STATISTICAL METHODS IN PROCESS MONITORING AND CONTROL

Douglas C. Montgomery
Arizona State University

Cheryl L. Jennings
Motorola, Inc.

Introduction

The quality of products and services is an important factor in most businesses today. Regardless of whether the consumer is an individual, a corporation, a military defense program, or a retail store, when the consumer is making purchase decisions, he or she is likely to consider quality of equal importance to cost and on-time delivery. Consequently, quality improvement has become a major concern to many corporations. This chapter is about statistical methods that are useful in quality control and improvement.

There are various definitions of quality. The classical definition is that quality means fitness for use. For example, you or I may purchase automobiles that we expect to be free of manufacturing defects and that should provide reliable and economical transportation, a retailer buys finished goods with the expectation that they are properly packaged and arranged for easy storage and display, or a manufacturer buys raw material and expects to process it with no rework or scrap. In other words, all consumers expect that the products and services they buy will meet their requirements. Those requirements define fitness for use. Quality or fitness for use is determined through the interaction of quality of design and quality of conformance. By quality of design we mean the different grades or levels of performance, reliability, serviceability, and function that are the result of deliberate engineering and management decisions. By quality of conformance, we mean the systematic reduction of variability and elimination of defects until every unit produced is identical and defect-free. Because both of these activities involve reducing variability in the key parameters that define fitness for use, a more modern definition of quality is focused on reduction of unnecessary variability in these parameters.

Statistical methods play a vital role in quality improvement. Some applications are outlined below:

1. In product design and development, statistical methods, including designed experiments, can be used to compare different materials, components, or ingredients, and to help determine both system and component tolerances. This application can significantly lower development costs, reduce overall development time, and decrease time to market.
2. Statistical methods can be used to determine the capability of a manufacturing process. Statistical process control can be used to systematically improve the capability of a process by reducing variability.
3. Experimental design methods can be used to characterize and optimize processes. This can lead to higher yields and lower manufacturing costs.
4. Statistical methods can be used to characterize the performance of measurement systems, which can lead to more informed decisions about product disposition.

5. Life testing provides reliability and other performance data about the product. This can lead to new and improved designs and products that have longer useful lives and lower operating and maintenance costs.

In this chapter we provide an introduction to the basic methods of statistical quality control that, along with experimental design, form the basis of a successful quality-improvement effort.

Methods of Statistical Quality Control

The field of statistical quality control can be broadly defined as those statistical and engineering methods that are used in measuring, monitoring, controlling, and improving quality. Statistical quality control is a relatively new field, dating back to the 1920s. Dr. Walter A. Shewhart of the Bell Telephone Laboratories was one of the early pioneers of the field. In 1924 he wrote a memorandum showing a modern control chart, one of the basic tools of statistical process control. Dr. W. Edwards Deming and Dr. Joseph M. Juran were instrumental in spreading statistical quality-control methods in the last 50 years.

Much of the interest in statistical quality control and improvement focuses on statistical process monitoring and control and experimental design. Many companies have extensive programs to implement these methods in their manufacturing, engineering, and other business organizations. Online statistical process control is a powerful tool for achieving process stability and improving capability through the reduction of variability.

It is customary to think of statistical process control (SPC) as a set of problem- solving tools that may be applied to any process. The major tools of SPC are:

1. Histogram
2. Pareto chart
3. Cause-and-effect diagram
4. Defect-concentration diagram
5. Control charts
6. Scatter diagram
7. Check sheet

Some prefer to include the experimental design methods discussed in a previous chapter as part of the SPC toolkit. We did not do so, because we think of SPC as an online approach to quality improvement using techniques founded on passive observation of the process, while design of experiments is an active approach in which deliberate changes are made to the process variables. As such, designed experiments are often referred to as offline quality control. Although the complete set of tools are important to the implementation of SPC, in this chapter we focus on only control charts.

