Effective Applications of Control Charts Using SAS Software

Robert N. Rodriguez
Sharad S. Prabhu
SAS Institute Inc.
Cary, NC 27513-8000

Short Course Presentation
22nd SAS Users Group International Conference
San Diego
March 16-19, 1997

Abstract

This short course presents methods for constructing control charts that use the SHEWHART procedure in the SAS System. The course highlights techniques for improving the visual clarity of control charts and for implementing statistical modifications that are broadly applicable, such as analysis of means and control charts for short runs. The course also demonstrates the use of the SHEWHART procedure in tandem with SAS procedures for statistical modeling as a flexible approach to special problems such as autocorrelation in process data. The audience should understand the principles of statistical process control (SPC) and have basic familiarity with SAS programming.

Objectives

The objectives of this course are to demonstrate the use of SAS procedures for constructing control charts that are broadly used in industry, and to present SAS solutions for graphical and statistical issues that often arise in practice.

These objectives are met by presenting examples based on data from a variety of industries (ranging from semiconductor manufacturing to health care); by providing simple SAS programs that are readily extended to large-scale applications; and by presenting online demonstrations.

Outline

The course is organized as follows:

- Introduction to the SHEWHART procedure
  - ◇ Brief review of the Shewhart chart
  - ◇ Basic syntax of the SHEWHART procedure
  - ◇ Structure of input data sets
  - ◇ Selection of appropriate rational subgroups
  - ◇ and the use of subgroup variables
- Control Limits
  - ◇ Computing, saving, and reusing standard control limits
  - ◇ Modifying control limits
  - ◇ Application to analysis of means
- Tests for Special Causes
  - ◇ Nelson's version of the Western Electric tests
  - ◇ Working with varying sample sizes
  - ◇ Special modifications of the tests
  - ◇ Computing average run lengths

- Graphical Enhancements of Control Charts
  - ◇ Multiple sets of control limits
  - ◇ Displays for stratified data
  - ◇ Boxplot and "multivari" displays
  - ◇ Axis scaling issues
  - ◇ Switch variables
- Break
- Statistical Modeling for Control Chart Applications

  Using the SHEWHART procedure in conjunction with SAS statistical modeling procedures. Applications to

  - ◇ charts for individual measurements using non-normal data
  - ◇ charts for means involving multiple components of variation
  - ◇ charts for short runs
  - ◇ charts adjusted for seasonality and autocorrelation

  Note: One or two of these topics will be covered depending on the interests of the audience.
- Creating a Customized Control Chart

  This section demonstrates the open-ended use of the SHEWHART procedure when both the chart statistic and the control limits are non-standard. A multivariate control chart technique drawn from the recent literature is implemented to illustrate the approach.
- Questions and Answers

  An appendix in the course notes provides a road map to SAS resources for SPC applications including other procedures for statistical quality improvement, the SQC Menu System, and recent additions to the SAS System for real-time SPC and development of front-end applications.

Target Population

This presentation is intended for statisticians, process engineers, and manufacturing engineers who apply control charts in a quality improvement or process control setting, and for SAS programmers who provide support for process control systems. The applications illustrated range from semiconductor manufacturing to health care quality improvement.

Participants should be familiar with the principles and methods of statistical process control, but these concepts will be reviewed briefly if necessary. Only a minimal knowledge of SAS programming is assumed (how to create and modify a SAS data set, and how to run a SAS procedure).

Learning Outcomes

After attending this presentation, participants should be able to use SAS software to

- construct standard control charts
- improve the interpretability of control charts through graphical enhancements
- implement control chart modifications for specialized applications of statistical process control

Instructors

Robert Rodriguez, Ph.D. (contact)
SAS Institute Inc.
SAS Campus Drive
Cary, NC 27513
Tel: 919-677-8000, ext. 7650
FAX: 919-677-4444

Robert Rodriguez is manager of the Statistical Quality Improvement and Linear Models R&D groups at SAS Institute. He serves on the Editorial Review Board of the *Journal of Quality Technology* and is an adjunct professor in the Department of Statistics at the University of North Carolina.

Sharad Prabhu, Ph.D.
SAS Institute Inc.
SAS Campus Drive
Cary, NC 27513
Tel: 919-677-8000, ext. 6663
FAX: 919-677-4444

Sharad Prabhu is a Technical Applications Developer in the Statistical Quality Improvement R&D group at SAS Institute. His areas of expertise are statistical process control and design of experiments. Sharad holds a PhD and MSE in Industrial Engineering from the Arizona State University. He has been a SAS developer for over two years.

Chapter 1

Review of Control Chart Concepts

This chapter is optional and reviews the principles of statistical process control and control chart terminology.

Motivation

- developed in 1924 by Walter A. Shewhart
- emphased in 1980s by W. Edwards Deming
- every process is subject to variability
- the natural variability can be quantified with a set of control limits
- variation exceeding these limits signals a change in the process

Applications

- manufacturing
- service industries (1990s)

Uses of Shewhart Charts

- distinguishing variation due to *special causes* from variation due to *common causes*
- monitoring an in-control process
- reducing variation by improving the system

Anatomy of a Shewhart Chart

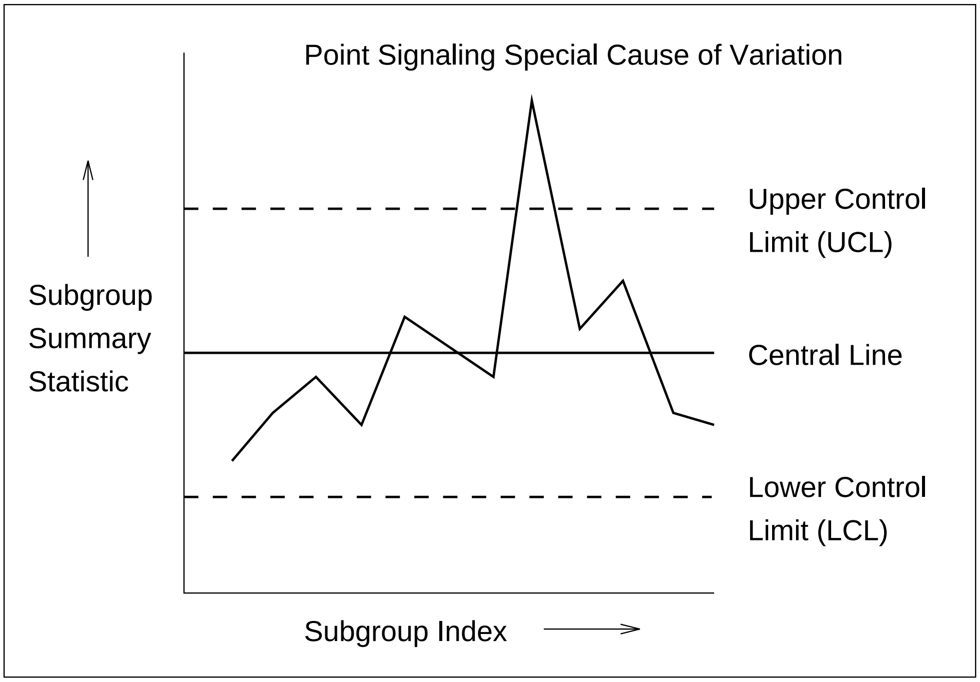


Figure 1.1. Prototype Shewhart Chart

- Each point represents a *summary statistic* computed from a sample of measurements
- The summary statistic is plotted along the *vertical axis*
- The samples are referred to as *rational subgroups* or *subgroup samples*. The organization of the data into subgroups is critical to the interpretation of the chart. Shewhart advocated selecting rational subgroups so that variation within subgroups is minimized and variation among subgroups is maximized; this makes the chart more sensitive to shifts in the process level. See Wheeler and Chambers (1986) or Montgomery (1991).

- The *horizontal axis* of a Shewhart chart identifies the subgroup samples.
- The *central line* on a Shewhart chart indicates the average (expected value) of the summary statistic when the process is in control.
- The *upper and lower control limits* (UCL and LCL) indicate the expected range of variation in the summary statistic when the process is in control.
- The control limits are also determined by the subgroup sample size because the standard error of the summary statistic is a function of sample size.
- Control limits can be estimated from the data being analyzed, or they can be based on "standard" (previously determined) values.
- *A point outside the control limits* signals the presence of a special cause of variation.
- Additionally, *tests for special causes* (*Western Electric rules*, *runs tests*) can signal an out-of-control condition.
- When the chart (correctly) signals the presence of a special cause, additional action is needed to diagnose and eliminate the problem.

Classification of Shewhart Charts

1. Charts for *variables* are used when the quality characteristic of a process is measured on a continuous scale.

   - $\bar{X}$ and R charts display subgroup means and ranges.
   - $\bar{X}$ and s charts display subgroup means and standard deviations.
   - Median and range charts display subgroup medians and ranges.
   - Charts for individual measurements and moving ranges display individual measurements and moving ranges of two or more successive measurements.

2. Charts for *attributes* are used when the quality characteristic of a process is measured by counting defects.

   - A p chart displays the proportion of defective items in a subgroup sample.
   - An np chart displays the number of defective items in a subgroup sample.
   - A u chart displays the number of defects per unit in a subgroup sample consisting of an arbitrary number of units.
   - A c chart displays the number of defects in a unit.

Chapter 2

Basic Syntax of SHEWHART Procedure

This chapter introduces syntax and input data set structures which will be referred to throughout the course.

For additional examples and details, see "Part 8. The Shewhart Procedure" in *SAS/QC® Software: Usage and Reference*, which is abbreviated as QCUR in these notes.

Basic $\bar{X}$ and R Charts

In the manufacture of silicon wafers, batches of five wafers are sampled, and their diameters are measured in millimeters. The following statements create a SAS data set named WAFERS, which contains the measurements for 25 batches:

```
data wafers;
   input batch @;
   do i=1 to 5;
      input diamtr @;
      output;
      end;
   drop i;
cards;
 1  35.00 34.99 34.99 34.98 35.00
 2  35.01 34.99 34.99 34.98 35.00
 3  34.99 35.00 35.00 35.00 35.00

 ...

24  35.00 35.00 34.99 35.01 34.98
25  34.99 34.99 34.99 35.00 35.00
;

title 'The Data Set WAFERS';
proc print data=wafers noobs;
run;
```

WAFERS is partially listed in Figure 2.1.

```
          The Data Set WAFERS

           BATCH    DIAMTR

              1      35.00
              1      34.99
              1      34.99
              1      34.98
              1      35.00
              2      35.01
              2      34.99
              2      34.99
              2      34.98
              2      35.00
              3      34.99
              3      35.00
              3      35.00
              3      35.00
              3      35.00
              .        .
              .        .
              .        .
```

Figure 2.1. Raw Measurements in Data Set WAFERS

Note that

- the observations are arranged in "strung-out form"
- BATCH is the *subgroup-variable*
- DIAMTR is the *process-variable*

The following statements create $\bar{X}$ and R charts for the diameter measurements.

```
title 'Mean and Range Charts for Diameters';

symbol v=dot;

proc shewhart data=wafers graphics;
   xrchart diamtr*batch;
run;
```

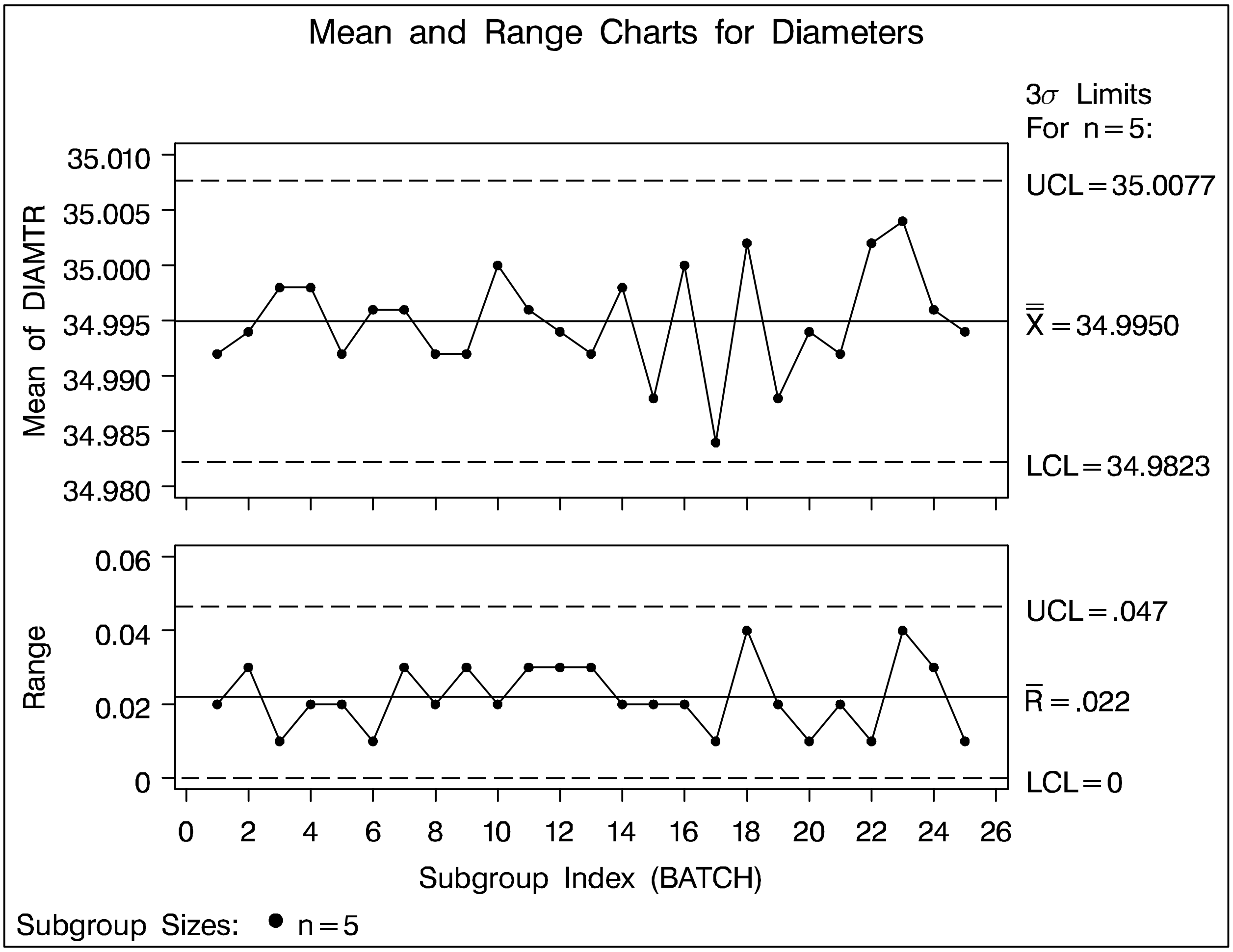


Figure 2.2. Basic $\bar{X}$ and R Charts

The control limits are 3σ limits by default.

Working with Summary Data

Data set WAFERSUM provides the wafer measurements in summarized form:

```
data wafersum;
   input batch diamtrx diamtrr;
   diamtrn = 5;
cards;
 1  34.992  0.02
 2  34.994  0.03
 3  34.998  0.01

 ...

24  34.996  0.03
25  34.994  0.01
;

proc print data=wafers noobs;
run;
```

```
 BATCH    DIAMTRX    DIAMTRR    DIAMTRN

    1      34.992      0.02        5
    2      34.994      0.03        5
    3      34.998      0.01        5
    4      34.998      0.02        5
    5      34.992      0.02        5
    .         .          .         .
    .         .          .         .
    .         .          .         .
```

Figure 2.3. Summary Measures in Data Set WAFERSUM

The following statements create $\bar{X}$ and R charts for the diameter measurements that are identical to the charts in Figure 2.2.

```
title 'Mean and Range Charts for Diameters';

symbol v=dot;

proc shewhart history=wafersum graphics;
   xrchart diamtr*batch;
run;
```

Note that

- DIAMTR is the common *prefix* for the names of three summary variables in WAFERSUM
- WAFERSUM is specified as a HISTORY= data set

Basic u Chart

- In manufacturing, a u chart is typically used to analyze the number of defects per inspection unit in samples that contain arbitrary numbers of units.
- In general, the events counted need not be "defects."
- A u chart is applicable when the counts are scaled by a measure of opportunity for the event to occur, and when the counts can be modeled by the Poisson distribution.

A health care provider uses a u chart to analyze the rate of cat scans performed each month by each of its clinics. Data for Clinic B are saved in a data set named CLINICB.

MONTH	NSCANB	MMSB	DAYS	NYRSB
JAN94	50	26838	31	2.31
FEB94	44	26903	28	2.09
MAR94	71	26895	31	2.32
APR94	53	26289	30	2.19
MAY94	53	26149	31	2.25
JUN94	40	26185	30	2.18
JUL94	41	26142	31	2.25
AUG94	57	26092	31	2.25
SEP94	49	25958	30	2.16
OCT94	63	25957	31	2.24
NOV94	64	25920	30	2.16
DEC94	62	25907	31	2.23
JAN95	67	26754	31	2.30
FEB95	58	26696	28	2.08
MAR95	89	26565	31	2.29

Figure 2.4. Data Set CLINICB

NSCANB is the number of cat scans performed each month, and MMSB is the number of members enrolled each month (in units of “member months”). DAYS is the number of days in each month.

NYRSB was computed to convert MMSB to units of “thousand members per year.”

```
data clinicb;
   set clinicb;
   nyrsb = mmsb * ( days / 30 ) / 12000;
   nyrsb = round( nyrsb, 0.01 );
run;
```

NYRSB provides the “measure of opportunity.”

The following statements create the u chart in Figure 2.5.

```
title 'U Chart for Cat Scans per 1,000 Members:'
      ' Clinic B';

proc shewhart data=clinicb graphics;
   uchart nscanb * month /
      subgroupn = nyrsb;

label nscanb = 'Rate per 1,000 Member-Years'
      month  = 'Month';
run;
```

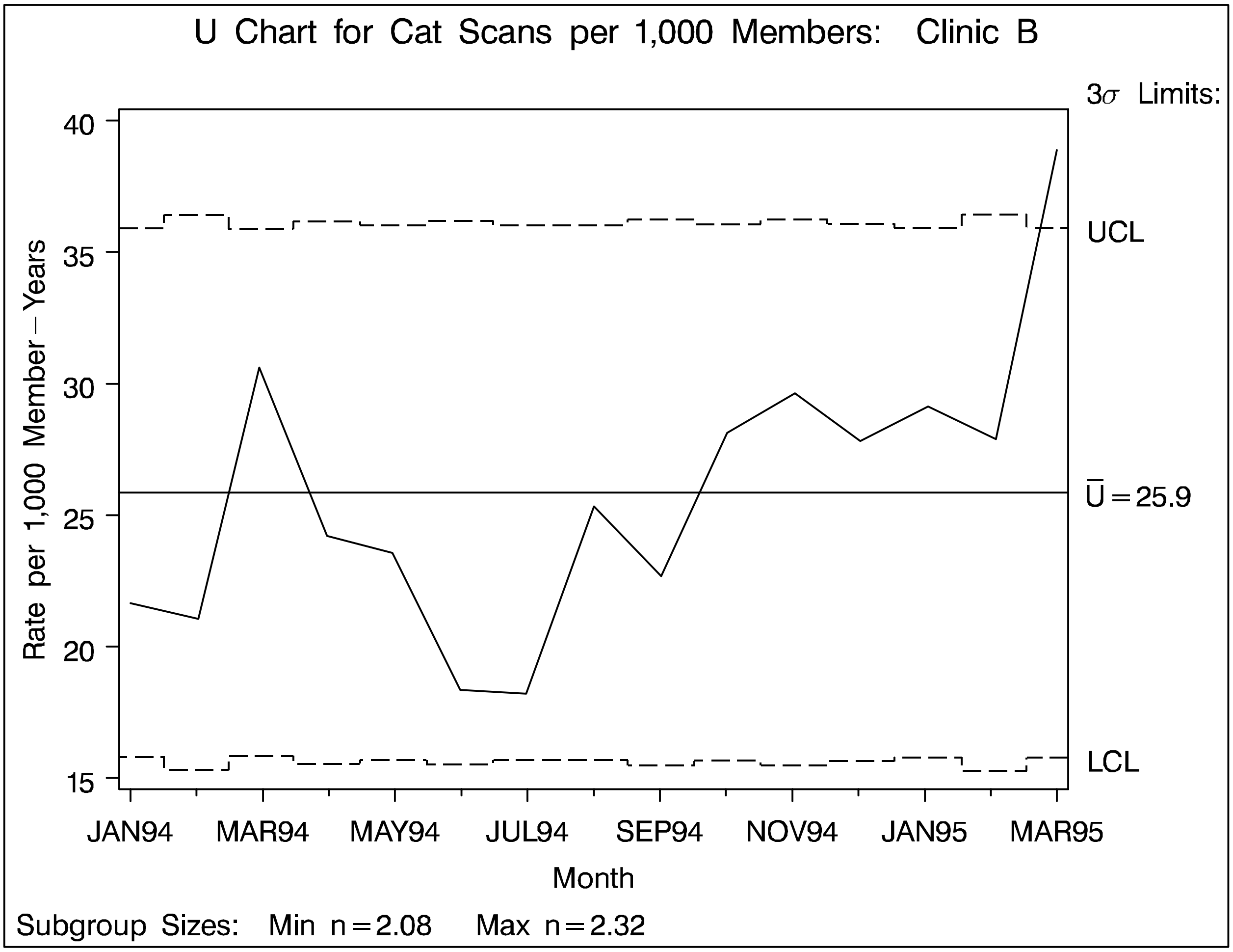


Figure 2.5. Basic u Chart

By default, the control limits are adjusted for the measure of opportunity.

Summary

You can create a basic Shewhart chart with as few as two SAS statements:

- the PROC SHEWHART statement, which starts the procedure and specifies the input SAS data set
- a chart statement, which requests the chart type with a keyword and the process and subgroup variables

Keyword	Chart(s) Displayed	"Getting Started" Page in QCUR
BOXCHART	box chart with optional trend chart	874
CCHART	c chart	924
IRCHART	individual and moving range charts	960
MCHART	median chart with optional trend chart	998
MRCHART	median and R charts	1032
NPCHART	np chart	1074
PCHART	p chart	1112
RCHART	R chart	1154
SCHART	s chart	1188
UCHART	u chart	1222
XCHART	$\bar{X}$ chart with optional trend chart	1258
XRCHART	$\bar{X}$ and R charts	1300
XSCHART	$\bar{X}$ and s charts	1346

Chapter 3

Control Limits

This chapter introduces methods for determining and managing control limits that are useful in standard applications involving many process variables that are changing over time. These methods are also useful for constructing nonstandard charts, as illustrated in later chapters.

Saving Control Limits

You can save control limits in a SAS data set; this enables you to apply the control limits to future data or modify the limits with a DATA step program.

```
title 'Control Limits for Wafer Diameters';
proc shewhart data=wafers;
   xrchart diamtr*batch / outlimits = waferlim
                          nochart;
run;
```

Control Limits for Wafer Diameters

VAR	_SUBGRP_	_TYPE_	_LIMITN_	_ALPHA_	_SIGMAS_	_LCLX_
DIAMTR	BATCH	ESTIMATE	5	.0026998	3	34.9823

MEAN	_UCLX_	_LCLR_	_R_	_UCLR_	_STDDEV_
34.9950	35.0077	0	0.022	0.046519	.0094586

Figure 3.1. SAS Data Set WAFERLIM

What do the variables represent?

Reading Pre-Established Control Limits

The following statements apply the control limits in Figure 3.1 to new process data in WAFERS2:

```
data wafers2;
   input batch @;
   do i=1 to 5;
      input diamtr @;
      output;
      end;
   drop i;
   cards;
26 34.99 34.99 35.00 34.99 35.00
27 34.99 35.01 34.98 34.98 34.97
28 35.00 34.99 34.99 34.99 35.01
...

44 35.00 35.00 34.98 35.00 34.99
45 34.99 34.99 35.00 34.99 34.99
;
```

```
title 'Mean and Range Charts for Diameters';
proc shewhart data=wafers2 limits=waferlim graphics;
   xrchart diamtr*batch;
run;
```

Note that you must specify the option READLIMITS in Release 6.09 and prior releases. See page 1417 of QCUR.

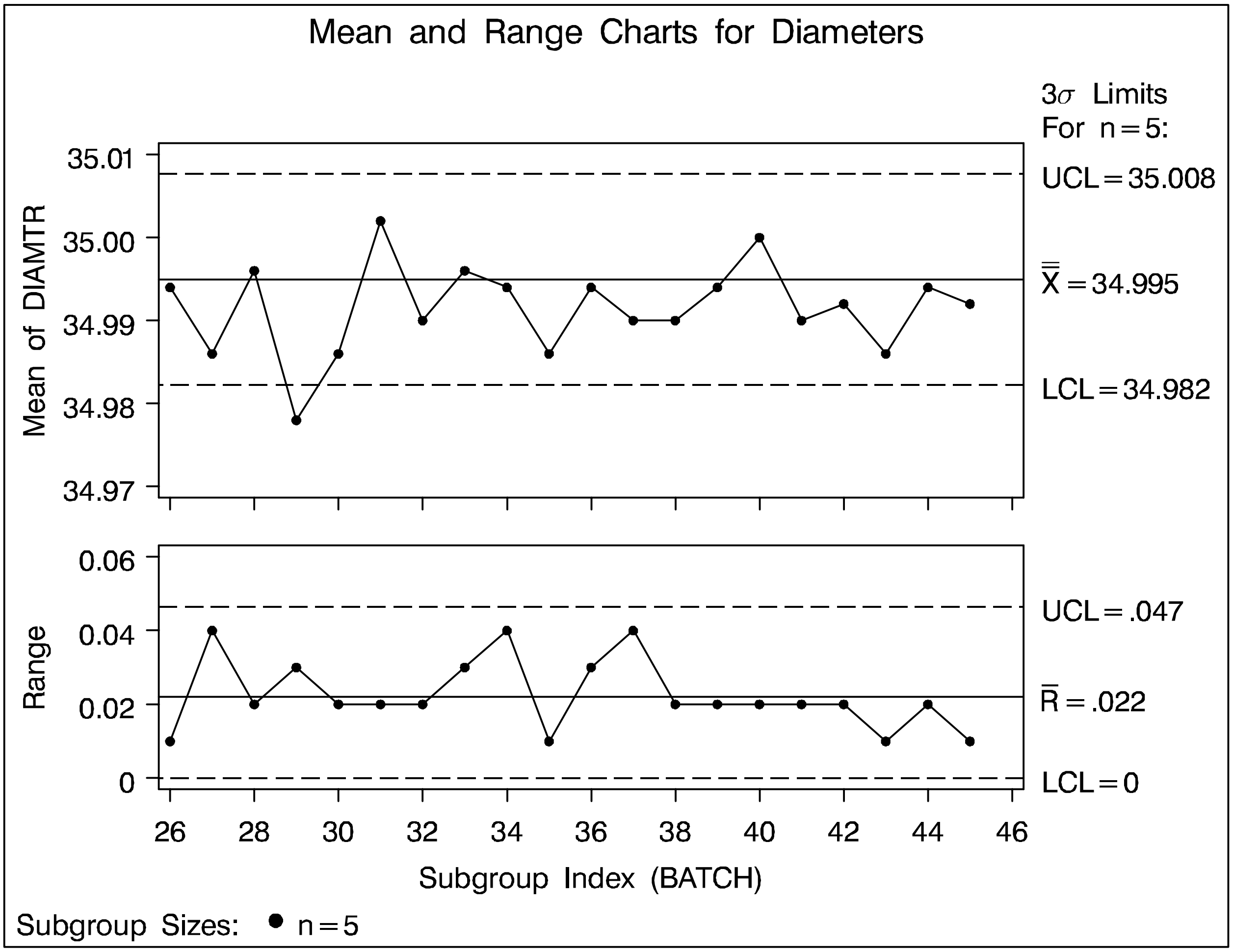


Figure 3.2. $\bar{X}$ and R Charts for Additional Wafer Data

Note the difference between

- the nominal sample size associated with the control limits
- the subgroup sample size(s)

Saving Varying Control Limits

What does an OUTLIMITS= data set look like when the control limits vary with subgroup sample size?

```
proc shewhart data=clinicb graphics;
   uchart nscanb * month /
      subgroupn = nyrsb
      outlimits = scanlim
      outtable  = scantab
      nochart;
run;
```

VAR	_SUBGRP_	_TYPE_	_LIMITN_	_ALPHA_	_SIGMAS_	_LCLU_	_U_	_UCLU_
NSCANB	MONTH	ESTIMATE	V	V	3	V	25.8559	V

Figure 3.3. SAS Data Set SCANLIM

Why is this information sufficient?

An OUTTABLE= data set saves both the control limits and the chart statistics for each subgroup.

VAR	MONTH	_SIGMAS_	_LIMITN_	_SUBN_	_LCLU_	_SUBU_	_U_	_UCLU_	_EXLIM_
NSCANB	JAN94	3	2.31	2.31	15.8191	21.6450	25.8559	35.8926	
NSCANB	FEB94	3	2.09	2.09	15.3040	21.0526	25.8559	36.4077	
NSCANB	MAR94	3	2.32	2.32	15.8407	30.6034	25.8559	35.8710	
NSCANB	APR94	3	2.19	2.19	15.5478	24.2009	25.8559	36.1640	
NSCANB	MAY94	3	2.25	2.25	15.6861	23.5556	25.8559	36.0256	
NSCANB	JUN94	3	2.18	2.18	15.5241	18.3486	25.8559	36.1876	
NSCANB	JUL94	3	2.25	2.25	15.6861	18.2222	25.8559	36.0256	
NSCANB	AUG94	3	2.25	2.25	15.6861	25.3333	25.8559	36.0256	
NSCANB	SEP94	3	2.16	2.16	15.4764	22.6852	25.8559	36.2353	
NSCANB	OCT94	3	2.24	2.24	15.6635	28.1250	25.8559	36.0483	
NSCANB	NOV94	3	2.16	2.16	15.4764	29.6296	25.8559	36.2353	
NSCANB	DEC94	3	2.23	2.23	15.6406	27.8027	25.8559	36.0711	
NSCANB	JAN95	3	2.30	2.30	15.7973	29.1304	25.8559	35.9144	
NSCANB	FEB95	3	2.08	2.08	15.2787	27.8846	25.8559	36.4330	
NSCANB	MAR95	3	2.29	2.29	15.7753	38.8646	25.8559	35.9364	UPPER

Figure 3.4. SAS Data Set SCANTAB

How do SCANLIM and SCANTAB compare?

In later sections we will exploit the use of OUTLIMITS= and OUTTABLE= data sets as *input* data sets.

How Control Limits Are Computed

Table 3.1. Limits for $\bar{X}$ and R Charts

Control Limits	
$\bar{X}$ Chart	LCL = lower limit = $\overline{\overline{X}} - k\hat{\sigma}/\sqrt{n_i}$ UCL = upper limit = $\overline{\overline{X}} + k\hat{\sigma}/\sqrt{n_i}$
R Chart	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	
$\bar{X}$ Chart	LCL = lower limit = $\overline{\overline{X}} - z_{\alpha/2}(\hat{\sigma}/\sqrt{n_i})$ UCL = upper limit = $\overline{\overline{X}} + z_{\alpha/2}(\hat{\sigma}/\sqrt{n_i})$
R Chart	LCL = lower limit = $D_{\alpha/2}\hat{\sigma}$ UCL = upper limit = $D_{1-\alpha/2}\hat{\sigma}$

You can provide parameters for the limits as follows:

- Specify k with the SIGMAS= option or with the variable _SIGMAS_ in a LIMITS= data set
- Specify α with the ALPHA= option or with the variable _ALPHA_
- Specify a constant nominal sample size $n_i \equiv n$ with the LIMITN= option or with the variable _LIMITN_
- Specify μ_0 with the MU0= option or with the variable _MEAN_
- Specify σ_0 with the SIGMA0= option or with the variable _STDDEV_

Application to Analysis of Means

Analysis of means is a graphical and statistical method for simultaneously comparing a group of k treatment means with their grand mean at a specified significance level α. Analysis of means is an extension to the Shewhart chart because it considers a group of sample means instead of one mean at a time in order to determine whether any of the sample means differ too much from the overall mean.

A health care system uses ANOM to compare medical/surgical admissions rates for a group of clinics. The data are saved in a SAS data set named MSADMITS.

```
        ID    COUNT95      MYRS95

        1A      1882     58.1003
        1K       600     18.7263
        1B       438     12.8933
        1D       318      6.8545
        3M       183      6.3708
        ...      ...         ...

        1F         7      0.2020
        1P         2      0.1692
```

Figure 3.5. Data Set MSADMITS

ID identifies the clinics, COUNT95 provides the number of admissions during 1995, and MYRS95 provides the number of 1,000 member-years, which serves as the “measure of opportunity” for admissions.

```
%anomsig( 0.01 /*alpha*/, 29 /*no. of groups*/ );

title 'Analysis of Medical/Surgical Admissions';

proc shewhart data=msadmits graphics;
   uchart count95*id /
      subgroupn = myrs95
      sigmas    = &sigmult
      cneedles  = yellow
      cinfill   = green
      lcllabel  = 'LDL'
      ucllabel  = 'UDL'
      turnhlabels
      nolegend;
label count95 = 'Admits per 1000 Member Years';
run;
```

See Rodriguez (1996, 1996b) for the ANOMSIG macro.

