute Inc.

Analyzing Quality Data with SAS® (Excerpt from Just Enough SAS®: A Quick-Start Guide to SAS® for Engineers)

Copyright © 2009, SAS Institute Inc., Cary, NC, USA

ISBN 978-1-60764-139-1

All rights reserved. Produced in the United States of America.

For a hard-copy book: No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

For a Web download or e-book: Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

U.S. Government Restricted Rights Notice: Use, duplication, or disclosure of this software and related documentation by the U.S. government is subject to the Agreement with SAS Institute and the restrictions set forth in FAR 52.227-19, Commercial Computer Software-Restricted Rights (June 1987).

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

1st electronic book, March 2009

SAS® Publishing provides a complete selection of books and electronic products to help customers use SAS software to its fullest potential. For more information about our e-books, e-learning products, CDs, and hard-copy books, visit the SAS Publishing Web site at support.sas.com/publishing or call 1-800-727-3228.

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

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



Chapter 9

Analyzing Quality Data with SAS

- 9.1 Introduction 2
- 9.2 Deciding Whether the Process Is In Control 5
- 9.3 Measuring Process Capability 10
- 9.4 Monitoring the Process 12
- 9.5 P-Charts for Fraction Failing 14
- 9.6 U-Charts for Defects per Unit 16
- 9.7 More Than Enough 18
- 9.8 Chapter Summary 20

9.1 Introduction

The terms *quality* and *reliability* are sometimes used interchangeably to refer to the relative “goodness” of manufactured products. In this book the term *quality data* refers to measurements taken on newly manufactured units, and *reliability data* refers to measurements of time to failure. For example, if your new laptop arrives with a dead screen, that’s a quality problem, but if it dies after three months of use, that’s a reliability problem. Quality data analysis is discussed in this chapter, and reliability data analysis is discussed in Chapter 10.

The three phases of statistical quality improvement are:

Phase 1: Bring the process into a state of *statistical control*.

- A process is said to be in *statistical control* if the process is stable and measurements taken on each batch of units can be reasonably assumed to come from the same statistical distribution. The steps to ensure a state of control are:
 1. Measure the variable of interest for several sequentially manufactured batches of data.
 2. Run tests for special causes to determine whether the process is in control.
 3. If the process is not in control, attempt to remove any special (or assignable) causes of variation, and then repeat the above steps.
 4. If the process is in control, proceed to Phase 2.

Phase 2: Ensure that the process is *capable* of meeting its specifications.

- A process is said to be *capable* if only an acceptably small fraction of units fail to meet the specification. If the process distribution is approximately normal, process capability is usually measured by a capability index such as C_{pk} , which is defined in Section 9.3. The steps to ensure process capability are:
 1. Compute the value of C_{pk} (or other capability index) for your stable process distribution, and compare it to your requirement (e.g. $C_{pk} > 1.33$).
 2. If the C_{pk} does not meet your requirement, improve the distribution by shifting the mean and/or removing any special causes of variation, and then re-measure the process capability.
 3. When the C_{pk} meets your requirement, proceed to Phase 3.

Phase 3: Continually monitor the process to ensure that it remains in control and capable. The steps used to monitor the process are:

1. Take measurements on periodic samples from the production units.
2. Plot the measurements on a control chart.
3. Whenever the process is out of control (i.e., fails one or more of the statistical tests), attempt to remove any special causes of variation.

This chapter illustrates some of the SAS procedures that can be used in support of each of these phases.

- PROC SHEWHART is used in the first phase to test for a state of control.
- PROC CAPABILITY is used in the second phase to measure process capability.
- PROC SHEWHART is used again in the third phase to monitor the process.

Table 9-1 lists all of the procedures in SAS/QC software, with a brief description of how each can be used. PROC ANOM is discussed in Section 5.6 and PROC RELIABILITY is discussed in Chapter 10. This chapter focuses on the use of PROC SHEWHART and PROC CAPABILITY to support the implementation of statistical process control.

The examples in this section cover the easy parts of statistical process control – creating the control charts, deciding whether the process is in control and computing the process capability index. The ISHIKAWA, PARETO, OPTEX and FACTEX procedures, as well as the ADX Interface for Design of Experiments, can be used to help with the more difficult tasks - improving the process until it is in control and capable of meeting the specification limits, and then diagnosing and eliminating special causes of variation when the process goes out of control.

Table 9-1: Procedures in SAS/QC Software

Statistical Quality Measurement and Improvement	
PROC SHEWHART	Creates Shewhart control charts to test whether a process is in control.
PROC CAPABILITY	Compares the distribution of process measurements to specifications and computes C_{pk} and other capability indices.
PROC CUSUM	Creates cumulative sum (cusum) control charts.
PROC MACONTROL	Creates uniformly or exponentially weighted moving average control charts.
Design of Experiments*	
PROC FACTEX	Constructs orthogonal factorial experimental designs.
PROC OPTEX	Searches for optimal experimental designs.
* Also consider the SAS ADX Interface for Design of Experiments. This is not a procedure but an interactive environment for creating experimental designs and analyzing the results of the experiments.	
Other	
PROC ISHIKAWA	Allows interactive creation of Ishikawa, or Fishbone, diagrams to display the factors which affect a particular quality characteristic or problem.
PROC PARETO	Creates Pareto charts showing the relative frequency of different classes of quality-related problems.
PROC ANOM	Compares treatment means and determines which are significantly different from the overall average.
PROC RELIABILITY	Analyzes event time data, for example, product failure or repair times or biological survival and recurrence times.

The examples in this chapter cover only a small fraction of the capabilities of PROC SHEWHART and PROC CAPABILITY. This should be enough to get you started, but please consult the SAS/QC documentation to explore the many other options you can choose to customize the kinds of charts, methods of analysis, and graphical output for these procedures.

9.1.1 Quality Data for the Examples

The JES.First, JES.Second, JES.Third and JES.Production data sets represent samples taken at the different phases of statistical process control.

- JES.First contains the measurements on the first set of 25 batches of units tested.
 - At this point the process is not in control.
- JES.Second contains the measurements on the second set of 25 batches.
 - At this point the process is in control, but not capable.
- JES.Third contains the measurements on the third set of 25 batches.
 - At this point the process is in control and capable.

- JES.Production contains measurements on samples from the production process.
 - At this point the process is assumed to be in control and capable and is being monitored to detect changes.

The table below shows the first 15 rows of the JES.First data set. The complete data set includes measurements on each of 10 units for each of 25 batches, for a total of 250 rows. The JES.Second, JES.Third, and JES.Production data sets have the same format.

JES.First					
Batch	Sample	Resistance	Fail	Result	Defects
1	1	15.04	0	Pass	0
1	2	14.87	0	Pass	0
1	3	14.44	0	Pass	2
1	4	14.50	0	Pass	0
1	5	15.39	0	Pass	0
1	6	16.66	0	Pass	0
1	7	10.60	1	Low Res	0
1	8	17.54	0	Pass	0
1	9	16.88	0	Pass	0
1	10	15.69	0	Pass	0
2	1	12.22	1	Low Res	0
2	2	17.07	0	Pass	1
2	3	13.00	0	Pass	0
2	4	20.18	0	Pass	0
2	5	16.54	0	Pass	0
(partial listing)					

- Each row contains a resistance measurement and a defect count for the corresponding unit.
- If the resistance measurement is outside the specification limits of 12.5-22.5:
 - Result is set to “Low Res” or “High Res”.
 - Fail is set to 1.
- A different type of control chart is appropriate for each of these three different types of measurement:
 - *Resistance* is a continuous variable which can be charted with a Mean Chart.
 - *Fail* is a binary variable which can be charted with a P-Chart.
 - *Defects* is a count per unit variable which can be charted with a U-Chart.