Introduction to Control Charts

Basic Principles

In any production process, regardless of how well-designed or carefully maintained it is, a certain amount of inherent or natural variability will always exist. This natural variability or "background noise" is the cumulative effect of many small, essentially unavoidable causes. When the background noise in a process is relatively small, we usually consider it an acceptable level of process performance. In the framework of statistical quality control, this natural variability is often called a "stable system of chance causes." A process that is operating with only chance causes of variation present (some authors use the terminology “common causes”) is said to be in statistical control. In other words, the chance causes are an inherent part of the process.

Other kinds of variability may occasionally be present in the output of a process. This variability in key quality characteristics usually arises from three sources: improperly adjusted machines, operator errors, or defective raw materials. Such variability is generally large when compared to the background noise, and it usually represents an unacceptable level of process performance. We refer to these sources of variability that are not part of the chance cause pattern as “assignable causes” (some authors use the terminology “special cause”). A process that is operating in the presence of assignable causes is said to be out of control.

Production processes will often operate in the in-control state, producing acceptable product for relatively long periods of time. Occasionally, however, assignable causes will occur, seemingly at random, resulting in a "shift" to an out-of-control state where a large proportion of the process output does not conform to requirements. A major objective of statistical process control is to quickly detect the occurrence of assignable causes or process shifts so that investigation of the process and corrective action may be undertaken before many nonconforming units are manufactured. The control chart is an online process- monitoring technique widely used for this purpose.

Control charts may also be used to estimate the parameters of a production process and, through this information, to determine the capability of a process to meet specifications. The control chart can also provide information that is useful in improving the process. Finally, remember, that the eventual goal of statistical process control is the reduction of unnecessary or harmful variability in the process. Although it may not be possible to eliminate variability completely, the control chart helps reduce it as much as possible.

A control chart is a graphical display of a quality characteristic that has been measured or computed from a sample versus the sample number or time. Often, the samples are selected at periodic intervals such as every hour. The chart contains a center line (CL) that represents the average value of the quality characteristic corresponding to the in-control state. (That is, only chance causes are present.) Two other horizontal lines, called the upper control limit (UCL) and the lower control limit (LCL), are also placed on the chart. These control limits are chosen so that if the process is in control, nearly all of the sample points will fall between them. In general, as long as the points plot within the control limits, the process is assumed to be in control, and no action is necessary. However, a point that plots outside of the control limits is interpreted as evidence that the process is out of control, and investigation and corrective action are required to find and eliminate the assignable cause or causes responsible for this behavior.

Even if all the points plot within the control limits, if they behave in a systematic or nonrandom manner, then this is an indication that the process is out of control. For example, if 18 of the last 20 points plotted above the center line but below the upper control limit and only two of these points plotted below the center line but above the lower control limit, we would be very suspicious that something was wrong. If the process is in control, the plotted points should have an essentially random pattern within the control limits. Tests designed to find sequences or nonrandom patterns can be applied to control charts as an aid in detecting out-of-control conditions. A particular nonrandom pattern usually appears on a control chart for a reason, and if that reason can be found and eliminated, process performance can be improved.

There is a close connection between control charts and hypothesis testing. Essentially, the control chart is a test of the hypothesis that the process is in a state of statistical control. A point plotting within the control limits is equivalent to failing to reject the hypothesis of statistical control, and a point plotting outside the control limits is equivalent to rejecting the hypothesis of statistical control.

We may give a general model for a control chart. Let W be a sample statistic that measures some quality characteristic of interest, and suppose that the mean of W is μ_W and the standard deviation of W is σ_W. Then the center line, the upper control limit, and the lower control limit become:

$$\begin{aligned} UCL &= \mu_W + k\sigma_W \\ CL &= \mu_W \\ LCL &= \mu_W - k\sigma_W \end{aligned} \tag{1}$$

where k is the "distance" of the control limits from the center line, expressed in standard deviation units. A common choice is $k = 3$. Dr. Walter A. Shewhart first proposed this general theory of control charts, and control charts developed according to these principles are often called Shewhart control charts.