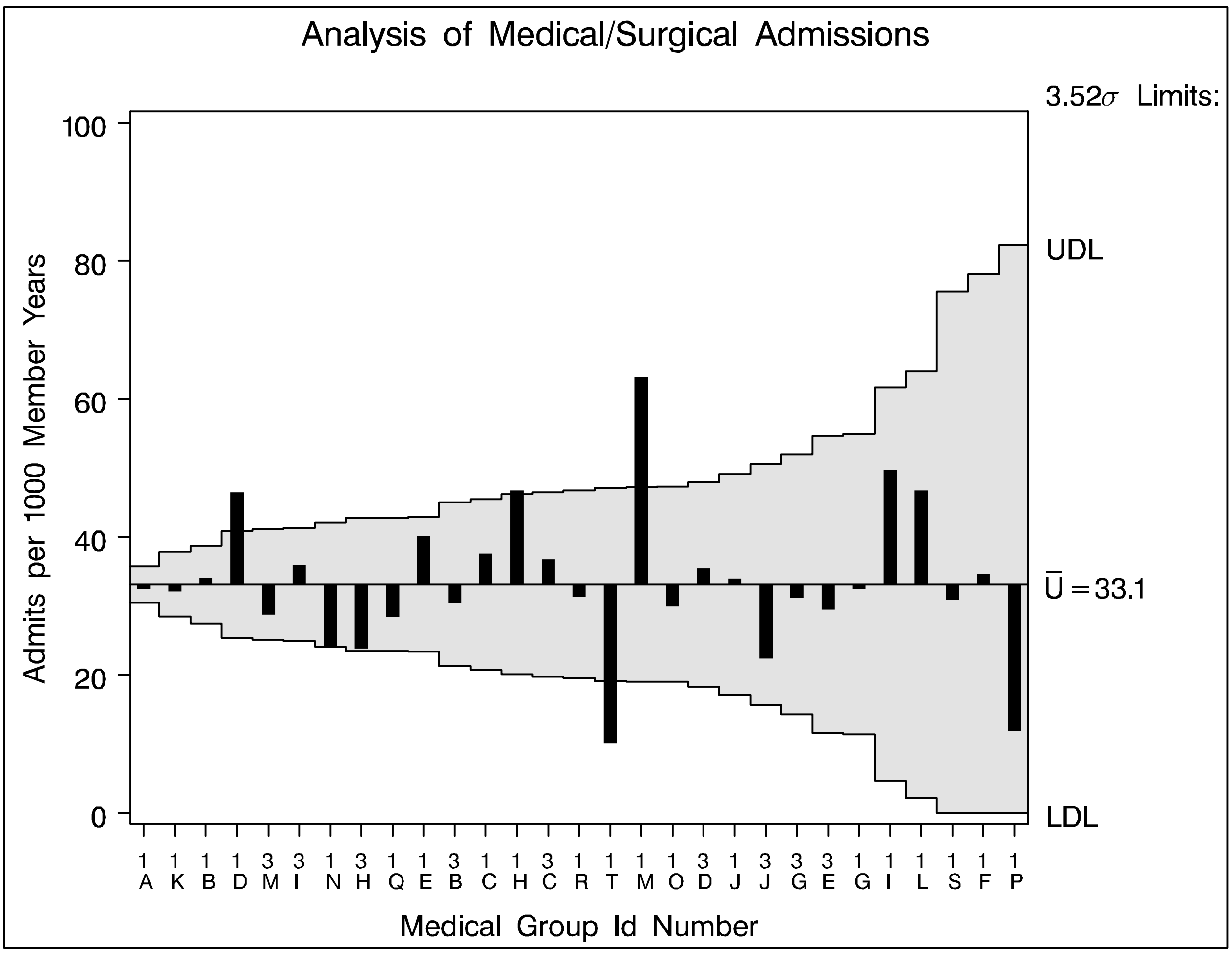


Figure 3.6. ANOM for Medical/Surgical Admissions Rates

The chart answers the question, “Do any of the clinics differ significantly from the system average in their rates of admission?”

Note that

- Analysis of means assumes that the system is statistically predictable, whereas a major reason for using a control chart is to bring the system into a state of statistical control; see Wheeler (1995).
- The decision limits UDL and LDL are not the same as the 3σ limits that the SHEWHART procedure would compute by default for a u chart. The reason is that control limits are applied to the rates *taken one at a time*, whereas the decision limits are applied to the rates *taken as a group*.
- Tests for special causes (next chapter) are not applicable in ANOM.

Summary

Control limits can be

- computed directly (standard formulas)
- saved in an OUTLIMITS= or OUTTABLE= data set
- read from a LIMITS= or TABLE= data set

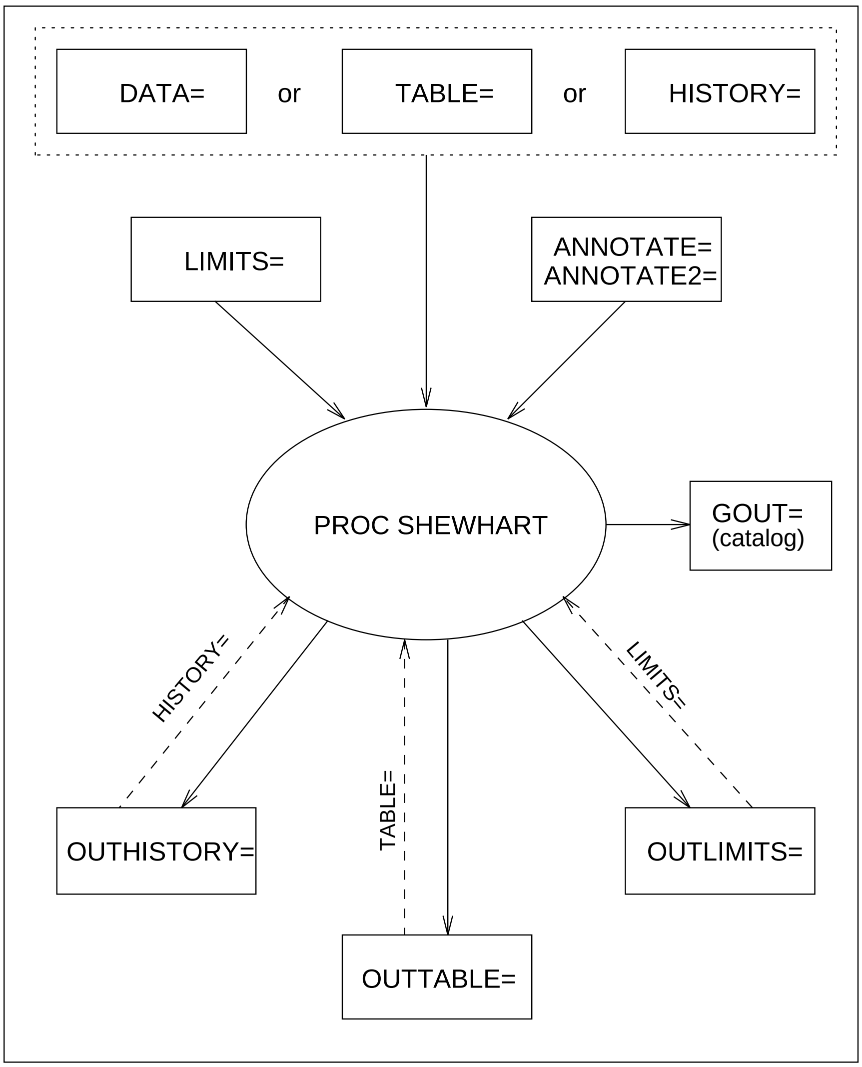


Figure 3.7. Input and Output Data Sets

Chapter 4

Tests for Special Causes

This chapter describes the tests for special causes that are supported by the SHEWHART procedure.

For further details, see Chapter 41 of QCUR beginning on page 1497.

Standard Tests

The SHEWHART procedure provides eight standard *tests for special causes*, also referred to as *supplementary rules*, *runs tests*, and *Western Electric rules*. These tests improve the sensitivity of the Shewhart chart to small changes in the process (as do cusum charts and moving average charts, which you can construct with the CUSUM and MACONTROL procedures).

The patterns detected by the eight standard tests are defined in Table 4.1 and Table 4.2, and they are illustrated in Figure 4.1 and Figure 4.2. All eight tests were developed for use with fixed 3σ limits. The tests are indexed according to the numbering sequence used by Nelson (1984, 1985).

You can request any combination of the eight tests by specifying the test *indexes* with the TESTS= option in the chart statement. For example:

```
proc shewhart data=wafers2 limits=waferlim graphics;
      xrchart diamtr*batch /
         tests = 1 to 4;
   run;
```

Table 4.1. Definitions of Tests 1 to 4

Test *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 or steadily decreasing
4	Fourteen points in a row alternating up and down

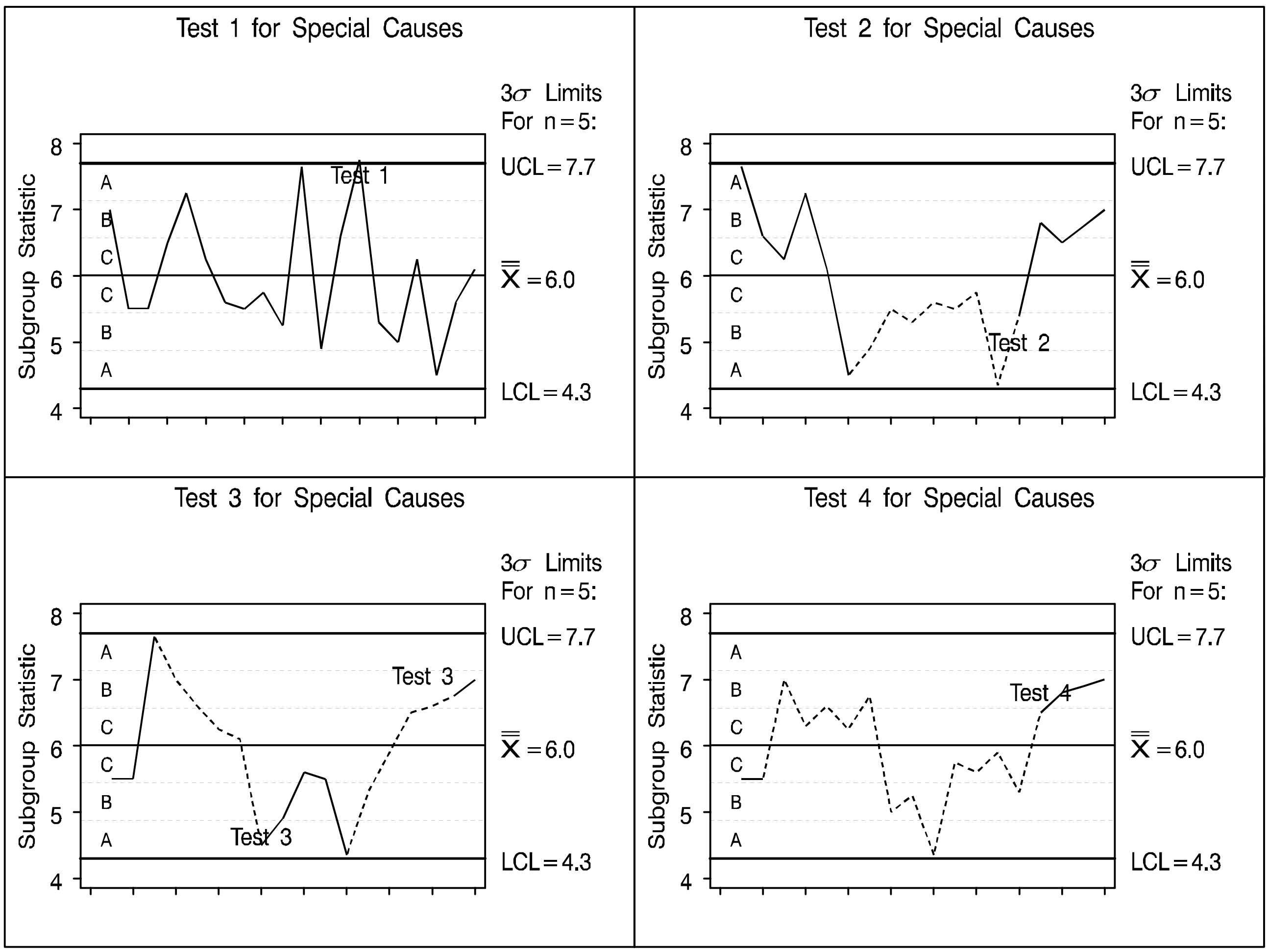


Figure 4.1. Examples of Tests 1 to 4

Table 4.2. Definitions of Tests 5 to 8

Test *Index*	Pattern Description
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

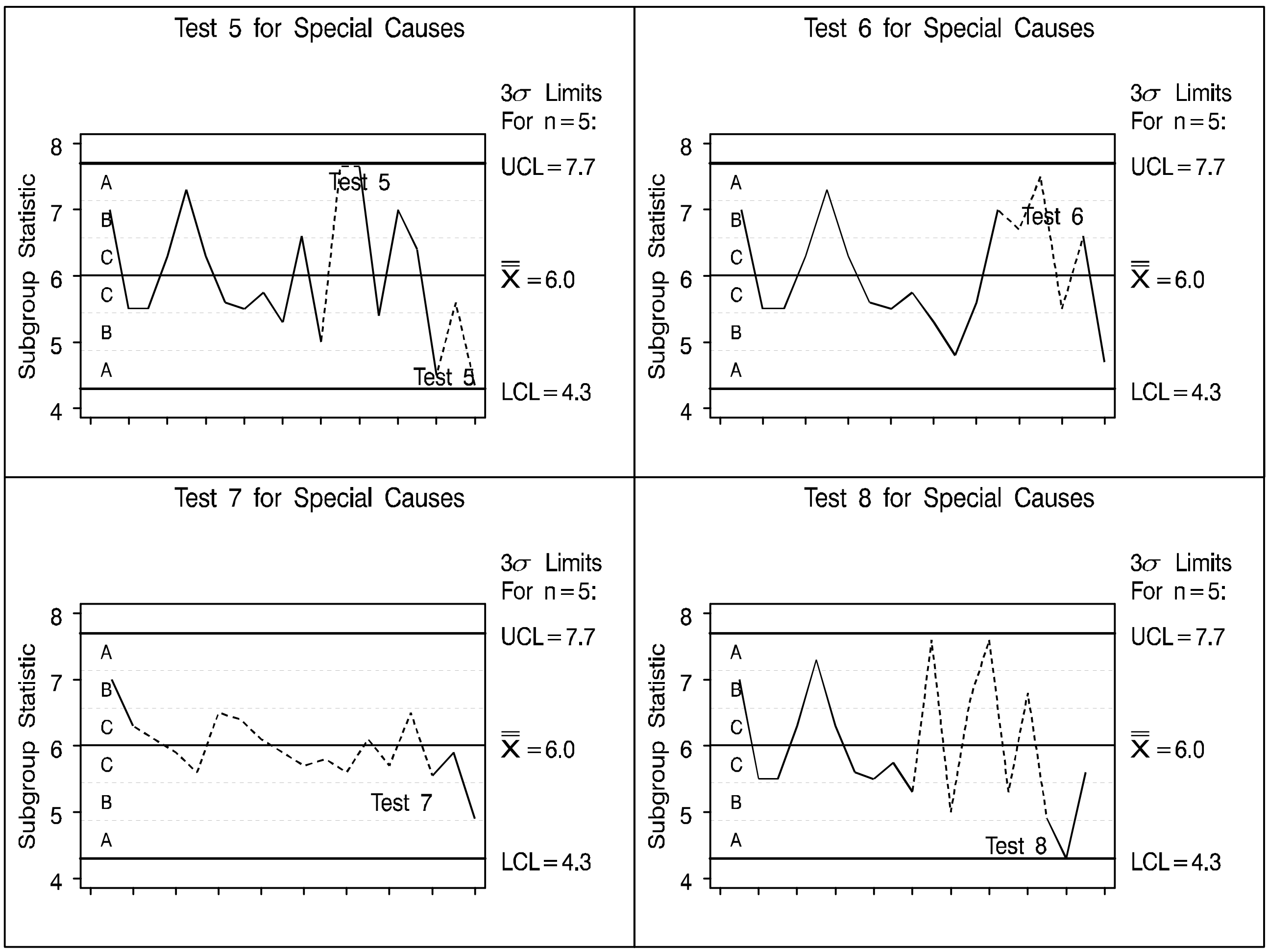


Figure 4.2. Examples of Tests 5 to 8

Interpreting Standard Tests for Special Causes

According to Nelson (1984, 1985):

- When a process is in statistical control, the chance of a false signal for each test is less than five in one thousand.
- Test 1 is positive if there is a shift in the process mean, if there is an increase in the process standard deviation, or if there is a "single aberration in the process such as a mistake in calculation, an error in measurement, bad raw material, a breakdown of equipment, and so on" (Nelson 1985).
- Test 2 signals a shift in the process mean.
- Test 3 signals a drift in the process mean. Nelson (1985) states that causes can include "tool wear, depletion of chemical baths, deteriorating maintenance, improvement in skill, and so on."
- Test 4 signals "a systematic effect such as produced by two machines, spindles, operators or vendors used alternately" (Nelson 1985).
- Tests 1, 2, 3, and 4 should be applied routinely; the combined chance of a false signal from one or more of these tests is less than one in a hundred. Nelson (1985) describes these tests as "a good set that will react to many commonly occurring special causes."

- In the case of charts for variables, the first four tests should be augmented by Tests 5 and 6 when earlier warning is desired. The chance of a false signal increases to two in a hundred.
- Tests 7 and 8 indicate stratification (observations in a subgroup have multiple sources with different means). Test 7 is positive when the observations in the subgroup always have multiple sources. Test 8 is positive when the subgroups are taken from one source at a time.

Nelson (1985) also comments that "the probabilities quoted for getting false signals should not be considered to be very accurate" since the probabilities are based on assumptions of normality and independence that may not be satisfied. Consequently, he recommends that the tests "should be viewed as simply practical rules for action rather than tests having specific probabilities associated with them." Nelson cautions that "it is possible, though unlikely, for a process to be out of control yet not show any signals from these eight tests."

Tests With Varying Sample Sizes

Nelson (1989, 1994) describes the use of standardization to apply the tests for special causes to data involving varying subgroup samples. This approach applies the tests to the standardized subgroup statistics, setting the control limits at ± 3 and the zone boundaries at ± 1 and ± 2. You can request this method with the TESTNMETHOD= option

```
title 'U Chart for Cat Scans per 1,000 Members:  Clinic B';
proc shewhart data=clinicb graphics;
   uchart nscanb * month /
      subgroupn   = nyrsb
      tests       = 1 to 4
      testlabel   = space
      testnmethod = standardize
      nohlabel
      nolegend
      ;
label nscanb = 'Rate per 1,000 Member-Years';
run;
```

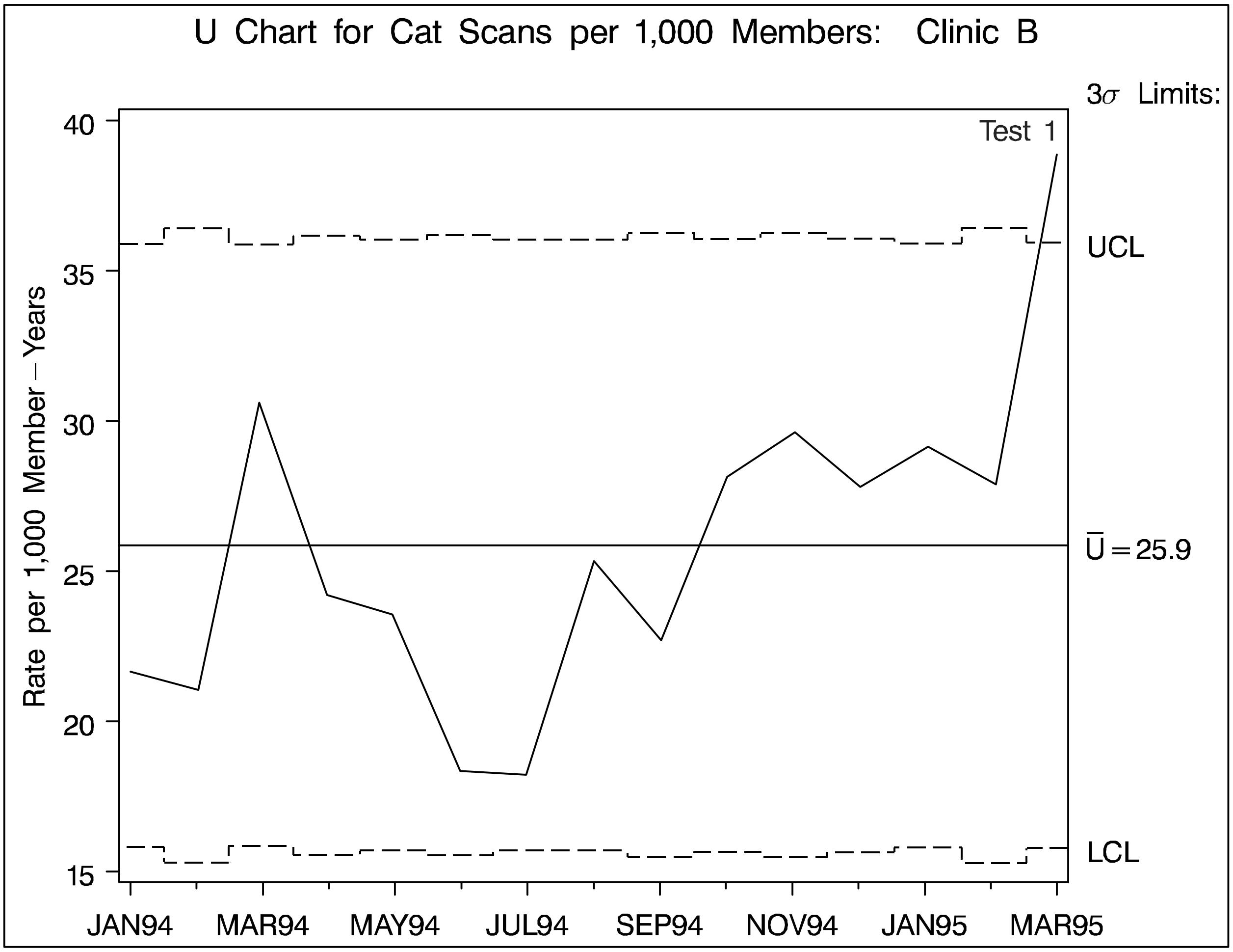


Figure 4.3. u Chart with Tests 1 to 4

Labeling Signaled Points with a Variable

You can use the TESTLABEL= option to display operator comments or other information that can aid in the identification of special causes.

SAMPLE	OFFSETX	OFFSETR	OFFSETN	COMMENT
1	19.80	3.8	5	
2	17.16	8.3	5	
3	20.11	6.7	5	
4	20.89	5.5	5	
5	20.83	2.3	5	
6	18.87	2.6	5	
7	20.84	2.3	5	
8	23.33	5.7	5	New Tool
9	19.21	3.5	5	
10	20.48	3.2	5	
11	22.05	4.7	5	
12	20.02	6.7	5	
13	17.58	2.0	5	
14	19.11	5.7	5	
15	20.03	4.1	5	
16	20.56	3.7	5	Recalibration
17	20.86	3.3	5	
18	21.10	5.6	5	Reset Tool
...	...	...	...	...

Figure 4.4. Data Set ASSEMBLY

```
title 'Analysis of Assembly Data';
proc shewhart history=assembly graphics;
   xrchart offset * sample / mu0        = 20
                             sigma0     = 2.24
                             limitn     = 5
                             alln
                             tests      = 1 to 4
                             testlabel  = ( comment )
                             ltests     = 2
                             split = '/' ;
   label offsetx = 'Avg Offset (cm)/Range';
run;
```

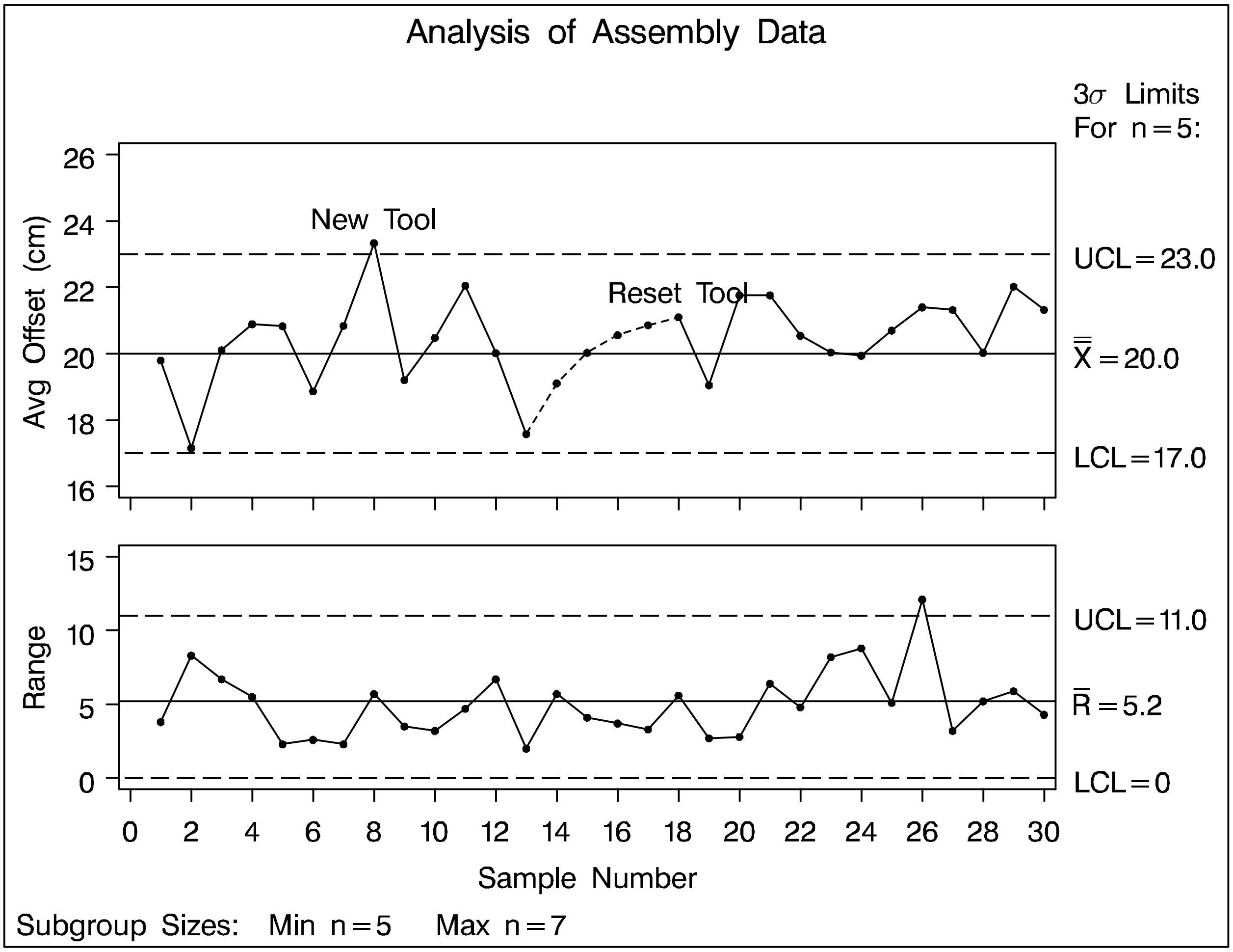


Figure 4.5. Labeling Points with a TESTLABEL= Variable

For details, see page 1506 of QCUR.

Generalized Tests for Special Causes

- You can apply Tests 1 to 8 to range charts (unusual applications in the semiconductor industry). See page 1512 of QCUR.
- You can request a search for k out of m points in a row in the interval (a, b). This generalizes Tests 1, 2, 5, and 6. See page 1513 of QCUR.
- A program (SHWARL2) for average run length computations is provided in the SAS/QC Sample Library. See Champ and Woodall (1987) and Wetherill and Brown (1991).
- Applications include the rules given by Westgard *et al.* (1981) for clinical chemistry.
- You can request a search for k points in a row increasing or decreasing. This generalizes Test 3. See page 1513 of QCUR.
- You can construct arbitrary pattern searches with a DATA step program and assign the results to the variable _TESTS_ in a TABLE= data set. See page 1516 of QCUR.

Chapter 5

Graphical Enhancements

This chapter illustrates techniques for making graphical modifications of control charts and for improving their visual clarity.

For further details, see Chapter 40 of QCUR beginning on page 1445.

Multiple Sets of Control Limits

This example illustrates the construction of a u chart for a situation where the process rate is known to have shifted, requiring the use of multiple sets of control limits.

A health care provider uses a u chart to report the rate of office visits performed each month by each of its clinics. The rate is computed by dividing the number of visits by the membership expressed in thousand-member years. Figure 5.1 shows data collected for Clinic E.

MONTH	_PHASE_	NVISITE	NYRSE	DAYS	MMSE
JAN94	Phase 1	1421	0.66099	31	7676
FEB94	Phase 1	1303	0.59718	28	7678
MAR94	Phase 1	1569	0.66219	31	7690
APR94	Phase 1	1576	0.64608	30	7753
MAY94	Phase 1	1567	0.66779	31	7755
JUN94	Phase 1	1450	0.65575	30	7869
JUL94	Phase 1	1532	0.68105	31	7909
AUG94	Phase 1	1694	0.68820	31	7992
SEP94	Phase 2	1721	0.66717	30	8006
OCT94	Phase 2	1762	0.69612	31	8084
NOV94	Phase 2	1853	0.68233	30	8188
DEC94	Phase 2	1770	0.70809	31	8223
JAN95	Phase 2	2024	0.78215	31	9083
FEB95	Phase 2	1975	0.70684	28	9088
MAR95	Phase 2	2097	0.78947	31	9168

Figure 5.1. Data Set CLINICE

NVISITE is the number of visits each month, and MMSE is the number of members enrolled each month (in units of "member months"). DAYS is the number of days in each month. NYRSE expresses MMSE in units of thousand members per year. _PHASE_ separates the data into two time phases (the system changed in September 1994).

```
title 'U Chart for Office Visits per 1,000'
      ' Members: Clinic E';
proc shewhart data=clinice graphics;
   uchart nvisite * month /
      subgroupn = nyrse
      cframe    = ligr
      cinfill   = yellow
      nohlabel
      nolegend
      ;
label nvisite = 'Rate per 1,000 Member-Years';
run;
```

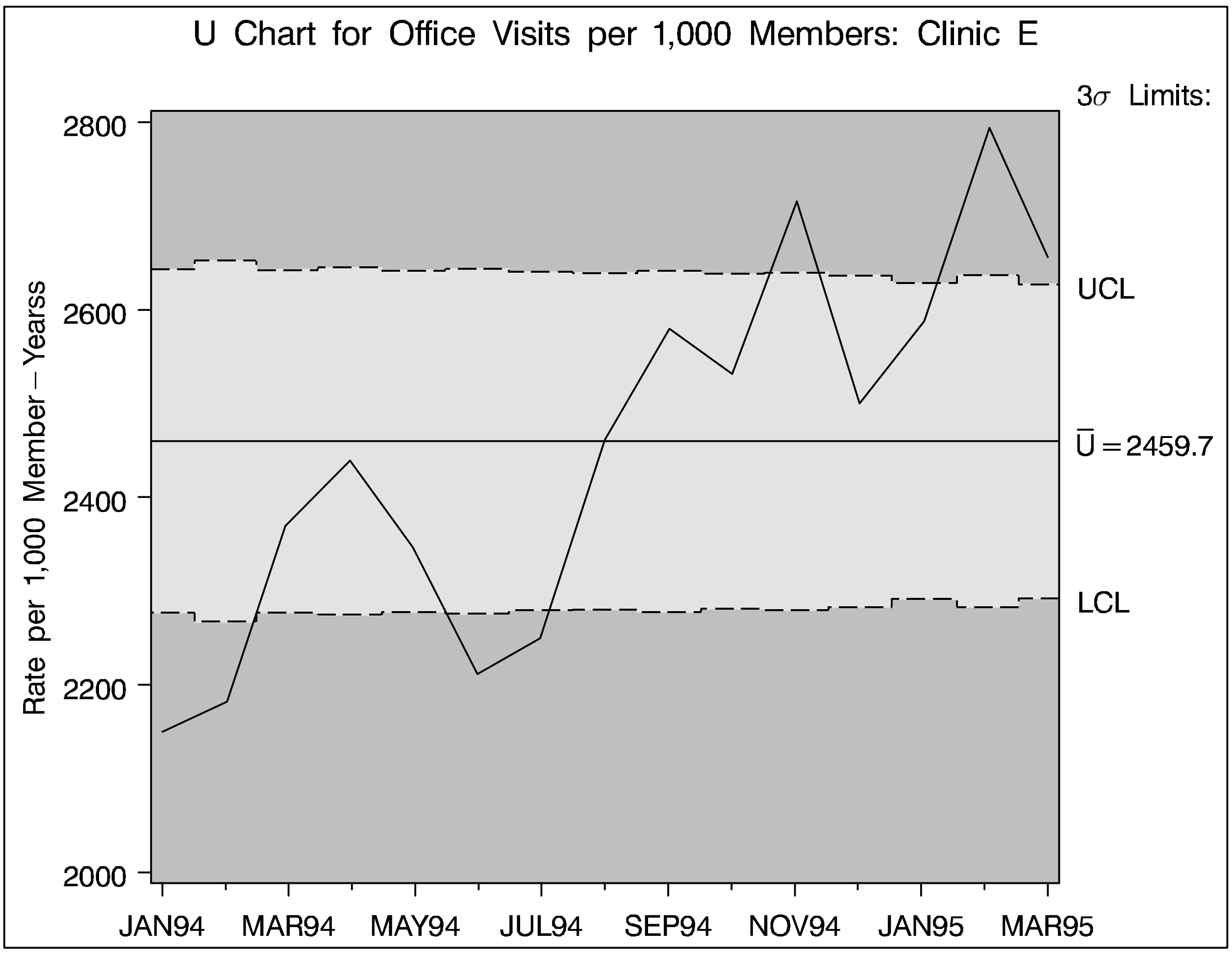


Figure 5.2. u Chart with Single Set of Limits

What is wrong with Figure 5.2?