PROC SHEWHART can create Mean control charts using raw data (i.e., one row per measurement) such as in the JES.First data set, or summary data containing statistics for each batch. The JES library includes summary data sets created with PROC MEANS for each set of raw data. The JES.First_Sum data set shown here is the summarized version of the JES.First data set.

JES.First_Sum				
Batch	ResistanceN	ResistanceX	ResistanceS	ResistanceR
1	10	15.16	1.92	6.94
2	10	15.28	3.68	10.66
3	10	14.72	2.56	8.45
4	10	16.80	2.19	7.00
5	10	13.76	3.91	13.47
6	10	14.69	1.57	4.67
7	10	16.04	4.83	13.96
8	10	15.32	3.11	11.04
9	10	17.13	3.51	11.31
10	10	19.42	1.79	5.60
11	10	18.15	2.90	8.83
12	10	17.27	3.23	11.06
13	10	17.16	3.39	11.08
14	10	16.96	2.81	8.46
15	10	18.24	2.42	6.56
16	10	16.97	3.08	9.48
17	10	17.11	1.92	7.01
18	10	18.30	2.37	6.39
19	10	17.66	2.99	9.18
20	10	18.33	3.93	10.88
21	10	18.20	3.18	9.55
22	10	18.80	3.34	8.46
23	10	18.52	3.27	11.34
24	10	19.59	3.72	10.73
25	10	19.99	2.79	8.90

In order to use a summary data set instead of the raw data, PROC SHEWHART requires that the variable names be constructed using the analysis or process variable name as the common prefix, followed by a one character code which identifies the statistic. In the example above:

- ResistanceN = the number of resistance measurements in the batch.
- ResistanceX = the mean of the resistance measurements in the batch.
- ResistanceS = the standard deviation of the resistance measurements in the batch.
- ResistanceR = the range of the resistance measurements in the batch.

The raw data sets are used in the examples in Section 9.2, and the summary data sets are used in the examples in Section 9.4.

Sections 9.2-9.4 go through the three phases of statistical quality improvement for the resistance measurements. Sections 9.5 and 9.6 illustrate the use of P-Charts and U-Charts for the Fail and Defects measurements, respectively.

9.2 Deciding Whether the Process Is In Control

In the first phase of the quality improvement process, measurements are taken on sequentially produced batches of units to determine whether the process is in a state of control. The JES.First data set contains this initial set of samples. You can check for control of a continuous variable by creating a Mean Chart, and applying tests for process stability. The code sample uses PROC SHEWHART to create a Mean control chart and perform tests for special causes based on the Western Electric rules.

```

GOPTIONS RESET=ALL;
ODS HTML PATH="&JES.SG/S_9" (URL=NONE) BODY="SPC.html";
ODS GRAPHICS ON / RESET IMAGENAME="F9_1_";
ODS OUTPUT XCHART=Sum_Tab; ❶
PROC SHEWHART DATA=JES.First; ❷
    XCHART Resistance*Batch /STDDEVIATIONS
    TESTS=1 TO 8 ❸
    ZONELABELS TABLEALL;
RUN;

```

❶ The ODS OUTPUT statement requests that summary statistics be saved to the Sum_Tab data set.

❷ The PROC SHEWHART statement calls for analysis of the JES.First data set. The XCHART statement requests a Mean Chart of Resistance, grouped by Batch. The STDDEVIATIONS option specifies that the sample standard deviation, rather than the range, be used to compute the control limits. The limits for the chart can be specified either as sigma limits or probability limits. Sigma limits are computed as a specified number of standard deviations from the mean, while probability limits are set to achieve a specified probability of false alarm when the process is in control. The default is 3 sigma limits, corresponding to a false alarm probability of 0.135%, based on the normal distribution.

❸ The TESTS = 1 TO 8 option requests the 8 standard tests for special causes (see Table 9-2). The ZONELABELS option requests that the A, B and C zones be displayed on the chart. The TABLEALL option sends a table of batch summary statistics to the Output window.

When you run the code, the Sum_Tab data set and the Mean Chart (Figure 9-1) are created. The Sum_Tab data set (not shown) is similar to the table written to the Output window, but does not include the test descriptions. The chart shows a very clear trend of increasing resistance, and several of the tests are positive, so the process is clearly not in control.

Table 9-2: Tests for Special Causes Using the Western Electric Rules

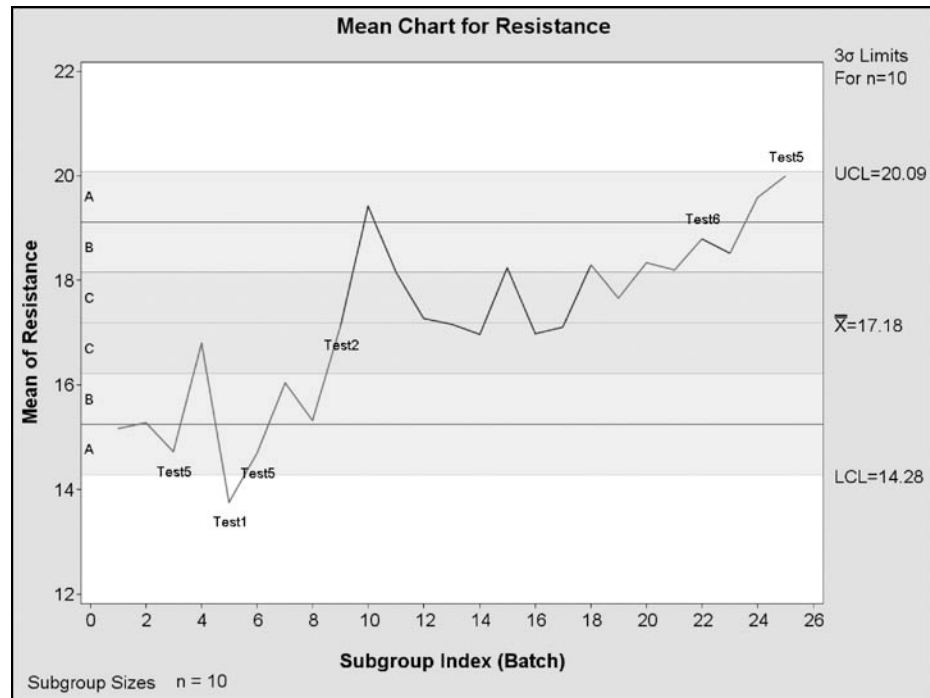
Index	Pattern Description
1	One point beyond Zone A (outside the control limits)
2	Nine points in a row in Zone C or beyond on one side of the central line
3	Six points in a row steadily increasing
4	Fourteen points in a row alternating up and down
5	Two out of three points in a row in Zone A or beyond
6	Four out of five points in a row in Zone B or beyond
7	Fifteen points in a row in Zone C on either or both sides of the central line
8	Eight points in a row on either or both sides of the central line with no points in Zone C
Zone Definitions	
Zone A	Distance from the center line (mean) is between 2 and 3 sigma
Zone B	Distance from the center line (mean) is between 1 and 2 sigma
Zone C	Distance from the center line (mean) is less than sigma

This is part of the output written to the Output window by the TABLEALL option.

