Chapter 16

Introduction to Statistical Quality Control (SQC)

- *Synonym*: *Statistical Process Control* (*SPC*)
- It is based on the premise that some degree of variability in manufacturing processes is inevitable.
- SQC is widely used for discrete parts manufacturing (e.g., automobiles, microelectronics) and product quality control.
- SQC References:
  1. Text, 4/e; Chapter 16. pp. 645-62.
  2. Seborg, Edgar and Mellichamp, *Process Dynamics and Control*, *2nd ed*., Chapter 21, Wiley, NY (2004)

  Note: Equation, figure, and table numbers that begin with "21" are from Ref. 2.

Introduction to SQC (continued)

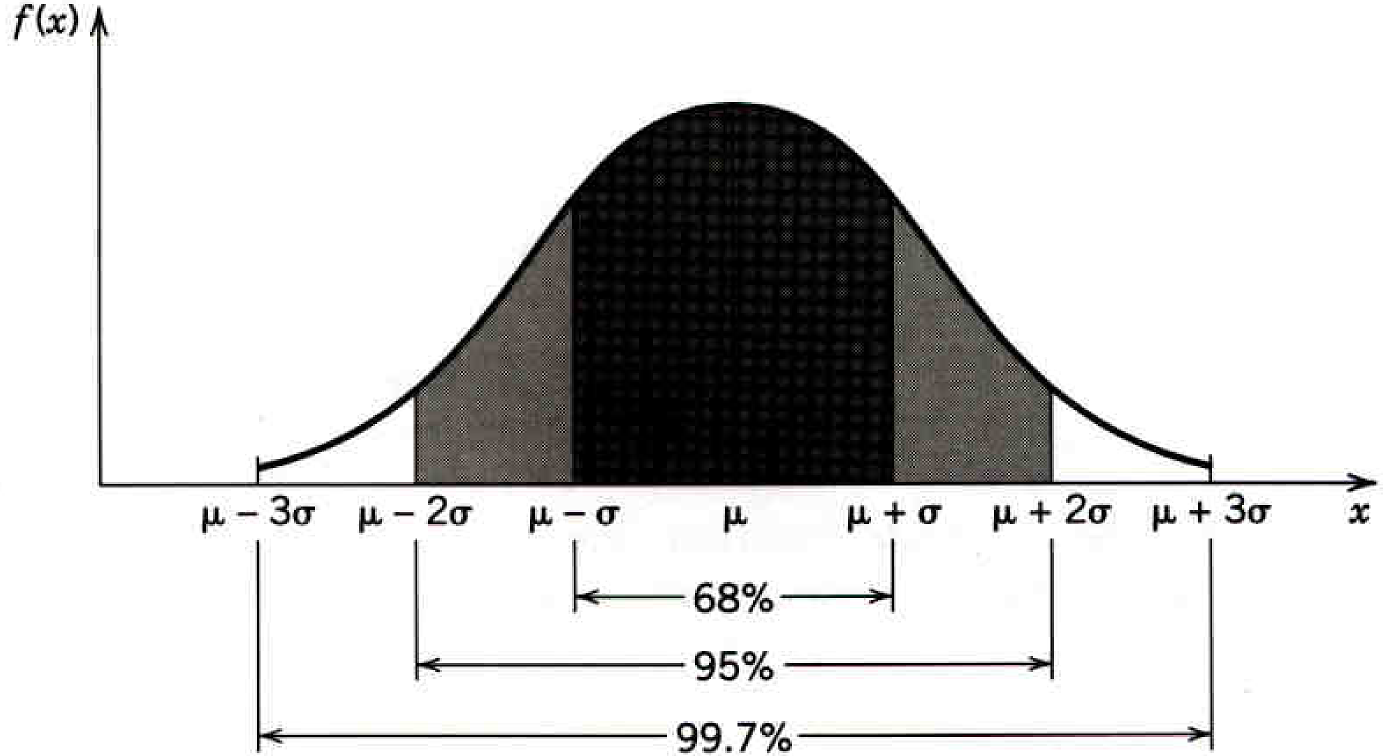
- Attempts to distinguish between two types of situations:
    1. *Normal operation* and "chance causes", that is, random variations.
    2. *Abnormal operation* due to a "special cause" (often unknown).
- SQC uses "Quality Control Charts" (also called, "Control Charts") to distinguish between normal and abnormal situations.
- Basic Model for SQC monitoring activity:

$$x = \mu + \varepsilon$$

where x is the measured value, μ is its (constant) mean, and ε is a random error.

Chapter 16

The Normal Distribution (review)



- It is important to distinguish between the theoretical mean, μ , and the sample mean, $\bar{x}$.

- If measurements of a variable are normally distributed, $N\left(\mu,\sigma^2\right)$, the sample mean is also normally distributed.

Control Charts

- In SQC, control charts are used to determine whether the process operation is normal or abnormal.
- Two general types of control charts:
  1. Charts for measures of central tendency (e.g., sample mean, individual measurements)
  2. Charts for measures of variability (e.g., standard deviation, range)

The $\bar{x}$ Control Chart

- The most widely used control chart is the $\bar{x}$ chart.
- This type of control chart is often referred to as a Shewhart Chart, in honor of the pioneering statistician, Walter Shewhart, who first developed it in the 1920s.

x

Example of an $\bar{x}$ Control Chart

Example 21.1

- A manufacturing plant produces 10,000 plastic bottles per day.
- Because the product is inexpensive and the plant operation is normally satisfactory, it is not economically feasible to inspect every bottle.
- Instead, a sample of *n* bottles is randomly selected and inspected each day. These *n* items are called a *subgroup*, and *n* is referred to as the *subgroup size*.
- The inspection includes measuring the toughness of *x* of each bottle in the subgroup and calculating the sample mean $\bar{x}$.