The following statements create distinct sets of control limits for each time phase.

```
proc shewhart data=clinice graphics;
   by _phase_;
   uchart nvisite * month /
      subgroupn = nyrse
      outlimits = vislimit
                  (rename=(_phase_=_index_))
      nochart;
run;
```

```
           Control Limits for Office Visit Data

                      _              _  _
                      S              L  _  S
    _                 U       _      I  A  I  _        _
    I         _       B       T      M  L  G  L        U
    N         V       G       Y      I  P  M  C        C
    D         A       R       P      T  H  A  L    _   L
    E         R       P       E      N  A  S  U    U   U
    X
    _         _       _       _      _  _  _  _    _   _

Phase 1  NVISITE  MONTH  ESTIMATE   V  V  3  V  2302.99  V
Phase 2  NVISITE  MONTH  ESTIMATE   V  V  3  V  2623.52  V
```

Figure 5.3. Data Set VISLIMIT

The following statements combine the data and control limits for both phases in a single u chart, shown in Figure 5.4.

```
title 'U Chart for Office Visits'
      ' per 1,000 Members: Clinic E';
proc shewhart data=clinice
              limits=vislimit graphics;
   uchart nvisite * month /
      subgroupn = nyrse
      cframe    = ligr
      cinfill   = yellow
      readindex = all
      readphase = all
      nohlabel
      nolegend
      phaselegend
      nolimitslegend;
label nvisite = 'Rate per 1,000 Member-Years';
run;
```

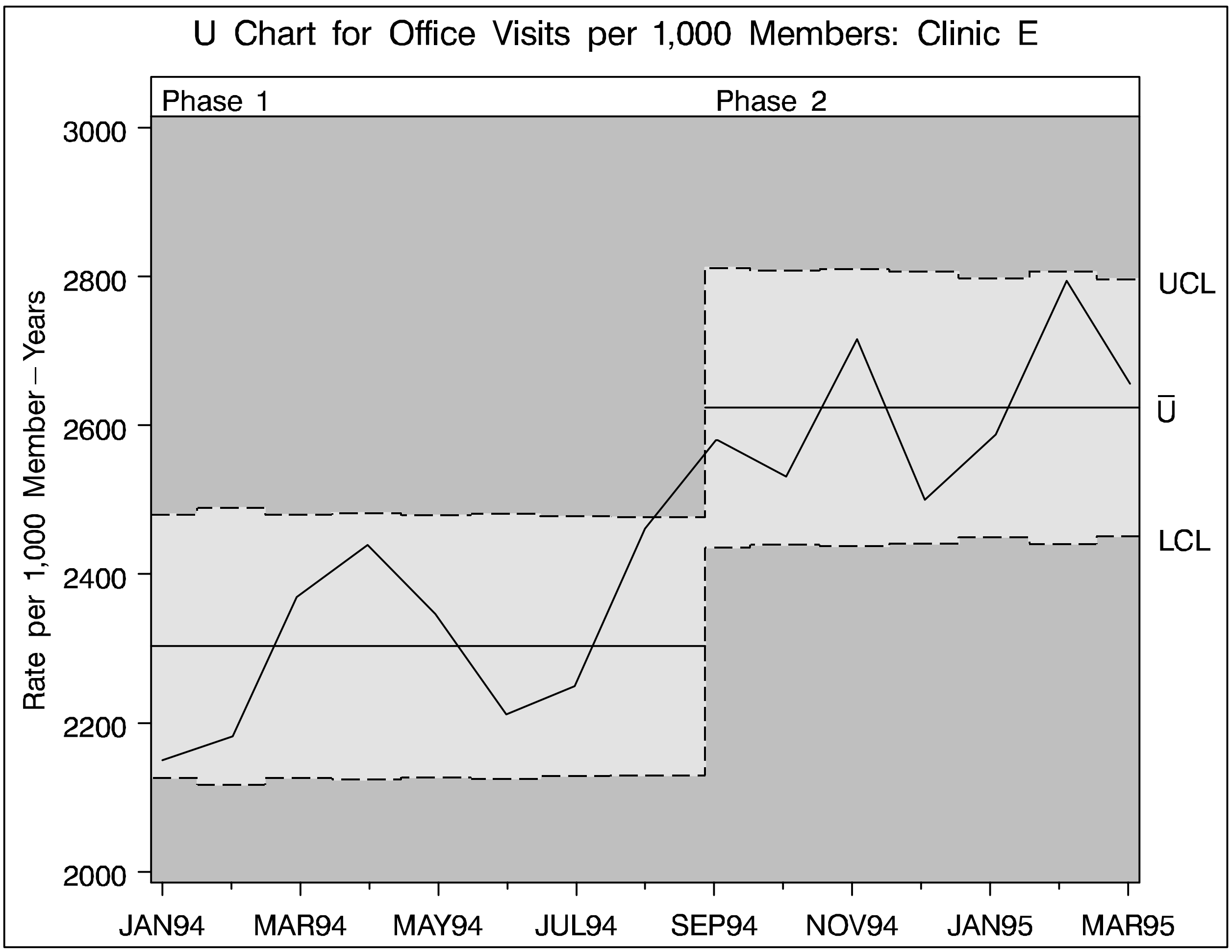


Figure 5.4. u Chart with Multiple Sets of Limits

Note that both sets of control limits in Figure 5.4 were estimated from the data with which they are displayed. You can, however, apply pre-established control limits from a LIMITS= data set to new data.

Also note that _PHASE_ is a reserved variable name.

See page 1458 of QCUR for various methods of constructing historical control charts with the READPHASE= and READINDEX= options.

Displays for Stratified Data

If the data for a Shewhart chart can be classified by factors relevant to the process (for instance, machines or operators), displaying the classification on the chart can facilitate the identification of causes of variation.

There are important differences between stratification and subgrouping.

- The data must always be classified into subgroups before a control chart can be produced.
- Subgrouping affects how control limits are computed from the data as well as the outcome of tests for special causes.
- The *subgroup-variable* is mandatory and classifies the data into subgroups.
- Stratification is optional and involves classification variables other than the *subgroup-variable*.
- Displaying stratification influences how the chart is interpreted, but it does not affect control limits or tests for special causes.

Stratification Based On Levels of a Classification Variable

MACHINE	SAMPLE	DAY	SHIFT	DIAMX	DIAMS	DIAMN
A386	1	1	1	4.32	0.39	6
A386	2	1	2	4.49	0.35	6
...						
A386	9	3	3	4.47	0.40	6
A455	10	4	1	4.42	0.37	6
A455	11	4	2	4.45	0.32	6
...						
A455	15	5	3	4.17	0.25	6
C334	16	8	1	4.15	0.28	6
C334	17	8	2	4.21	0.33	6
...						
C334	24	10	3	4.23	0.14	6
A386	25	11	1	4.27	0.28	6
A386	26	11	2	4.70	0.45	6
A386	27	11	3	4.51	0.45	6
...						
A386	33	15	3	4.52	0.33	6

Figure 5.5. Data Set PARTS

```
symbol1 value=star     color=black height=3 pct;
symbol2 value=circle   color=black height=3 pct;
symbol3 value=triangle color=black height=3 pct;

title 'Control Chart for Diameter '
      'Stratified by Machine';

proc shewhart history=parts graphics;
   xchart diam*sample=machine /
      symbollegend=legend1;

label sample = 'Sample Number'
      diamx  = 'Average Diameter' ;
legend1 label=('Machine') frame;
run;
```

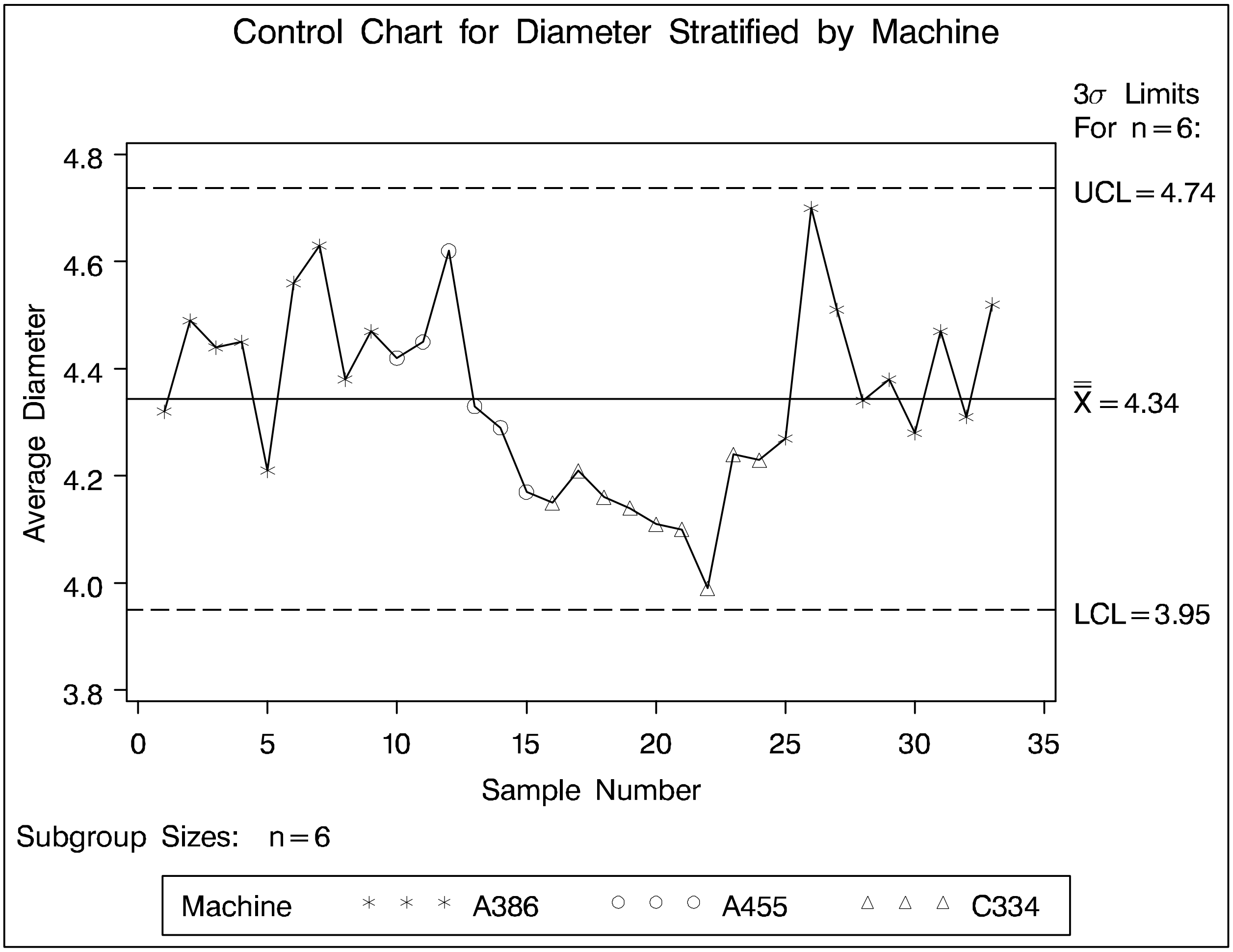


Figure 5.6. Stratification Using Symbols

What can we learn from this display?

For details, see page 1449 of QCUR.

Stratification in Blocks of Observations

```
data parts2;
   length cmachine $8;
   set parts;
   if      machine='A386' then cmachine='ligr' ;
   else if machine='A455' then cmachine='megr' ;
   else if machine='C334' then cmachine='white';
   else                        cmachine='dagr' ;

proc shewhart history=parts2 graphics;
   xchart diam*sample (machine day) /
      stddeviations
      nolegend
      blockpos   = 3
      cblockvar  = cmachine;
   label sample  = 'Sample Number'
         diamx   = 'Average Diameter'
         day     = 'Date of Production in June'
         machine = 'Machine in Use';
run;
```

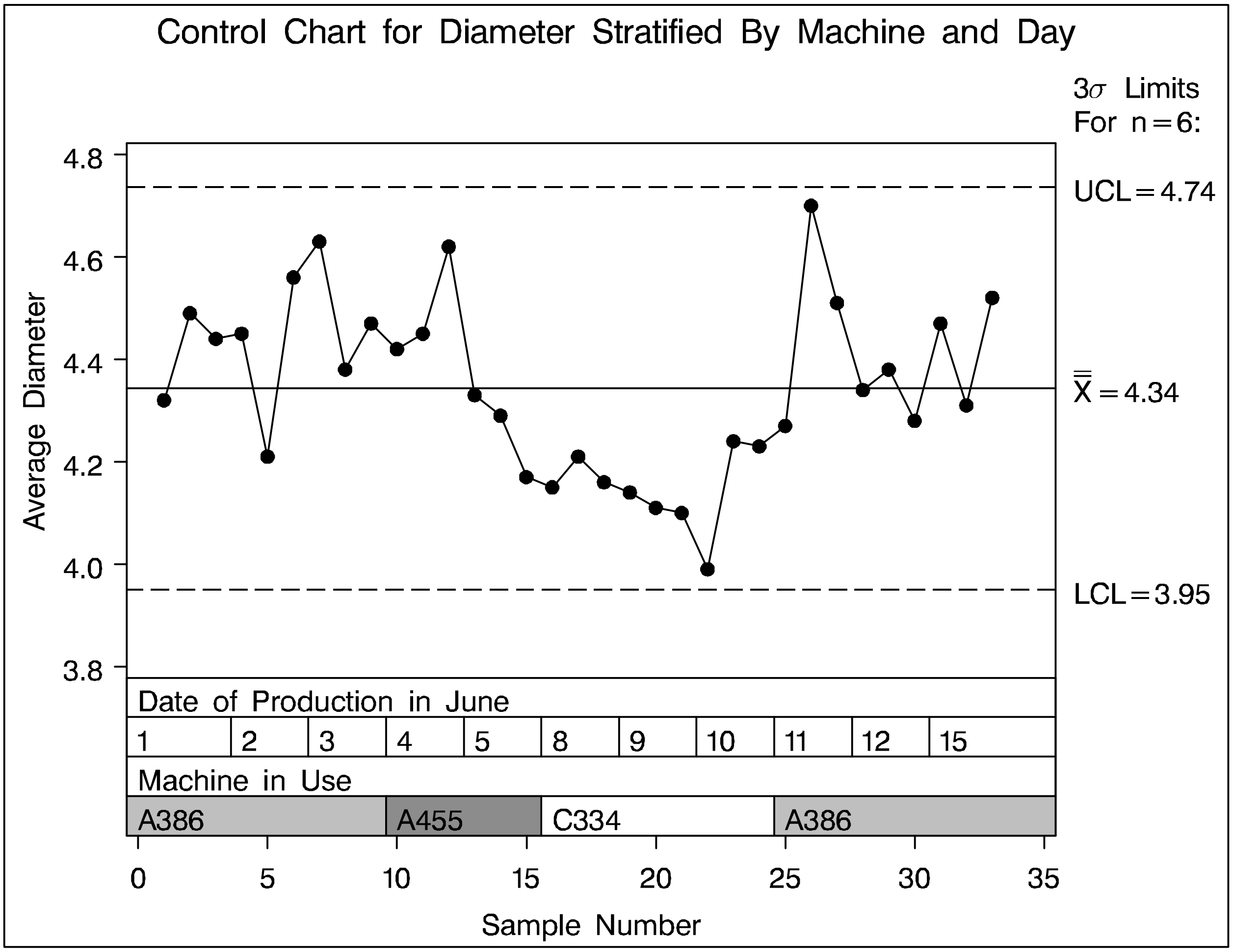


Figure 5.7. Stratification Using Blocks

For details, see page 1450 of QCUR.

Boxplot Displays

The data set DELAYS contains flight departure delays (in minutes) recorded daily for eight consecutive days:

DELAY	DAY	FLIGHT
12	01MAR90	1
4	01MAR90	2
2	01MAR90	3
2	01MAR90	4
15	01MAR90	5
8	01MAR90	6
0	01MAR90	7
11	01MAR90	8
0	01MAR90	9
0	01MAR90	10
...	...	...
4	08MAR90	23
1	08MAR90	24
1	08MAR90	25

Figure 5.8. Data Set DELAYS

The following statements create an $\bar{X}$ chart for the average delays superimposed with box-and-whisker plots of all the delays for each day.

```
title 'Analysis of Airline Departure Delays';
symbol v=plus;

proc shewhart graphics data=delays;
   boxchart delay * day /
      stddevs
      nohlabel
      interval      = day
      boxstyle      = schematic
      boxwidthscale = 1 ;
run;
```

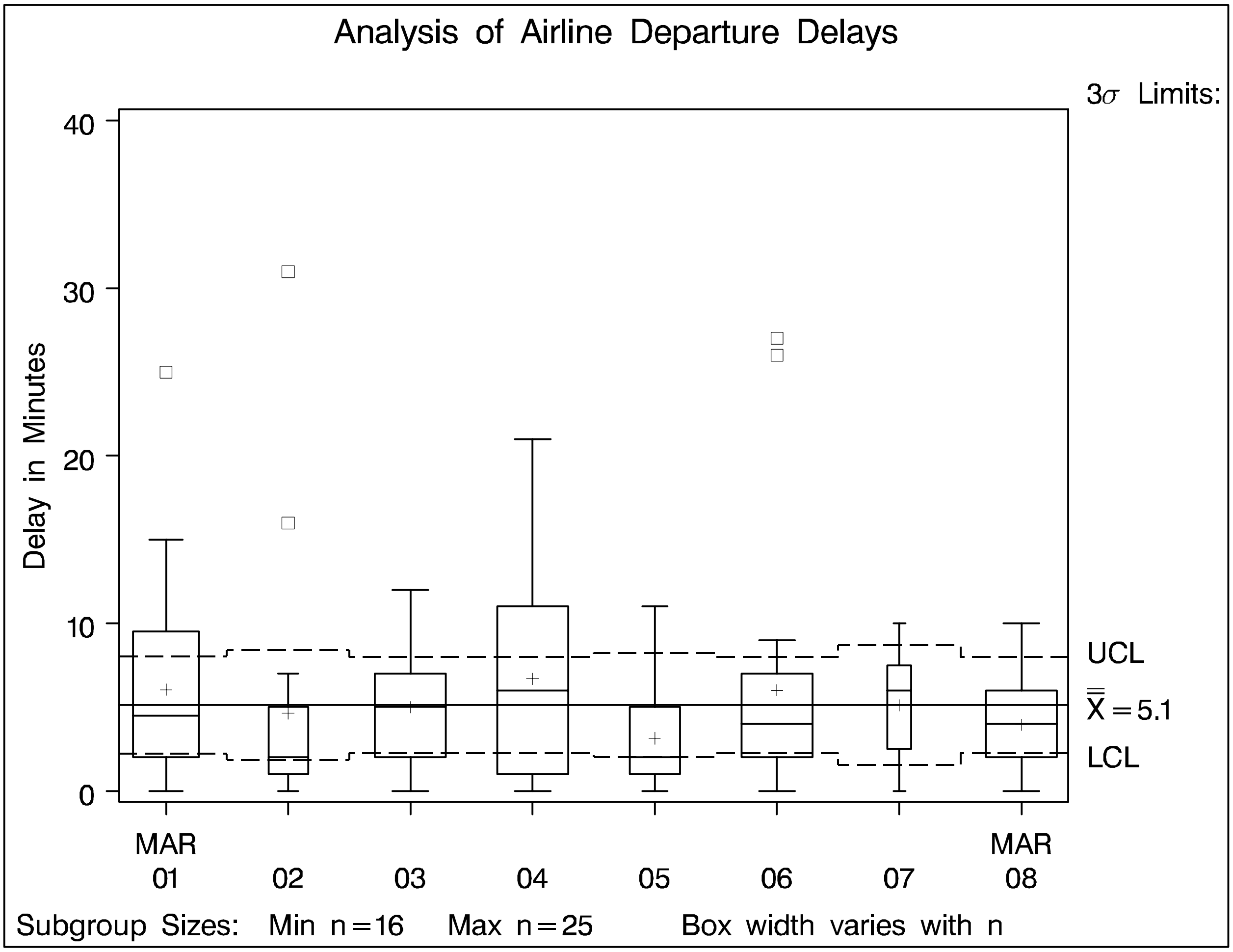


Figure 5.9. Box Chart

You can also use the BOXCHART statement to create a wide variety of box-and-whisker displays *without* control limits.

Tip: Be sure to specify NOLIMITS and STDDEVS.

For details, see page 911 of QCUR.

```
title 'Analysis of Airline Departure Delays';

proc shewhart graphics data=times ;
   boxchart delay * day /
      boxstyle = schematicid
      cboxes   = black
      cboxfill = ligr
      interval = day
      stddevs
      nolimits
      nohlabel
      nolegend
      notches;

   id reason;
   label delay   = 'Delay in Minutes';
   run;
```

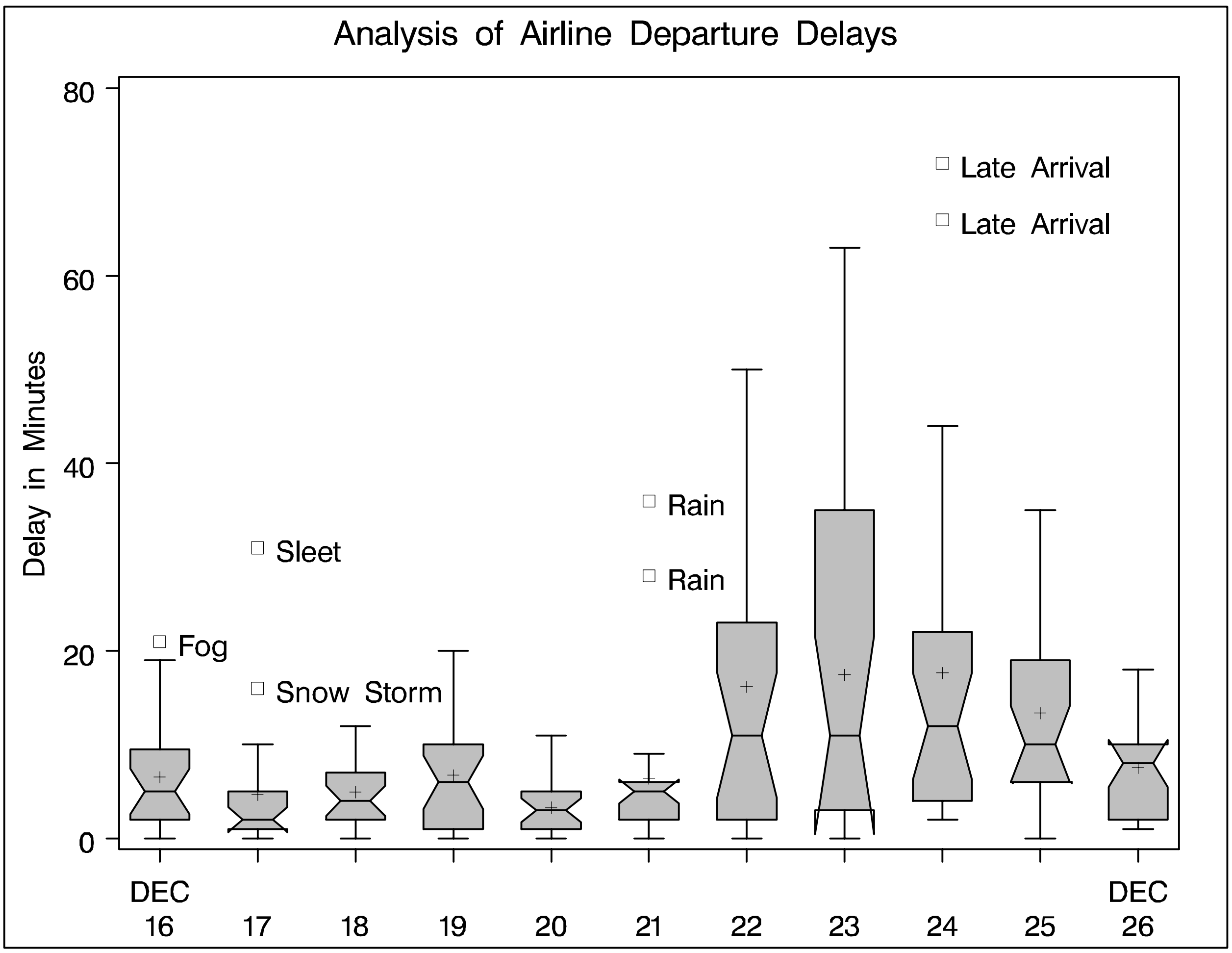


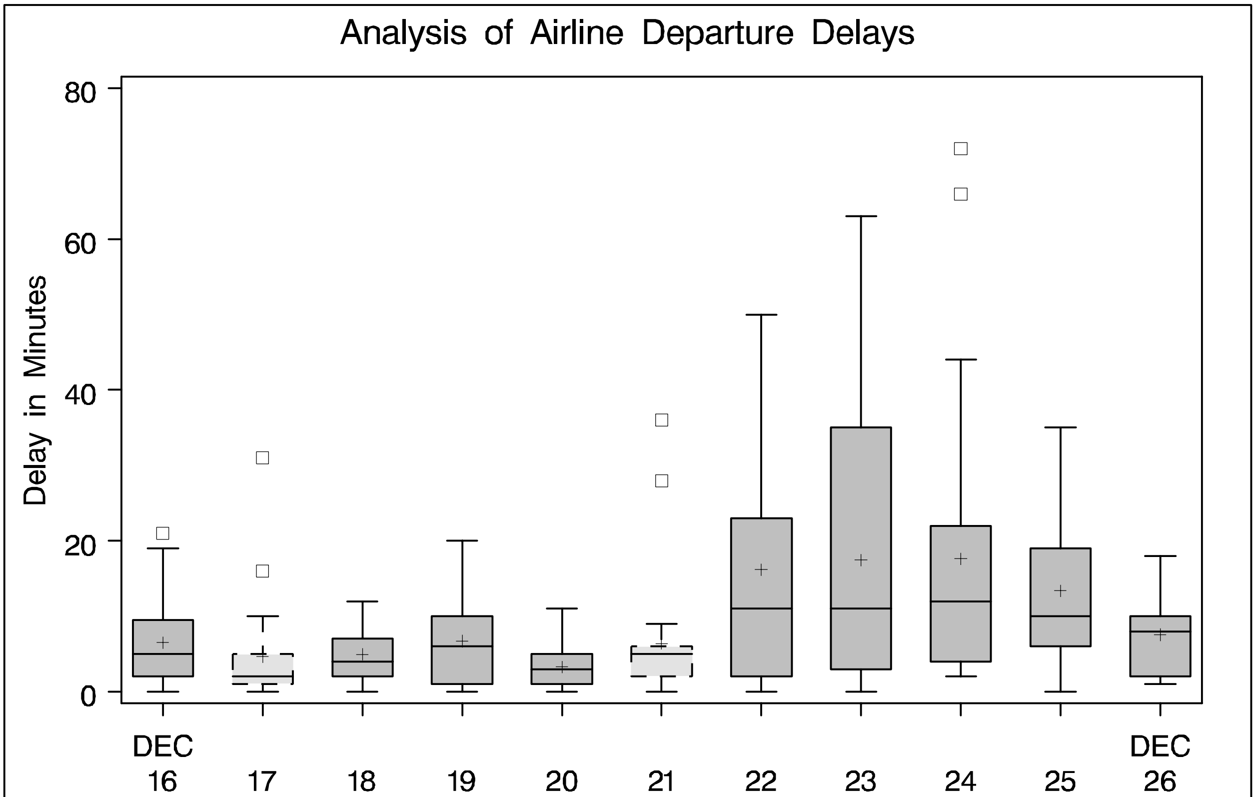
Figure 5.10. Box-and-Whisker Display

The next example illustrates how you can add a table of summary statistics to a boxplot display.

```
proc univariate noprint data=times;
   var delay;
   by day;
   output out=stats n=n mean=avg skewness=skew;

data stats;
   set stats;
   label tabn   = 'N'
         tabavg = 'Mean'
         tabskw = 'Skewness';
   length tabn tabavg tabrng $ 4 color $ 8;
   tabn   = put( n,    best4. );
   tabavg = put( avg,  best4. );
   tabskw = put( skew, 4.2    );

   if ( skew > 2.0 ) then do;
      color = 'yellow';
      lnstyle = 20;
      end;
   else do;
      color = 'ligr';
      lnstyle = 1;
      end;

data times;
   merge times stats;
   by day;
run;
```

In the next statements, the formatted summary variables are read as “block variables” by the SHEWHART procedure. Note that the colors and line styles of the boxes are used to indicate high skewness.

```
title  'Analysis of Airline Departure Delays' ;

symbol v=plus;
proc shewhart graphics data=times;
   boxchart delay * day ( tabskw tabavg tabn ) /
      llimits       = 1
      blockrep
      blocklabelpos = left
      blocklabtype  = 3
      blockpos      = 4
      stddevs
      nolimits
      nohlabel
      nolegend
      boxstyle      = schematic
      cboxfill      = ( color )
      cboxes        = black
      lboxes        = ( lnstyle  );
label delay = 'Delay in Minutes';
run;
```



N	24	23	25	25	24	25	25	23	25	25	25
Mean	6.54	4.65	4.96	6.72	3.29	6.4	16.2	17.4	17.6	13.4	7.56
Skew	1.12	2.96	0.56	0.64	1.00	2.94	0.81	1.02	1.80	0.79	0.48

Figure 5.11. Adding a Table of Summary Statistics

Multi-Vari Displays

"Multi-vari" charts are used in a variety of industries to analyze several types of variation:

- within-sample (*e.g.*, position within wafer)
- within-batch (*e.g.*, wafer within batch)
- batch-to-batch

The following example (contributed by Leslie Fowler at Motorola) illustrates the construction of "multi-vari" displays.