Means Chart Summary for Resistance						
-----3 Sigma Limits with n=10 for Mean-----						
Batch	Subgroup Sample Size	Lower Limit	Subgroup Mean	Average Mean	Upper Limit	Special Tests Signaled
1	10	14.280573	15.161478	17.183435	20.086296	
2	10	14.280573	15.279966	17.183435	20.086296	
3	10	14.280573	14.721389	17.183435	20.086296	5
4	10	14.280573	16.803839	17.183435	20.086296	
5	10	14.280573	13.758538	17.183435	20.086296	1 6
(___ Rows for Batches 10-21 Omitted)						
22	10	14.280573	18.797901	17.183435	20.086296	6
23	10	14.280573	18.521382	17.183435	20.086296	
24	10	14.280573	19.585652	17.183435	20.086296	
25	10	14.280573	19.994593	17.183435	20.086296	5

Test Descriptions	
Test 1	One point beyond Zone A (outside control limits)
Test 2	Nine points in a row on one side of center line
Test 5	Two out of three points in a row in Zone A or beyond
Test 6	Four out of five points in a row in Zone B or beyond

Figure 9-1: Using the XCHART Statement in PROC SHEWHART



Because the initial process is found to be out of control, action is taken to remove special causes of variation, and then a second set of 25 batches of units are produced with the new process and measured as before. The next bit of code creates a Mean and Standard Deviation Chart from the new data set, JES.Second. This is similar to the code in the previous example, except that an XSCHART statement is used instead of XCHART.

```
/*==== Mean Chart for Second 25 Batches ====*/
ODS GRAPHICS ON / RESET IMAGENAME="F9_2_";
PROC SHEWHART DATA=JES.Second;
  XSCHART Resistance*Batch /
  TESTS=1 TO 8
  TABLEALL ZONELABELS;
RUN;
```

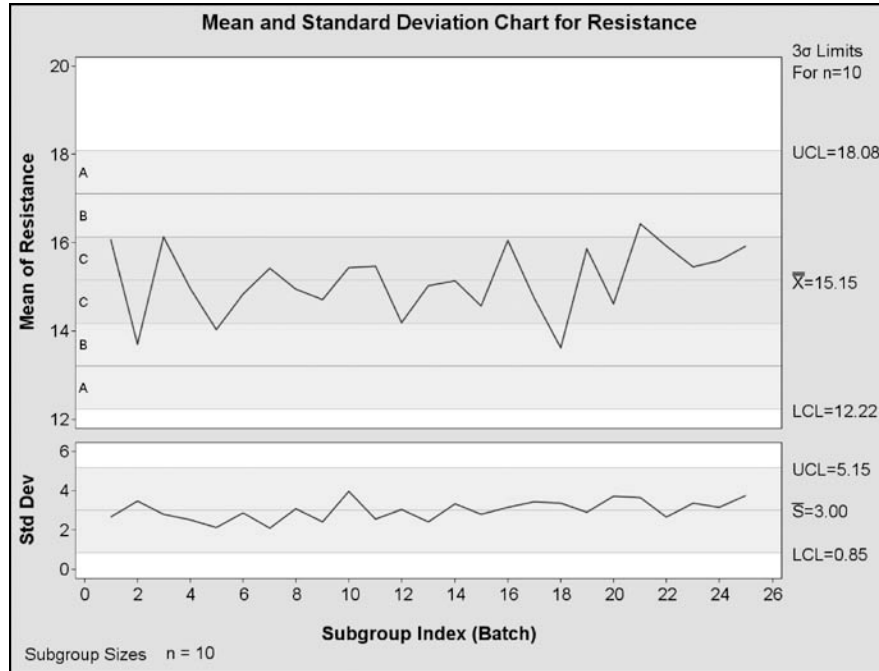
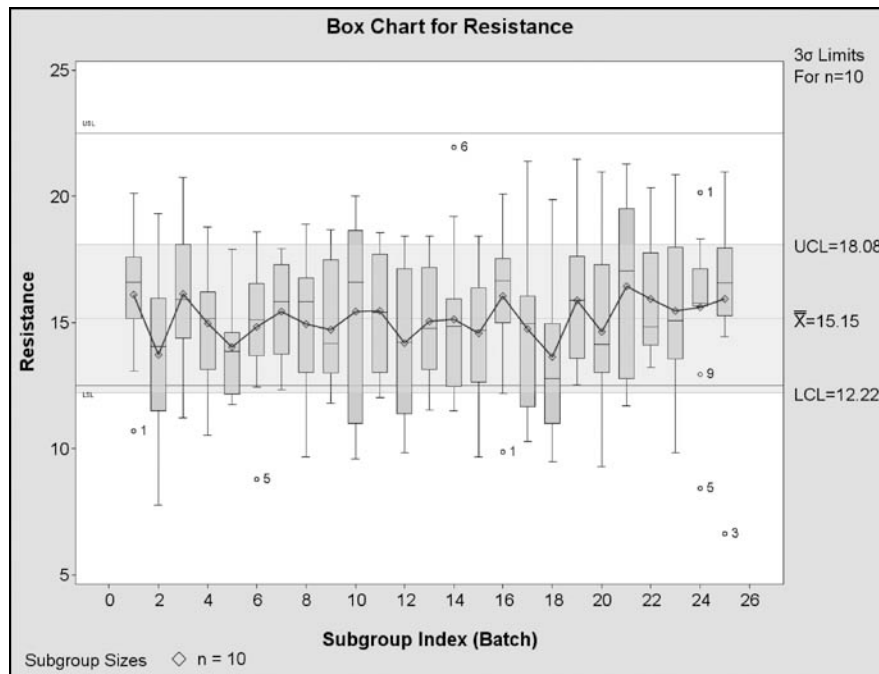
In this case, you can see from the resulting plot (Figure 9-2) that none of the eight tests are positive, so you can assume that the process is in control, with a predictable distribution, and move on to Phase 2.

Next you need to compare this distribution to the product specifications for resistance to see if the process is capable of meeting the specification limits. In Section 9.3, PROC CAPABILITY is used to assess the capability of the process, but first a box plot is created to provide a view of process trend vs. the specifications. You have already seen four ways to create box plots, using PROC BOXPLOT in Section 5.5, PROC ANOM in Section 5.6, PROC GPLOT in Section 6.2.2.7 and PROC SGPLOT in Section 8.3.10. The BOXCHART statement of PROC SHEWHART, which creates a box plot superimposed on a Mean Chart, uses many of the same options available for box plots created with PROC ANOM or PROC BOXPLOT.

```
/*==== Box Plots for Second 25 Batches ====*/
ODS GRAPHICS ON / RESET IMAGENAME="F9_3_";
PROC SHEWHART DATA=JES.Second;
  BOXCHART Resistance*Batch / ❶
  BOXCONNECT BOXSTYLE=SCHEMATICID ❷
  VREF=12.5 22.5 VREFLABELS='LSL' 'USL'; ❸
  ID Sample; ❹
RUN;
```

- ❶ The BOXCHART statement requests a box plot of Resistance by Batch. In a BOXCHART statement, the sample standard deviation is used by default, so the STDDEVIATIONS option is not needed.
- ❷ The BOXCONNECT option requests a line joining the means for each batch. The BOXSTYLE=SCHEMATIC option requests a schematic box plot in which the outliers are labeled with the value of an ID variable. See Section 5.5 for a discussion of Schematic and Skeletal box plots.
- ❸ The VREF and VREFLABELS options add reference lines at the specification limits.
- ❹ The ID option specifies the variable used to label the outlier points.