The control chart is a device for describing exactly what is meant by statistical control; as such, it may be used in a variety of ways. In many applications, it is used for online process monitoring. That is, sample data are collected and used to construct the control chart, and if the sample values of the sample statistic W fall within the control limits and do not exhibit any systematic pattern, we say the process is in control at the level indicated by the chart. Note that "sigma" refers to the standard deviation of the statistic plotted on the chart and not the standard deviation of the quality characteristic. Note that we may be interested here in determining both whether the past data came from a process that was in control and whether future samples from this process indicate statistical control.

The most important use of a control chart is to improve the process by reducing variability. Generally, most processes do not operate in a state of statistical control so the routine and attentive use of control charts will identify assignable causes. If these causes can be eliminated from the process, variability will be reduced and the process will be improved. However, the control chart will only detect assignable causes. Management, operator, and engineering action will usually be necessary to eliminate the assignable cause. An out-of-control action plan for responding to control chart signals is vital. In identifying and eliminating assignable causes, it is important to find the underlying root cause of the problem and to attack it. A cosmetic solution will not result in any real, longterm process improvement. Developing an effective system for corrective action is an essential component of an effective SPC implementation.

We may also use the control chart as an estimating device. That is, from a control chart that exhibits statistical control, we may estimate certain process parameters, such as the mean, standard deviation, and fraction nonconforming or fallout. These estimates may then be used to determine the capability of the process to produce acceptable products. Process capability studies have considerable impact on many management decision problems that occur over the product cycle, including make-or-buy decisions, plant and process improvements that reduce process variability, and contractual agreements with customers or suppliers regarding product quality.

Control charts may be classified into two general types. Many quality characteristics can be measured and expressed as numbers on some continuous scale of measurement. In such cases, it is convenient to describe the quality characteristic with a measure of central tendency and a measure of variability. Control charts for central tendency and variability are collectively called variables control charts. The $\overline{X}$ control chart is the most widely used chart for monitoring central tendency, whereas charts based on either the sample range or the sample standard deviation are used to control process variability. Many quality characteristics are not measured on a continuous scale or even a quantitative scale. In these cases, we may judge each unit of product as either conforming or nonconforming on the basis of whether or not it possesses certain attributes, or we may count the number of nonconformities (defects) appearing on a unit of product. Control charts for such quality characteristics are called attributes control charts.

Design of a Control Chart

To illustrate these ideas, we give a simplified example of a control chart. In the manufacture of automobile engine piston rings, the inside diameter of the rings is a critical quality characteristic. The process mean inside ring diameter is 74 mm, and it is known that the standard deviation of ring diameter is 0.01 mm. Every hour a random sample of five rings is taken, the average ring diameter of the sample is computed, and is plotted on the chart. Because this control chart utilizes the sample average or sample mean to monitor the process mean, it is called an $\overline{X}$ control chart.

Consider how the control limits are determined. Since the process average is 74 mm, and the process standard deviation is 0.01 mm if samples of size n = 5 are taken, the standard deviation of the sample average $\overline{X}$ is $\sigma_{\overline{X}} = \sigma\sqrt{n} = 0.01/\sqrt{5} = 0.0045$. Therefore, if the process is in control with a mean diameter of 74 mm, by using the central limit theorem to assume that $\overline{X}$ is approximately normally distributed, we would expect approximately 99.73 percent of the sample mean diameters $\overline{X}$ to fall between 74 + 3(0.0045) and 74 - 3(0.0045). Therefore, the upper and lower control limits become UCL = 74.0135 and LCL = 73.9865. These are the 3-sigma control limits referred to earlier.