PHASE	WAFER	POSITION	PARM
LOTA	01	L	0.96974
LOTA	01	B	0.97660
LOTA	01	C	0.96857
LOTA	01	T	0.97984
LOTA	01	R	1.00020
LOTA	02	L	1.07275
LOTA	02	B	1.02878
LOTA	02	C	1.01871
LOTA	02	T	1.06391
LOTA	02	R	1.07683
LOTA	03	L	0.87202
LOTA	03	B	0.85437
LOTA	03	C	0.97721
LOTA	03	T	0.91621
LOTA	03	R	0.90385
LOTB	01	L	0.98629
...	...	...	...
LOTG	03	T	1.06365
LOTG	03	R	1.13751

Figure 5.12. Data Set PARM

```
title 'Multi-Vari Display for Measured Parameter';

proc shewhart data=parm graphics;
   boxchart parm*wafer /
      cboxes             = black
      cphasebox          = black
      cphaseboxfill      = ligr
      cphasemeanconnect = black
      boxstyle           = pointsjoin
      phasemeansymbol   = dot
      readphase          = all
      phaselegend
      nolegend
      nohlabel
      stddevs
      nolimits;
run;
```

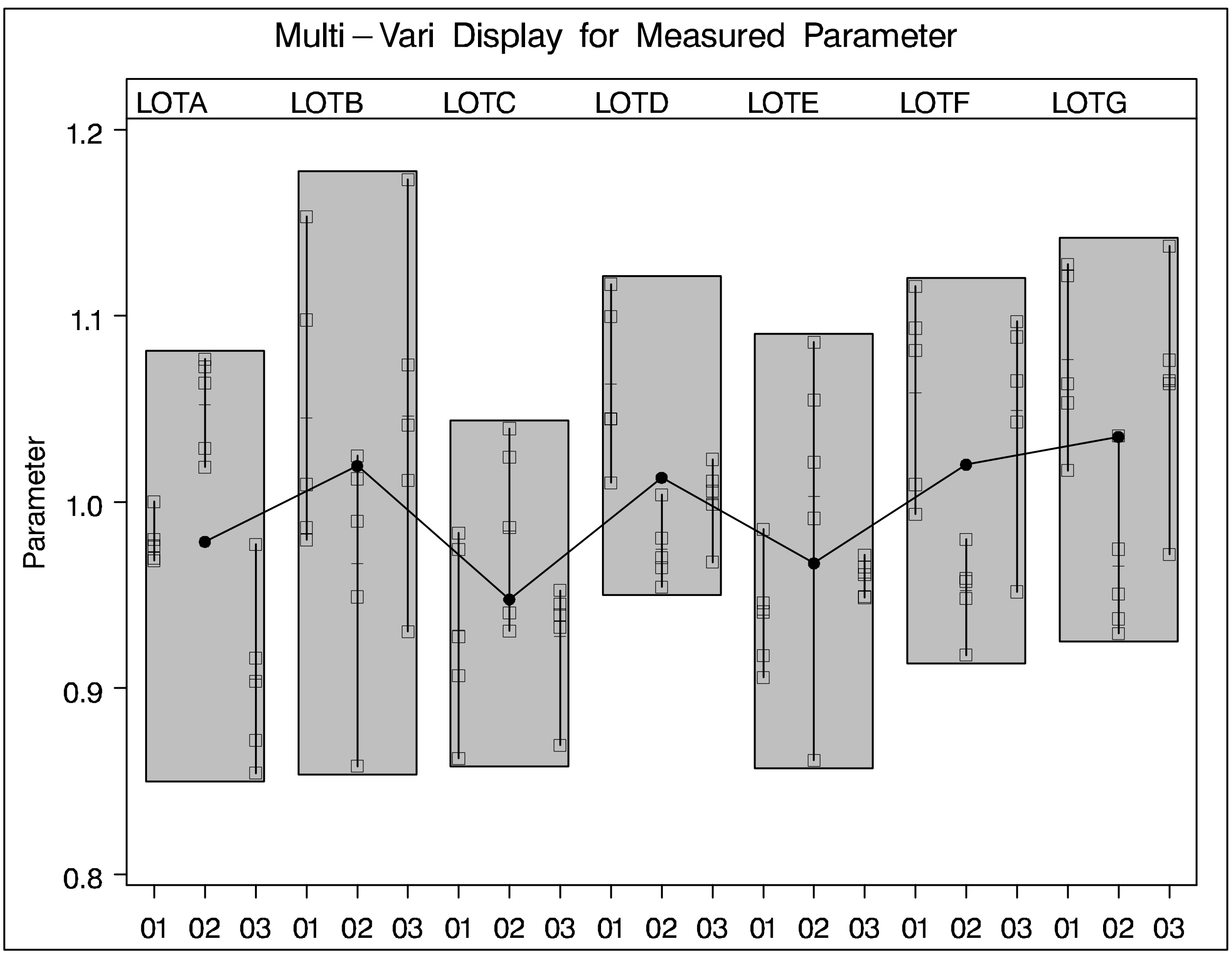


Figure 5.13. BOXSTYLE=POINTSJOIN Option

```
title 'Multi-Vari Display for Measured Parameter';

proc shewhart data=parm graphics;
   boxchart parm*wafer /
      cboxes              = black
      cphasebox           = black
      cphaseboxfill       = ligr
      cphasemeanconnect = black
      boxstyle            = pointsid
      phasemeansymbol   = dot
      readphase           = all
      phaselegend
      nolegend
      nohlabel
      stddevs
      nolimits;
id position;
run;
```

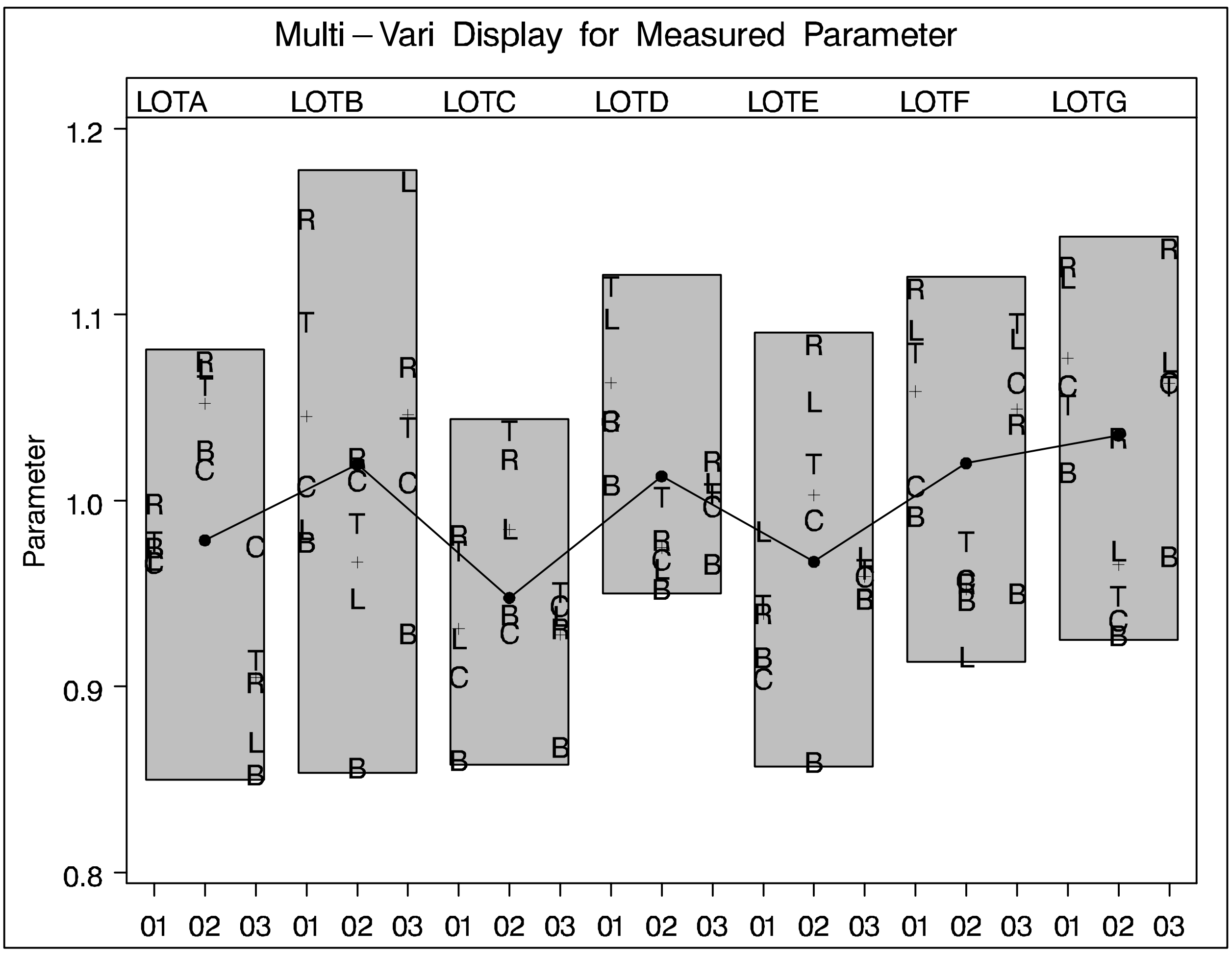


Figure 5.14. BOXSTYLE=POINTSID Option

Scaling Issues

Scale compression can be a problem when extreme out-of-control points are plotted.

```
title 'Control Chart for New Copper Tubes';
symbol v=plus;

proc shewhart data=newtubes graphics;
   xrchart diameter*batch /
      mu0    = 70
      sigma0 = 0.75 ;
run;
```

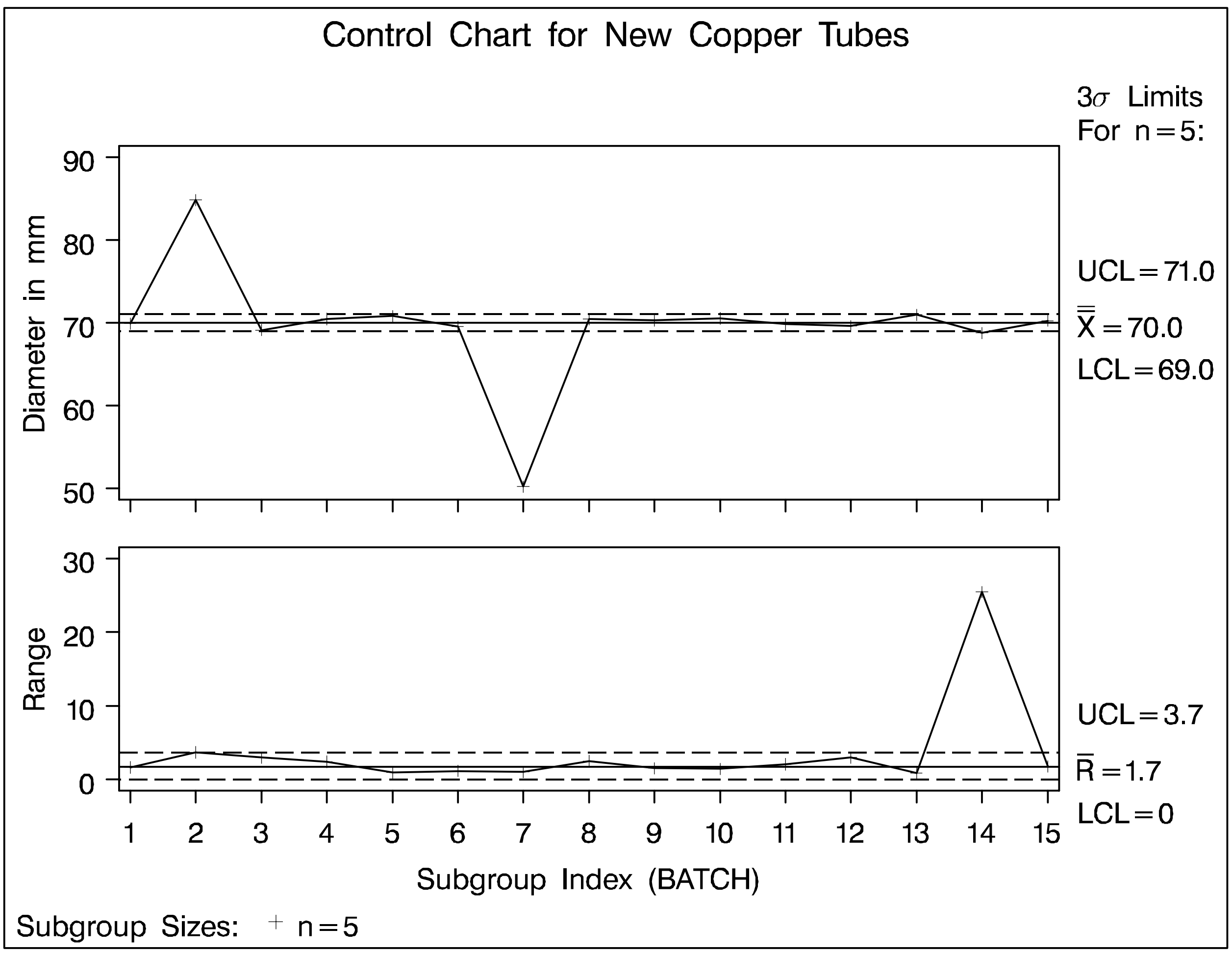


Figure 5.15. $\bar{X}$ and R Charts Without Clipping

You can use the CLIPFACTOR= option as follows:

```
title 'Control Chart for New Copper Tubes';
symbol v=plus;

proc shewhart data=newtubes graphics;
   xrchart diameter*batch /
      mu0        = 70
      sigma0     = 0.75
      clipfactor = 1.5;
run;
```

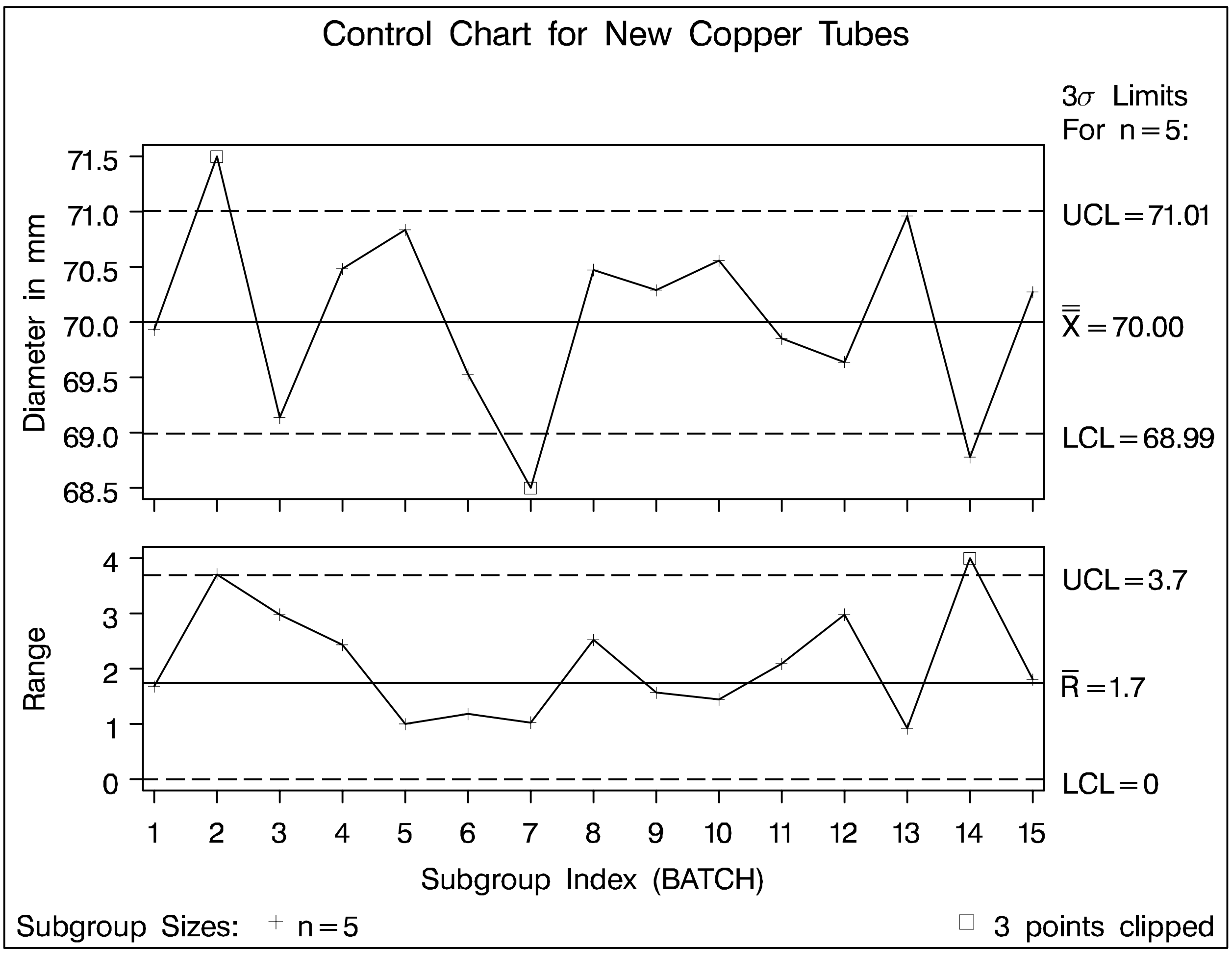


Figure 5.16. $\bar{X}$ and R Charts With Clipping

For details, see page 1481 of QCUR.

Selecting Subgroups for Computation and Display

In many control chart applications, it is necessary to distinguish between

- the subgroups used to compute the control limits (or to estimate the control limit parameters)
- the subgroups that are displayed on the control chart

There are a variety of methods for making this distinction:

- read the control limits from a LIMITS= data set
- use a WHERE statement

```
proc shewhart data=bottles graphics;
   where day <= '31JAN94'D;
   pchart ncracks * day /
      subgroupn = nbottles
      outlimits = botlim;
run;
```

- use the switch variables _COMP_ and _DISP_ in the input data set

```
data bottles;
   length _comp_ _disp_ $ 1;
   set bottles;
   if      day  = '13JAN94'D then _comp_ = 'n';
   else if day  = '14JAN94'D then _comp_ = 'n';
   else if day <= '31JAN94'D then _comp_ = 'y';
   else                           _comp_ = 'n';
   if      day <= '31JAN94'D then _disp_ = 'n';
   else                           _disp_ = 'y';
run;

title 'Analysis of February Production';
proc shewhart graphics data=bottles;
   pchart ncracks * day / subgroupn = nbottles
                          nolegend
                          nohlabel;
   label ncracks = 'Proportion With Cracks';
run;
```

For details, see page 1489 of QCUR.

Chapter 6

Statistical Modeling for Control Chart Applications

While the Shewhart chart is remarkably versatile, it is not the best solution for every SPC application. There is a growing awareness among quality engineers that standard control charts are inappropriate or of limited value in a number of manufacturing situations:

- “What can I do about autocorrelation in my process data?”
- “Is there a way to adjust control charts for multiple sources of variation?”
- “How can I do short run process control?”
- “Are the standard control limits appropriate for nonnormal data?”

These questions are subjects of current research and debate. Here, the goal is to mention some of the approaches that have been proposed and illustrate how they can be implemented with short SAS programs.

The examples in this chapter use SAS procedures for statistical modeling in conjunction with the SHEWHART procedure. This combination is highly effective in a variety of nonstandard SPC applications.

Note: The examples are readily extended to SAS programs that can handle large numbers of processes. Likewise, they can be incorporated in customized point-and-click interfaces developed with SAS/AF® software.

Autocorrelation in Process Data

- recognized as a natural phenomenon in process industries
- as automated data collection becomes prevalent in parts industries, it is possible to recognize autocorrelation that was previously undetected
- the distinction between parts and process industries is becoming blurred in areas such as computer chip manufacturing
- see Box and Kramer (1992), Schneider and Pruett (1994) and Woodall (1993)

The standard Shewhart "model" is

$$x_t = \mu + \epsilon_t$$

where ϵ_t is a random displacement or error from the process mean μ. The errors are typically assumed to be statistically independent in control chart derivations. When measurements are autocorrelated, the result can be too many false signals, and users sometimes comment that "the control limits are too tight."

The following data are from Montgomery and Mastrangelo (1991).

```
                Chemical Data

                  T      XT

                  1      96
                  2      90
                  3      89
                  4      89
                  .       .
                  .       .
                  .       .
                 96      81
                 97      83
                 98      83
                 99      86
                100      88
```

Figure 6.1. SAS Data Set CHEMICAL

The following code produces Figure 6.2:

```
title 'Individual Measurements Chart';

proc shewhart graphics data=chemical;
   irchart xt*t / npanel = 100
                  needles
                  split  = '/';
   label xt = 'Observed/Mvg Rng';
run;
```

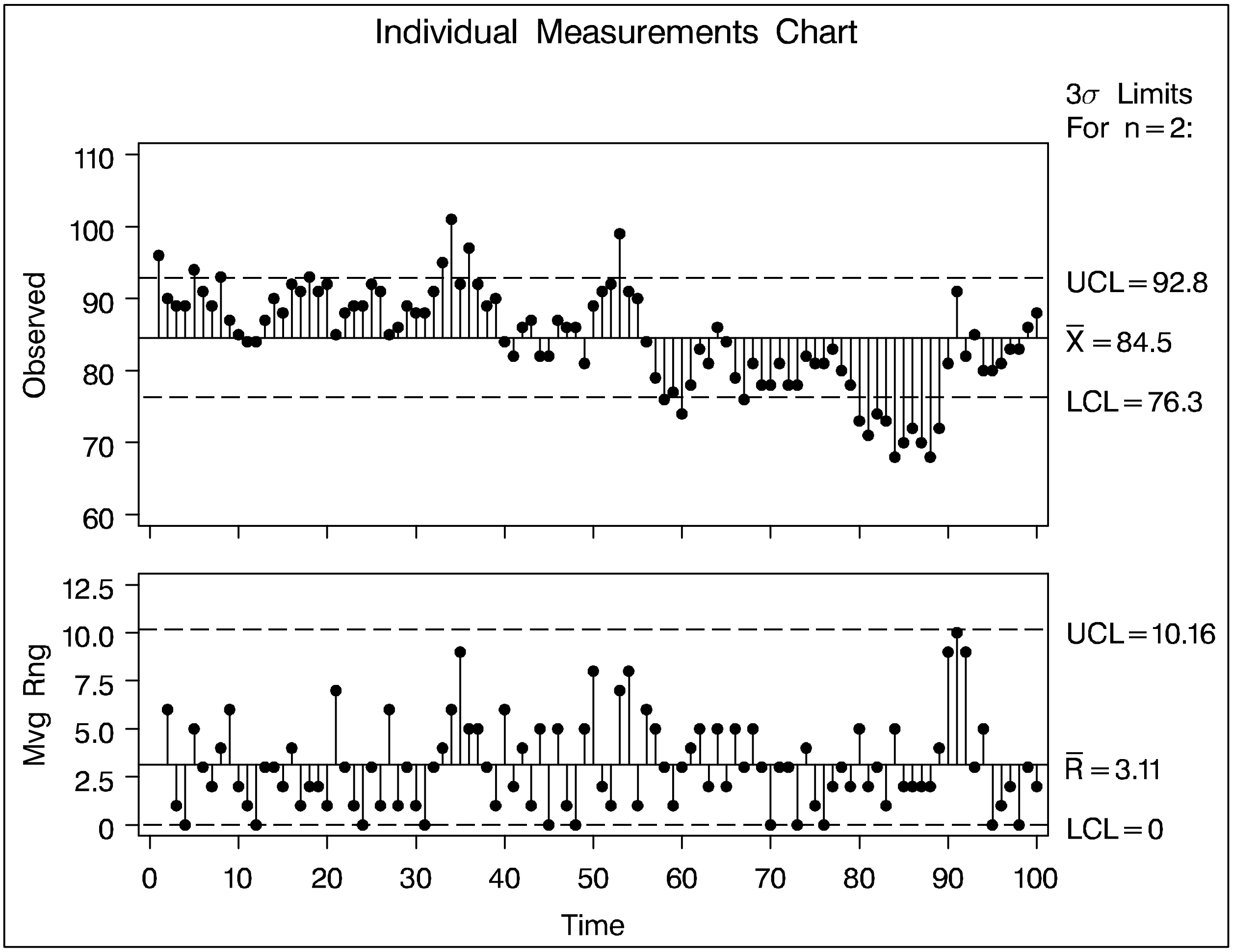


Figure 6.2. Conventional Shewhart Chart

You can use the ARIMA procedure for diagnosis and modeling of autocorrelation.

```
   proc arima data=chemical;
      identify var = xt;
   run;
```

See *SAS/ETS® User's Guide, Version 6, Second Edition.*

```
                     Autocorrelations

Lag  Corr     -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
  0  1.00000  |                    |********************|
  1  0.83026  |                .   |*****************   |
  2  0.71837  |              .     |**************      |
  3  0.61948  |             .      |************        |
  4  0.51169  |            .       |**********          |
  5  0.42596  |           .        |*********           |
  6  0.38114  |           .        |********.           |
  7  0.35989  |           .        |******* .           |
  8  0.36446  |           .        |******* .           |
  9  0.37982  |          .         |******** .          |
 10  0.34653  |          .         |*******  .          |
 11  0.34841  |          .         |*******  .          |
 12  0.35445  |          .         |*******  .          |
 13  0.34922  |         .          |*******   .         |
 14  0.37081  |         .          |*******   .         |
 15  0.34752  |         .          |*******   .         |
                           "." marks two standard errors
```

Figure 6.3. Autocorrelation Plot for Chemical Data

Figure 6.3 indicates that the data are highly autocorrelated with a lag one autocorrelation of 0.83.

```
                Partial Autocorrelations

 Lag Correlation -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
   1    0.83026  |                .   |*****************   |
   2    0.09346  |                .   |** .                |
   3    0.00385  |                .   |   .                |
   4   -0.07340  |                .  *|   .                |
   5   -0.00278  |                .   |   .                |
   6    0.09013  |                .   |** .                |
   7    0.08781  |                .   |** .                |
   8    0.10327  |                .   |** .                |
   9    0.07240  |                .   |*  .                |
  10   -0.11637  |                . **|   .                |
  11    0.08210  |                .   |** .                |
  12    0.07580  |                .   |** .                |
  13    0.04429  |                .   |*  .                |
  14    0.11661  |                .   |** .                |
  15   -0.10446  |                . **|   .                |
```

Figure 6.4. Partial Autocorrelation Plot

Figure 6.4 suggests a first-order autoregressive model, commonly referred to as an AR(1) model:

$$\tilde{x}_t \equiv x_t - \mu = \phi_0 + \phi_1 \tilde{x}_{t-1} + \epsilon_t$$

You can fit this model with the ARIMA procedure:

```
proc arima data=chemical;
   identify var = xt;
   estimate p = 1 method = ml;
run;
```

The equation of the fitted model is

$$\tilde{x}_t = 13.05 + 0.847\tilde{x}_{t-1}$$

```
              Maximum Likelihood Estimation

                               Approx.
      Parameter   Estimate    Std Error   T Ratio  Lag
      MU          85.28375      2.32973     36.61   0
      AR1,1        0.84694      0.05221     16.22   1

      Constant Estimate  = 13.0532881

      Variance  Estimate = 14.2767606
      Std Error Estimate = 3.77846008
      AIC                = 552.894156
      SBC                = 558.104497
      Number of Residuals=        100
```

Figure 6.5. Fitted AR(1) Model

There are differing views on dealing with autocorrelation:

1. Wheeler (1991b) argues that the usual control limits are contaminated "only when the autocorrelation becomes excessive (say 0.80 or larger)" and that, in these situations, the "running record [is] very coherent and therefore very easy to understand."

2. others suggest removing autocorrelation from the data and constructing a chart for the residuals; see Alwan and Roberts (1988).

3. The "automatic process control" or "engineering process control" approach views dependence as a phenomenon to be exploited rather than removed. See MacGregor (1987,1990), MacGregor, Hunter, and Harris (1988), Montgomery *et al.* (1994), and Box and Kramer (1992).

The following example illustrates the second viewpoint and is based on Montgomery and Mastrangelo (1991). In the chemical data example, the residuals can be computed as forecast errors and saved in a SAS data set named RESULTS.

```
proc arima data=chemical;
   identify var=xt;
   estimate p=1 method=ml;
   forecast out=results id=t;
run;
```

RESULTS saves the one-step-ahead forecasts as the variable FORECAST, and it also contains the original variables XT and T. You can create a Shewhart chart for the residuals by using RESULTS as input to the SHEWHART procedure:

```
title 'Residual Analysis Using AR(1) Model';
proc shewhart data=results graphics;
   xchart xt*t / trendvar = forecast
                 xsymbol  = xbar
                 npanel   = 100
                 ypct1    = 40
                 split    = '/'
                 nolegend;
   label xt='Residual/Forecast';
run;
```

The lower chart in Figure 6.6 plots FORECAST, and the upper chart plots the residuals (XT - FORECAST) together with their 3σ limits.

Note that the TRENDVAR= option requests this display (otherwise, a standard individual measurements chart for XT would be produced by default).

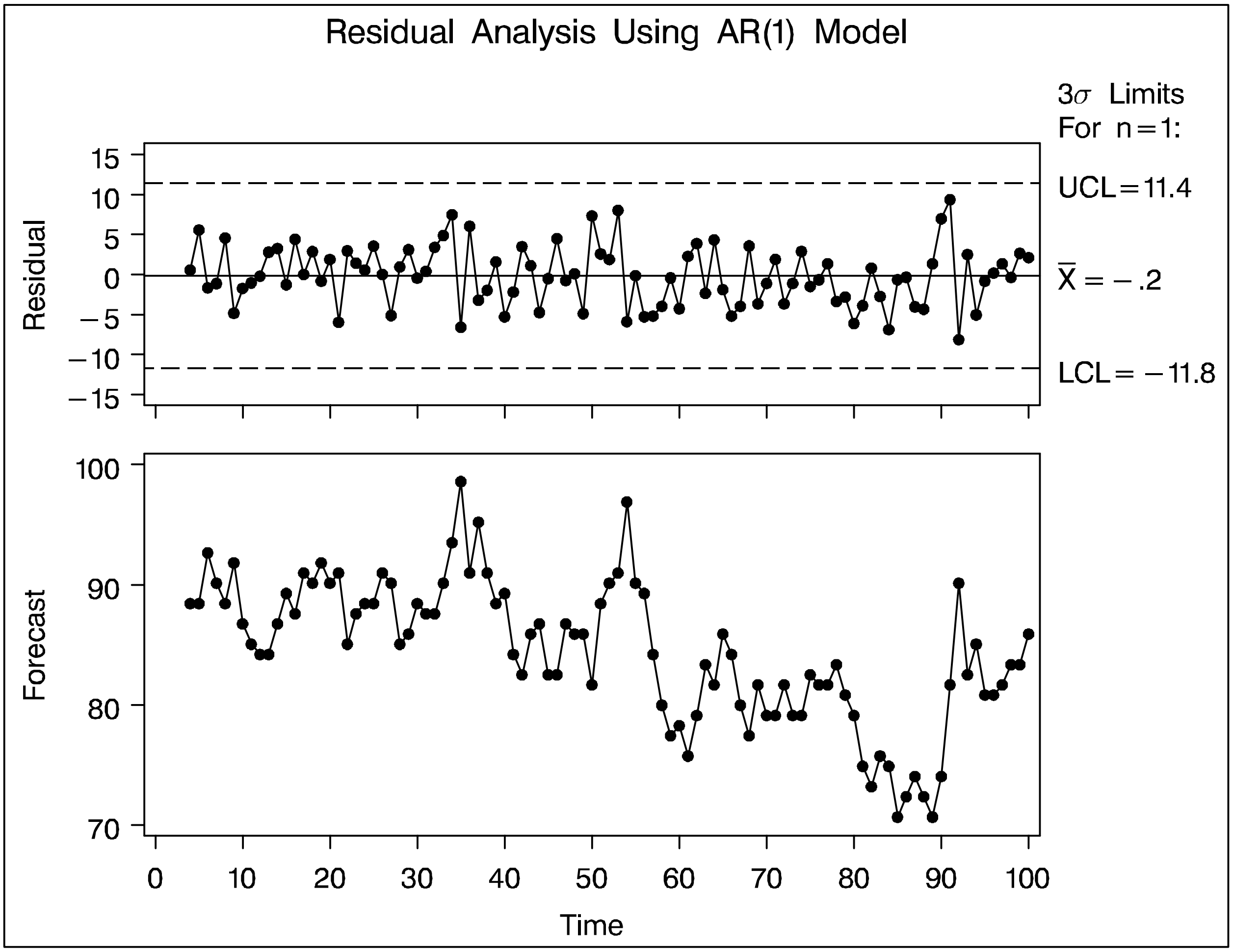


Figure 6.6. Residuals from AR(1) Model

The upper chart in Figure 6.6 resembles Figure 2 of Montgomery and Mastrangelo (1991), who conclude that the process is in control. In addition, they suggest fitting an exponentially weighted moving average (EWMA) *model* to the data and using this model as the basis for an "EWMA center line control chart."

Recall that the EWMA *statistic* plotted on a conventional EWMA control chart is defined as

$$z_t = \lambda x_t + (1 - \lambda) z_{t-1}$$

The EWMA chart (which you can construct with the MACONTROL procedure) assumes the x_t are independent. However, in the context of autocorrelated process data, the EWMA statistic z_t plays a different role; it is the optimal one-step-ahead forecast for a process that can be modeled by an ARIMA(0,1,1) model

$$x_t = x_{t-1} + \epsilon_t - \theta \epsilon_{t-1}$$

provided that $\lambda = 1 - \theta$. See Hunter (1986). This statistic is also a good predictor when the process can be described by a subset of ARIMA models for which the process is "positively autocorrelated and the process mean does not drift too quickly."

You can fit an ARIMA(0,1,1) model to the chemical data with the following statements:

```
proc arima data=chemical;
   identify var=xt(1);
   estimate q=1 method=ml noint;
   forecast out=ewma id=t;
run;
```

The forecast values and their standard errors, together with the original measurements, are saved in a data set named EWMA. The following statements create a Shewhart chart for the residuals from the fitted ARIMA(0,1,1) model.

```
data ewma; set ewma(firstobs=2 obs=100);

title 'Residual Analysis Using ARIMA(0,1,1) '
      'Model';
symbol v=dot;

proc shewhart data=ewma graphics;
   xchart xt*t / trendvar = forecast
                 xsymbol  = xbar
                 npanel   = 100
                 ypct1    = 40
                 split    = '/'
                 nolegend;
   label xt = 'Residual/Forecast';
run;
```

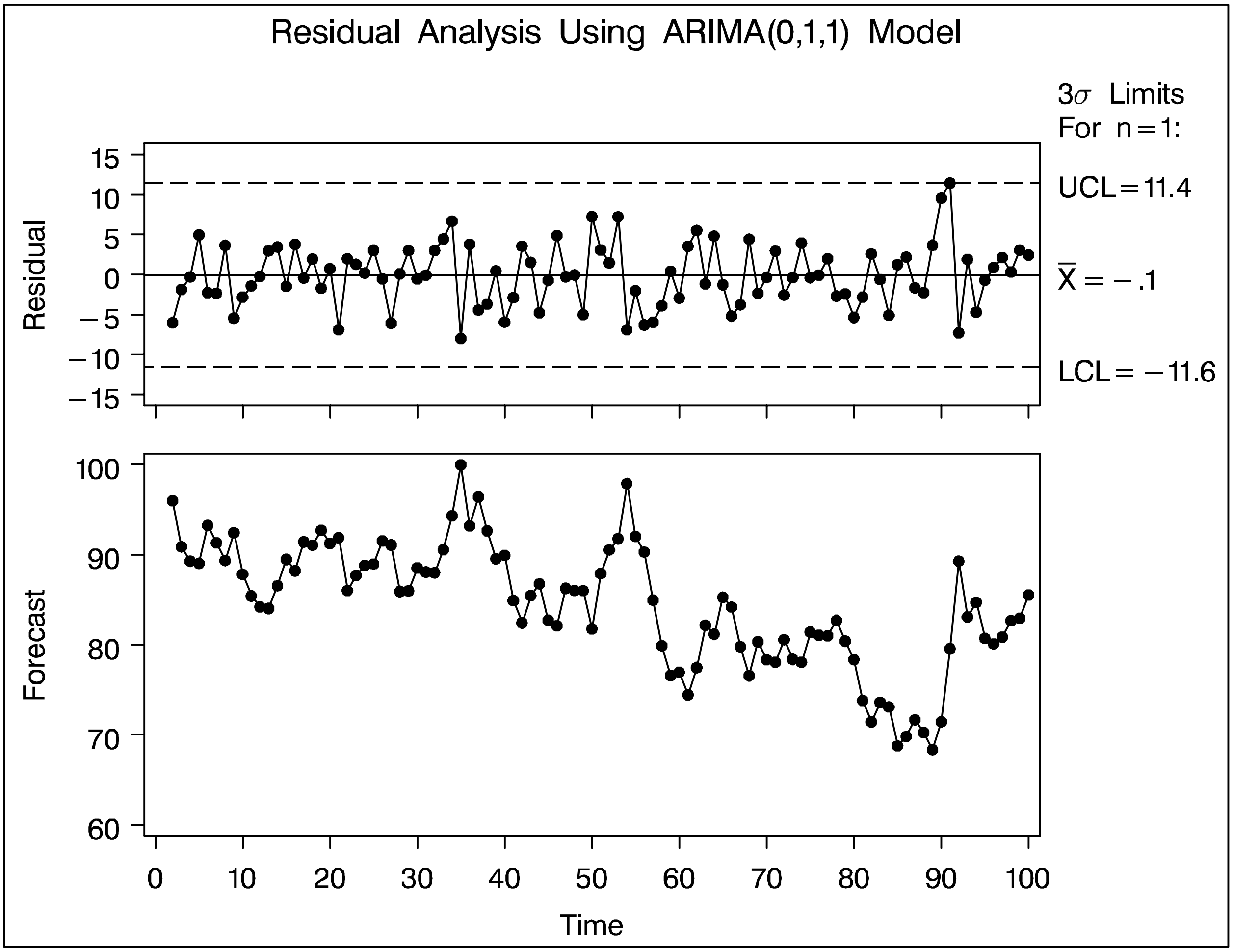


Figure 6.7. Residuals from ARIMA(0,1,1) Model

Note the similarity between Figure 6.7, and Figure 6.6.