The resulting box plot (Figure 9-3) includes the same mean line and control limits as the Mean Chart in Figure 9-2. Note that the control limits apply to the mean values only, and not to any other features of the box plot. It is, however, fair to compare the box plots to the specification limits, LSL and USL. You can see that a large percentage of the units are below the lower specification limit, so the process, though in a state of control, is not yet capable of meeting the specifications. In Section 9.3 PROC CAPABILITY is used to compute the C_{pk} capability index, which provides a more quantitative measure of the process capability.

Figure 9-2: Using the XSCHEART Statement in PROC SHEWHART**Figure 9-3:** Using the BOXCHART Statement in PROC SHEWHART

9.3 Measuring Process Capability

When the process is in a state of control, the measurements on each batch are considered to be random samples from the same distribution, so you can therefore use the aggregate data from all batches to estimate the process capability index, C_{pk} , defined by:

$$C_{pk} = \min\left(\frac{USL - \mu}{3\sigma}, \frac{\mu - LSL}{3\sigma}\right)$$

where μ and σ are the process mean and standard deviation. C_{pk} measures the distance of the process mean, μ , from the nearest specification limit, in units of 3σ . If the measured variable is normally distributed, then a C_{pk} of 1 means that the nearest specification limit is 3σ from the mean, so you can expect 0.135% of the units to be beyond the limit. If $C_{pk} = 1.33$, then the specification limit is 4σ from the mean, and only 0.003% of the units will be beyond the limit. A process is generally considered to be:

- *Incapable* if $C_{pk} < 1$
- *Marginally Capable* if $1 < C_{pk} < 1.33$
- *Capable* if $C_{pk} > 1.33$

However, some would recommend putting the boundary at 1.67 (5σ) instead of 1.33 (4σ) in these definitions.

The code below uses PROC CAPABILITY to compute C_{pk} for the second set of 25 batches.

```
/*==== Process Capability - Second 25 Batches =====*/
ODS GRAPHICS ON / RESET IMAGENAME="F9_4_";
PROC CAPABILITY DATA=JES.Second OUTTABLE=Second_Tab; ❶
  VAR Resistance; ❷
  SPEC TARGET=17.5 LSL=12.5 USL=22.5 ❸
  CLEFT=Red CRIGHT=Red;
  HISTOGRAM / NORMAL CFILL=GRAYD0 ; INSET N MEAN STD CPK; ❹
RUN;
```

- ❶ The PROC CAPABILITY statement requests that the analysis be run on the JES.Second data set. The OUTTABLE option requests that summary statistics be saved to the Second_Tab data set.
- ❷ The VAR statement specifies that Resistance is the variable to be analyzed.
- ❸ The SPEC statement specifies the target value and specification limits for Resistance. The CLEFT and CRIGHT options specify the fill colors to be used for areas outside of the specification limits.
- ❹ The HISTOGRAM statement requests a histogram of the data. The NORMAL option requests a normal density fit to the histogram. The CFILL option specifies the fill color for the histogram. The INSET statement specifies statistics to be displayed in an inset on the plot.

When you run the code, the Second_Tab data set and the plot shown in Figure 9-4 are created. The Second_Tab data set also includes many other statistics, including skewness, kurtosis, percentiles and other capability indices.

Second_Tab (selected columns)						
Obs	_VAR_	_NOBS_	_MEAN_	_STD_	_MEDIAN_	_CPK_
1	Resistance	250	15.1502	2.98951	15.1932	0.29550

Because the C_{pk} is only 0.295, the process is not yet capable, so actions are taken to increase the mean and reduce the variability to improve the C_{pk} . Then a third set of 25 batches of units are produced with the new process and measured as before. Figure 9-5 shows the result of running PROC CAPABILITY on the new data, which is in the JES.Third data set. The process is now marginally capable, with a C_{pk} of 1.1, so you can proceed to Phase 3, monitoring the process.

Figure 9-4: Using PROC CAPABILITY on JES.Second - The Second Set of 25 Batches

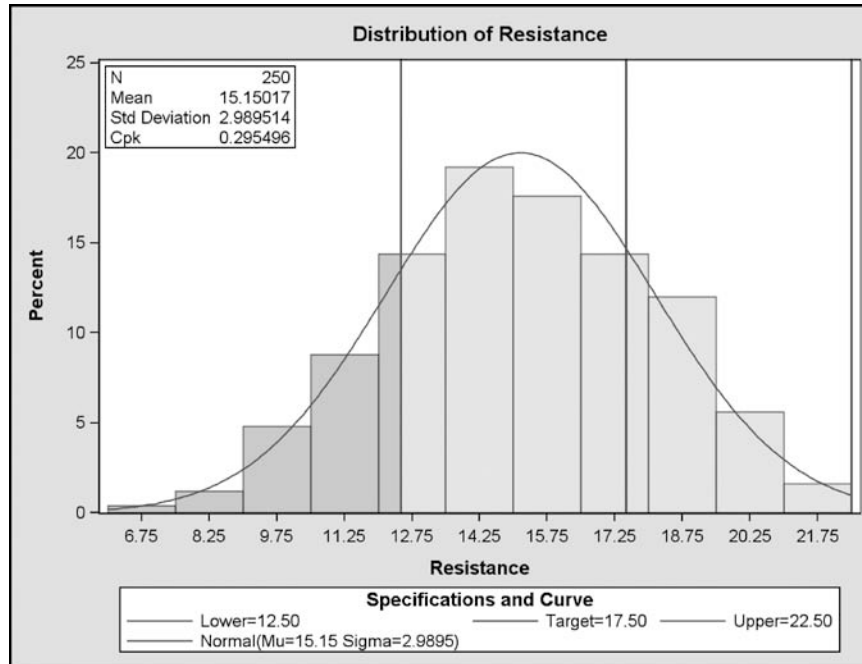
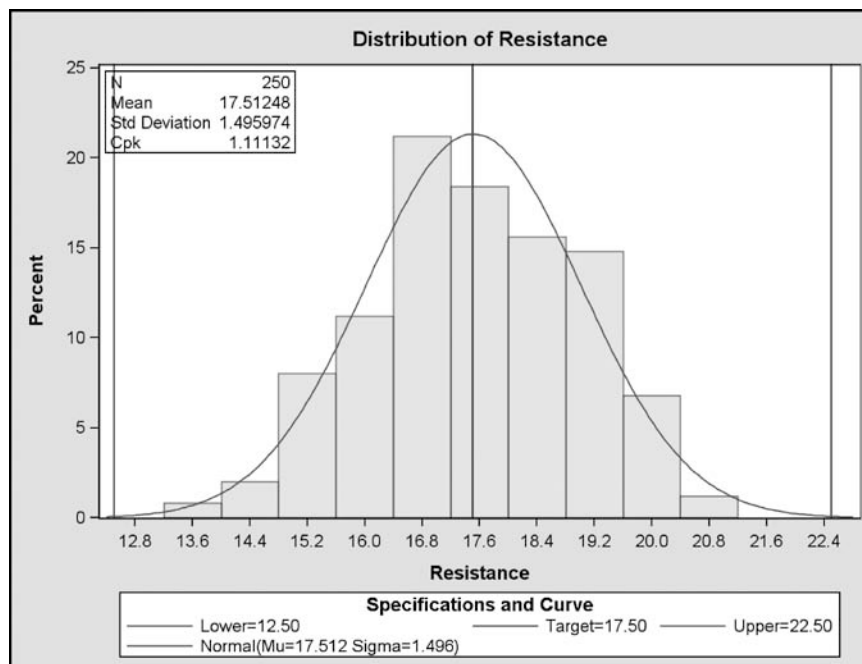