Note that the use of 3-sigma limits implies that the type I error level for the control chart is approximately 0.0027; that is, the probability that the point plots outside the control limits when the process is in control is 0.0027. The width of the control limits is inversely related to the sample size n for a given multiple of sigma. Choosing the control limits is equivalent to setting up the critical region for testing the null hypothesis that the process mean is equal to 74 versus a two-sided alternative. Essentially, the control chart tests this hypothesis repeatedly at different points in time.

In designing a control chart, we must specify both the sample size to use and the frequency of sampling. In general, larger samples will make it easier to detect small shifts in the process. When choosing the sample size, we must keep in mind the size of the shift that we are trying to detect. If we are interested in detecting a relatively large process shift, then we use smaller sample sizes than those that would be employed if the shift of interest were relatively small. We must also determine the frequency of sampling. The most desirable situation from the point of view of detecting shifts would be to take large samples very frequently; however, this is usually not economically feasible. The general problem is one of allocating sampling effort. That is, either we take small samples at short intervals or larger samples at longer intervals. Current industry practice tends to favor smaller, more frequent samples, particularly in high-volume manufacturing processes, or where a great many types of assignable causes can occur. Furthermore, as automatic sensing and measurement technology develops, it is becoming possible to greatly increase frequencies. Ultimately, every unit can be tested as it is manufactured. This capability will not eliminate the need for control charts because the test system will not prevent defects. The increased data will increase the effectiveness of process control and improve quality.

Rational Subgroups

A fundamental idea in the use of control charts is to collect sample data according to what Shewhart called the rational subgroup concept. Generally, rational subgroups or samples should be selected so that to the extent possible, the variability of the observations within a subgroup should include all the chance or natural variability and exclude the assignable variability. Then, the control limits will represent bounds for all the chance variability and not the assignable variability. Consequently, assignable causes will tend to generate points that are outside of the control limits, while chance variability will tend to generate points that are within the control limits.

When control charts are applied to production processes, the time order of production is a logical basis for rational subgrouping. Even though time order is preserved, it is still possible to form subgroups erroneously. If some of the observations in the subgroup are taken at the end of one 8-hour shift and the remaining observations are taken at the start of the next 8-hour shift, then any differences between shifts might not be detected. Time order is frequently a good basis for forming subgroups because it allows us to detect assignable causes that occur over time.

Two general approaches to constructing rational subgroups are used. In the first approach, each subgroup consists of units that were produced at the same time (or as closely together as possible). This approach is used when the primary purpose of the control chart is to detect process shifts. It minimizes variability due to assignable causes within a sample, and it maximizes variability between samples if assignable causes are present. It also provides better estimates of the standard deviation of the process in the case of variables control charts. This approach to rational subgrouping essentially gives a "snapshot" of the process at each point in time where a sample is collected.

In the second approach, each sample consists of units of product that are representative of all units that have been produced since the last sample was taken. Essentially, each subgroup is a random sample of all process output over the sampling interval. This method of rational subgrouping can be used when the control chart is employed to make decisions about the acceptance of all units of product that have been produced since the last sample. In fact, if the process shifts to an out-of-control state and then back in control again between samples, the first method of rational subgrouping defined above can be ineffective against these types of shifts, and so the second method must be used.

When the rational subgroup is a random sample of all units produced over the sampling interval, considerable care must be taken in interpreting the control charts. If the process mean drifts between several different levels during the interval between samples, the range of the observations within a sample may be relatively large. It is the within- sample variability that determines the width of the control limits on an $\bar{X}$ chart, so this practice will result in wider limits on the chart. This makes it harder to detect shifts in the mean. In fact, we can often make any process appear to be in statistical control just by stretching out the interval between observations in the sample. It is also possible for shifts in the process average to cause points on a control chart for the range or standard deviation to plot out of control, even though no shift in process variability has taken place.

There are other bases for forming rational subgroups. For example, suppose a process consists of several machines that pool their output into a common stream. If we sample from this common stream of output, it will be very difficult to detect whether or not some of the machines are out of control. A logical approach to rational subgrouping in this scenario is to apply control chart techniques to the output for each individual machine. Sometimes this concept needs to be applied to different heads on the same machine, different workstations, different operators, and so forth.