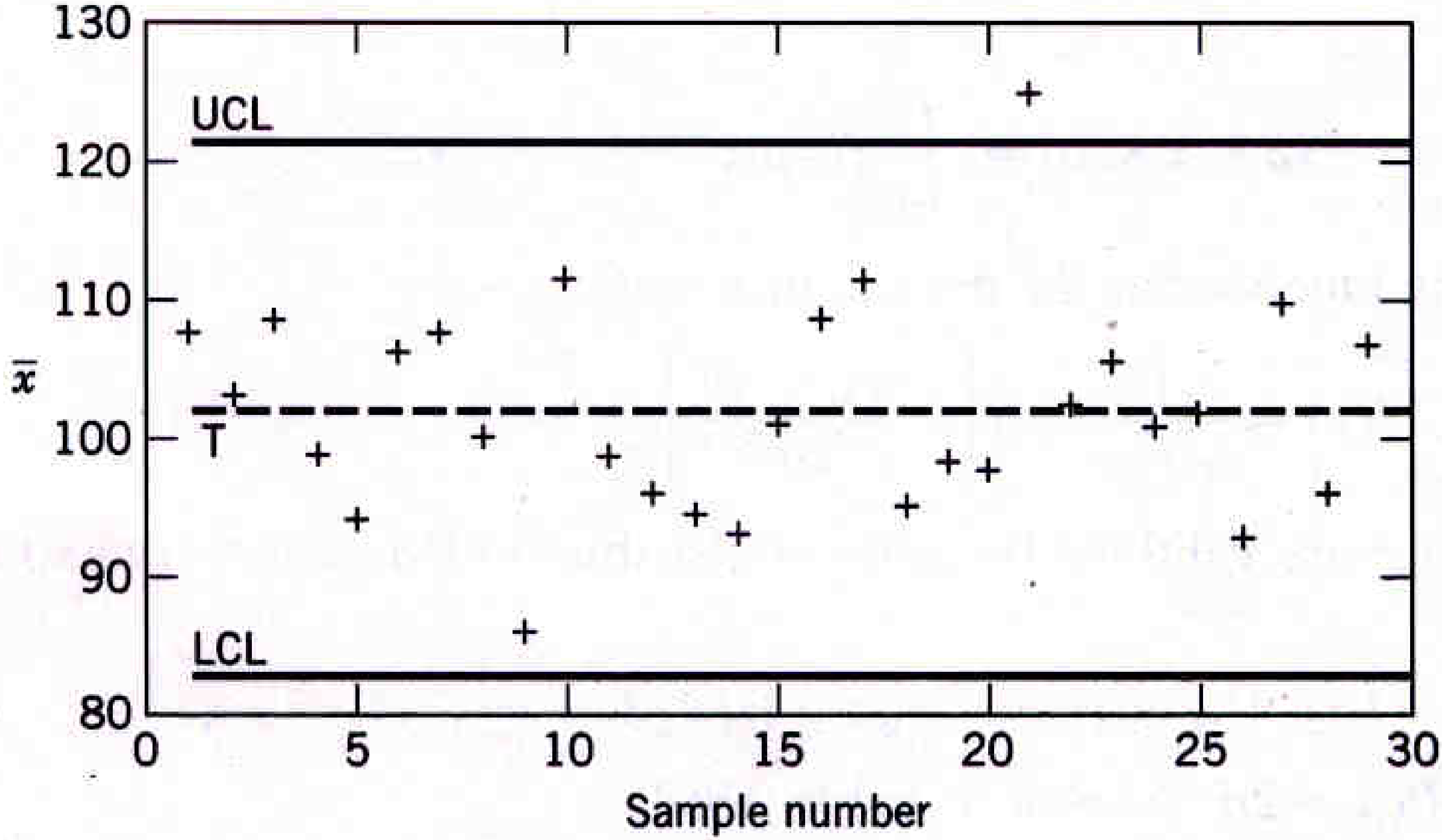


Figure 21.4 The $\bar{x}$ control chart for Example 21.1.

Chapter 16

- The $\overline{x}$ control chart in Fig. 21.4 displays data for a 30-day period. The control chart has a *target* (*T*), an *upper control limit* (*UCL*), and a *lower control limit* (*LCL*).
- The target (or *centerline*) is the desired (or *expected*) value for $\overline{x}$ while the region between UCL and LCL defines the range of normal variability, as discussed below.
- If all of the $\overline{x}$ data are within the control limits, the process operation is considered to be normal or "in a state of control". Data points outside the control limits are considered to be abnormal, indicating that the process operation is out of control.
- This situation occurs for the twenty-first sample. A single measurement located slightly beyond a control limit is not necessarily a cause for concern.
- But frequent or large chart violations should be investigated to determine a special cause.