Figure 9-5: Using PROC CAPABILITY on JES.Third - The Third Set of 25 Batches



9.4 Monitoring the Process

After the process is in control and at least marginally capable, you need to monitor the process to ensure that it stays that way. The same control charts and tests for special causes that are used to test for a state of control can also be used to detect changes that cause the process to go out of control. You should use the control limits determined during the previous phase to judge the stability of the production process. The next code sample creates a Mean Chart for the JES.Third data set, and saves the control limits to a data set. This example uses the summary data set, JES.Third_Sum, instead of the raw data set, JES.Third, just to illustrate the syntax for running PROC SHEWHART with summary data. The relationship between the raw and summarized data sets is explained in Section 9.1.1.

```
/*==== Mean Chart for Resistance - Third 25 Batches =====*/
ODS GRAPHICS ON / RESET IMAGENAME="F9_6_";
PROC SHEWHART HISTORY=JES.Third_Sum; ❶
    XCHART Resistance*Batch / STDDEVIATIONS OUTLIMITS=ResLim ❷
    TESTS=1 TO 8 ZONELABELS TABLEALL;
RUN;
```

- ❶ This run of PROC SHEWHART uses the summarized data records in the JES.Third_Sum data set. Using the HISTORY= option instead of DATA= tells SAS that this is a summary data set.
- ❷ Note that in this case, *Resistance* is not the name of a variable in the data set. It is the prefix for the three variable names used to construct the chart, ResistanceN, ResistanceX and ResistanceS, which are described in Section 9.1.1. The OUTLIMITS=ResLim option requests that the computed control limits and other statistics be saved to the ResLim data set. The OUTLIMITS option works exactly the same way if you use raw data instead of summary data.

This table shows selected columns in the ResLim data set, including the statistics used to create the control limits for the production data.

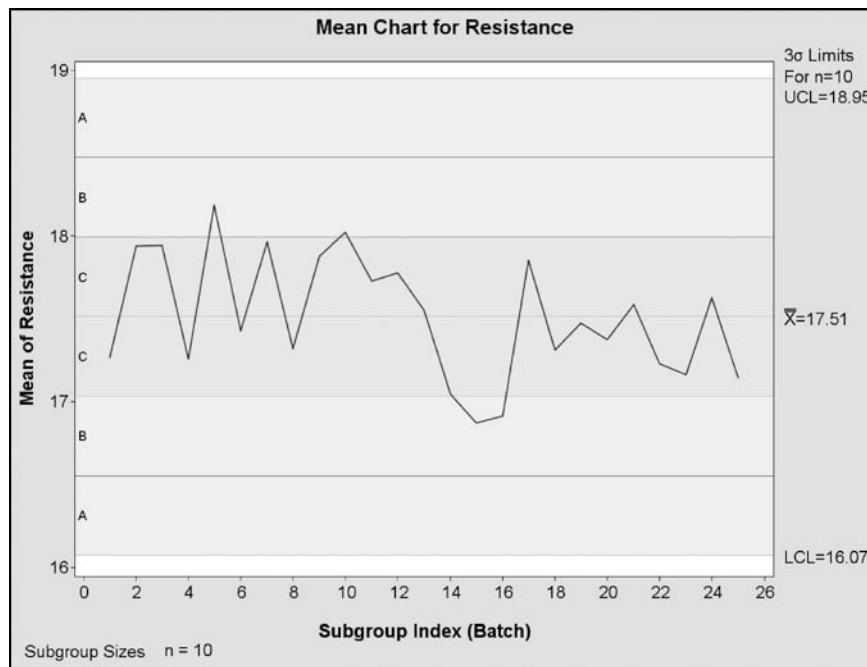
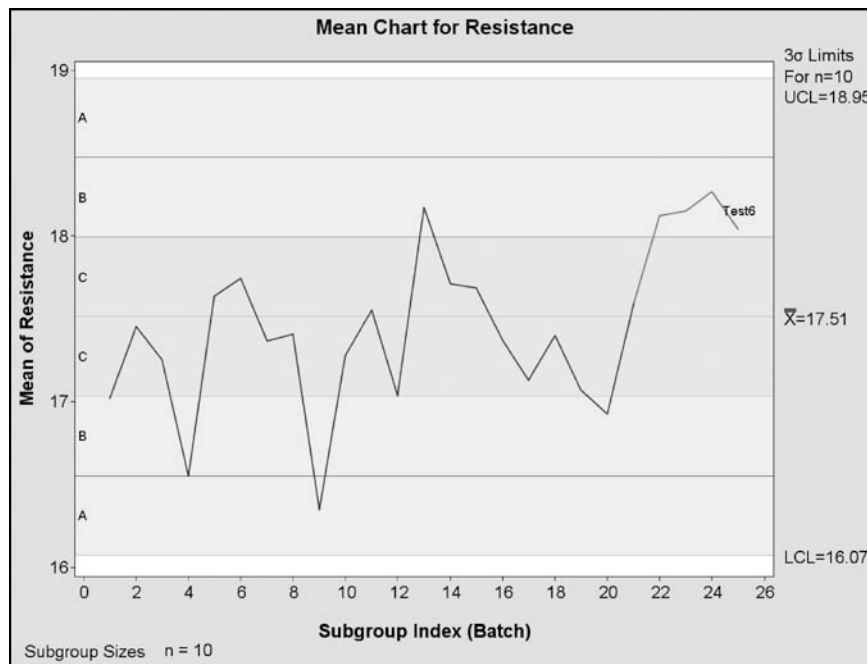
ResLim (selected columns)						
VAR	_SUBGRP_	_LIMITN_	_LCLX_	_MEAN_	_UCLX_	_STDDEV_
Resistance	Batch	10	16.0712	17.5125	18.9537	1.51922

The next bit of code creates a Mean Chart for the production data using the control limits created above.

```
/*==== Mean Chart for Resistance - Production Samples =====*/
ODS GRAPHICS ON / RESET IMAGENAME="F9_7_";
PROC SHEWHART HISTORY=JES.Production_Sum LIMITS=ResLim; ❶
    XCHART Resistance*Batch /
    TESTS=1 TO 8 ZONELABELS TABLEALL;
RUN;
```

- ❶ The LIMITS=ResLim option tells SAS to use the control limits from the ResLim data set. By default, the batch means are only plotted for batches with sample size equal to _LIMITN_ in the ResLim data set (all batches in this example), but you can use the ALLN option to force plotting of all means, and the NMARKERS option to use a different symbol to identify batches with different sample sizes.