'The rational subgroup concept is very important. The proper selection of samples requires careful consideration of the process, with the objective of obtaining as much useful information as possible from the control chart analysis.

Analysis of Patterns on Control Charts

A control chart may indicate an out-of-control condition when one or more points fall beyond the control limits, or when the plotted points exhibit some nonrandom pattern of behavior. If the points plotted on the chart are truly random, we should expect a relatively even distribution of them above and below the center line. A type of nonrandom pattern often observed on a control chart is a run. In general, we define a run as a sequence of observations of the same type. In addition to runs up and runs down, we could define the types of observations as those above and below the center line, respectively, so that two points in a row above the center line would be a run of length 2. A run of length 8 or more points has a very low probability of occurrence in a random sample of points. Consequently, any type of run of length 8 or more is often taken as an indication that some type of disturbance or assignable cause is present.

The problem is one of pattern recognition, that is, recognizing systematic or nonrandom patterns on the control chart and identifying the reason for this behavior. The ability to interpret a particular pattern in terms of assignable causes requires experience and knowledge of the process. That is, we must not only know the statistical principles of control charts, but we must also have a good understanding of the process.

The Western Electric Handbook (1956) suggests a set of decision rules for detecting nonrandom patterns on control charts. Specifically, the Western Electric rules would conclude that the process is out of control if either

1. One point plots outside 3-sigma control limits.
2. Two out of three consecutive points plot beyond a 2-sigma limit.
3. Four out of five consecutive points plot at a distance of 1-sigma or beyond from the center line.
4. Eight consecutive points plot on one side of the center line.

These rules very effective for enhancing the sensitivity of control charts. Rules 2 and 3 apply to one side of the center line at a time. That is, a point above the upper 2-sigma limit followed immediately by a point below the lower 2- sigma limit would not generate an out-of-control signal. Because the 1-, 2-, and 3-sigma bands form zones on the control chart, the Western Electric rules are sometimes called the zone rules.

Sensitizing rules such as the Western Electric rules for control charts are not without controversy. While they do increase sensitivity to small shifts in process parameters, they also greatly increase the risk of a false alarm. When control charts are used in the initial stage of an SPC implementation and the processes are almost always out of control, this is less of an issue than when control charts are used for monitoring a relatively stable process. When the process is operating in control, sensitizing rules can result in many unnecessary false alarms and can be disruptive to both manufacturing and other ongoing process improvements efforts. Any application of supplemental rules should be accompanied by specific out of control action plans indicating how to respond to specific rule violations. Many computer software packages for SPC turn the Western Electric rules (and others as well) on routinely, regardless of the application situation. For more information on sensitizing rules and control charts, see Montgomery (2001).

$\bar{X}$ and *R* Control Charts

When dealing with a quality characteristic that can be expressed as a measurement, it is customary to monitor both the mean value of the quality characteristic and its variability. Control over the average quality is maintained by the control chart for averages, usually called the $\bar{X}$ chart. Process variability can be controlled by either a range chart (*R* chart) or a standard deviation chart (*S* chart), depending on how the population standard deviation is estimated. We will discuss only the *R* chart.

In general, the process mean and standard deviation are unknown and are estimated on the basis of preliminary samples, taken when the process is thought to be in control. We recommend the use of at least 20 to 25 preliminary samples. Suppose *m* preliminary samples are available, each of size *n*. Typically, *n* will be 4, 5, or 6; these relatively small sample sizes are widely used in practice and often arise from the construction of rational subgroups. Let the sample mean for the ith sample be $\bar{X}_i$. Then we estimate the mean of the population by the grand mean

$$\bar{\bar{X}} = \sum_{i=1}^{m} \bar{X}_i \tag{2}$$

Thus, we may take $\bar{\bar{X}}$ as the center line on the $\bar{X}$ control chart.