Control Chart Development

- The first step in devising a control chart is to select a set of representative data for a period of time when the process operation is believed to be normal, that is, when the process is in a state of control.
- Suppose that these test data consist of *N* subgroups that have been collected on a regular basis (for example, hourly or daily) and that each subgroup consists of *n* randomly selected items.
- Let x_{ij} denote the *j-th* measurement in the *i-th* subgroup. Then, the subgroup sample means can be calculated:

$$\overline{x}_i \triangleq \frac{1}{n}\sum_{j=1}^{n} x_{ij} \quad (i = 1,2,\ldots, N) \qquad (21\text{-}7)$$

The *grand mean* $\bar{\bar{x}}$ is defined to be the average of the subgroup means:

$$\bar{\bar{x}} \triangleq \frac{1}{N} \sum_{i=1}^{N} \bar{x}_i \tag{21-8}$$

The general expressions for the control limits are

$$UCL \triangleq T + c\hat{\sigma}_{\bar{x}} \tag{21-9}$$

$$LCL \triangleq T - c\hat{\sigma}_{\bar{x}} \tag{21-10}$$

where $\hat{\sigma}_{\bar{x}}$ is an estimate of the standard deviation for $\bar{x}$ and c is a positive integer; typically, $c = 3$.

- The choice of $c = 3$ and Eq. 21-6 imply that the measurements will lie within the control chart limits 99.73% of the time, for normal process operation.
- The target T is usually specified to be either $\bar{\bar{x}}$ or the desired value of $\bar{x}$.

- The estimated standard deviation $\hat{\sigma}_{\overline{x}}$ can be calculated from the subgroups in the test data by two methods: (1) the standard deviation approach, and (2) the range approach.
- By definition, the *range R* is the difference between the maximum and minimum values.
- Consequently, we will only consider the standard deviation approach.

The average sample standard deviation $\overline{s}$ for the N subgroups is:

$$\overline{s} \triangleq \frac{1}{N} \sum_{i=1}^{N} s_i \qquad (21\text{-}11)$$

Chapter 16

where the standard deviation for the *ith* subgroup is

$$s_i \triangleq \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(x_{ij}-\overline{x}_i\right)^2} \qquad (21\text{-}12)$$

If the x data are normally distributed, then $\hat{\sigma}_{\overline{x}}$ is related to $\overline{s}$ by

$$\hat{\sigma}_{\overline{x}} = \frac{1}{c_4\sqrt{n}}\overline{s} \qquad (21\text{-}13)$$

where c_4 is a constant that depends on n and is tabulated in Table X, pg. 673 of the text.

Table X Factors for Constructing Variables Control Charts

	Factor for Control Limits						
	$\overline{X}$ Chart			R Chart		S Chart	
n^*	A_1	A_2	d_2	D_3	D_4	c_4	n
2	3.760	1.880	1.128	0	3.267	0.7979	2
3	2.394	1.023	1.693	0	2.575	0.8862	3
4	1.880	.729	2.059	0	2.282	0.9213	4
5	1.596	.577	2.326	0	2.115	0.9400	5
6	1.410	.483	2.534	0	2.004	0.9515	6
7	1.277	.419	2.704	.076	1.924	0.9594	7
8	1.175	.373	2.847	.136	1.864	0.9650	8
9	1.094	.337	2.970	.184	1.816	0.9693	9
10	1.028	.308	3.078	.223	1.777	0.9727	10
11	.973	.285	3.173	.256	1.744	0.9754	11
12	.925	.266	3.258	.284	1.716	0.9776	12
13	.884	.249	3.336	.308	1.692	0.9794	13
14	.848	.235	3.407	.329	1.671	0.9810	14
15	.816	.223	3.472	.348	1.652	0.9823	15
16	.788	.212	3.532	.364	1.636	0.9835	16
17	.762	.203	3.588	.379	1.621	0.9845	17
18	.738	.194	3.640	.392	1.608	0.9854	18
19	.717	.187	3.689	.404	1.596	0.9862	19
20	.697	.180	3.735	.414	1.586	0.9869	20
21	.679	.173	3.778	.425	1.575	0.9876	21
22	.662	.167	3.819	.434	1.566	0.9882	22
23	.647	.162	3.858	.443	1.557	0.9887	23
24	.632	.157	3.895	.452	1.548	0.9892	24
25	.619	.153	3.931	.459	1.541	0.9896	25

*$n > 25$: $A_1 = 3/\sqrt{n}$ where n = number of observations in sample.

The *s* Control Chart

- In addition to monitoring average process performance, it is also advantageous to monitor process variability.
- The variability within a subgroup can be characterized by its range, standard deviation, or sample variance.
- Control charts can be developed for all three statistics but our discussion will be limited to the control chart for the standard deviation, the *s control chart*.
- The centerline for the *s* chart is $\bar{s}$, which is the average standard deviation for the test set of data. The control limits are

$$UCL = \bar{s} + 3\frac{\bar{s}}{c_4}\sqrt{1-c_4^2}$$

$$LCL = \bar{s} - 3\frac{\bar{s}}{c_4}\sqrt{1-c_4^2} \qquad (16\text{-}16)$$

Example 21.2

In semiconductor processing, the photolithography process is used to transfer the circuit design to silicon wafers. In the first step of the process, a specified amount of a polymer solution, *photoresist*, is applied to a wafer as it spins at high speed on a turntable. The resulting photoresist thickness x is a key process variable. Thickness data for 25 subgroups are shown in Table 21.2. Each subgroup consists of three randomly selected wafers. Construct $\bar{x}$ and s control charts for these test data and critcially evaluate the results.