The following statements construct the EWMA center line control chart by plotting the forecasts from the ARIMA(0,1,1) model as the center "line" and using the standard errors of prediction to compute upper and lower control limits.

```
data ewmatab;
   length _var_ $ 8 ;
   set ewma
     (rename=(forecast=_mean_ xt=_subx_));
   _var_    = 'xt';
   _sigmas_ = 3;
   _limitn_ = 1;
   _lclx_   = _mean_ - 3 * std;
   _uclx_   = _mean_ + 3 * std;
   _subn_   = 1;
run;

title 'EWMA Center Line Control Chart';
symbol v=dot;

proc shewhart table=ewmatab graphics;
   xchart xt*t / npanel   = 100
                 llimits  = 1
                 xsymbol  = 'Center'
                 nolegend;
   label _subx_ = 'Observed';
run;
```

FORECAST and STD are the forecasts and standard errors in the data set EWMA created earlier by the ARIMA procedure.

EWMATAB is read by the SHEWHART procedure as a TABLE= input data set. Why?

Recall that variables in a TABLE= data set have reserved names, and for this reason FORECAST and XT are temporarily renamed as _MEAN_ and _SUBX_.

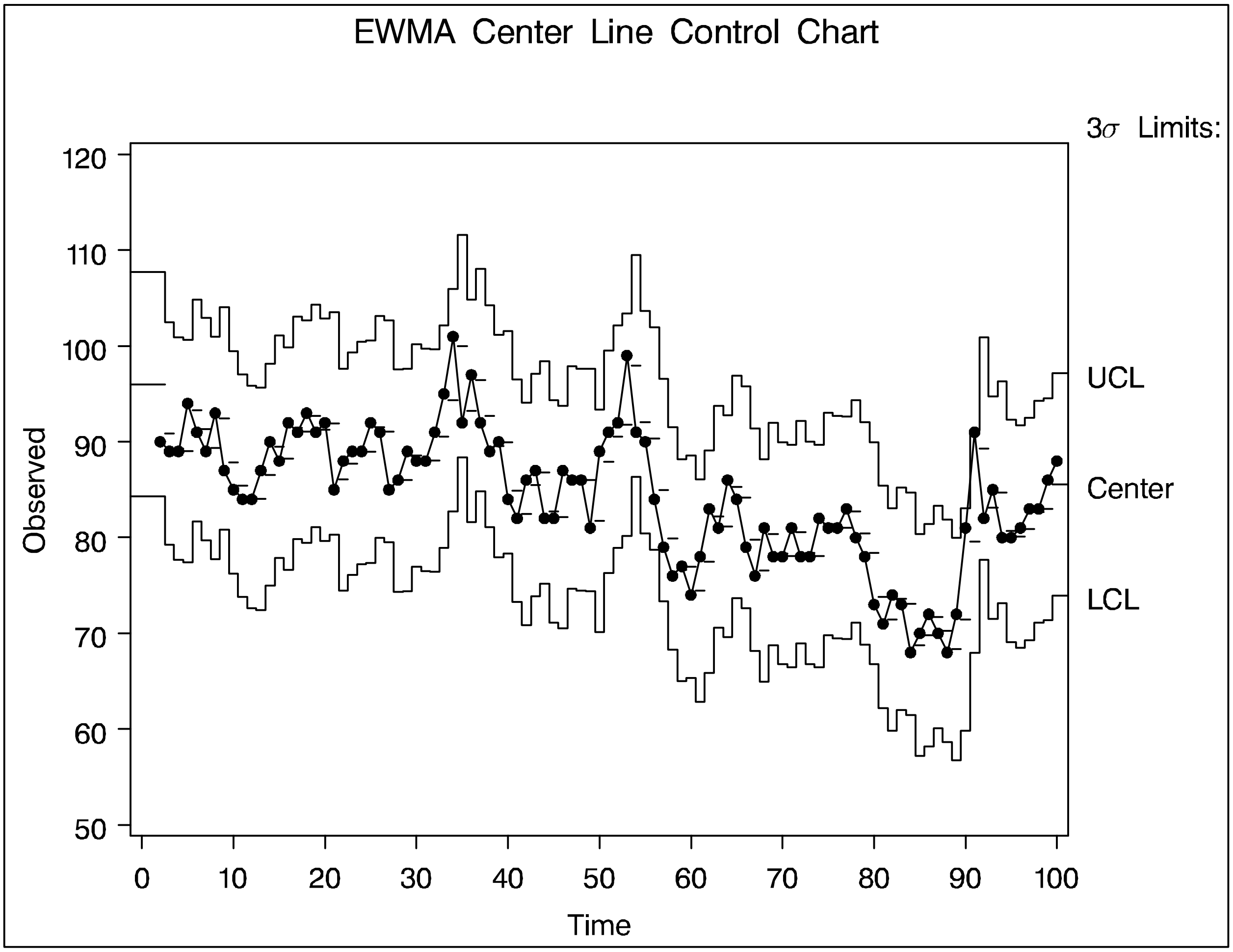


Figure 6.8. EWMA Moving Center Line Chart

Again, the conclusion is that the process is in control. Although Figure 6.6 and Figure 6.8 are not the only displays that might be constructed, they illustrate the tandem of the ARIMA and SHEWHART procedures in applications involving autocorrelated data.

Multiple Components of Variation

This section considers another form of departure from the Shewhart model. Here, measurements are *independent* from one subgroup sample to the next, but there are multiple components of variation for each measurement.

A company that manufactures polyethylene film monitors statistical control of an extrusion process that produces a continuous sheet of film. At periodic intervals of time, samples are taken at four locations (referred to as lanes) along a cross section of the sheet, and a test measurement is made of each sample. The test values are saved in a SAS data set named FILM.

SAMPLE	LANE	TESTVAL
1	A	93
1	B	87
1	C	92
1	D	78
2	A	87
2	B	83
2	C	79
2	D	77
.	.	.
.	.	.
.	.	.

Figure 6.9. Data Set FILM

After removal of outliers, the following statements create box plots for the data:

```
proc sort data=film; by lane;

title 'Variation Within Lane';
proc shewhart data=film graphics;
   boxchart testval*lane /
      boxstyle = schematicid
      hoffset  = 5
      idsymbol = dot
      cboxfill = red
      stddevs
      nolimits
      nolegend;
   id sample;
run;
```

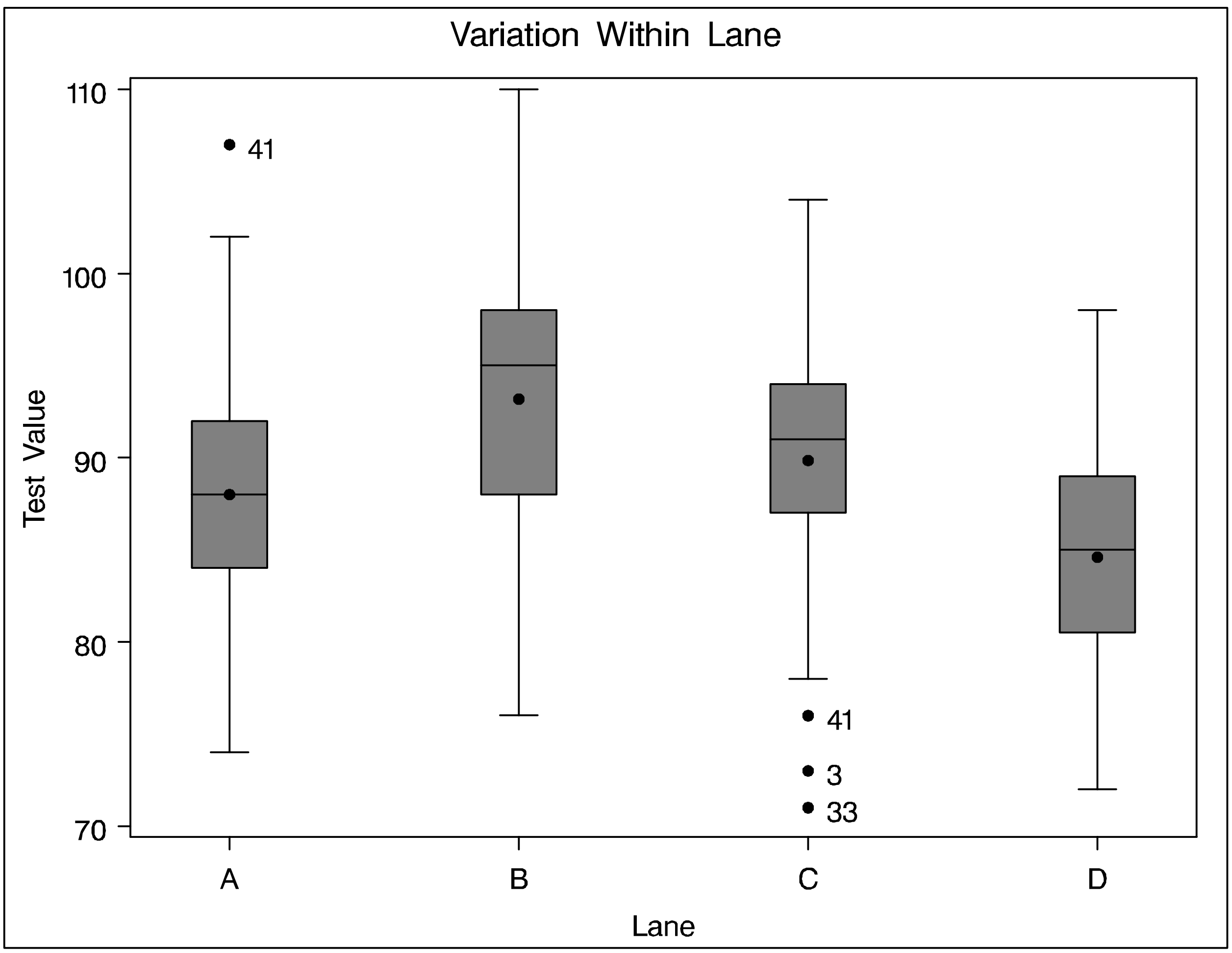


Figure 6.10. Boxchart of Test Values in FILM

It is tempting to create conventional $\bar{X}$ and R charts for the test values grouped by the variable SAMPLE.

```
proc sort data=film; by sample;

title 'Shewhart Chart for Means and Ranges';
proc shewhart data=film graphics;
   xrchart testval*sample /
      split    = '/'
      npanel   = 60
      limitn   = 4
      nolegend
      alln;
label testval = 'Avg Test Value/Range';
run;
```

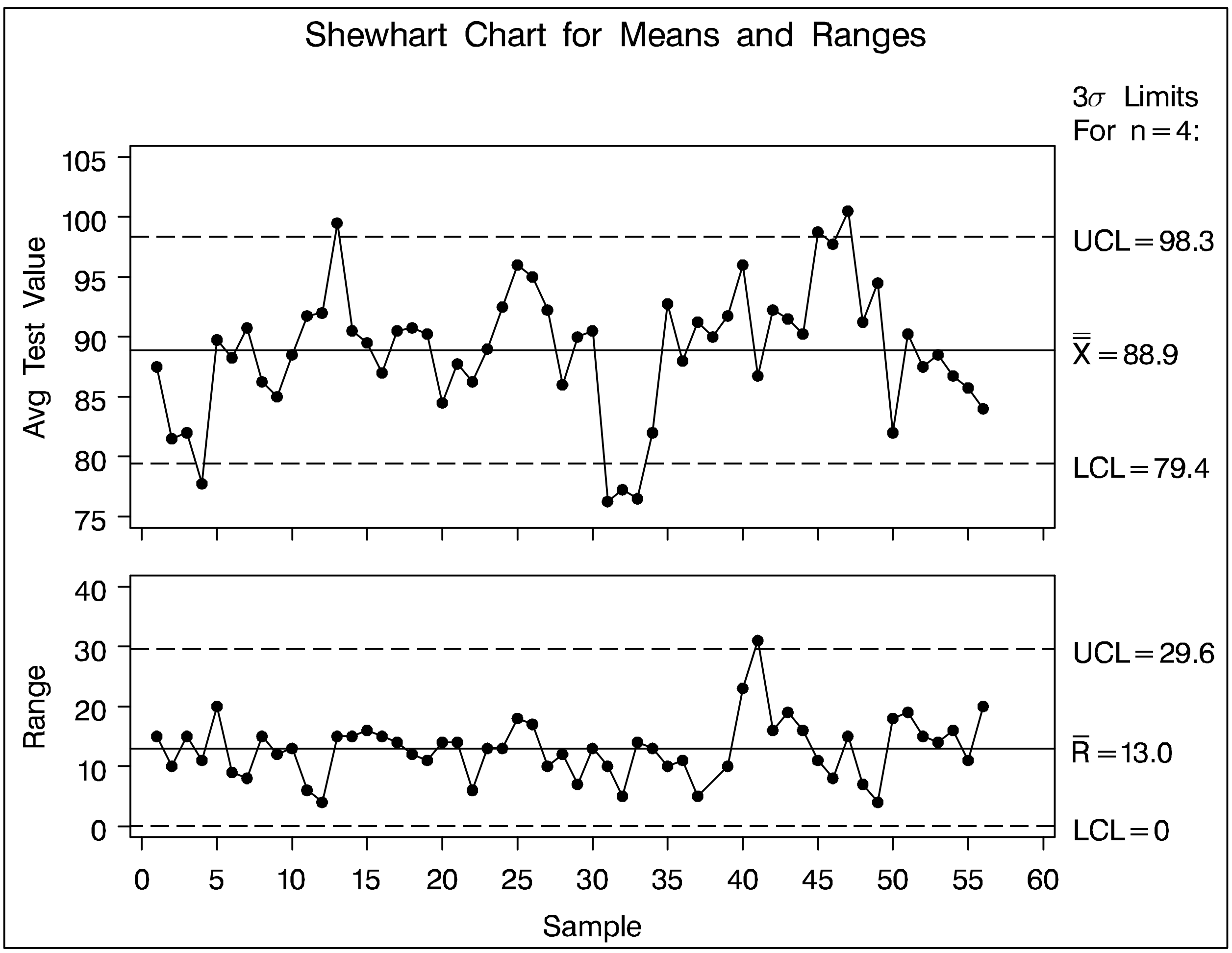


Figure 6.11. Conventional $\bar{X}$ and R Charts

The number of out-of-control points in Figure 6.11 suggests that the process is not in control. However, since the process is known to be stable and the data have been screened for outliers, it is suspected that the basic Shewhart model is not adequate.

This model assumes that sampling variation, also referred to as *within-group* variation, is the only source of variation. Writing x_{ij} for the *j*th measurement within the *i*th subgroup, the usual model for the $\bar{X}$ and R charts is

$$x_{ij} = \mu + \sigma_W \epsilon_{ij}$$

for $i = 1, 2, \ldots k$ and $j = 1, 2, \ldots n$. Here, there are $n = 4$ measurements in a "subgroup" and there are $k = 56$ subgroup samples. The ϵ_{ij} are assumed to be independent with zero mean and unit variance, and σ_W^2 is the within-subgroup variance.

For the film manufacturing process, this model is not adequate because there is additional variation due to changes in temperature, pressure, raw material, and other factors. A more useful model is

$$x_{ij} = \mu + \sigma_B \omega_i + \sigma_W \epsilon_{ij}$$

where σ_B^2 is the between-subgroup variance, the ω_i are independent with zero mean and unit variance, and the ω_i

are independent of the ϵ_{ij}. See Chapter 3 of Wetherill and Brown (1991).

To plot the averages

$$\overline{x}_{i.} \equiv \sum_{j=1}^{n} x_{ij}/n$$

on a control chart, we need estimates of

$$\begin{aligned} E(\overline{x}_{i.}) &= \mu \\ \text{Var}(\overline{x}_{i.}) &= \sigma_B^2 + (\sigma_W^2/n) \end{aligned}$$

The center line should be located at $\hat{\mu}$, and 3σ limits should be located at

$$\hat{\mu} \pm 3\sqrt{\widehat{\sigma_B^2} + \widehat{\sigma_W^2}/n}$$

where $\widehat{\sigma_B^2}$ and $\widehat{\sigma_W^2}$ estimate the variance components.

You can estimate the variance components with the MIXED procedure (among other procedures). See *SAS/STAT® Software: Changes and Enhancements through Release 6.12* for details on the MIXED procedure.

```
proc mixed data=film;
   class sample;
   model testval = / s;
   random sample;
   make 'solutionf' out=sf;
   make 'covparms'  out=cp;
run;
```

The results are shown in Figure 6.12. The estimates are $\widehat{\sigma_B^2} = 19.25$, $\widehat{\sigma_W^2} = 39.68$, and $\hat{\mu} = 88.90$.

```
           Covariance Parameter Estimates (REML)

Cov Parm           Ratio      Estimate     Std Error       Z

SAMPLE        0.48516517   19.25257992    5.67829660    3.39
Residual      1.00000000   39.68252702    4.35915223    9.10


           Covariance Parameter Estimates (REML)

           Pr > |Z|

             0.0007
             0.0001


                  Solution for Fixed Effects

Parameter      Estimate     Std Error    DDF        T  Pr > |T|

INTERCEPT   88.89629708    0.72504535     55   122.61    0.0001
```

Figure 6.12. Partial Output from the MIXED Procedure

The following statements merge the MIXED output data sets into a data set named NEWLIM that has the structure of a LIMITS= input data set for the SHEWHART procedure.

```
   data sf; set sf; rename _est_=est;
   data cp; set cp sf; keep est;
   proc transpose data=cp out=newlim;

   data newlim;
      set newlim;
      drop _name_ _label_ col1-col3;
      length _var_ _subgrp_ _type_ $8;
      _var_    = 'testval';
      _subgrp_ = 'sample';
      _type_   = 'estimate';
      _limitn_ = 4;
      _mean_   = col3;
      _stddev_ = sqrt(4*col1 + col2);
      output;
   run;
```

LIMITN is assigned the value n, _MEAN_ is assigned the value $\hat{\mu}$, and _STDDEV_ is assigned the value

$$\hat{\sigma}_{\text{adj}} \equiv \sqrt{4\widehat{\sigma^2_B} + \widehat{\sigma^2_W}}$$

In the next statements the SHEWHART procedure reads these estimates and displays the $\bar{X}$ and R charts shown in Figure 6.13. The control limits for the $\bar{X}$ chart are displayed at $\hat{\mu} \pm 3\hat{\sigma}_{\text{adj}}/\sqrt{n}$.

```
title 'Control Chart With Adjusted Limits';
symbol v=dot;

proc shewhart data=film limits=newlim graphics;
   xrchart testval*sample / npanel = 60 ;
run;
```

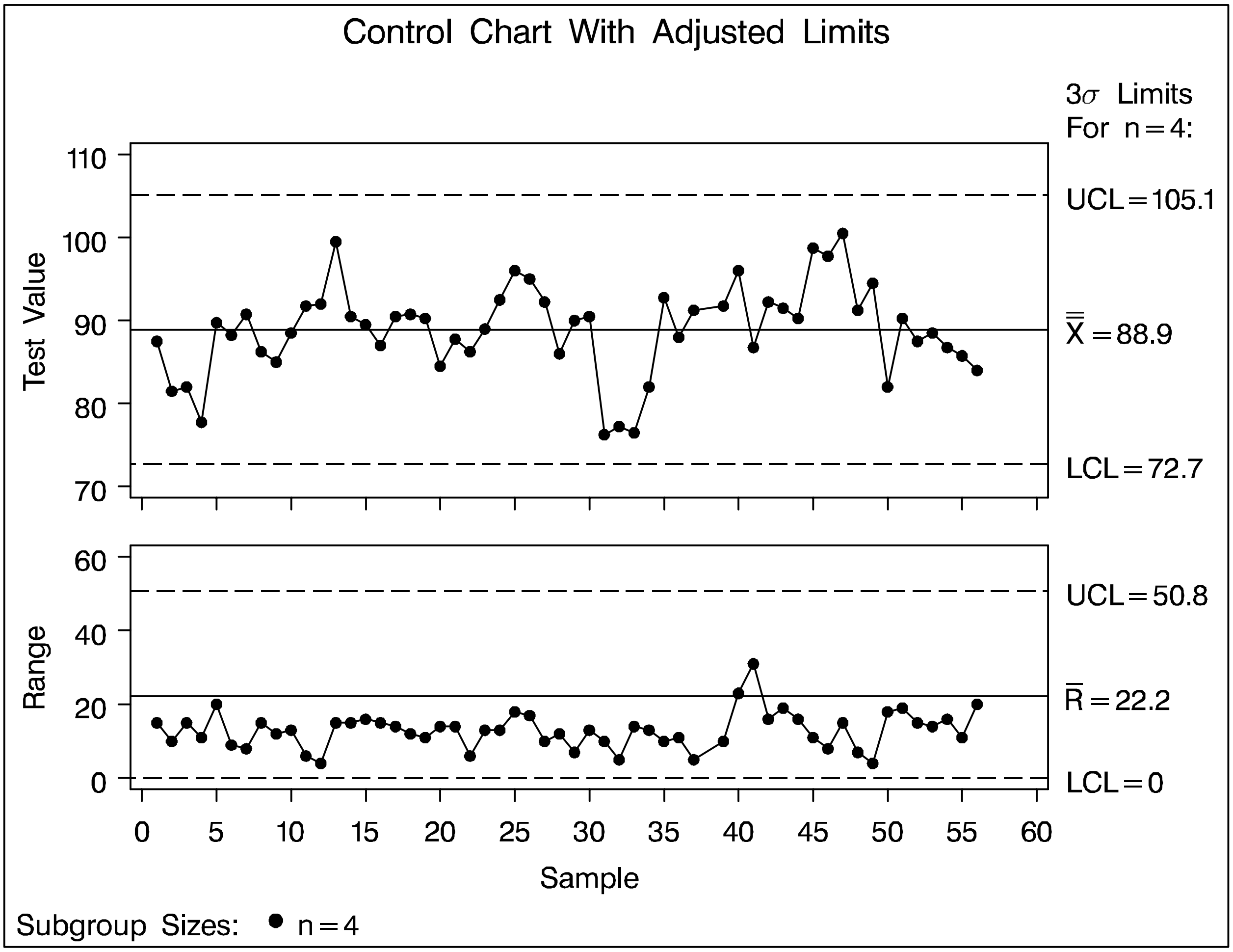


Figure 6.13. Derived Control Limits

Figure 6.13 correctly indicates that the variation in the process is due to common causes.

A simple alternative to Figure 6.13 is a conventional "individual measurements" chart for the subgroup *means*, as follows:

```
proc means noprint data=film;
   by sample;
   output out=filmx mean=lanemean;

title 'Individuals Chart for Means';
symbol v=dot;

proc shewhart data=filmx graphics;
   irchart lanemean*sample /
      split  = '/'
      npanel = 60
      nochart2;
label lanemean = 'Avg Value/Mvg Rng';
run;
```

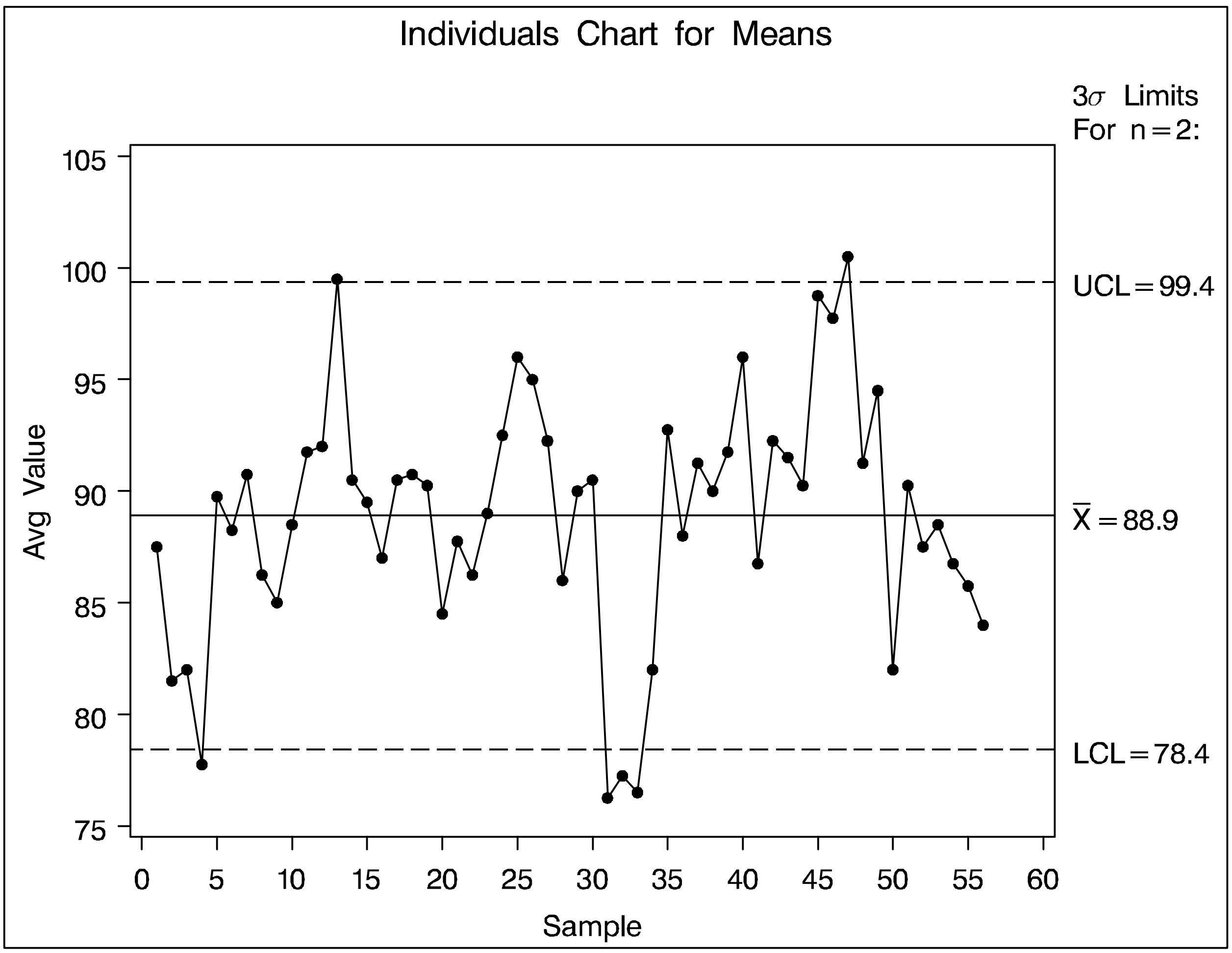


Figure 6.14. Individuals Chart for Averages

Note that

- Figure 6.14 is only roughly equivalent to Figure 6.13 because the standard error for the means in Figure 6.14 is determined by averaging the moving ranges of the means and dividing by d_2.
- If you replace this standard error with a value determined from the sample standard deviation of the means then the control limits will be identical.
- Assuming the process is in control, the variance components approach is preferable because it yields separate estimates of the components due to lane and sample, as well as a number of hypothesis tests (these require assumptions of normality).
- You can use the variance components approach to extend the model to a variety of effects.

The following statements fit a mixed model in which LANE is a fixed effect and SAMPLE is a random effect.

```
proc mixed data=film;
   class sample lane;
   model testval = lane;
   random sample;
   lsmeans lane;
run;
```

Cov Parm	Ratio	Estimate	Std Error	Z	Pr > |Z|
SAMPLE	0.82943785	22.52245916	5.66313893	3.98	0.0001
Residual	1.00000000	27.15388406	3.01371956	9.01	0.0001

Covariance Parameter Estimates (REML)

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
LANE	3	162	26.37	0.0001

Least Squares Means

Level	LSMEAN	Std Error	DDF	T	Pr > |T|
LANE A	88.04445518	0.94862578	162	92.81	0.0001
LANE B	93.22627337	0.94862578	162	98.28	0.0001
LANE C	89.89900064	0.94862578	162	94.77	0.0001
LANE D	84.60714286	0.94184795	162	89.83	0.0001

Figure 6.15. Partial Output from the MIXED Procedure

Since the lane effect is significant, it might be wise to maintain separate control charts for each lane.

Short Run Process Control

Short run process control also requires special consideration of process variability.

When conventional Shewhart charts are used to establish statistical control, the initial control limits are typically based on between 25 and 30 subgroup samples. However, this amount of data is often not available in manufacturing situations where product changeover occurs frequently.

A variety of methods have been introduced for analyzing data from a process that alternates between short runs of multiple products. The two most common methods are:

- *difference from nominal*. A product-specific nominal value is subtracted from each measured value, and the differences (together with appropriate control limits) are charted. It is assumed that the nominal value represents the central location of the process (ideally estimated with historical data) and that the process variability is *constant* across products.
- *standardization*. Each measured value is standardized with a product-specific nominal and standard deviation values. This approach is followed when the process variability is *not constant* across products.

Note: Quesenberry (1991a, 1991b) introduced the *Q chart* for short (or long) production runs, which standardizes and normalizes the data using probability integral transformations. SAS examples are provided by Rodriguez and Bynum (1992).

Difference From Nominal

A metal extrusion process is used to make three slightly different models of the same component. Three product types (M1, M2, and M3) are produced in small quantities because the process is expensive and time-consuming.

Diameter measurements and nominal values for several short runs are saved in a data set named OLD shown in Figure 6.16. Samples 1 to 30 are to be used to estimate the common σ for the differences from nominal.

Note: See Wheeler (1991a) for a similar example.

SAMPLE	PRODTYPE	NOMINAL	DIAMETER	DIFF
1	M3	14.8	13.99	-0.81
2	M3	14.8	14.69	-0.11
3	M3	14.8	13.86	-0.94
4	M3	14.8	14.32	-0.48
5	M3	14.8	13.23	-1.57
6	M1	15.0	17.55	2.55
7	M1	15.0	14.26	-0.74
8	M1	15.0	14.62	-0.38
9	M1	15.0	12.97	-2.03
10	M2	15.5	16.18	0.68
11	M2	15.5	15.29	-0.21
12	M2	15.5	16.20	0.70
13	M3	14.8	13.89	-0.91
14	M3	14.8	12.71	-2.09
15	M3	14.8	14.32	-0.48
16	M3	14.8	15.35	0.55
17	M2	15.5	15.08	-0.42
18	M2	15.5	14.72	-0.78
19	M2	15.5	14.79	-0.71
20	M2	15.5	15.27	-0.23
21	M2	15.5	15.95	0.45
22	M1	15.0	14.78	-0.22
23	M1	15.0	15.19	0.19
24	M1	15.0	15.41	0.41
25	M1	15.0	16.26	1.26
26	M3	14.8	16.68	1.88
27	M3	14.8	15.60	0.80
28	M3	14.8	14.86	0.06
29	M3	14.8	16.67	1.87
30	M3	14.8	14.35	-0.45

Figure 6.16. Data Set OLD

For now, assume that the variability in the process is constant across product types. First, compute $\hat{\sigma} = \bar{R}/d_2$ for each product type.

```
proc sort data=old; by prodtype;

proc shewhart data=old;
   irchart diff*sample /
      nochart
      outlimits=baselim;
   by prodtype;
run;
```

```
                 Control Limits By Product Type

PRODTYPE _VAR_ _SUBGRP_  _TYPE_  _LIMITN_  _ALPHA_ _SIGMAS_

   M1    DIFF   SAMPLE  ESTIMATE     2    .0026998     3
   M2    DIFF   SAMPLE  ESTIMATE     2    .0026998     3
   M3    DIFF   SAMPLE  ESTIMATE     2    .0026998     3

 _LCLI_    _MEAN_    _UCLI_ _LCLR_   _R_     _UCLR_  _STDDEV_

-3.13153   0.12924  3.39001    0   1.22646  4.00628   1.08692
-1.78485  -0.06551  1.65382    0   0.64669  2.11243   0.57311
-3.22326  -0.19034  2.84258    0   1.14076  3.72633   1.01097
```

Figure 6.17. Data Set BASELIM

The following statements compute an overall estimate $\hat{\sigma}$, which is saved in a LIMITS= data set named DIFFLIM.

```
proc means data=baselim noprint;
   var _r_;
   output out=difflim (keep=_r_) mean=_r_;

data difflim;
   set difflim;
   drop _r_;
   length _var_ _subgrp_ $ 8;
   _var_    = 'diff';
   _subgrp_ = 'sample';
   _mean_   = 0.0;
   _stddev_ = _r_ / d2(2);
   _limitn_ = 2;
   _sigmas_ = 3;
run;
```

Why is _MEAN_ set to zero?

Control Limit Parameters For Differences

VAR	_SUBGRP_	_MEAN_	_STDDEV_	_LIMITN_	_SIGMAS_
diff	sample	0	0.89034	2	3

Figure 6.18. Data Set DIFFLIM

Now, you can construct short run control charts for new data. Suppose that diameters for an additional 30 parts (numbered 31 to 60) are measured and saved in a SAS data set named NEW.