We may estimate the standard deviation σ from either the sample standard deviation or the range of the observations within each sample. Since it is more frequently used in practice, we confine our discussion to the range method. The sample size is relatively small, so there is little loss in efficiency in estimating a from the sample ranges. The relationship between the range R of a sample from a normal population with known parameters and the standard deviation of that population is needed. Since R is a random variable, the quantity $W = R/\sigma$, called the relative range, is also a random variable. The parameters of the distribution of W have been determined for any sample size n. The mean of the distribution of W is called d_2, and the standard deviation of W is called d_3. Therefore,

$$\mu_R = d_2\sigma$$
$$\sigma_R = d_3\sigma$$

Let R_i be the range of the ith sample and let

$$\bar{R}_i = \sum_{i=1}^{m} R_i \tag{3}$$

be the average range. Thus $\bar{R}$ is an estimate of μ_R and an estimate of the process standard deviation is

$$\hat{\sigma} = \frac{\bar{R}}{d_2} \tag{4}$$

Therefore, the upper and lower control limits and the center line of an $\bar{X}$ control chart are

$$\begin{aligned} UCL &= \bar{\bar{X}} + \frac{3}{d_2\sqrt{n}}\bar{R} = \bar{\bar{X}} + A_2\bar{R} \\ CL &= \bar{\bar{X}} \\ LCL &= \bar{\bar{X}} - \frac{3}{d_2\sqrt{n}}\bar{R} = \bar{\bar{X}} - A_2\bar{R} \end{aligned} \tag{5}$$

The constant A_2 is tabulated in many statistical quality control texts; for example, see Montgomery (2001).

The parameters of the R chart may be easily determined. Since

$$\hat{\sigma}_R = d_3\sigma = d_3\frac{\bar{R}}{d_2} \tag{6}$$

we may write the R chart parameters as

$$
\begin{aligned}
UCL &= \bar{R} + \frac{3d_3}{d_2}\bar{R} = \bar{R}\left(1 + \frac{3d_3}{d_2}\right) = D_4\bar{R} \\
CL &= \bar{R} \\
LCL &= \bar{R} - \frac{3d_3}{d_2}\bar{R} = \bar{R}\left(1 - \frac{3d_3}{d_2}\right) = D_3\bar{R}
\end{aligned} \tag{7}
$$

The constant D_3 and D_4 are tabulated in many statistical quality control texts; for example, see Montgomery (2001).

When preliminary samples are used to construct limits for control charts, these limits are customary treated as trial values. Therefore, the *m* sample means and ranges should be plotted on the appropriate charts, and any points that exceed the control limits should be investigated. If assignable causes for these points are discovered, they should be eliminated and new limits for the control charts determined. In this way, the process may be eventually brought into statistical control and its inherent capabilities assessed. Other changes in process centering and dispersion may then be contemplated. Also, we often study the *R* chart first because if the process variability is not constant over time the control limits calculated for the $\bar{X}$ chart can be misleading.

An Example of $\bar{X}$ and *R* Control Charts

A component part for a jet aircraft engine is manufactured by an investment casting process. The vane opening on this casting is an important functional parameter of the part. We will illustrate the use of $\bar{X}$ and *R* control charts to assess the statistical stability of this process. Table 1 presents 20 samples of five parts each. The values given in the table have been coded by using the last three digits of the dimension; that is, 31.6 should be 0.50316 inch. The quantities $\bar{\bar{X}}$ = 33.3 and $\bar{R}$ = 5.8 are shown at the foot of Table 1. The value of A_2 for samples of size 5 is 0.577. Then the trial control limits for the $\bar{\bar{X}}$ chart are

$$
\begin{aligned}
\bar{\bar{X}} \pm A_2\bar{R} &= 33.32 \pm 0.577(5.8) = 33.32 \pm 3.35 \\
UCL &= 36.67 \\
LCL &= 29.97
\end{aligned}
$$