Solution

The following sample statistics can be calculated from the data in Table 21.2: $\bar{\bar{x}} = 199.8$ Å, $\bar{s} = 10.4$ Å. For $n = 3$ the required constant from Table X is $c_4 = 0.8862$. Then the $\bar{x}$ and s control limits can be calculated from Eqs. 21-9 to 21-13 and 16-16.

The traditional value of $c = 3$ is selected for Eqs. (21-9) and (21-10). The resulting control limits are labeled as the "original limits" in Fig. 21.5.

Figure 21.5 indicates that sample #5 lies beyond both the UCL for both the $\overline{x}$ and s control charts, while sample #15 is very close to a control limit on each chart. Thus, the question arises whether these two samples are "outliers" that should be omitted from the analysis. Table 21.2 indicates that sample #5 includes a very large value (260.0), while sample #15 includes a very small value (150.0). However, unusually large or small numerical values by themselves do not justify discarding samples; further investigation is required.

Suppose that a more detailed evaluation has discovered a specific reason as to why measurements #5 and #15 should be discarded (e.g., faulty sensor, data misreported, etc.). In this situation, these two samples should be removed and control limits should be recalculated based on the remaining 23 samples.

These modified control limits are tabulated below as well as in Fig. 21.5.

	Original Limits	Modified Limits (omit samples #5 and #15)
$\overline{x}$ Chart Control Limits		
UCL	220.1	216.7
LCL	179.6	182.2
***s* Chart Control Limits**		
UCL	26.6	22.7
LCL	0	0

Table 21.2 Thickness Data (in Å) for Example 21.2

No.	x Data			$\overline{x}$	s	No.	x Data			$\overline{x}$	s
1	209.6	20.76	211.1	209.4	1.8	14	202.9	210.1	208.1	207.1	3.7
2	183.5	193.1	202.4	193.0	9.5	15	198.6	195.2	150.0	181.3	27.1
3	190.1	206.8	201.6	199.5	8.6	16	188.7	200.7	207.6	199.0	9.6
4	206.9	189.3	204.1	200.1	9.4	17	197.1	204.0	182.9	194.6	10.8
5	260.0	209.0	212.2	227.1	28.6	18	194.2	211.2	215.4	206.9	11.0
6	193.9	178.8	214.5	195.7	17.9	19	191.0	206.2	183.9	193.7	11.4
7	206.9	202.8	189.7	199.8	9.0	20	202.5	197.1	211.1	203.6	7.0
8	200.2	192.7	202.1	198.3	5.0	21	185.1	186.3	188.9	186.8	1.9
9	210.6	192.3	205.9	202.9	9.5	22	203.1	193.1	203.9	200.0	6.0
10	186.6	201.5	197.4	195.2	7.7	23	179.7	203.3	209.7	197.6	15.8
11	204.8	196.6	225.0	208.8	14.6	24	205.3	190.0	208.2	201.2	9.8
12	183.7	209.7	208.6	200.6	14.7	25	203.4	202.9	200.4	202.2	1.6
13	185.6	198.9	191.5	192.0	6.7						