SAMPLE	PRODTYPE	DIFF	NOMINAL	DIAMETER
31	M2	-0.81	15.5	14.69
32	M2	-0.11	15.5	15.39
33	M2	-0.94	15.5	14.56
34	M2	-0.48	15.5	15.02
35	M2	-1.57	15.5	13.93
36	M1	2.55	15.0	17.55
...	...	...	...	...
56	M2	1.88	15.5	17.38
57	M3	0.80	14.8	15.60
58	M3	0.06	14.8	14.86
59	M3	1.87	14.8	16.67
60	M3	-0.45	14.8	14.35

Figure 6.19. Data Set NEW

The following statements construct a short run control chart for the measurements in NEW:

```
title 'Chart for Difference from Nominal';
symbol v=dot;

proc shewhart data=new (rename=(prodtype=_phase_))
   limits=difflim graphics;
   irchart diff * sample /
      readphases = all
      phaseref
      phasebreak
      phaselegend
      split='/';
   label diff = 'Difference/Mvg Rng';
run;
```

Why is PRODTYPE temporarily renamed to _PHASE_?

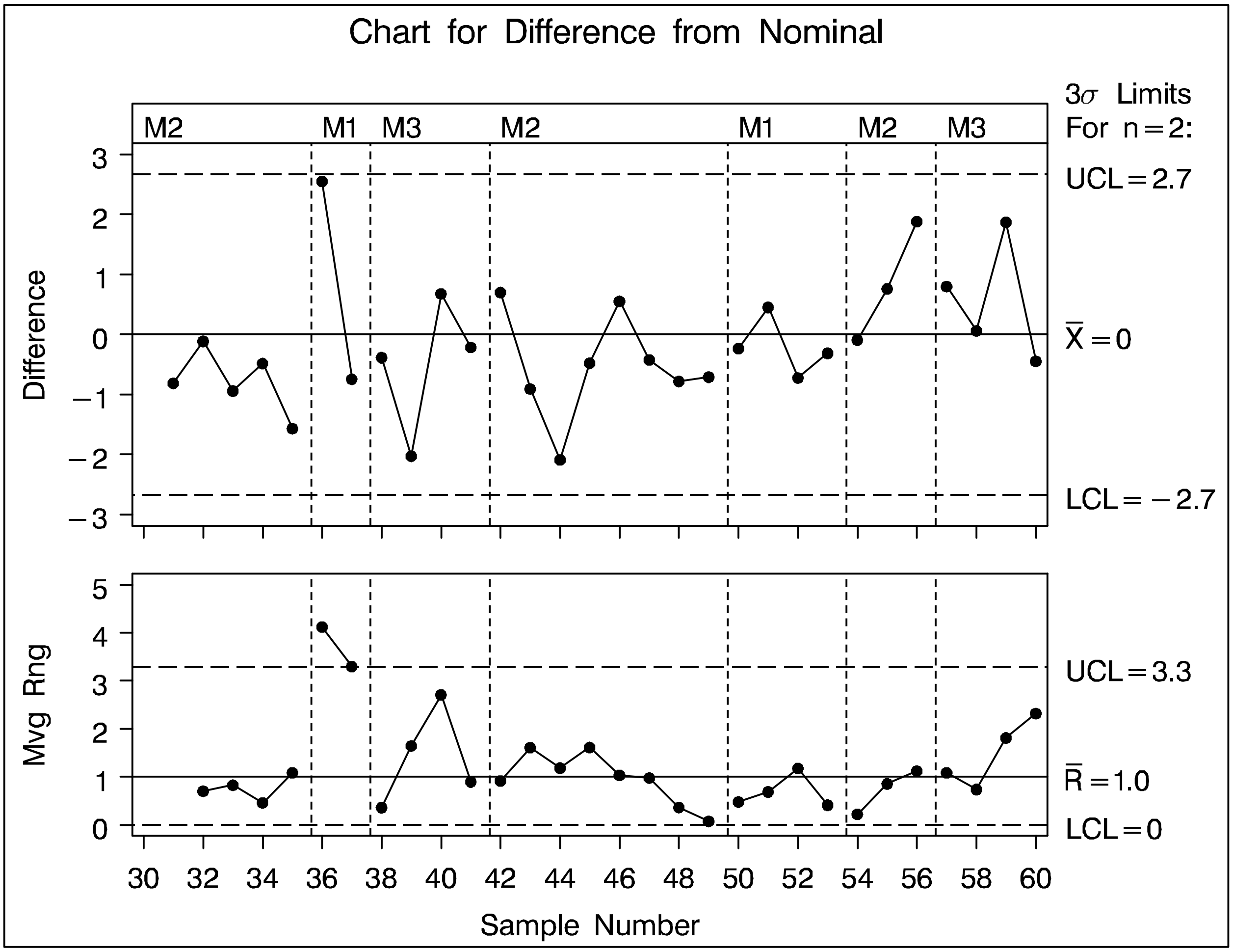


Figure 6.20. Short Run Control Chart

It may be useful to replace the moving range chart with a plot of the nominal values.

```
data difflim;
   set difflim;
   _var_    = 'diameter';
   _limitn_ = 1;
run;

title 'Differences and Nominal Values';

proc shewhart data=new limits=difflim graphics;
   xchart diameter * sample ( prodtype ) /
      readlimits
      nolimitslegend
      nolegend
      split = '/'
      blockpos = 3
      blocklabtype  = scaled
      blocklabelpos = left
      xsymbol       = xbar
      trendvar     = nominal;
label diameter = 'Difference/Nominal'
      prodtype = 'Product';
run;
```

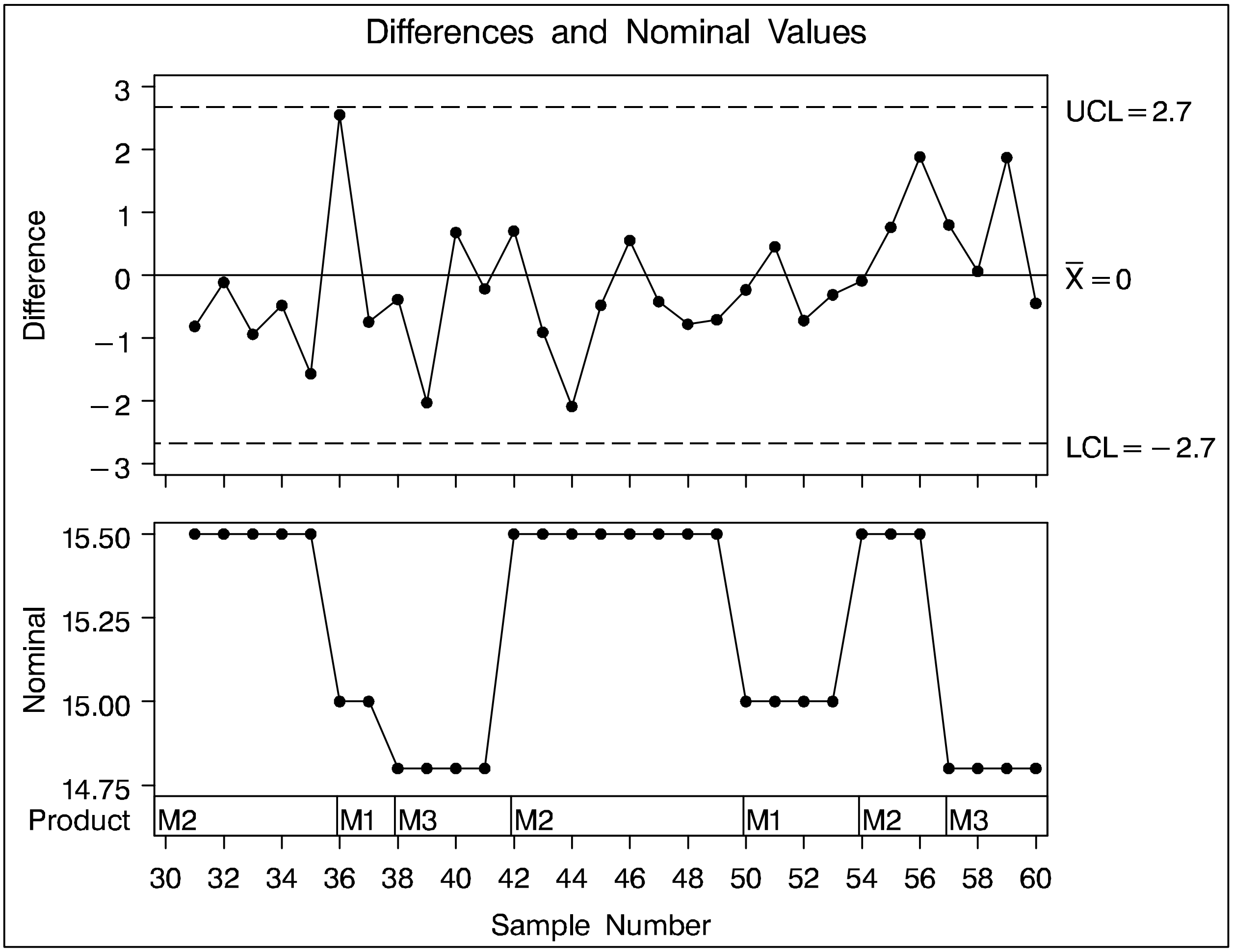


Figure 6.21. Short Run Control Chart with Nominal Values

Are the Variances Constant?

The difference-from-nominal chart should be accompanied by a test that checks whether the variances for each product type are statistically indistinguishable (homogeneous).

```
proc glm data=old;
   class prodtype;
   model diameter = prodtype;
   means prodtype / hovtest = levene
                    hovtest = obrien;
run;
```

```
              General Linear Models Procedure

      Levene's Test for Equality of DIAMETER Variance
        ANOVA of Squared Deviations from Group Means

                           Sum of       Mean
Source            DF      Squares     Square      F Value   Pr > F

PRODTYPE           2       7.4490     3.7245       1.4129   0.2609
Error             27      71.1738     2.6361


      O'Brien's Test for Equality of DIAMETER Variance
        ANOVA of O'Brien's Spread Variable, W = 0.5

                           Sum of       Mean
Source            DF      Squares     Square      F Value   Pr > F

PRODTYPE           2       9.4061     4.7030       1.2715   0.2967
Error             27      99.8666     3.6988
```

Figure 6.22. Tests for Homogeneity of Variance

The large P-values support the hypothesis of homogeneity.

Standardization

When the variances are *not* constant, various authors recommend standardizing the differences from nominal and charting them with control limits at ± 3.

You can use the product-specific estimates of σ in BASELIM to standardize the differences from nominal in NEW:

```
proc sort data=baselim; by prodtype;
proc sort data=new;     by prodtype;

data new;
   keep sample prodtype z diameter
        diff nominal _stddev_;
   label sample = 'Sample Number'
   format diff 5.2 ;
   merge baselim new(in = a);
      by prodtype; if a;
   z = diff / _stddev_ ;

proc sort data=new; by sample;
run;
```

The following statements create a standardized chart.

```
title 'Standardized Chart';
symbol v=dot;
proc shewhart data=new graphics;
   irchart z*sample (prodtype) /
      blocklabtype  = scaled
      mu0           = 0
      sigma0        = 1
      split         = '/';
   label prodtype = 'Product Classification'
         z        = 'Std Difference/Mvg Rng';
run;
```

Why are MU0=0 and SIGMA0=1 specified?

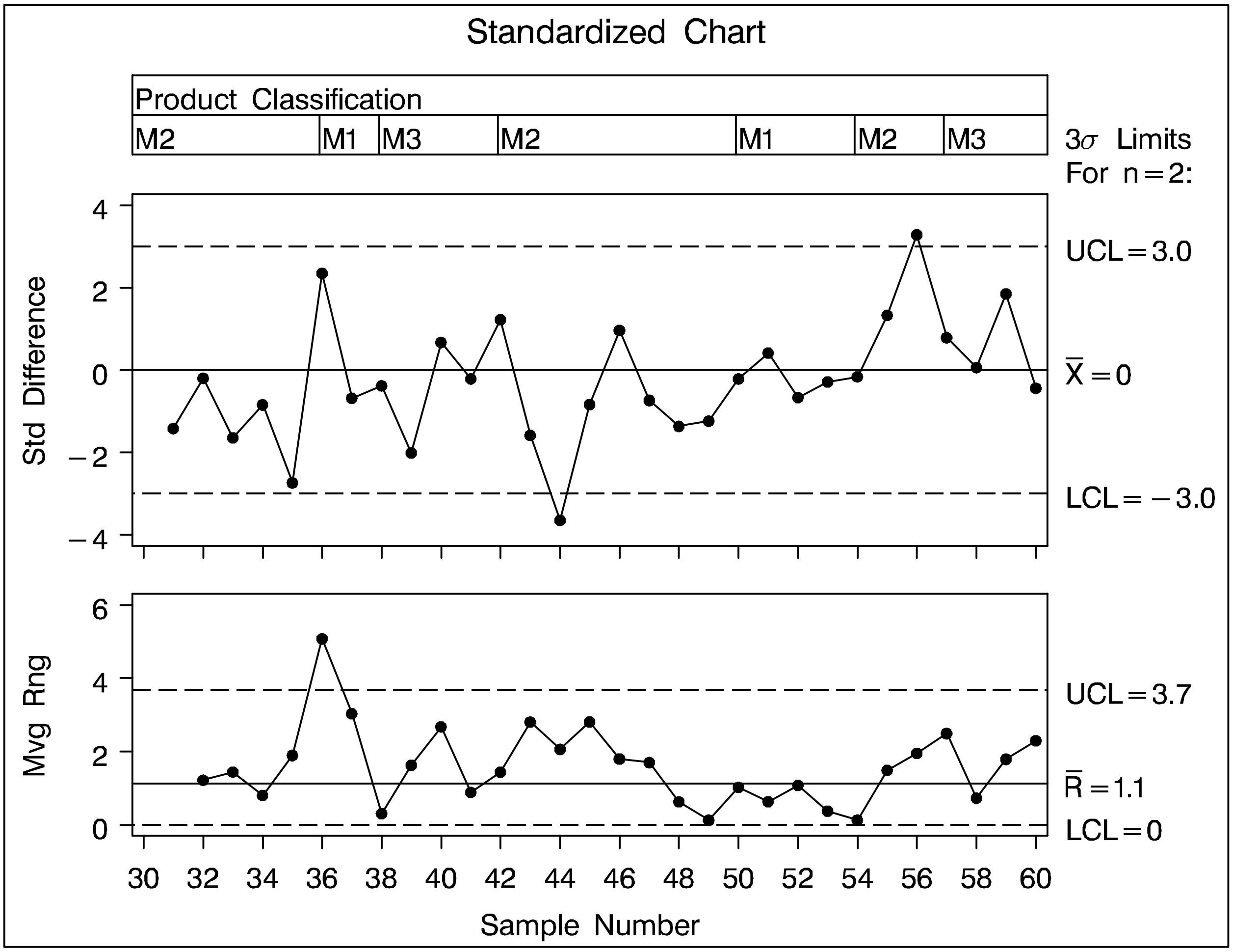


Figure 6.23. Standardized Difference Chart

Non-Normal Process Data

It is well known that Shewhart charts for means work well whether the measurements are normally distributed or not; see Schilling and Nelson (1976) and Wheeler (1995). On the other hand, the interpretation of charts for individual measurements (X charts) can be affected by departures from normality, which are common in SPC applications.

If subgrouping is not possible and nonnormality is evident, alternatives include

- transforming the data to normality
- modifying the usual limits based on a suitable distribution model for the data

The time taken by staff members to answer a phone was measured, and the delays (in minutes) were saved as values of a variable named TIME in a SAS data set named CALLS.

RECNUM	TIME
1	3.2
2	3.1
3	3.1
4	2.9
5	2.8
.	.
.	.
.	.
49	2.6
50	2.9

Figure 6.24. Data Set CALLS

```
title 'Standard Analysis of Individual '
      'Delays';

proc shewhart data=calls graphics ;
   irchart time * recnum /
      rtmplot   = schematic
      outlimits = delaylim
      cboxfill  = grey
      nochart2;
run;
```

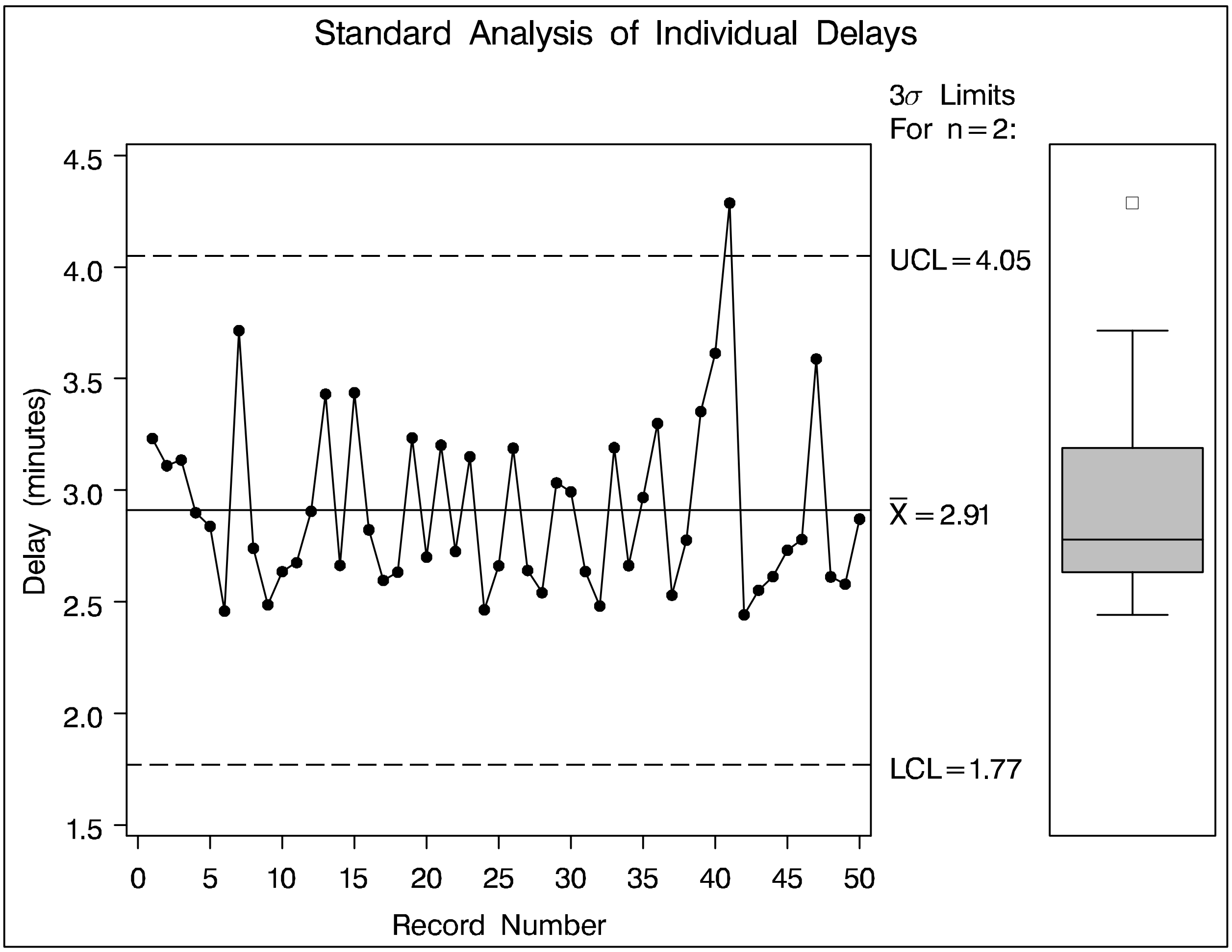


Figure 6.25. Standard Control Limits for Delays

Is the process in control?

VAR	_SUBGRP_	_TYPE_	_LIMITN_	_ALPHA_
TIME	RECNUM	ESTIMATE	2	.0026998

SIGMAS	_LCLI_	_MEAN_	_UCLI_	_STDDEV_
3	1.77001	2.91035	4.05070	0.38011

Figure 6.26. Data Set DELAYLIM

The control limits in DELAYLIM can be replaced with percentiles from a fitted lognormal distribution.

```
title 'Lognormal Fit for Delay Distribution';

proc capability data=calls graphics  ;
   histogram time /
      lognormal( threshold = 2.3 )
      cfill  = ligr
      outfit = lnfit
      nolegend ;
   inset n = 'Number of Calls'
         lognormal( sigma='Shape' (4.2)
                    zeta ='Scale' (5.2)
                    theta ) /
         pos = ne;
run;
```

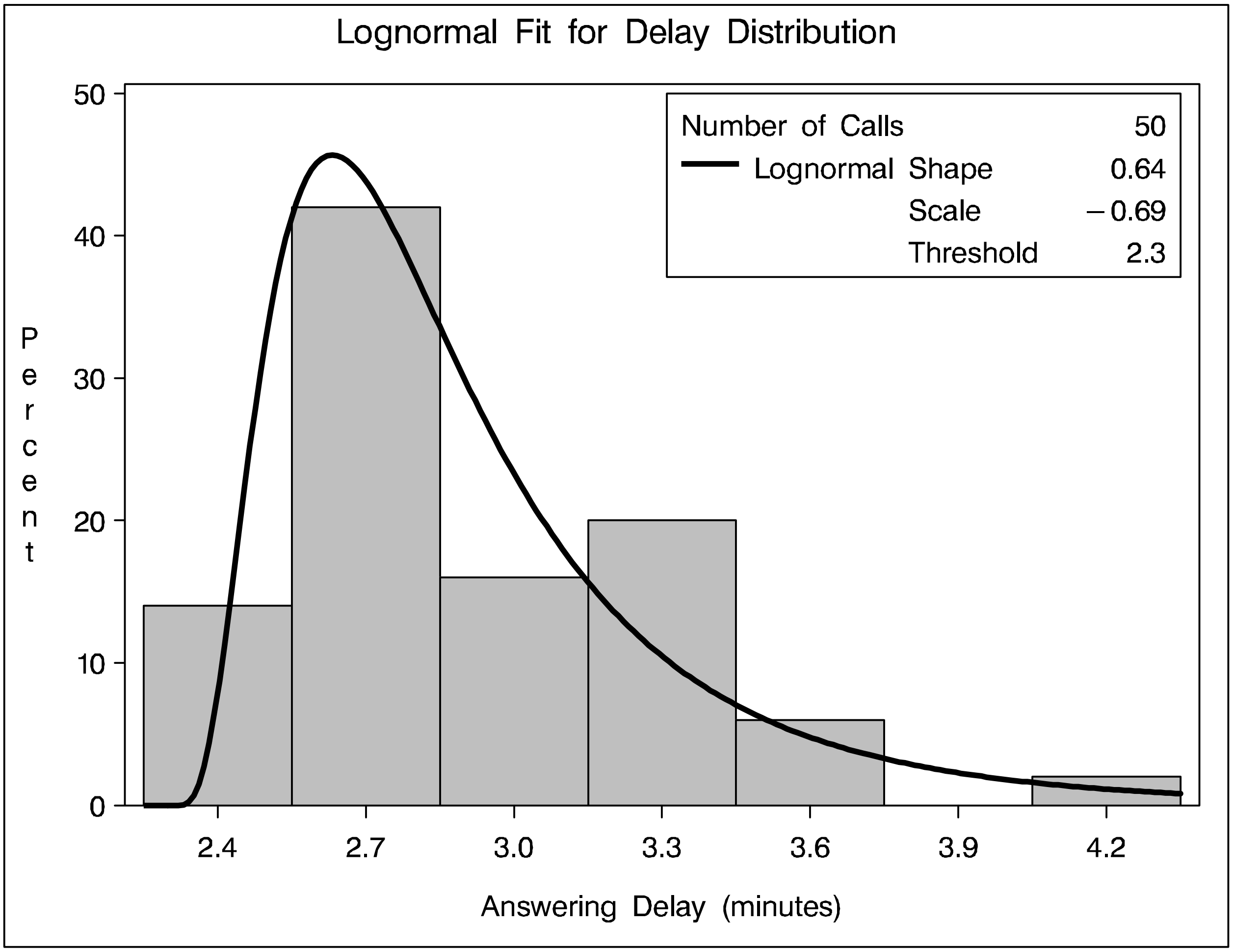



Figure 6.27. Distribution of Delays

Parameters	
Threshold (Theta)	2.3
Scale (Zeta)	-0.6891925
Shape (Sigma)	0.64121248
Est Mean	2.91655007
Est Std Dev	0.4396814
Chi-Square Goodness-of-Fit	
Chi-Square	6.81723374
Df	4
Pr > Chi-Square	0.1459

EDF Goodness-of-Fit	
A-D (A-Square)	0.34761167
Pr > A-Square	0.4765
C-von M (W-Square)	0.05863868
Pr > W-Square	0.4106
Kolmogorov (D)	0.09208575
Pr > D	>.15

Figure 6.28. Fit Summary for Lognormal Model

The large P-values for the goodness-of-fit tests are evidence that the lognormal model provides a good fit.

VAR	_CURVE_	_LOCATN_	_SCALE_	_SHAPE1_	_MIDPTN_
TIME	LNORMAL	2.3	-0.68919	0.64121	4.2
ADASQ	_ADP_	_CVMWSQ_	_CVMP_	_KSD_	_KSP_
0.34761	0.47648	0.058639	0.41060	0.092086	0.15

Figure 6.29. Data Set LNFIT

The αth lognormal percentile is $P_\alpha = \exp(\sigma\Phi^{-1}(\alpha) + \zeta)$

```
data delaylim;
   merge delaylim lnfit;
   drop _sigmas_ ;
   _lcli_ = _locatn_ +
      exp(_scale_+probit(0.5*_alpha_)*_shape1_);
   _ucli_ = _locatn_ +
      exp(_scale_+probit(1-0.5*_alpha_)*_shape1_);
   _mean_ = _locatn_ +
      exp(_scale_+0.5*_shape1_*_shape1_);
run;
```

```
title 'Lognormal Control Limits for Delays';
proc shewhart data=calls limits=delaylim graphics;
   irchart time * recnum /
      rtmplot  = schematic
      cboxfill = ligr
      nochart2;
run;
```

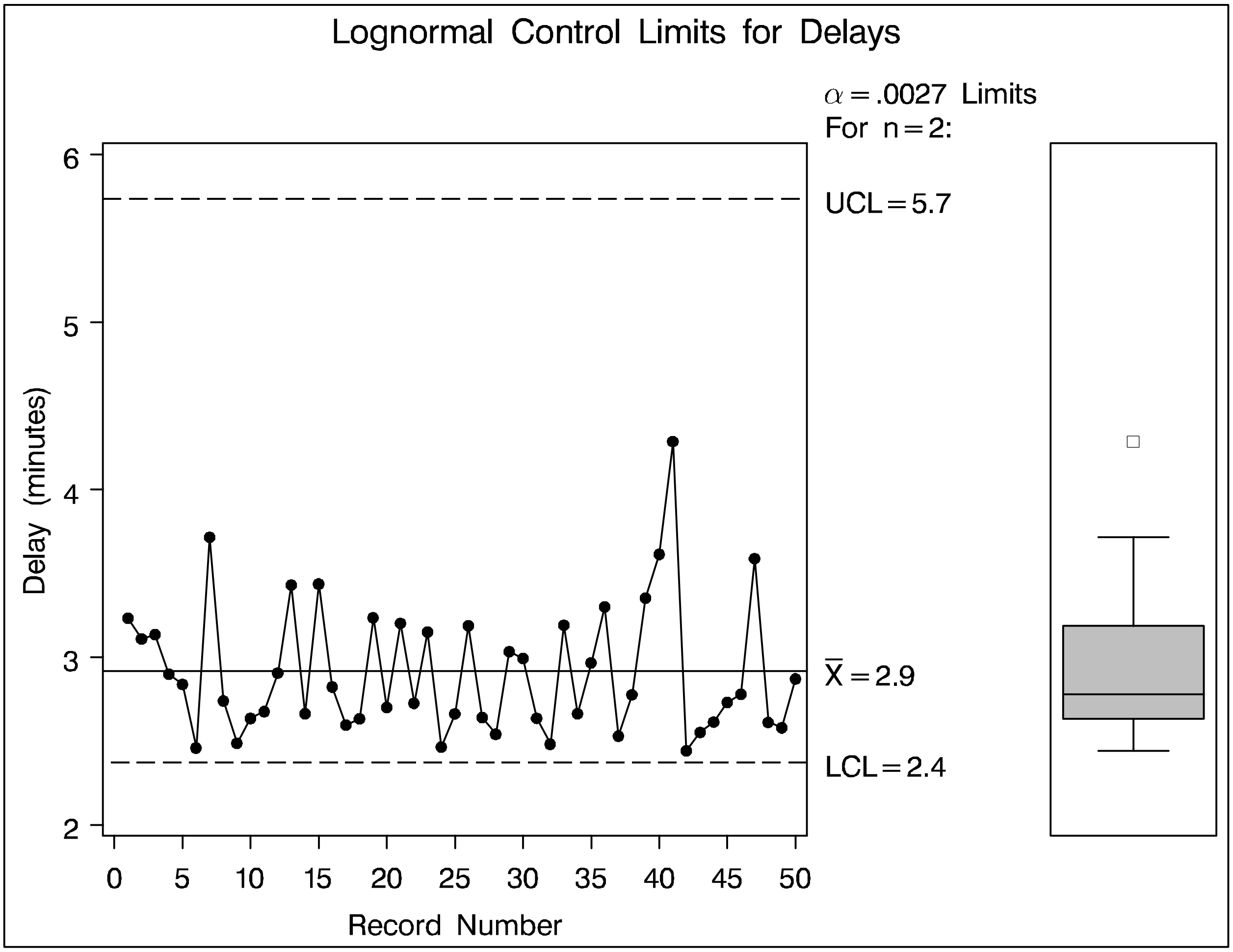


Figure 6.30. Adjusted Control Limits for Delays

Clearly the process is in control, and the control limits are appropriate for the data. The particular probability level $\alpha = 0.0027$ associated with these limits is somewhat immaterial. Other values of α could be specified with the ALPHA= option in the original IRCHART statement.