The resulting plot is shown in Figure 9-7. Note that the control limits are the same as the limits in Figure 9-6, which was created with the first code sample using the JES.Third_Sum data set. You can see that the process is out of control because four out of five points in a row are in Zone B or beyond (Test 6 of Table 9-2). At this point you need to determine the cause of the apparent process shift, and take action to bring the process back into control.

Figure 9-6: Using the HISTORY Option of PROC SHEWHART**Figure 9-7:** Using the LIMITS Option of PROC SHEWHART

9.5 P-Charts for Fraction Failing

The Mean and Standard Deviation Charts discussed in Sections 9.2 and 9.4 are generally appropriate for continuous variables, such as resistance, where the subgroup means can be assumed to follow the normal distribution. But quality data is often captured as the outcome of a pass/fail test, or a defect count for each unit. For these kinds of data, the subgroup means are not well approximated by the normal distribution unless the sample sizes are very large. It is usually more appropriate to use a P-Chart, based on the binomial distribution, for pass/fail data or a U-Chart, based on the Poisson distribution, for defect per unit data. You can create these charts using the PCHART and UCHART statements of PROC SHEWHART. The PCHART statement is illustrated in this section, and the UCHART statement is illustrated in Section 9.6.

Like the Means Chart, the P-Chart can be used in different ways during the different phases of quality improvement described in Section 9.1. You can use the P-Chart during Phase 1 to test whether the process is in control, and again in Phase 3 to monitor the process. The example in this section is focused on the monitoring phase. Continuing the previous example, the Resistance variable is assumed to be in control, and the process is now being monitored by a simple pass/fail test on a large sample, which is less expensive than the more precise measurement of resistance. The test results are in the Fail variable of the JES.Production_2 data set, which is similar to the JES.First data set shown in Section 9.1.1, except that the sample size for each batch is 1000 instead of 10.

PROC SHEWHART requires that the input data for P-Charts be in summary form, where each observation has summary data for one subgroup, including the number of units tested and the number or proportion of nonconforming units. This code creates the Fail_Tab data set with the required summary data.

```
/*===== P Chart for Fraction Failing - First 25 Batches =====*/
PROC MEANS DATA=JES.Production_2 NWAY NOPRINT;
  CLASS Batch;
  OUTPUT OUT=Fail_Tab N=Num_Test SUM(Fail)=Num_Fail;
RUN;
```

Fail_Tab				
Batch	_TYPE_	_FREQ_	Num_Test	Num_Fail
1	1	1000	1000	1
2	1	1000	1000	1
3	1	1000	1000	0
4	1	1000	1000	4
5	1	1000	1000	3
(Batches 6-20 Omitted)				
21	1	1000	1000	9
22	1	1000	1000	8
23	1	1000	1000	9
24	1	1000	1000	5
25	1	1000	1000	8

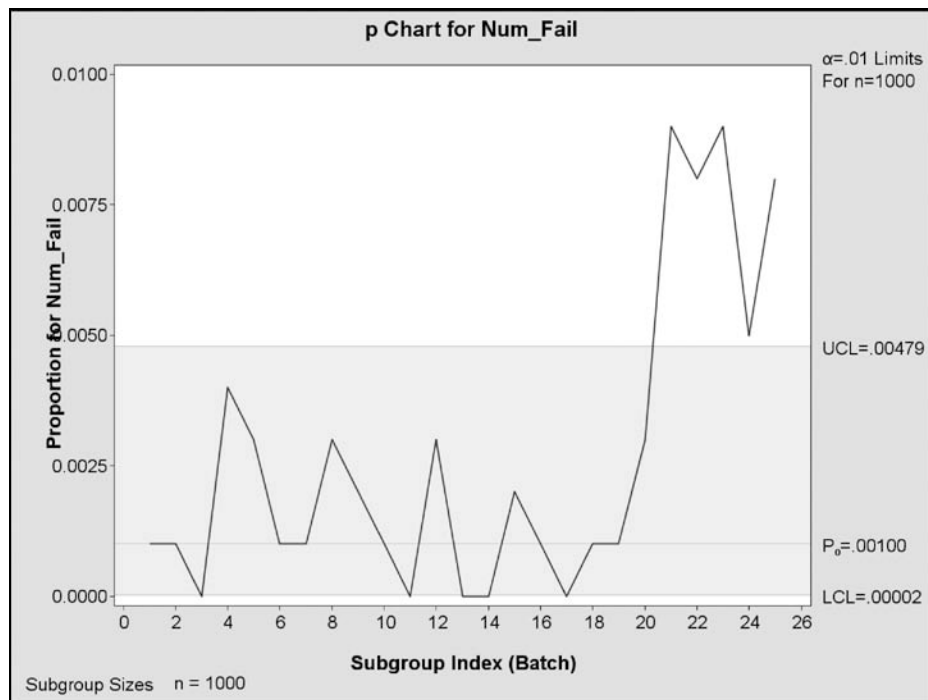
As for the Means Chart (see Section 9.2), the limits for the P-Chart can be specified either as sigma limits or probability limits. Sigma limits are computed as a specified number of standard deviations from the mean, while probability limits are set to achieve a specified probability of false alarm when the process is in control. When the P-Chart is used in Phase 1, the central line is usually computed as the overall average of the data, but in Phase 3 it is more appropriate to specify the known value of P, based on the previous process average. For this example it is assumed that the known process average is 0.1%, and a P-Chart is constructed to monitor the process and detect any changes.

The next code sample uses PROC SHEWHART to create the required P-Chart using probability limits.

```
ODS GRAPHICS ON / RESET IMAGENAME="F9_8_";
PROC SHEWHART DATA=Fail_Tab;
  PCHART Num_Fail*Batch / SUBGROUPN=Num_Test ❶
  P0 = .001 PSYMBOL=P0 ❷
  ALPHA=.01; ❸
RUN;
```

- ❶ The PCHART statement specifies a P-Chart for Num_Fail by Batch. The SUBGROUPN option specifies that the Num_Test variable contains the number of units tested in each batch.
- ❷ The P0=.001 option sets the known value of P for the chart. If this option is omitted, P is computed as the average fraction failing in the sample. The PSYMBOL option specifies that the central line be labeled 'P₀' to indicate that it is a user-specified value and not the sample average.
- ❸ The ALPHA=.01 option specifies that the chart limits should be probability limits, with a false alarm probability of 0.01. You can use SIGMAS=K to request K-sigma limits instead. If neither the ALPHA nor SIGMAS options are specified, 3-sigma limits are used by default.

Figure 9-8: Using the PCHART Statement with PROC SHEWHART



Beginning with Batch 21, several points are above the upper control limit, so the process is clearly out of control and steps should be taken to understand and eliminate the cause of the change.