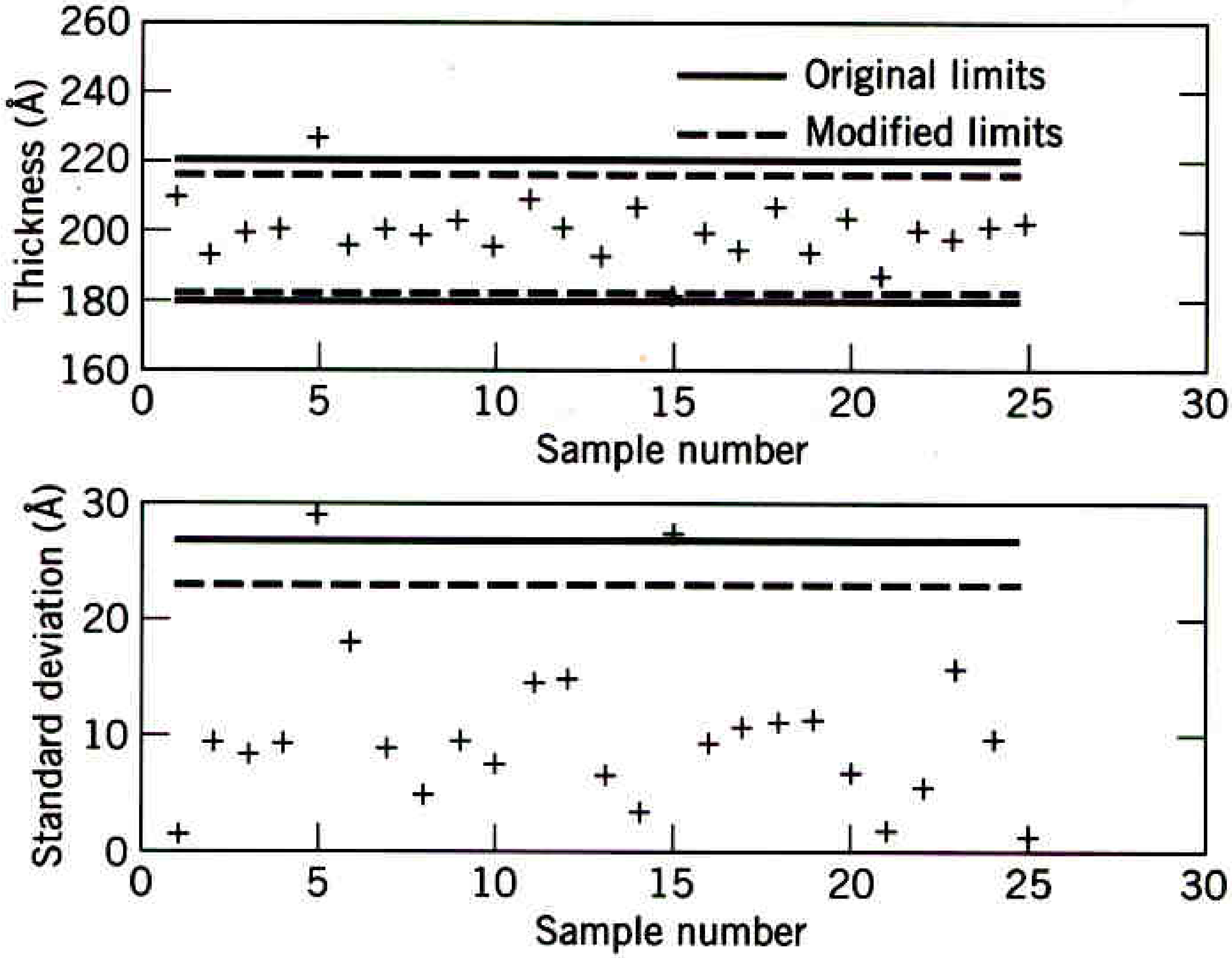


Figure 21.5 The $\overline{x}$ and s control charts for Example 21.2.

Theoretical Basis for Quality Control Charts

The traditional SQC methodology is based on the assumption that the natural variability for "in control" conditions can be characterized by random variations around a constant average value,

$$x(k) = x^* + \varepsilon(k) \qquad (21\text{-}16)$$

where $x(k)$ is the measurement at time k, x^* is the true (but unknown) value, and $\varepsilon(k)$ is an additive error. Traditional control charts are based on the following assumptions:

1. Each additive error, $\{\varepsilon(k),\ k = 1, 2, \ldots\}$, is a zero mean, random variable that has the same normal distribution, $N\left(0, \sigma^2\right)$.
2. The additive errors are statistically independent and thus uncorrelated. Consequently, $\varepsilon(k)$ does not depend on $\varepsilon(j)$ for $j \neq k$.

Chapter 16

3. The true value of x^* is constant.

4. The subgroup size n is the same for all of the subgroups.

The second assumption is referred to as the *independent, identically, distributed* (*IID*) assumption. Consider an individuals control chart for x with x^* as its target and "3σ control limits":

$$UCL \triangleq x^* + 3\sigma \qquad (21\text{-}17)$$

$$LCL \triangleq x^* - 3\sigma \qquad (21\text{-}18)$$

- These control limits are a special case of Eqs. 21-9 and 21.10 for the idealized situation where σ is known, $c = 3$, and the subgroup size is $n = 1$.
- The typical choice of $c = 3$ can be justified as follows.
- Because x is $N\left(0,\sigma^2\right)$, the probability p that a measurement lies outside the 3σ control limits can be calculated from Eq. 21-6: $p = 1 - 0.9973 = 0.0027$.

- Thus on average, approximately 3 out of every 1000 measurements will be outside of the 3σ limits.
- The average number of samples before a chart violation occurs is referred to as the *average run length* (ARL).
- For the normal ("in control") process operation,

$$\text{ARL} \triangleq \frac{1}{p} = \frac{1}{0.0027} = 370 \qquad (21\text{-}19)$$

- Thus, a Shewhart chart with 3σ control limits will have an average of one control chart violation every 370 samples, even when the process is *in a state of control*.
- Industrial plant measurements are not normally distributed.
- However, for large subgroup sizes ($n > 25$), $\bar{x}$ is approximately normally distributed even if x is not, according to the famous *Central Limit Theorem*.

- Fortunately, modest deviations from "normality" can be tolerated.
- In industrial applications, the control chart data are often *serially correlated* because the current measurement is related to previous measurements.
- Standard control charts such as the $\overline{x}$ and s charts can provide misleading results if the data are serially correlated.
- But if the degree of correlation is known, the control limits can be adjusted accordingly (Montgomery, 2001).

Pattern Tests and the Western Electric Rules

- We have considered how abnormal process behavior can be detected by comparing individual measurements with the $\overline{x}$ and s control chart limits.
- However, the pattern of measurements can also provide useful information.

- A wide variety of pattern tests (also called *zone rules*) can be developed based on the IID and normal distribution assumptions and the properties of the normal distribution.

- For example, the following excerpts from the *Western Electric Rules* indicate that the process is *out of control* if one or more of the following conditions occur:

1. One data point is outside the 3σ control limits.
2. Two out of three consecutive data points are beyond a 2σ limit.
3. Four out of five consecutive data points are beyond a 1σ limit and on one side of the center line.
4. Eight consecutive points are on one side of the center line.