Chapter 7

Construction of Multivariate Control Charts

- Simultaneous measurement of p variables
- What if the variables are correlated?
- Diagnosis is more complex than for a univariate chart

Introduction

Denote the i^{th} measurement on the j^{th} variable as X_{ij} for $i = 1, 2, \ldots, n$, where n is the number of measurements, and $j = 1, 2, \ldots, p$.

$$\bar{X}_j = \tfrac{1}{n} \Sigma_{i=1}^{n} X_{ij} \;\;, \;\; \mathbf{x}_i = \begin{bmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{ip} \end{bmatrix}, \;\; \bar{\mathbf{x}}_n = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix}$$

$$\mathbf{S}_n = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}}_n)(\mathbf{x}_i - \bar{\mathbf{x}}_n)'$$

Standard practice is to construct a chart for a statistic T_i^2 of the form

$$T_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}}_n)' \mathbf{S}_n^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_n)$$

Example of Multivariate Control Chart

Example from a start-up phase (See Tracy, Young, and Mason (1992)) of a chemical process based on the three measured quality characteristics:

Percent of impurities (IMPURE)
Temperature (TEMP)
Concentration (CONC)

Data are saved in data set STARTUP.

```
data startup;
    input sample impure temp conc;
    label sample = 'Sample Number'
          impure = 'Impurities'
          temp   = 'Temperature'
          conc   = 'Concentration' ;
    cards;
  1  14.92  85.77  42.26
  2  16.90  83.77  43.44
  3  17.38  84.46  42.74
  4  16.90  86.27  43.60
  5  16.92  85.23  43.18
  6  16.71  83.81  43.72
  7  17.07  86.08  43.33
  8  16.93  85.85  43.41
  9  16.71  85.73  43.28
 10  16.88  86.27  42.59
 11  16.73  83.46  44.00
 12  17.07  85.81  42.78
 13  17.60  85.92  43.11
 14  16.90  84.23  43.48
 ;
```

Compute Principal Components

Obtain Sample Size:

```
proc means data=startup noprint ;
   var impure temp conc;
   output out=means n=n;

data startup;
   if _n_ = 1 then set means;
   set startup;
   p        = 3;
   _subn_   = 1;
   _limitn_ = 1;
   run;
```

Compute Principal Components:

```
proc princomp data=startup out=prin
   outstat=scores std cov;
   var impure temp conc;
run;
```

Compute T_i^2 and its Exact Control Limits

```
data prin (rename=(tsquare=_subx_));
   length _var_ $ 8 ;
   drop prin1 prin2 prin3 _type_ _freq_;
   set prin;
   comp1   = prin1*prin1;
   comp2   = prin2*prin2;
   comp3   = prin3*prin3;
   tsquare = comp1 + comp2 + comp3;
   _var_   = 'tsquare';
   _alpha_ = 0.05;
   _lclx_  = ((n-1)*(n-1)/n)*
             betainv(_alpha_/2, p/2, (n-p-1)/2);
   _mean_  = ((n-1)*(n-1)/n)*
             betainv(0.5, p/2, (n-p-1)/2);
   _uclx_  = ((n-1)*(n-1)/n)*
             betainv(1-_alpha_/2, p/2, (n-p-1)/2);
   label tsquare = 'T Squared'
         comp1   = 'Comp 1'
         comp2   = 'Comp 2'
         comp3   = 'Comp 3';
run;
```

Gnanadesikan and Kettenring (1972) use a result of Wilks (1962) that T_i^2 is exactly distributed as a multiple of a variable with a beta distribution. Specifically,

$$T_i^2 \sim \frac{(n-1)^2}{n} \, B\left(\frac{p}{2}, \frac{n-p-1}{2}\right)$$

T2 Chart For Chemical Example

VAR	N	SAMPLE	IMPURE	TEMP	CONC	P	_SUBN_	_LIMITN_	COMP1
tsquare	14	1	14.92	85.77	42.26	3	1	1	0.79603
tsquare	14	2	16.90	83.77	43.44	3	1	1	1.84804
tsquare	14	3	17.38	84.46	42.74	3	1	1	0.33397
tsquare	14	4	16.90	86.27	43.60	3	1	1	0.77286
tsquare	14	5	16.92	85.23	43.18	3	1	1	0.00147
tsquare	14	6	16.71	83.81	43.72	3	1	1	1.91534
tsquare	14	7	17.07	86.08	43.33	3	1	1	0.58596
tsquare	14	8	16.93	85.85	43.41	3	1	1	0.29543
tsquare	14	9	16.71	85.73	43.28	3	1	1	0.23166
tsquare	14	10	16.88	86.27	42.59	3	1	1	1.30518
tsquare	14	11	16.73	83.46	44.00	3	1	1	3.15791
tsquare	14	12	17.07	85.81	42.78	3	1	1	0.43819
tsquare	14	13	17.60	85.92	43.11	3	1	1	0.41494
tsquare	14	14	16.90	84.23	43.48	3	1	1	0.90302

COMP2	COMP3	_SUBX_	_ALPHA_	_LCLX_	_MEAN_	_UCLX_
10.1137	0.01606	10.9257	0.05	0.24604	2.44144	7.13966
0.0162	0.17681	2.0410	0.05	0.24604	2.44144	7.13966
0.1538	5.09491	5.5827	0.05	0.24604	2.44144	7.13966
0.3289	2.76215	3.8640	0.05	0.24604	2.44144	7.13966
0.0165	0.01919	0.0372	0.05	0.24604	2.44144	7.13966
0.0645	0.27362	2.2534	0.05	0.24604	2.44144	7.13966
0.4079	0.44146	1.4354	0.05	0.24604	2.44144	7.13966
0.1729	0.73939	1.2077	0.05	0.24604	2.44144	7.13966
0.0001	0.44483	0.6766	0.05	0.24604	2.44144	7.13966
0.0004	0.86364	2.1692	0.05	0.24604	2.44144	7.13966
0.0274	0.98639	4.1717	0.05	0.24604	2.44144	7.13966
0.0823	0.87976	1.4003	0.05	0.24604	2.44144	7.13966
1.6153	0.30167	2.3320	0.05	0.24604	2.44144	7.13966
0.0001	0.00010	0.9032	0.05	0.24604	2.44144	7.13966

Figure 7.1. Data Set PRIN

T_i^2 Chart

```
symbol value=dot;
title 'T' m=(+0,+0.5) '2'
          m=(+0,-0.5) ' Chart For Chemical Example';
proc shewhart table=prin graphics;
   xchart tsquare*sample /
      xsymbol = mu
      nolegend ;
run;
```

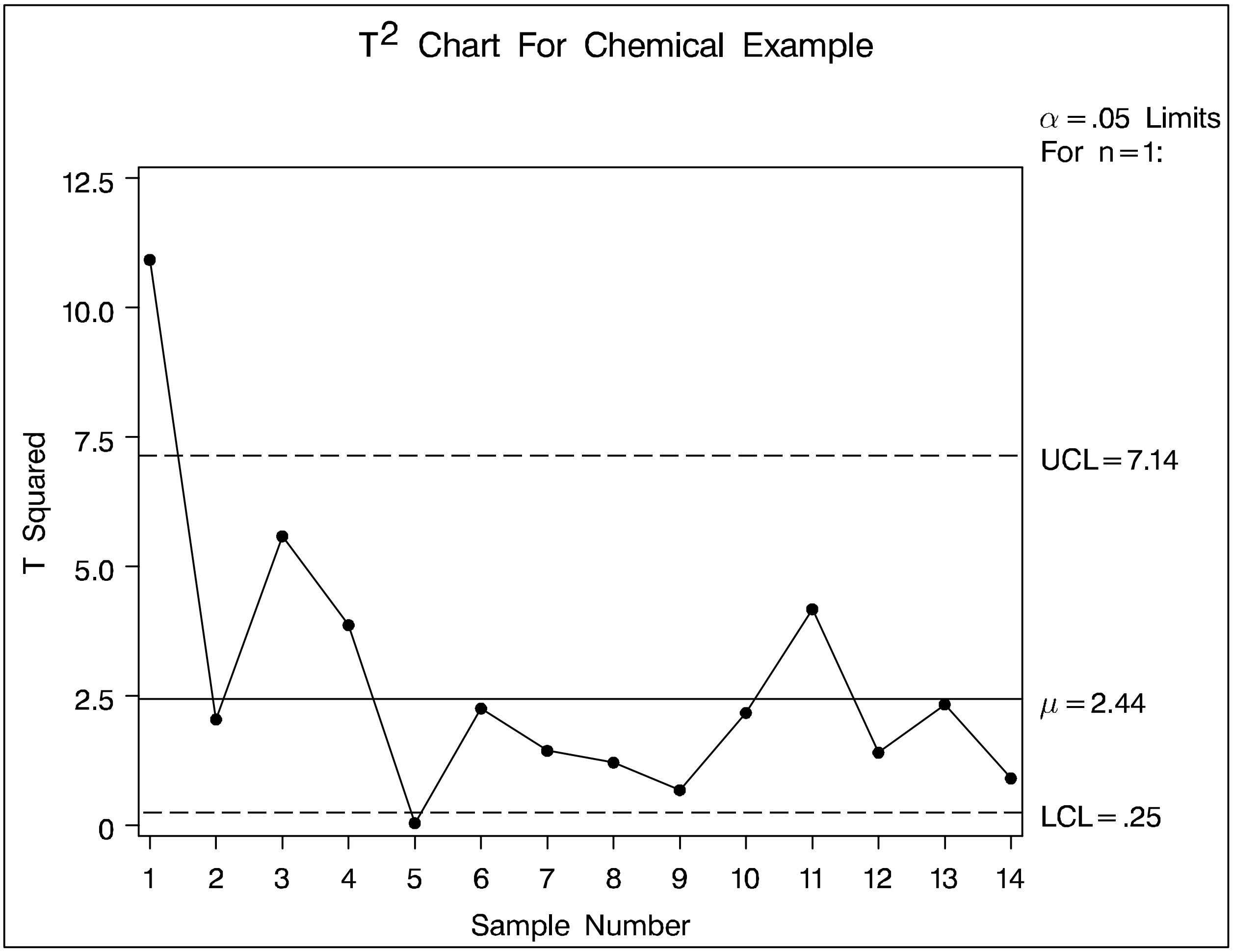
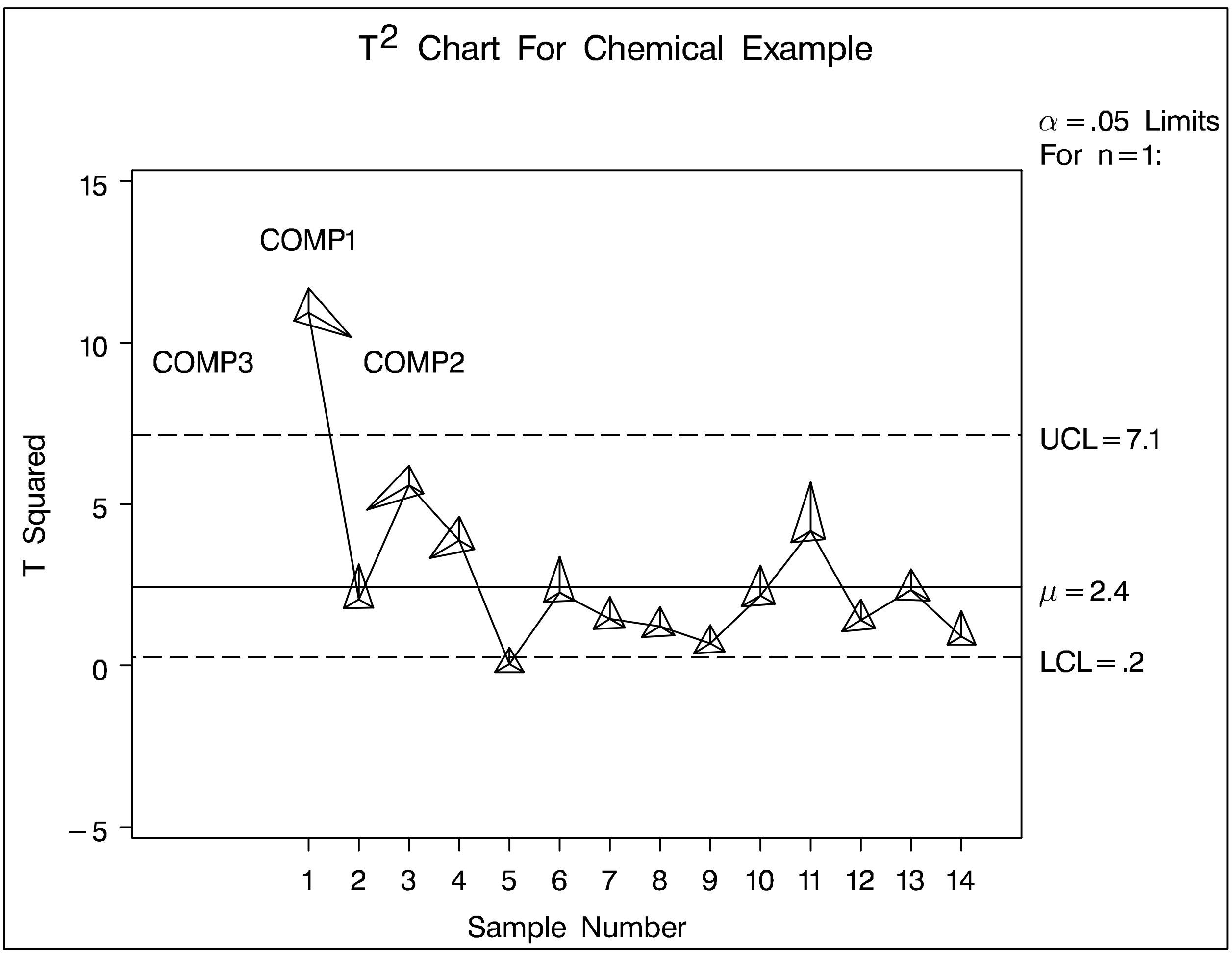


Figure 7.2. Multivariate Control Chart for Chemical Process

Principal Component Contributions

```
symbol value=none;
proc shewhart table=prin graphics;
   xchart tsquare*sample /
      starvertices  = (comp1 comp2 comp3)
      startype      = wedge
      cstars        = black
      starlegend    = none
      starlabel     = first
      staroutradius = 4
      npanelpos     = 14
      xsymbol       = mu
      nolegend ;
run;
```



Diagnosing a Multivariate Chart

- T^2 chart cannot differentiate between mean and variance shift

- T^2 chart signal not immediately interpretable

- Recent papers address these problems and present alternative approaches:
  - -- Mason, Tracy, and Young (1995, 1996)
  - -- Wade and Woodall (1993)
  - -- Hawkins (1991, 1993)
  - -- Doganaksoy, Faltin, and Tucker (1991)
  - -- *etc.*.

Here we simply illustrate the method of Hawkins (1991).

Regression Adjustment of Variables

For shifts in some of the variables, Hawkins (1991) suggests control charts based on the vector of scaled residuals from the regression of each variable on all others to improve the diagnostic power of multivariate T^2 charts following a chart signal.

$$\mathbf{Z} = \mathbf{A}(\mathbf{X}-\mu_0)$$

where

Z is a p-component vector of residuals whose component Z_i (scalar) is a residual when X_i is regressed on all other components of **X** rescaled to unit variance

X is a p-component vector such that $\mathbf{X} \sim N(\mu, \Sigma)$

A is a transformation matrix = $[diag(\Sigma_0^{-1})]^{-1/2}\Sigma_0^{-1}$

$\mu = \mu_{\mathbf{0}}$ and $\Sigma = \Sigma_0$ when process is in-control

Switch Drum Example

Hawkins (1991) use example from Flury and Riedwyl (1988) for five dimensions ($X_1 - X_5$) of switch drums.

OBS	X1	X2	X3	X4	X5
1	20	8	13	9	7
2	20	12	17	12	11
3	15	7	9	7	2
4	16	12	14	13	8
5	20	12	16	13	9
6	16	10	15	10	8
..	..	..	..	..	..
49	20	11	14	12	11
50	18	11	14	12	10

Figure 7.3. In-Control Data Saved in ICONTROL

X1	X2	X3	X4	X5
17.96	10.30	13.76	11.08	8.26

Figure 7.4. Means

X1	X2	X3	X4	X5
1.8622	1.7053	1.7090	1.8718	2.2114

Figure 7.5. Standard Deviations

Out-of Control Data For Switch Drum

Data are simulated from the same multinormal distribution as the in-control data, but a shift of one-quarter standard deviation is introduced into X_5 after observation 35. Also, the marginal std. dev. of X_1 was increased by half at the same instant.

OBS	X1	X2	X3	X4	X5
1	17.265	11.788	15.101	13.903	10.465
2	17.384	6.996	11.552	7.253	6.641
3	16.517	10.277	11.724	13.013	9.111
4	14.997	10.682	12.087	11.457	6.320
5	17.633	9.348	12.672	10.475	5.481
6	16.041	11.320	13.957	11.474	8.176
..	...	...	...	...	...
49	16.188	9.140	13.284	10.991	9.126
50	22.047	10.824	14.796	10.872	9.264

Figure 7.6. Out-of-Control Data Saved in OCONTROL

Using SAS/IML® to Compute **Z**

```
proc iml;

  /*--- compute Mu0 from the in-control data ---*/
   use icontrol;
   read all var{x1 x2 x3 x4 x5};
   summary  var{x1 x2 x3 x4 x5}
            stat{mean} opt{noprint save};
   mu0 = shape(x1||x2||x3||x4||x5,50,5);

  /*--- construct the transformation matrix A ---*/
   read all var{x1 x2 x3 x4 x5};
   x   = x1||x2||x3||x4||x5;
   cor = corr(x);
   a = inv(sqrt(diag(inv(cor))))*inv(cor);

  /*--- read the out-of-control data ---*/
   use ocontrol;
   read all var{x1 x2 x3 x4 x5};
   x = x1||x2||x3||x4||x5;

   z = (x-mu0)*a;

  /*--- save the Z matrix to a output data set ---*/
   create zdata from z [colname=('z1':'z5')];
   append from z;

   close;
   run;
   quit;
```

Figure 7.7. Construct the Z matrix

Pre-process Data Sets

```
data xdata;
   set ocontrol;
   x1 = (x1-17.96)/1.8622;
   x2 = (x2-10.30)/1.7053;
   x3 = (x3-13.76)/1.7090;
   x4 = (x4-11.08)/1.8718;
   x5 = (x5- 8.26)/2.2114;
   run;

data xdata;
   drop i;
   do i = 1 to 5;
      set xdata;
      batch = 5*_n_;
      output;
   end;
   run;
```

Figure 7.8. Standardize the **X** Raw Data and Subgroup

```
proc standard data=zdata mean=0 std=1 out=zdata;
run;

data zdata;
   drop i;
   do i = 1 to 5;
      set zdata;
      batch = 5*_n_;
      output;
   end;
   run;
```

Figure 7.9. Standardize the **Z** Data and Subgroup

Control Chart for X_1

```
symbol v=dot;
title 'Mean and Range Charts for X1';
proc shewhart data=xdata graphics;
   xrchart x1*batch;
   run;
```

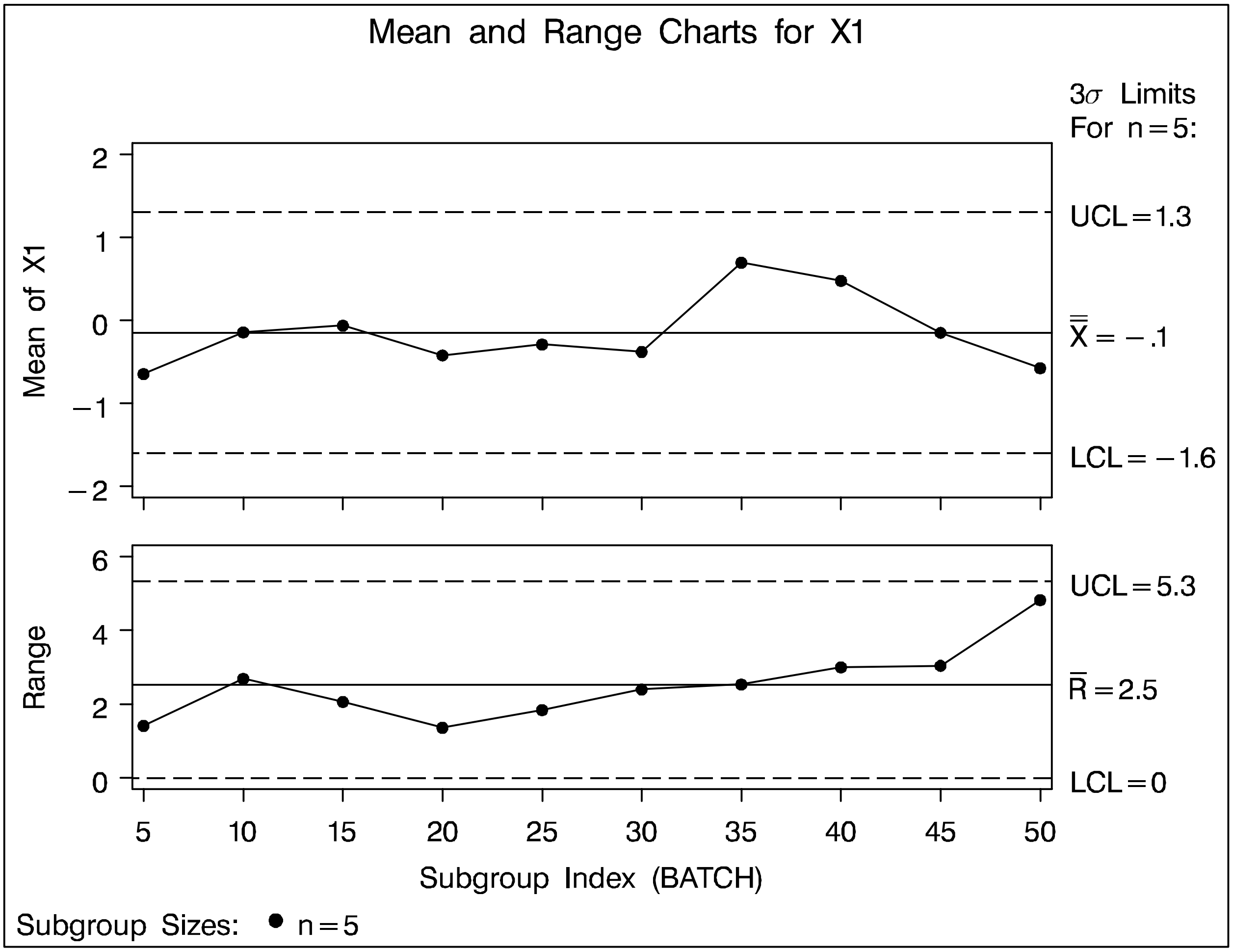


Figure 7.10. $\bar{X}$ and R Charts for X_1

Control Chart for Z_1

```
symbol v=dot;
title 'Mean and Range Charts for Z1';
proc shewhart data=zdata graphics;
   xrchart z1*batch;
   run;
```

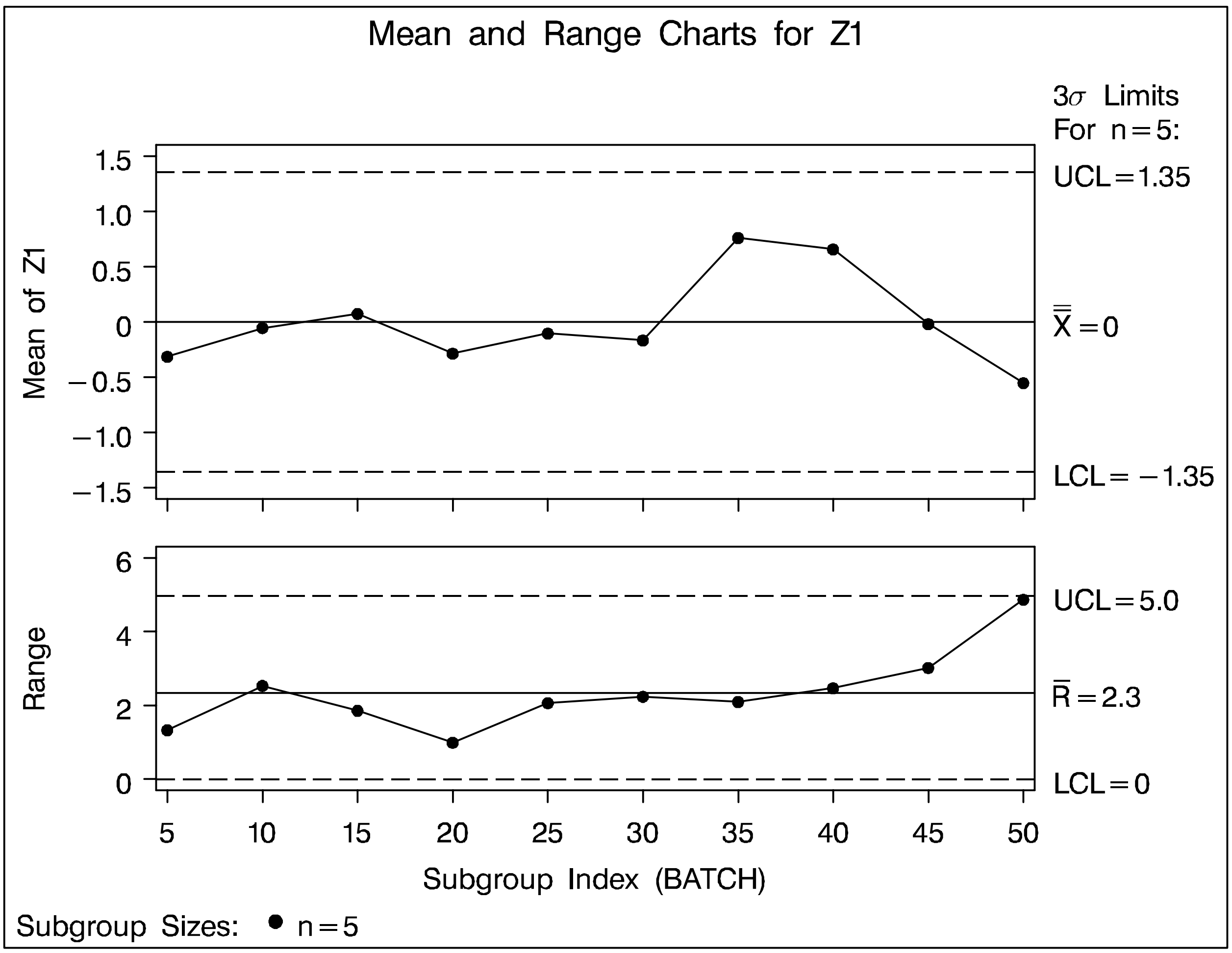


Figure 7.11. $\bar{X}$ and R Charts for Z_1

Control Chart for X_5

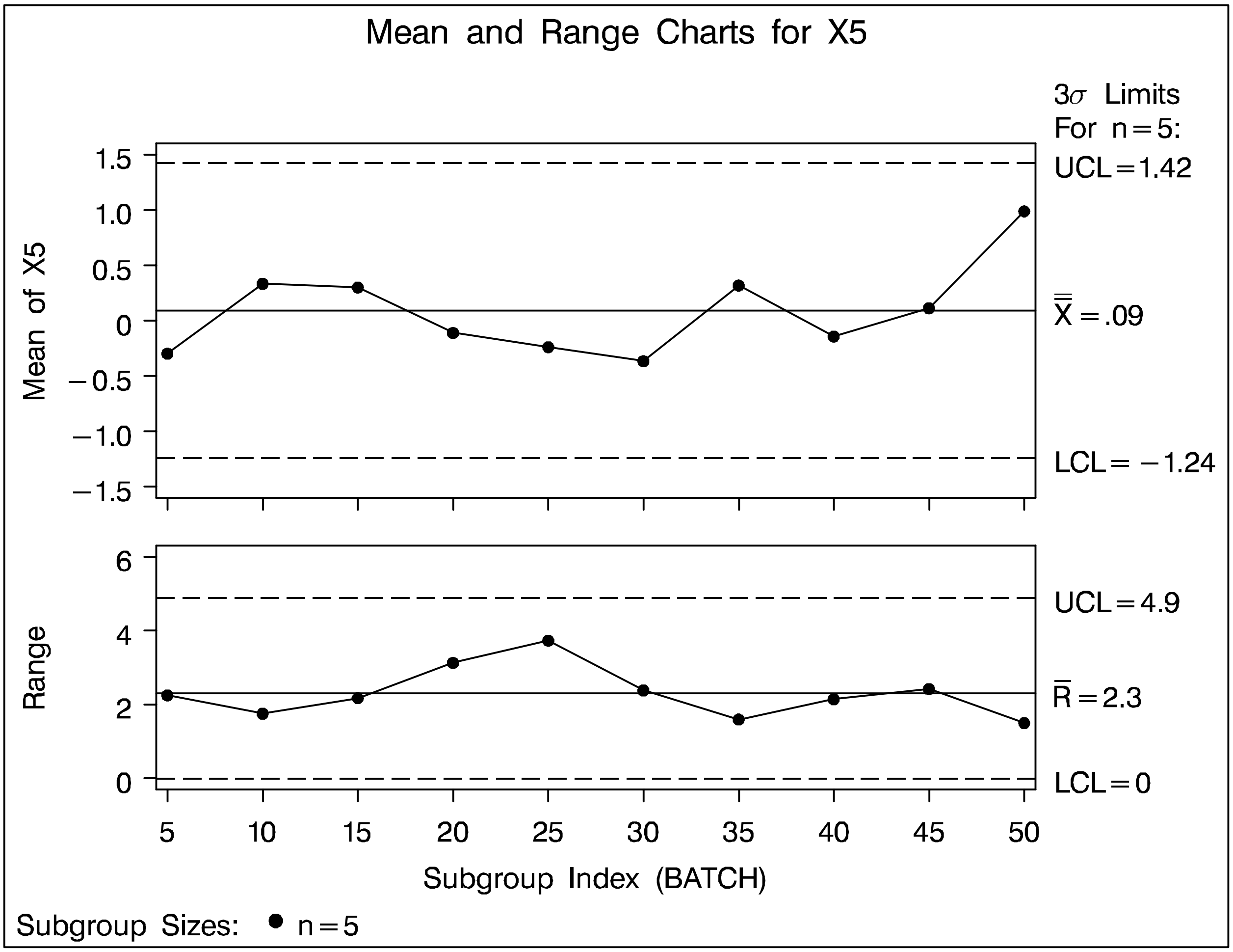


Figure 7.12. $\bar{X}$ and R Charts for X_5

Control Chart for Z_5

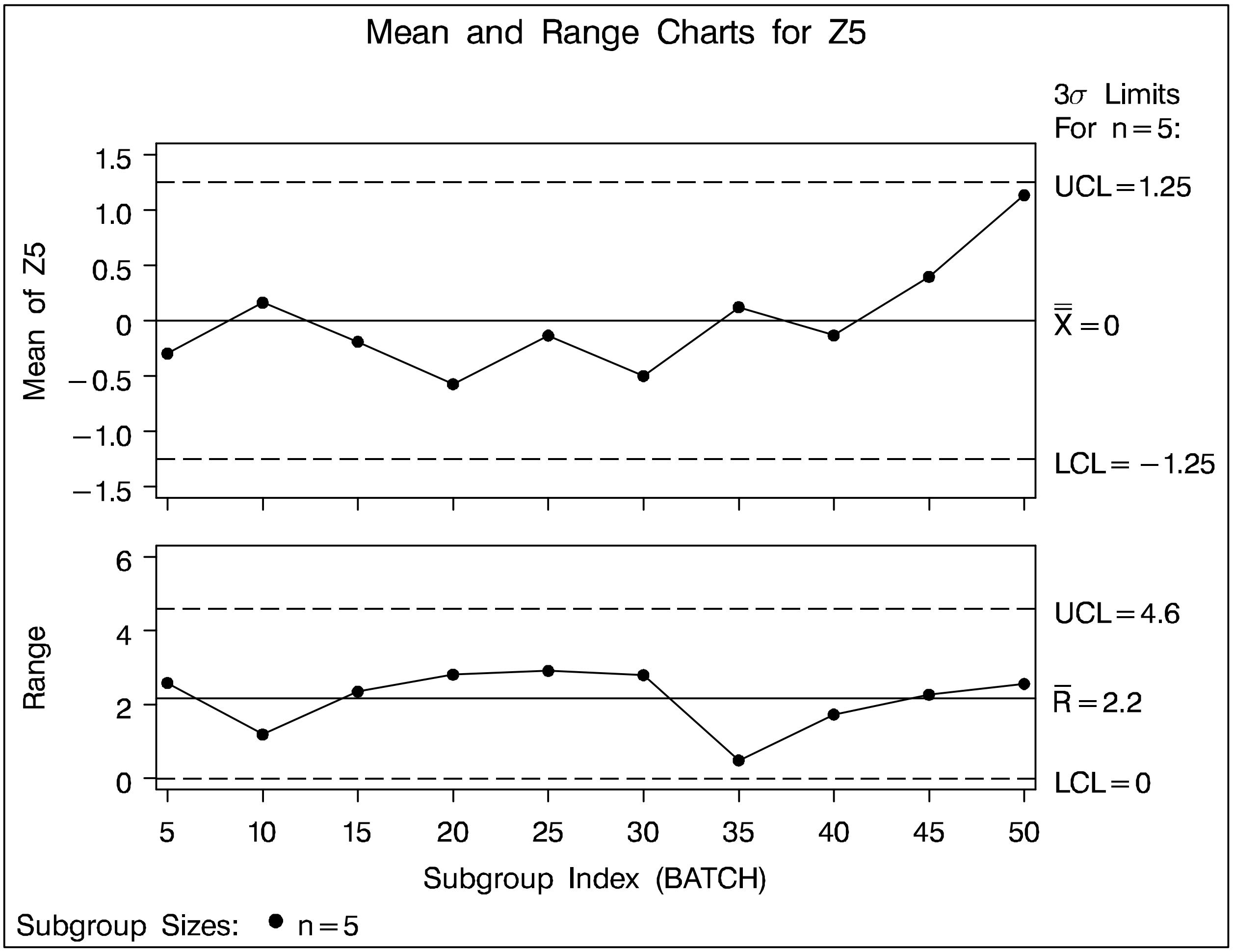


Figure 7.13. $\bar{X}$ and R Charts for Z_5

Chapter 8

Resources

General References

Al-Assaf, A. F. and Schmele, J. A. (1993), *The Textbook of Total Quality in Healthcare*, Delray Beach, Florida: St. Lucie Press.

Al-Salti, M. and Statham, A. (1994), "A Review of the Literature on the Use of SPC in Batch Production," *Quality and Reliability Engineering International*, 10, 49−62.

Alwan, L. C. and Roberts, H. V. (1988), "Time Series Modeling for Statistical Process Control," *Journal of Business and Economic Statistics*, 6, 87−95.

Balestracci, D. and Barlow, J. L. (1994), *Quality Improvement: Practical Applications for Medical Group Practice*, Englewood, Colorado: Center for Research in Ambulatory Health Care Administration.

Benneyan, J. (1995), "Applications of Statistical Process Control (SPC) to Improve Health Care," conference presentation, *Healthcare Information and Mangement Systems Society*.

Benneyan, J. and Kaminsky, F. C. (1995), "Another View on How to Measure Health Care Quality," *Quality Progress*, 28(2), 120−124.

Berwick, D. M. (1989), "Continuous Improvement as an Ideal in Health Care," *New England Journal of Medicine*, 320, 53−56.

Berwick, D. M. (1991), "Controlling Variation in Healthcare: A Consultation from Walter Shewhart," *Medical Care*, 29, 1212−1225.

Berwick, D. M. (1992), "Can Quality Management Really Work in Health Care?," *Quality Progress*.

Berwick, D. M., Godfrey, A. B., and Roessner, J. (1991), *Curing Health Care*, San Francisco: Jossey-Bass.

Box, G. E. P. and Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.

Box, G. E. P. and Kramer, T. (1992), "Statistical Process Monitoring and Feedback Adjustment—A Discussion," *Technometrics*, 34, 51−285 (with discussion).

Champ, S. W. and Woodall, W. H. (1987), "Exact Results for Shewhart Control Charts With Supplementary Runs Rules," *Technometrics*, 29, 393−401.

Doganaksoy, N., Faltin, F. W., and Tucker, W. T. (1991), "Identification of Out-of-Control Quality Characteristics in a Multivariate Manufacturing Environment," *Communications in Statistics–Theory and Methods*, 20, 2775−2790.

Farnum, N. R. (1992), "Control Charts for Short Runs: Nonconstant Process and Measurement Error," *Journal of Quality Technology*, 24, 138−144.

Flury, B. and Riedwyl, H. (1988), *Multivariate Statistics: A Practical Approach*, London: Chapman & Hall.

Fulenwider, D. O. (1988), "Using SAS Software for the Analysis of Means," *SAS Users Group International: Proceedings of the Thirteenth Annual Conference*, 1212−1219.

Gnanadesikan, R. and Kettenring, J. R. (1972), “Robust Estimates, Residuals, and Outlier Detection with Multiresponse Data,” *Biometrics,* 28, 81−124.

Hawkins, D. M. (1991), “Multivariate Quality Control Based on Regression-Adjusted Variables,” *Technometrics,* 33, 61−75.

Hawkins, D. M. (1993), “Regression Adjustment for Variables in Multivariate Quality Control,” *Journal of Quality Technology*, 25, 170−182.

Hillier, F. S. (1969), “$\bar{X}$- and R-Chart Control Limits Based On a Small Number of Subgroups,” *Journal of Quality Technology*, 1, 17−26.

Hunter, J. S. (1986), “The Exponentially Weighted Moving Average,” *Journal of Quality Technology*, 18, 203−210.