9.6 U-Charts for Defects per Unit

The U-Chart, which is based on the assumption that the response variable has a Poisson distribution, is usually a good choice when the variable being controlled is the count of defects, or nonconformities, per unit. Because the structure of the data used for P-Charts and U-Charts is so similar - number of units tested and number of failures or defects for each batch - there is sometimes confusion about which chart to use. The answer depends on whether a single unit can 'fail' more than once. If the test is pass/fail, and the number of failures per unit is either zero or one, then the sample data follows the binomial distribution, and the P-Chart should be used. But if a single unit can have more than one defect, for example if the number of defective parts on each circuit board are counted, then the sample data can be represented by the Poisson distribution, and a U-Chart should be used.

Following these rules, the Fail and Defects variables in the JES.First data set (see Section 9.1.1) should use the P-Chart and U-Chart, respectively. The example in this section creates a U-Chart for the Defects variable in the JES.Production_2 data set, which, as noted in Section 9.5, is similar to JES.First but with a batch size of 1000 units. As in Section 9.5, the process is assumed to be in control, with a known defect rate, and the U-Chart is being used to monitor the process.

PROC SHEWHART requires that the input data for U-Charts be in summary form, where each observation has summary data for one subgroup, including the number of units inspected and the total number of defects found on all units in the subgroup. This code sample creates the Defect_Tab data set with the required summary data.

```
/*===== U Chart for Defects per Unit - First 25 Batches =====*/
PROC MEANS DATA=JES.Production_2 NWAY NOPRINT;
  CLASS Batch;
  OUTPUT OUT=Defect_Tab N=Num_Inspect SUM(Defects)=Num_Defects;
RUN;
```

Defect_Tab (first 10 rows)				
Batch	_TYPE_	_FREQ_	Num_Inspect	Num_Defects
1	1	1000	1000	14
2	1	1000	1000	10
3	1	1000	1000	10
4	1	1000	1000	22
5	1	1000	1000	12
6	1	1000	1000	10
7	1	1000	1000	14
8	1	1000	1000	8
9	1	1000	1000	10
10	1	1000	1000	9

The limits for the U-Chart can also be specified either as sigma limits, computed as a specified number of standard deviations from the mean, or probability limits, set to achieve a specified probability of false alarm when the process is in control. The central line of the U-Chart can be computed from the sample data, or specified by the user based on requirements or a previously measured process average. For this example it is assumed that the known process average is 0.1%, and a U-Chart is constructed to monitor the process and detect any changes.

The code below uses PROC SHEWHART to create a U-Chart using probability limits.

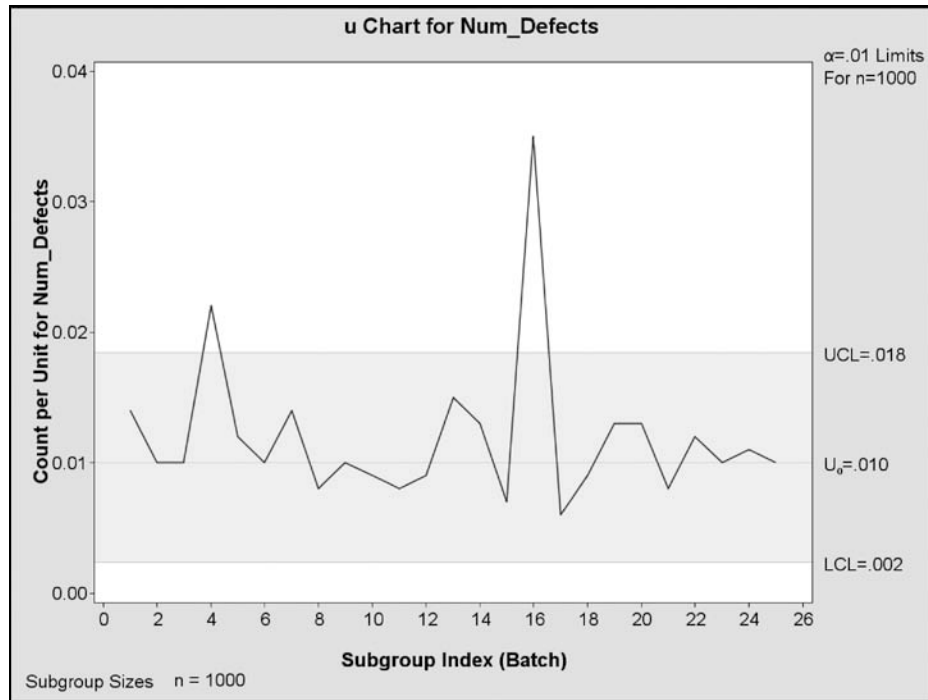
```

ODS GRAPHICS ON / RESET IMAGENAME="F9_9_";
PROC SHEWHART DATA=Defect_Tab;
  UCHART Num_Defects*Batch / SUBGROUPN=Num_Inspect ❶
  U0=.01 ❷
  USYMBOL=U0
  ALPHA = .01; ❸
RUN;

```

- ❶ The UCHART statement specifies a U-Chart for Num_Defects by Batch. The SUBGROUPN specifies that the Num_Inspect variable contains the number of units inspected in each batch.
- ❷ The U0=.01 option sets the target value of U for the chart. If this option is omitted, U is computed as the average number of defects per unit in the sample. The USYMBOL option specifies that the central line be labeled 'U₀' to indicate that it is a user-specified value and not the sample average.
- ❸ The ALPHA=.01 option specifies that the chart limits be probability limits, with a false alarm probability of 1%. You can use the SIGMAS=K option to request K-sigma limits instead. If neither the ALPHA nor SIGMAS options are specified, 3-sigma limits are used by default.

Figure 9-9: Using the UCHART Statement with PROC SHEWHART



The points for batches 4 and 16 are beyond the control limits, so the process is not in control, and action should be taken to eliminate the causes of variation. Note that it is the *process* that is out of control, not the individual batches, so it is not correct to conclude that the process is “in control again” after batch 16, although this type of reasoning is often used as an excuse for inaction.

9.7 More Than Enough

The Shewhart charts described in the previous sections are the control charts most widely known and used in industry, but SAS/QC also includes procedures for creating other types of control charts. The MACRONCONTROL procedure can be used to create *Moving Average* charts and *Exponentially Weighted Moving Average* charts, and the CUSUM procedure can be used to create *Cumulative Sum*, or *Cusum*, charts. You can learn more about these procedures from the SAS online documentation. This section includes a very brief introduction to the CUSUM procedure.

Cusum control charts are used for the same purpose as Shewhart control charts in Phase 3 of the quality improvement process – to detect significant process shifts by examining the sample means of a series of subgroups. But instead of comparing each subgroup mean to the control limits independently, a Cusum Chart plots the cumulative sum of the subgroup means, suitably normalized, and tests for a significantly increasing or decreasing trend. The cusum statistic is defined recursively as

$$S_0 = 0$$

$$S_{t+1} = S_t + \frac{\bar{X}_t - \mu_0}{\sigma_0 / \sqrt{n}}$$

where μ_0 and σ_0 are the known process mean and standard deviation, and n is the subgroup size. The theoretical advantage of a Cusum Chart over a Means Chart is that it can more quickly detect small shifts in the process mean because each value of the cusum statistic includes data from all previous subgroups. Therefore, the effect of a small shift can accumulate over several samples, and provide enough statistical evidence to trip the alarm before any one subgroup mean has exceeded its control limit. The difficulty, of course, is that the Cusum Chart is more difficult to create, and MUCH more difficult to explain. The CUSUM procedure can at least remove the difficulty in constructing the charts.