- Pattern tests can be used to augment Shewhart charts.

- Although Shewhart charts with 3σ limits can quickly detect large process changes, they are ineffective for small, sustained process changes (for example, changes smaller than 1.5σ)
- An alternative control chart has been developed to detect small changes, the CUSUM control chart.
- They also can detect large process changes (for example, 3σ shifts), but detection is usually somewhat slower than for Shewhart charts.

CUSUM Control Chart

- The *cumulative sum* (*CUSUM*) is defined to be a running summation of the deviations of the plotted variable from its target.
- If the sample mean is plotted, the cumulative sum, $C(k)$, is

$$C(k)=\sum_{j=1}^{k}\left(\overline{x}(j)-T\right) \qquad (21\text{-}20)$$

where T is the target for $\bar{x}$.

- During normal process operation, $C(k)$ fluctuates around zero.
- But if a process change causes a small shift in $\bar{x}$, $C(k)$ will drift either upward or downward.
- The CUSUM control chart was originally developed using a graphical approach based on *V-masks*.
- However, for computer calculations, it is more convenient to use an equivalent algebraic version that consists of two recursive equations,

$$C^{+}(k) = \max\left[0, \bar{x}(k) - (T + K) + C^{+}(k-1)\right] \quad (21\text{-}21)$$

$$C^{-}(k) = \max\left[0, (T - K) - \bar{x}(k) + C^{-}(k-1)\right] \quad (21\text{-}22)$$

where C^{+} and C^{-} denote the sums for the high and low directions and K is a constant, the *slack parameter*.

- The CUSUM calculations are initialized by setting $C^+(0) = C^-(0) = 0$.
- A deviation from the target that is larger than K increases either C^+ or C^-.
- A control limit violation occurs when either C^+ or C^- exceeds a specified control limit (or *threshold*), H.
- After a limit violation occurs, that sum is reset to zero or to a specified value.
- The selection of the threshold H can be based on considerations of average run length.
- Suppose that we want to detect whether the sample mean $\bar{x}$ has shifted from the target by a small amount, δ .
- The slack parameter K is usually specified as $K = 0.5\delta$.

- For the ideal situation where the normally distributed and IID assumptions are valid, ARL values have been tabulated for specified values of δ, K, and H (Ryan, 2000; Montgomery, 2001).

Table 21.3 Average Run Lengths for CUSUM Control Charts

Shift from Target (in multiples of $\sigma_{\overline{x}}$)	ARL for $H = 4\sigma_{\overline{x}}$	ARL for $H = 5\sigma_{\overline{x}}$
0	168.	465.
0.25	74.2	139.
0.50	26.6	38.0
0.75	13.3	17.0
1.00	8.38	10.4
2.00	3.34	4.01
3.00	2.19	2.57

Multivariate Statistical Techniques

- For common SPC monitoring problems, two or more quality variables are important, and they can be highly correlated.
- For example, ten or more quality variables are typically measured for synthetic fibers.
- For these situations, multivariable SPC techniques can offer significant advantages over the single-variable methods discussed in Section 21.2.
- In the statistics literature, these techniques are referred to as *multivariate methods*, while the standard Shewhart and CUSUM control charts are examples of *univariate methods*.
- Well-known multivariate monitoring methods:

  a. *Hotelling's T^2 statistic* – traditional method

  b. *Principal Component Analysis (PCA)* – better method

Example 21.5

- The effluent stream from a wastewater treatment process is monitored to make sure that two process variables, the biological oxidation demand (BOD) and the solids content, meet specifications.
- Representative data are shown in Table 21.4. Shewhart charts for the sample means are shown in parts (a) and (b) of Fig. 21.8.
- These univariate control charts indicate that the process appears to be in-control because no chart violations occur for either variable. However, the bivariate control chart in Fig. 21.8c indicates that the two variables are highly correlated because the solids content tends to be large when the BOD is large and vice versa.
- When the two variables are considered together, their joint confidence limit (for example, at the 99% confidence level) is an ellipse, as shown in Fig. 21.8c. Sample # 8 lies well beyond the 99% limit, indicating an out-of-control condition.

By contrast, this sample lies within the Shewhart control chart limits for both individual variables.

This example has demonstrated that univariate SPC techniques such as Shewhart charts can fail to detect abnormal process behavior when the process variables are highly correlated.

By contrast, the abnormal situation was readily apparent from the multivariate analysis.

Table 21.4 Wastewater Treatment Data

Sample Number	BOD (mg/L)	Solids (mg/L)	Sample Number	BOD (mg/L)	Solids (mg/L)
1	17.7	1380	16	16.8	1345
2	23.6	1458	17	13.8	1349
3	13.2	1322	18	19.4	1398
4	25.2	1448	19	24.7	1426
5	13.1	1334	20	16.8	1361
6	27.8	1485	21	14.9	1347
7	29.8	1503	22	27.6	1476
8	9.0	1540	23	26.1	1454
9	14.3	1341	24	20.0	1393
10	26.0	1448	25	22.9	1427
11	23.2	1426	26	22.4	1431
12	22.8	1417	27	19.6	1405
13	20.4	1384	28	31.5	1521
14	17.5	1380	29	19.9	1409
15	18.4	1396	30	20.3	1392