Jackson, J. E. (1980), “Principal Components and Factor Analysis: Part I —Principal Components,” *Journal of Quality Technology*, 12, 201–213.

Jackson, J. E. (1991), *A User’s Guide to Principal Components*, New York: John Wiley & Sons, Inc.

Laffel, G. and Blumenthal, D. (1989), “The Case for Using Industrial Quality Mangement Science in Health Care Organizations,” *Journal of the American Medical Association*, 262, 2869−2873.

Longo, D. R. and Bohr, D. (1991), *Quantitative Methods in Quality Management: A Guide for Practitioners,* Chicago: American Hospital Association.

MacGregor, J. (1987), “Interfaces Between Process Control and Online Statistical Process Control,” *Computing and Systems Technology Division Communications*, 10, 9−20.

MacGregor, J. (1990), “A Different View of the Funnel Experiment,” *Journal of Quality Technology*, 22, 255−259.

MacGregor, J., Hunter, J. S., and Harris, T. (1988), “SPC Interfaces.” Short course notes.

Mason, R.L., Tracy N. D., and Young, J. C. (1995), “Decomposition of T^2 for Multivariate Control Chart Techniques,” *Journal of Quality Technology,* 27, 99−108.

Mason, R.L., Tracy N. D., and Young, J. C. (1996), “Monitoring a Multivariate Step Process,” *Journal of Quality Technology*, 28, 39−50.

Montgomery, D. C. (1991), *Introduction to Statistical Quality Control, Second Edition*, New York: John Wiley & Sons, Inc.

Montgomery, D. C. and Mastrangelo, C. M. (1991), “Some Statistical Process Control Methods for Autocorrelated Data,” *Journal of Quality Technology,* 23, 179−204 (with discussion).

Montgomery, D. C., Keats, J. B., Runger, G. C. and Messina, W. S. (1994), “Integrating Statistical Process Control and Engineering Process Control,” *Journal of Quality Technology,* 26, 79−87.

Neave, H. R. (1990), *The Deming Dimension*, Knoxville, TN: SPC Press, Inc.

Nelson, L. S. (1974), “Factors for the Analysis of Means,” *Journal of Quality Technology*, 6, 175−181.

Nelson, L. S. (1983), “Exact Critical Values for Use with the Analysis of Means,” *Journal of Quality Technology,* 15, 40−44.

Nelson, L. S. (1984), "The Shewhart Control Chart—Tests For Special Causes," *Journal of Quality Technology*, 15, 237−239.

Nelson, L. S. (1985), "Interpreting Shewhart $\bar{X}$ Control Charts," *Journal of Quality Technology*, 17, 114−116.

Nelson, L. S. (1989), "Standardization of Shewhart Control Charts," *Journal of Quality Technology*, 21, 287−289.

Nelson, L. S. (1994), "Shewhart Control Charts With Unequal Subgroup Sizes," *Journal of Quality Technology*, 26, 64−67.

Nelson, P. R. (1982), "Exact Critical Values for Use with the Analysis of Means," *Communications in Statistics*, A11, 699−709.

Nelson, P. R. (1988), "Application of the Analysis of Means," *SAS Users Group International: Proceedings of the Thirteenth Annual Conference*, 225−230.

Ott, E. R. (1975), *Process Quality Control: Troubleshooting and Interpretation of Data*, New York: McGraw-Hill.

Plsek, P. E. (1992), "Introduction to Control Charts," *Quality Management in Health Care*, 1, 65−74.

Quesenberry, C. P. (1991a), "SPC Q Charts for Start-Up Processes and Short or Long Runs," *Journal of Quality Technology*, 23, 213−224.

Quesenberry, C. P. (1991b), "SPC Q Charts for a Binomial Parameter p: Short or Long Runs," *Journal of Quality Technology*, 23, 239−246.

Quesenberry, C. P. (1993), "The Effect of Sample Size on Estimated Effects," *Journal of Quality Technology*, 25, 237−247.

Ramig, P. R. (1983), "Applications of the Analysis of Means," *Journal of Quality Technology*, 15, 19−25.

Rodriguez, R. N. (1994), "Recent Issues in Statistical Process Control: SAS Solutions Using Statistical Modeling Procedures," *SAS Users Group International: Proceedings of the Nineteenth Annual Conference*, 1151−1168.

Rodriguez, R. N. (1996), "Health Care Applications of Statistical Process Control: Examples Using the SAS System," *SAS Users Group International: Proceedings of the Twenty-First Annual Conference*, 1381−1396.

Ryan, T. P. (1989), *Statistical Methods for Quality Improvement*, New York: John Wiley & Sons, Inc.

Schilling, E. G. and Nelson, P. R. (1976), "The Effect of Non-Normality on the Control Limits of $\bar{X}$ Charts," *Journal of Quality Technology*, 8, 183−187.

Schneider, H. and Pruett, J. M. (1994), "Control Charting Issues in the Process Industries," *Quality Engineering*, 6, 347−373.

Spoeri, R. K. (1991), "The Emerging Use of Measurement and Statistics in Health Care Quality and Productivity Improvement," presentation at the 1991 Joint Statistical Meetings, Atlanta, Georgia.

Staker, L. V. (1995), “Using Statistical Process Control to Improve Clinical Outcomes in a Primary Care Practice,” presentation at the 7th Annual National Forum on Quality Improvement in Health Care.

Tracy, N. D., Young, J. C. and Mason, R. L. (1992), “Multivariate Control Charts for Individual Observations,” *Journal of Quality Technology*, 24, 88–95.

VanderVeen, L. M. (1992), “Statistical Process Control: A Practical Application for Hospitals,” *Journal for Healthcare Quality*, 14, 20−29.

Wade, W. H. and Woodall, M. M. (1993), “A Review and Analysis of Cause-Selecting Control Charts,” *Journal of Quality Technology,* 25, 161−169.

Wadsworth, H. M., Stephens, K. S., and Godfrey, A. B. (1986), *Modern Method for Quality Control and Improvement*, New York: John Wiley & Sons.

Westgard, J. O., Barry, P. L., Hunt, M. R., and Groth, T. (1981), “A Multi-Rule Shewhart Chart for Quality Control in Clinical Chemistry,” *Clinical Chemistry*, 27, 493−501.

Wetherill, G. B. and Brown, D. B. (1991), *Statistical Process Control: Theory and Practice*, London: Chapman and Hall.

Wheeler, D. J. (1991a), *Short Run SPC*, Knoxville, Tennessee: SPC Press, Inc.

Wheeler, D. J. (1991b), “Shewhart’s Chart: Myths, Facts, and Competitors,” *45th Annual Quality Congress Transactions*, American Society for Quality Control. 533−538.

Wheeler, D. J. (1993), *Understanding Variation: The Key to Managing Chaos*, Knoxville, TN: SPC Press, Inc.

Wheeler, D. J. (1995), *Advanced Topics in Statistical Process Control: The Power of Shewhart’s Chart*, Knoxville, TN: SPC Press, Inc.

Wheeler, D. J. and Chambers, D. S. (1986), *Understanding Statistical Process Control,* Knoxville, Tennessee: Statistical Process Controls, Inc.

Wilks, S. S. (1962), *Mathematical Statistics,* New York: John Wiley & Sons, Inc.

Woodall, W. H. (1993), “Autocorrelated Data and SPC,” *ASQC Statistics Division Newsletter*, 13, 18−21.

SAS Institute References

Rodriguez, R. N. and Bynum, R. A. (1992), *Examples of Short Run Process Control Methods With the SHEWHART Procedure in SAS/QC Software*. Unpublished manuscript available from the authors.

SAS Institute Inc. (1990a), *SAS/GRAPH Software: Reference, Version 6, First Edition, Volumes 1 and 2*, Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1990b), *SAS/STAT User’s Guide, Version 6, Fourth Edition, Volumes 1 and 2*, Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1993), *SAS/ETS User’s Guide, Version 6, Second Edition*, Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1993), *SAS/QC Software: ADX Menu System for Design of Experiments*, Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1995a), *SAS/AF Software: FRAME Class Dictionary, Version 6, First Edition*, Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1995b), *SAS/ETS Software: Time Series Forecasting System, Version 6, First Edition*, Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1995c), *SAS/QC Software: SQC Menu System for Quality Improvement, Version 6, Second Edition*, Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1995d), *SAS/QC Software: Usage and Reference, Version 6, Volumes 1 and 2*, Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1996), *SAS/STAT Software: Changes and Enhancements through Release 6.11*, Cary, NC: SAS Institute Inc.

References Available Online

Rodriguez, R. N. (1996a), "Recent Issues in Statistical Process Control: SAS Solutions Using Statistical Modeling Procedures."

Rodriguez, R. N. (1996b), "Health Care Applications of Statistical Process Control: Examples Using the SAS System."

Acknowledgement

We are grateful to Martin S. King of the Applications Division at SAS Institute Inc. for his assistance in the preparation of the notes for this workshop.

Appendix 1

Overview of SAS/QC Software

Overview of SAS/QC Software

SAS/QC software, a component of the SAS System, provides a comprehensive set of tools for statistical quality improvement. You can use these tools to

- organize quality improvement efforts
- design industrial experiments for product and process improvement
- apply Taguchi methods for quality engineering
- establish statistical control of a process
- maintain statistical control and reduce variation
- analyze process capability
- develop and evaluate acceptance sampling plans

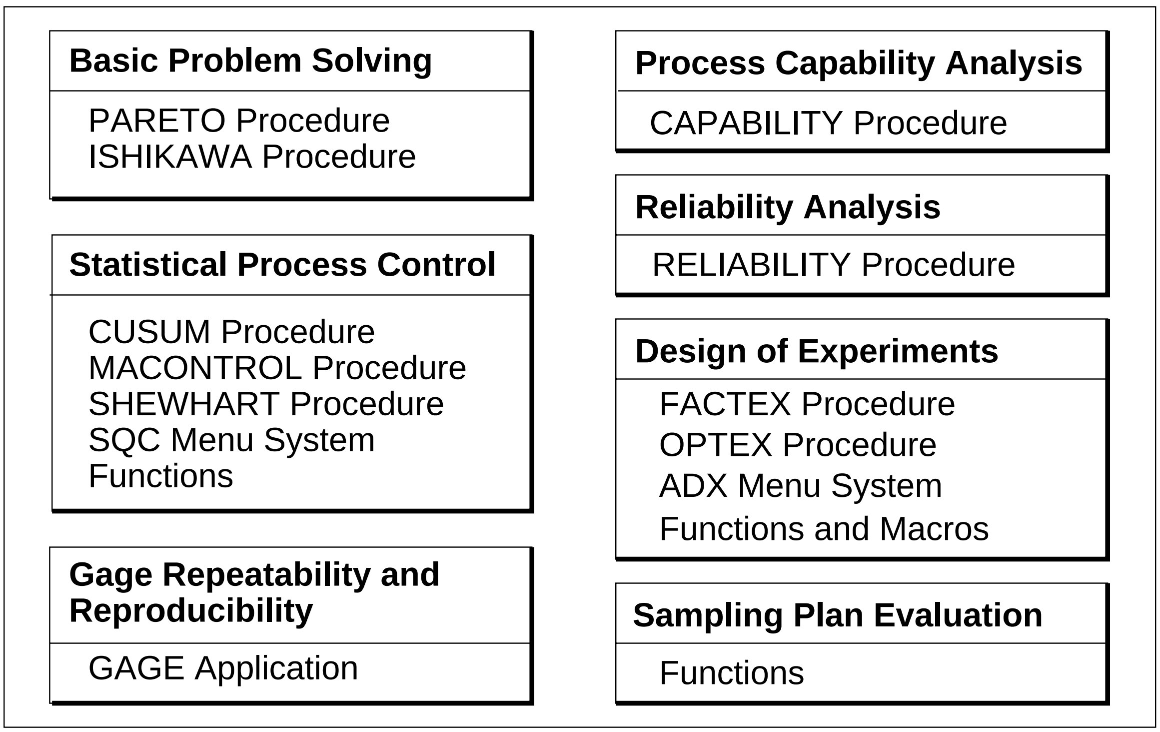
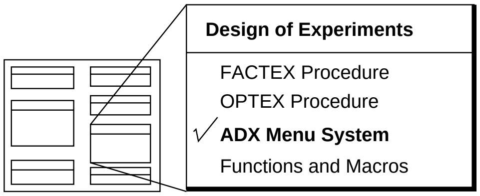


Figure 1. Components of SAS/QC Software

There are two main types of tools in SAS/QC software: menu systems and procedures.

- The menu systems are complete, full-screen oriented environments for statistical quality improvement applications. Unlike the procedures, the menu systems require no knowledge of SAS programming. Internally, however, the menu systems translate the user's selections into SAS statements that are then submitted for execution.
- The procedures in SAS/QC software offer greater flexibility and power than the menu systems. To use a procedure, you must have a basic knowledge of the SAS language and the syntax of the procedure. You can run the procedures in a batch program or interactively with the SAS Display Manager System.

ADX Menu System for Design of Experiments



The ADX Menu System provides facilities for designing and analyzing standard experiments and is intended primarily for quality engineers, researchers, and other non-statisticians.

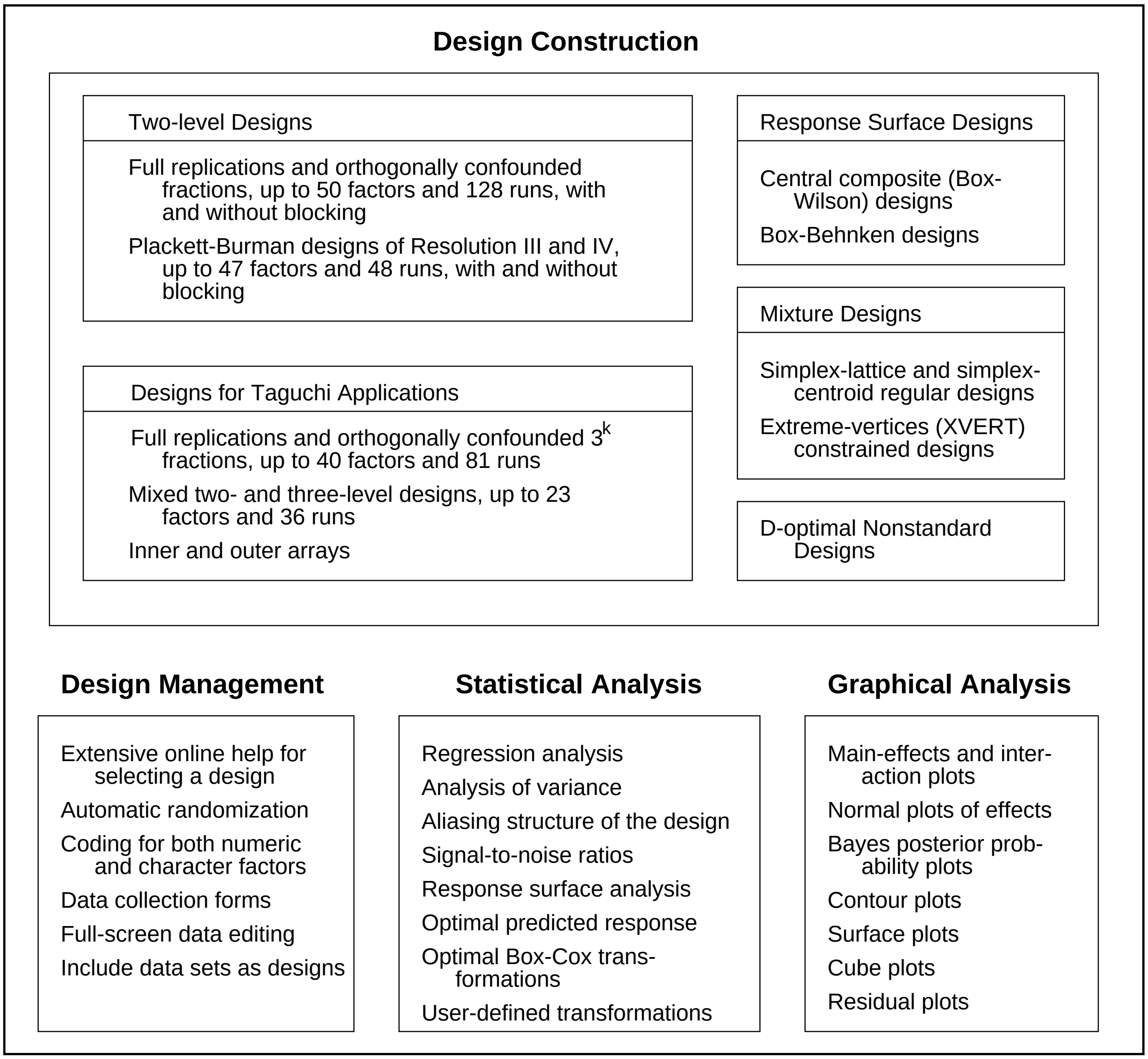
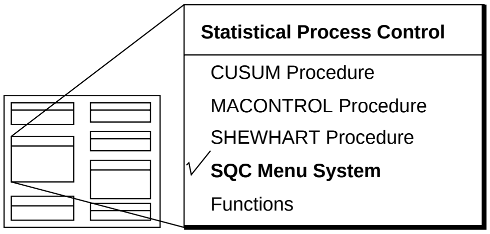


Figure 2. General Design and Analysis Facilities

Note: The ADX Menu System is documented in *SAS/QC Software: ADX Menu System for Design of Experiments, Version 6, First Edition*.

SQC Menu System for Statistical Quality Control



The SQC Menu System provides facilities for standard statistical quality control applications and is intended for quality analysts, quality control managers, and other non-statisticians.

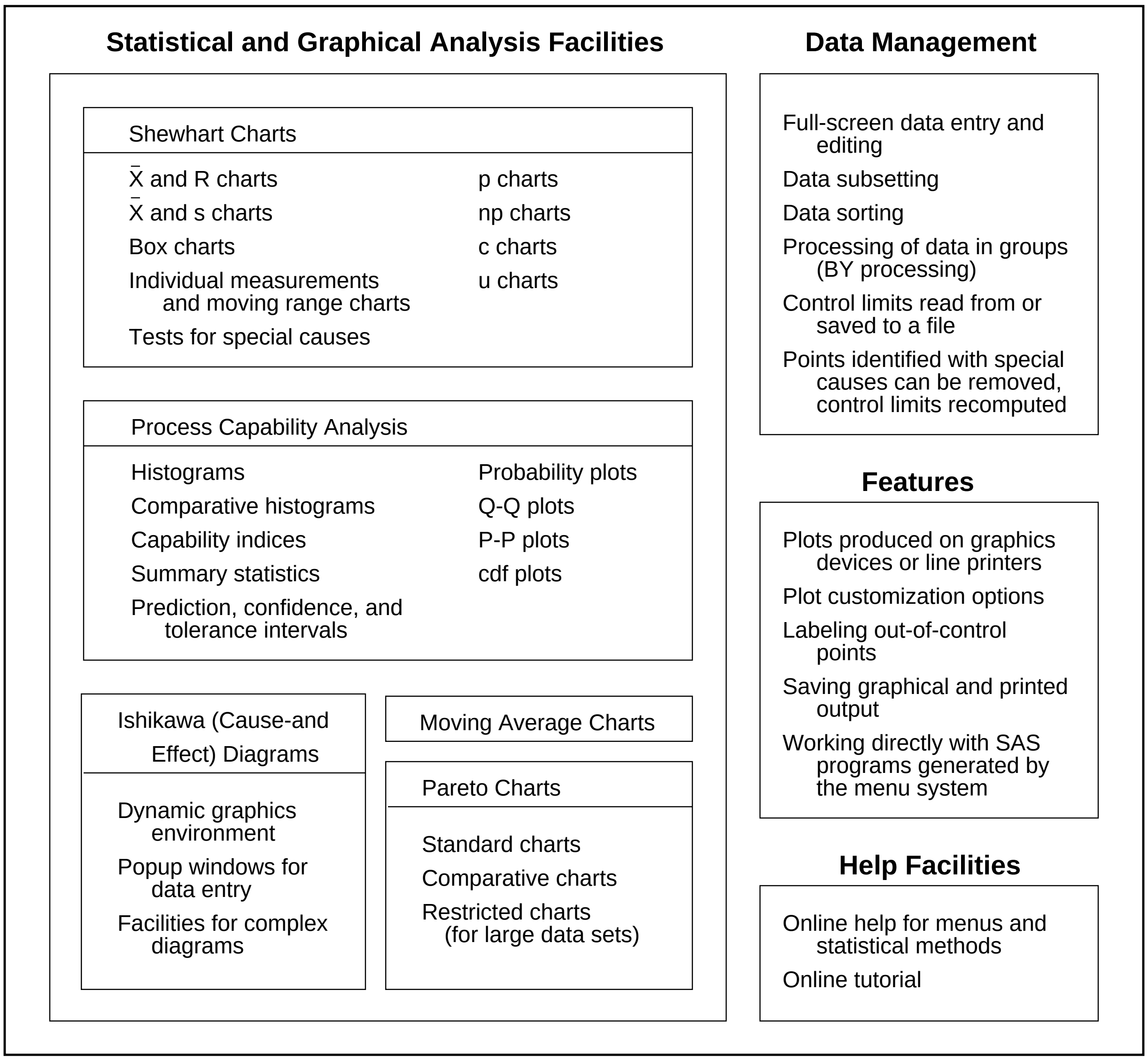
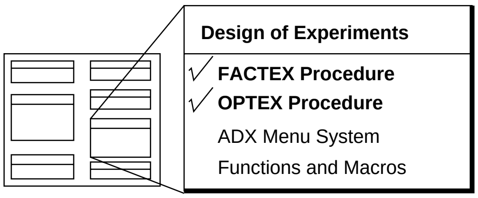


Figure 3. Overview of the SQC Menu System

Note: The SQC Menu System is documented in *SAS/QC Software: SQC Menu System, Version 6, First Edition*.

Procedures for Design of Experiments



SAS/QC software provides two procedures for the design of experiments. The FACTEX procedure constructs factorial experimental designs, which are useful for studying the effects of various factors on a response. The OPTEX procedure searches for optimal designs in situations in which standard designs are not available.

FACTEX Procedure

Overview

Constructs orthogonally confounded full and fractional factorial designs, with and without blocking, including 2^k orthogonal array designs

Constructs q^k factorial designs, where q is a prime. These can be used with the SAS DATA step to construct many complex designs, such as mixed-level factorial and balanced incomplete block designs.

Features

No limitation on the number of factors or factor levels

Appropriate confounding rules are given for model constraints

Designs can be chosen to have
- minimum size
- maximum resolution
- maximum block size
- minimum aberration

OPTEX Procedure

Overview

Searches for D-optimal or A-optimal designs

Intended for situations in which a standard design is not appropriate. For example,
- Not all combinations of factor levels are feasible
- Resources restrict the number of runs
- The experimentation region is irregular
- The model is nonlinear or nonstandard linear

Features

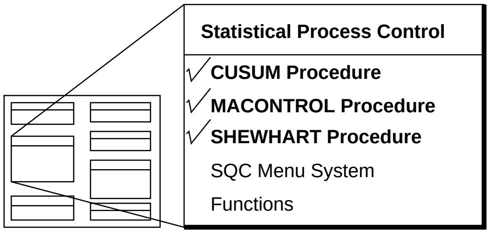
Add points to an existing design
Augment preexisting covariates
Evaluate an existing design
Improve an initial design

Various search algorithms and selection criteria are available

Flexible GLM-type models are available

Figure 4. Overview of the Experimental Design Procedures

Procedures for Control Chart Analysis



SAS/QC software provides three procedures for the creation and analysis of control charts. The SHEWHART procedure creates all commonly encountered Shewhart charts for variables and attributes. The CUSUM procedure creates cumulative sum control charts. The MACONTROL procedure creates moving average charts.

Control Chart Procedures

Shewhart Charts

- $\bar{X}$ and R charts
- $\bar{X}$ and s charts
- Box charts
- Individual measurements and moving range charts
- Tests for special causes
- p charts
- np charts
- c charts
- u charts

Cusum Charts

- Cumulative sum control charts
  - One-sided scheme (decision interval)
  - Two-sided scheme (V-mask)

Moving Average Charts

- Uniformly weighted moving average charts
- Exponentially weighted moving average charts

Nonstandard Control Charts

- Trend charts for time-dependent data
- Star charts for multivariate process data

Features

Control Limits

- Computed from the data
- Computed from standard values
- Computed as probability limits
- Read from an input data set
- Limits vary with subgroup sample size or fixed limits can be specified

Process Standard Deviation

- Estimated by one of a variety of methods
- Specified as a known value

Data Sets

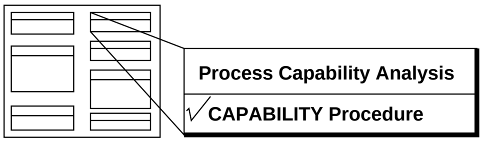
- Specify various input data sets
- Create various output data sets

Graphical Enhancement Features

- Displaying multiple sets of control limits
- Displaying stratification of process data
- Classifying subgroups into blocks
- Highlighting and labeling out-of-control points
- Specifying date or time format for subgroups
- Labeling extreme points in box charts
- Specifying number of subgroups per page
- Adding titles, notes, and footnotes
- Customizing axes and control limit labels
- Specifying colors, fonts, and line types

Figure 5. Overview of Control Chart Analysis Procedures

Procedure for Process Capability Analysis



SAS/QC software provides one procedure for the analysis of process capability. The CAPABILITY procedure compares the distribution of output from an in-control process to the specification limits of the process to determine the consistency with which the specification limits can be met.

CAPABILITY Procedure

Overview

Provides graphical and statistical tools for process capability analysis

Statistical Methods

- Descriptive statistics
- Capability indices: Cp, Cpk, CPL, CPU, Cpm, k
- Confidence, tolerance, and prediction intervals
- Goodness-of-fit based on ecdf
- Estimation of areas outside specification limits

Other Graphical Features

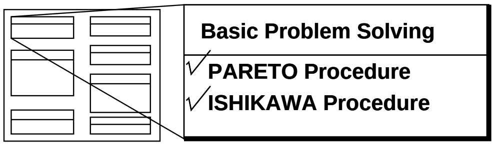
- Inset boxes with summary statistics
- Customized axes and legends
- Titles, footnotes, notes
- Colors, fonts, symbol markers

Graphical Methods

- Histograms, superimposed with
  - specification limits
  - fitted pdf curves
  - kernel density estimates
  - summary statistics
- Comparative histograms, superimposed with
  - specification limits
  - fitted pdf curves
  - kernel density estimates
  - summary statistics
- cdf plots, superimposed with
  - specification limits
  - fitted theoretical distribution curves
- Probability plots, superimposed with
  - specification limits
- Quantile-quantile (Q-Q) plots, superimposed with specification limits
- Probability-probability (P-P) plots
- Available distributions:
  - normal
  - lognormal
  - exponential
  - Weibull
  - gamma
  - beta

Figure 6. Overview of Process Capability Analysis

Procedures for Basic Quality Problem Solving



SAS/QC software provides two procedures for basic quality problem solving. The PARETO procedure creates charts that display the relative frequency of problems in a process or operation. The ISHIKAWA procedure creates a cause-and-effect or fishbone diagram, which displays factors that affect a quality characteristic or problem.

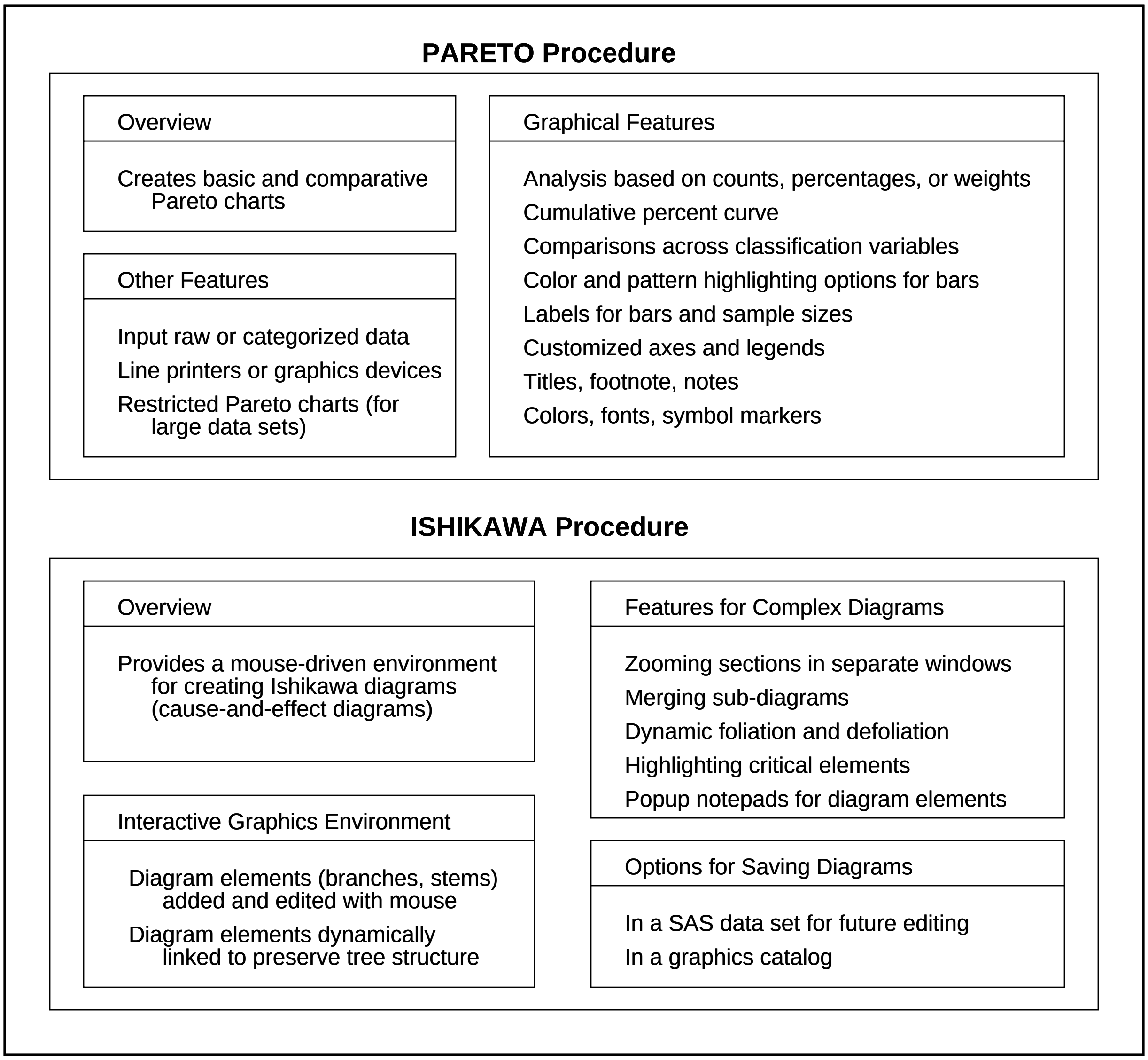
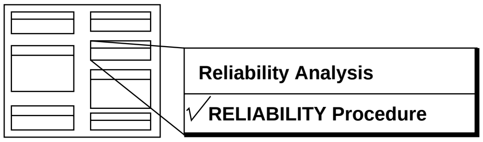


Figure 7. Overview of Quality Problem Solving Procedures

Procedure for Reliability Analysis



SAS/QC software provides one procedure for reliability analysis. The RELIABILITY procedure performs graphical and statistical analysis of component lifetime data and system repair data.

RELIABILITY Procedure

Overview

Provides graphical and statistical tools for lifetime data analysis and for repair data analysis.

Statistical Analysis

- Regression models (including accelerated life test models) for censored data
- ML estimates of distribution parameters, percentiles, and reliability functions
- Asymptotic normal and likelihood ratio CIs for distribution parameters and percentiles; asymptotic normal CIs for the reliability function
- Weibayes analysis
- Nonparametric estimates and CIs for the mean cumulative function for cost or number of repairs
- Analysis for multiple failure modes

Graphical Analysis

- Probability plots
- Plots of data and fitted relation for life versus stress for accelerated life test data
- Nonparametric plots of the mean cumulative function and associated confidence intervals
- Inset boxes with summary statistics

Distributions

Binomial	Loglogistic
Exponential	Lognormal
Extreme Value	Normal
Gamma	Poisson
Logistic	Weibull

Other Graphical Features

- Customized axes and legends
- Titles, footnotes, notes
- Colors, fonts, symbol markers

Figure 8. Overview of Reliability Analysis

Appendix 2

Generating Graphics Output

Generating Graphics Output

The graphics examples in these notes are generated with the following macro:

```
%macro gout(gname);
   filename GSASFILE "&gname..ps";
   goptions dev=pslepsf gaccess=sasgaedt gsfname=GSASFILE
          hsize   = 6.5 in
          vsize   = 5 in
          htext   = 3.0 pct
          htitle  = 3.5 pct
          vorigin = 0 in
          horigin = 1 in
          display  gsfmode=replace
          border ftext=swiss lfactor=3;
   %mend;
```

You can include this macro in your AUTOEXEC.SAS file or at the beginning of your program. The output is a PostScript file. You can specify other types of files with the DEV= option, and you can change the size of the figure with the HSIZE= and VSIZE= options.

The following statements create a file called EXAMPLE1.PS that contains the charts shown in Figure 2.2 on page 14.

```
%gout(example1);

title 'Mean and Range Charts for Diameters';
symbol v=dot;

proc shewhart data=wafers graphics;
   xrchart diamtr*batch;
run;

goptions reset=all dev=xcolor;
```

The DEV= option in the GOPTIONS statement specifies that further output is to be sent to the monitor. The appropriate value for the DEV= option depends on your hardware. See *SAS/GRAPH Software: Reference, Version 6, First Edition, Volumes 1 and 2,* for more information on the GOPTIONS statement.