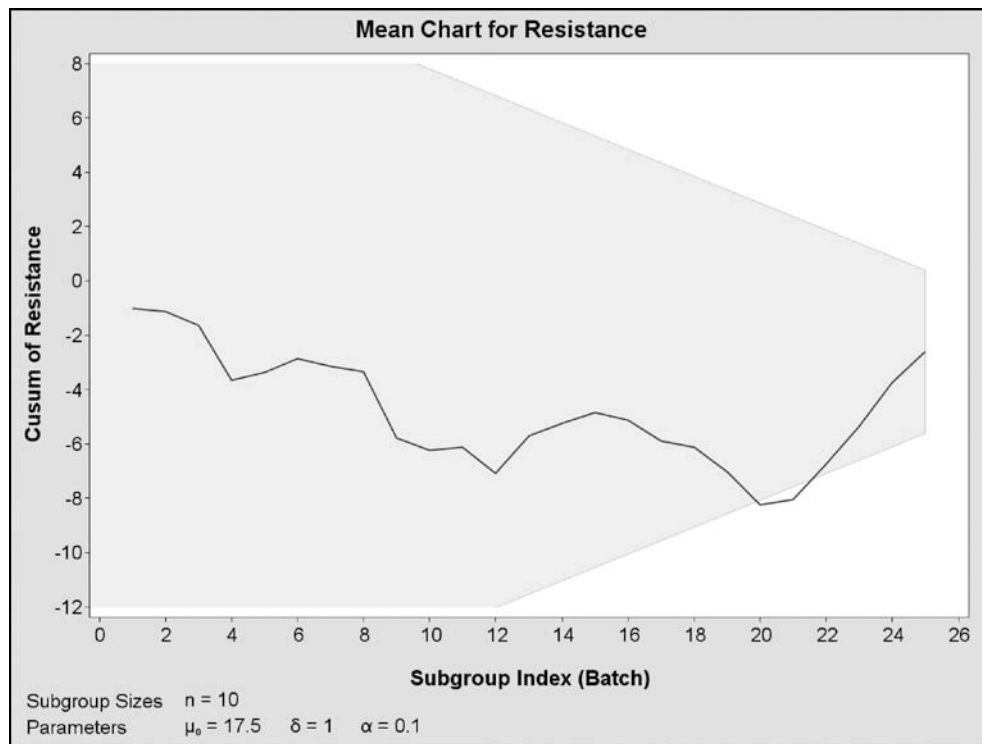
This example creates a Cusum Chart for the JES.Production data that can be compared to the Means Chart of the same data created in Section 9.4 and shown in Figure 9-7.

```
ODS GRAPHICS ON / RESET IMAGENAME="F9_10_";
PROC CUSUM DATA=JES.Production; ❶
  XCHART Resistance*Batch / ❷
  MU0=17.5 SIGMA0=1.5 ❸
  DELTA=1 ALPHA = .1 ❹
  VAXIS=-12 TO 8 BY 2; ❺
RUN;
```

- ❶ The PROC CUSUM statement requests analysis of the JES.Production data set.
- ❷ The XCHART statement requests a Cusum Chart of Resistance, with subgroups determined by the Batch variable.
- ❸ The MU0 and SIGMA0 options specify the known process mean and standard deviation. These are the process averages computed in Section 9.4 for the JES.Third data set, and saved to the ResLim data set.
- ❹ The DELTA option specifies the size of the mean shift to be detected, as a multiple of the standard error. The ALPHA option specifies the probability of a Type 1 error, or false alarm.
- ❺ The VAXIS option specifies the vertical axis for the plot.

Figure 9-10 shows the graph created by the sample code.

Figure 9-10: Cusum Chart Created with the XCHART Statement of PROC CUSUM



The plotted line is the cusum statistic $S(t)$. The shaded area, or *V-Mask*, is the acceptance or continuation region, and is positioned so that the most recent point (Batch 25) is at the center of the vertical edge on the right. The process is considered to be out of control if any of the plot points are outside the V-Mask. In this example you can see that the points for batches 20 and 21 are outside the V-Mask, so a significant mean shift has been detected. The chart is basically looking backwards from the most recent subgroup to try to detect a trend in the previous several subgroups, and the dimensions of the V-Mask are computed to detect a shift of the size specified by the DELTA option as quickly as possible, with the false alarm probability specified by the ALPHA option.

If you compare this plot with Figure 9-7, you can see that the Cusum Chart has in fact fulfilled its promise by detecting the mean shift before any of the points on the Means Chart exceeded the upper control limit. On the other hand, the Means Chart did trigger an alarm at the same time based on one of the Western Electric rules, because four out of five points in a row are in Zone B or beyond.

9.8 Chapter Summary

9.8.1 Recap

After finishing this chapter you should know how to:

- Use PROC SHEWHART to create control charts from SAS data sets, including:
 - X-Bar and R (or Sigma) Charts for continuous variables.
 - P-Charts for binomial variables, such as fraction failing.
 - U-Charts for Poisson variables, such as defect counts.
- Use the OUTLIMITS option to save control limits created by PROC SHEWHART.
- Use the LIMITS option to specify the control limits to be used by PROC SHEWHART.
- Use the TESTS option to determine whether the process is in a state of control.
- Use a BOXCHART statement to create box plots of your process data.
- Use PROC CAPABILITY to evaluate the capability of your process.

9.8.2 For More Information

A good place to start is the “Getting Started” sections for PROC CAPABILITY and PROC SHEWHART in the SAS/QC online documentation.

Books

Duncan, Acheson J. 1986. *Quality Control and Industrial Statistics, Fifth Edition*, Homewood, IL.: Irwin.

Grant, Eugene L., and Richard S. Leavenworth. 1988. *Statistical Quality Control, Sixth Edition*, New York: McGraw-Hill.

Montgomery, Douglas C. 1996. *Introduction to Statistical Quality Control, Third Edition*, New York: John Wiley & Sons.

van Dobben de Bruyn, Cornelis S. 1968. *Cumulative Sum Tests: Theory and Practice*, London: Griffin.

Papers

Nelson, Lloyd S. 1984. “The Shewhart Control Chart—Tests for Special Causes.” *Journal of Quality Technology* 16, 237–239.

Nelson, Lloyd S. 1985. “Interpreting Shewhart \bar{X} Control Charts.” *Journal of Quality Technology* 17, 114–116.

9.8.3 Exercises

Your JES library contains data sets JES.Fourth, JES.Fifth and JES.Sixth, which are similar to JES.First, and also the summary data sets JES.Fourth_Sum, JES.Fifth_Sum and JES.Sixth_Sum, which are similar to JES.First_Sum. For each of these data sets:

- Plot a Mean Chart for Resistance to determine whether the process is in control.
 - Do this using the raw data (e.g., JES.Fourth) and also using the summary data (e.g., JES.Fourth_Sum).
- If the process is in control:
 - Use PROC CAPABILITY to estimate the process capability.
 - Re-run the Mean Chart using the OUTLIMITS option to capture the control limits.
 - Re-run the Mean Charts for the other data sets using these control limits.