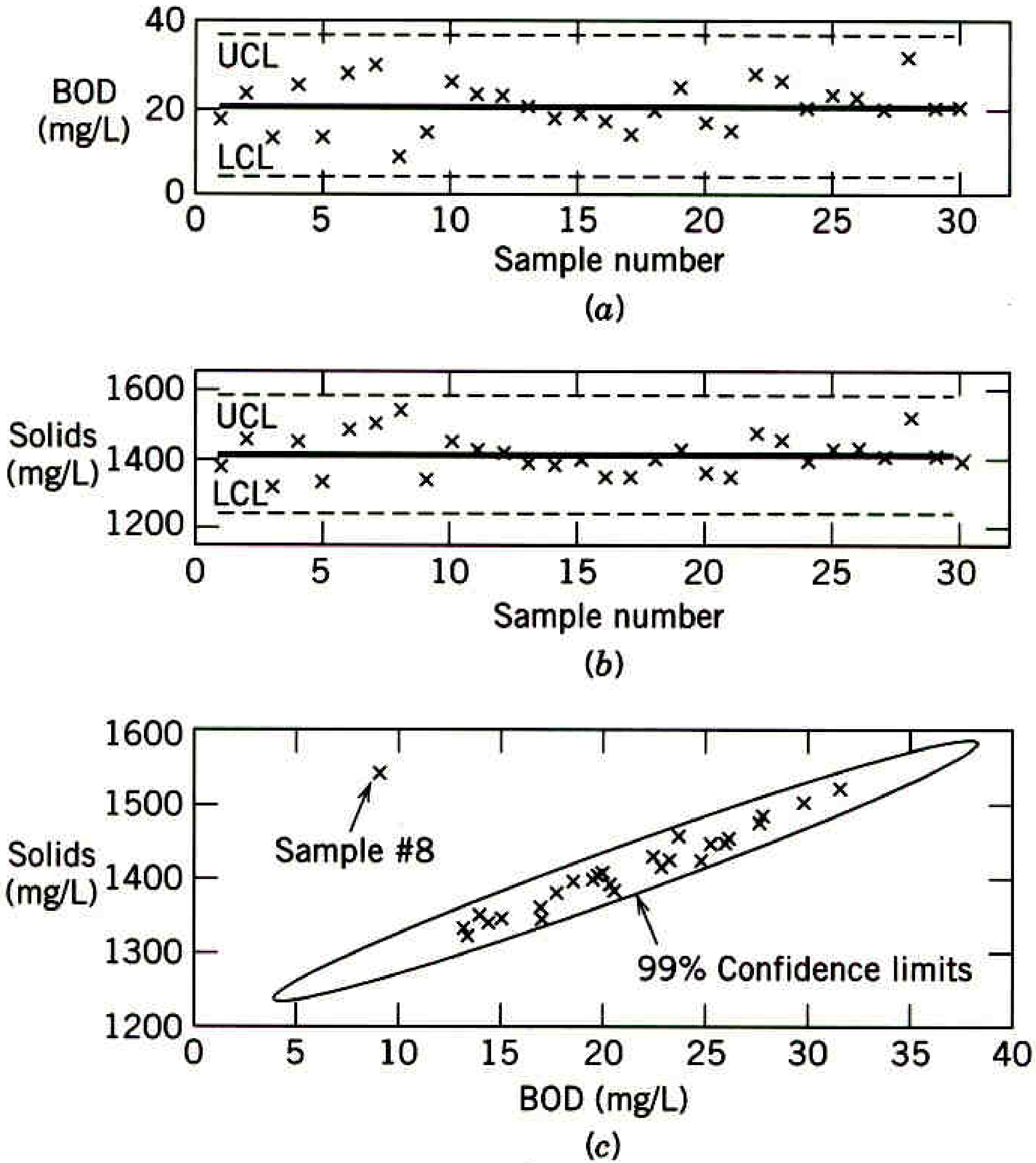


Figure 21.8 Confidence regions for Example 21.5 univariate (a) and (b), bivariate (c).

Hotelling's T^2 Statistic

- Suppose that it is desired to use SPC techniques to monitor *p* variables, which are correlated and normally distributed.
- Let $\mathbf{x}$ denote the column vector of these *p* variables,

$$\mathbf{x} = col\,[x_1, x_2, \ldots, x_p].$$

- At each sampling instant, a subgroup of *n* measurements is made for each variable.
- The subgroup sample means for the *k*th sampling instant can be expressed as a column vector:

$$\bar{x}(k) = col\left[\bar{x}_1(k), \bar{x}_2(k), \ldots, \bar{x}_p(k)\right]$$

- Multivariate control charts are traditionally based on *Hotelling's T^2 statistic*,

$$T^2(k) \triangleq n\left[\bar{x}(k) - \bar{\bar{x}}\right]^T S^{-1}\left[\bar{x}(k) - \bar{\bar{x}}\right] \qquad (21\text{-}27)$$

where $T^2(k)$ denotes the value of the T^2 statistic at the kth sampling instant.

- The vector of grand means $\bar{\bar{\mathbf{x}}}$ and the covariance matrix $\boldsymbol{S}$ are calculated for a test set of data for *in-control* conditions.
- By definition S_{ij}, the (i,j)-element of matrix $\boldsymbol{S}$, is the sample covariance of x_i and x_j:

$$S_{ij} \triangleq \frac{1}{N}\sum_{k=1}^{N}\left[\bar{x}_i(k)-\bar{\bar{x}}_i\right]^T\left[\bar{x}_j(k)-\bar{\bar{x}}_j\right] \qquad (21\text{-}28)$$

- In Eq. (21-28) N is the number of subgroups and $\bar{\bar{x}}_i$ denotes the grand mean for $\bar{x}_i$.
- Note that T^2 is a scalar, even though the other quantities in Eq. 21-27 are vectors and matrices.
- The inverse of the sample covariance matrix, $\boldsymbol{S}^{-1}$, scales the p variables and accounts for correlation among them.

- A multivariate process is considered to be out-of-control at the kth sampling instant if $T^2(k)$ exceeds an upper control limit, UCL.
- (There is no target or lower control limit.)

Example 21.6

Construct a T^2 control chart for the wastewater treatment problem of Example 21.5. The 99% control chart limit is $T^2 = 11.63$. Is the number of T^2 control chart violations consistent with the results of Example 21.5?

Solution

The T^2 control chart is shown in Fig. 21.10. All of the T^2 values lie below the 99% confidence limit except for sample #8. This result is consistent with the bivariate control chart in Fig. 21.8c.

Chapter 16

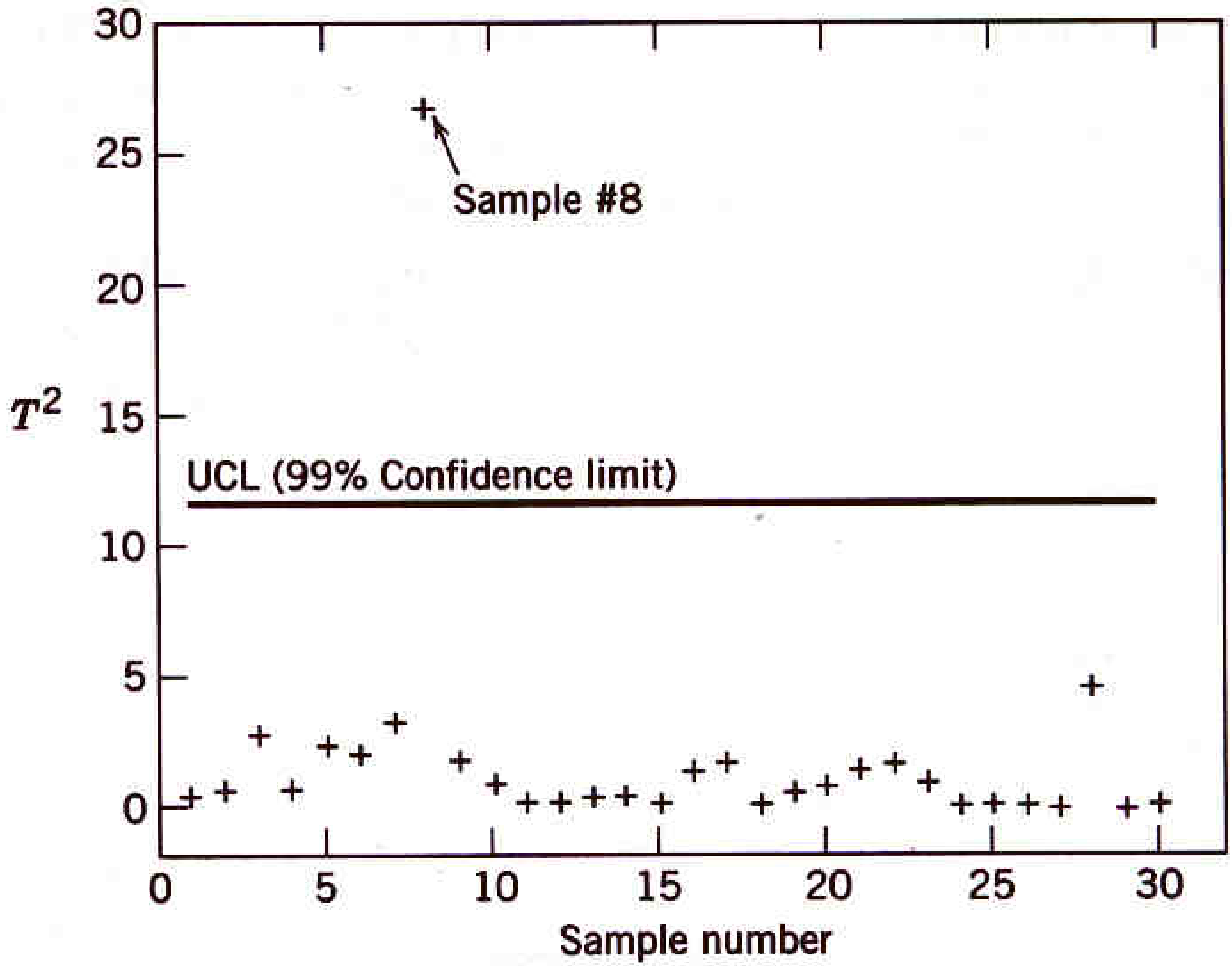


Figure 21.10 T^2 control chart for Example 21.5.