For the *R* chart, the trial control limits are

$$
\begin{aligned}
UCL &= D_4\bar{R} = (2.115)5.8 = 12.27 \\
LCL &= D_3\bar{R} = (0)5.8 = 0
\end{aligned}
$$

The control charts, which were generated by Minitab, are shown in Figure 1. Notice that samples 6, 8, 11, and 19 are out of control on the $\bar{X}$ chart and sample 9 is out of control on the *R* chart. If assignable causes can be determined for these points, then we can discard the data from these five samples and recalculate or revise the control limits. Suppose that these assignable causes can be found. The new control charts with the revised limits computed without samples 6, 8, 9, 11, and 19 are shown in Figure 2.

Table 1 Vane Opening Measurements

x_1	x_2	x_3	x_4	x_5	Average	Range
33	29	31	32	33	31.6	4
33	31	35	37	31	33.4	6
35	37	33	34	36	35.0	4
30	31	33	34	33	32.2	4
33	34	35	33	34	33.8	2
38	37	39	40	38	38.4	3
30	31	32	34	31	31.6	4
29	39	38	39	39	36.8	10
28	33	35	36	43	35.0	15
38	33	32	35	32	34.0	6
28	30	28	32	31	29.8	4
31	35	35	35	34	34.0	4
27	32	34	35	37	33.0	10
33	33	35	37	36	34.8	4
35	37	32	35	39	35.6	7
33	33	27	31	30	30.8	6
35	34	34	30	32	33.0	5
32	33	30	30	33	31.6	3
25	27	34	27	28	28.2	9
35	35	36	33	30	33.8	6
					$\bar{\bar{X}} = 33.32$	$\bar{R} = 5.8$

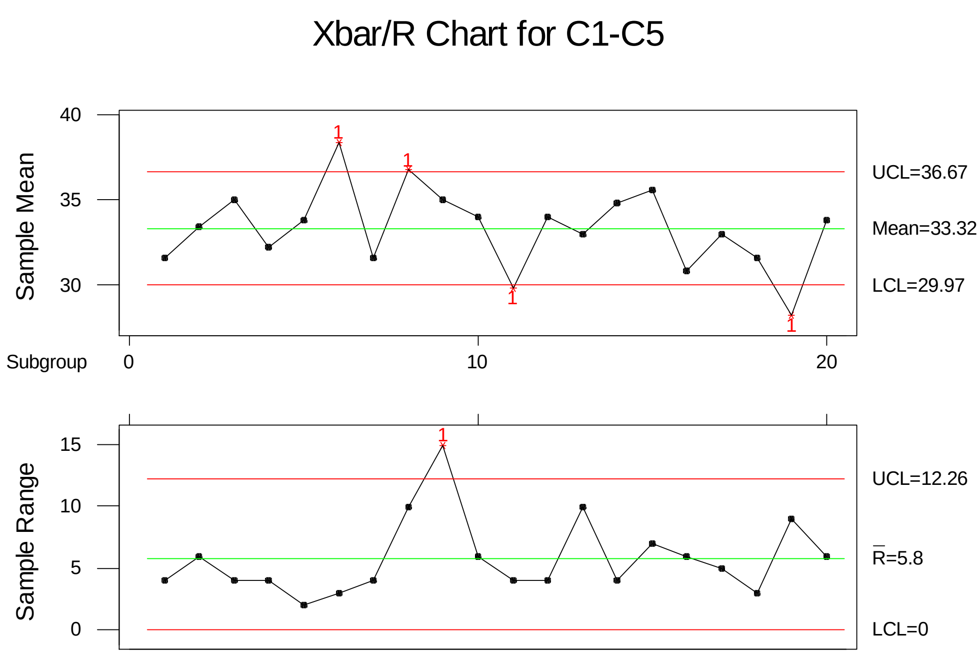

Figure 1. Minitab $\bar{X}$ and R Control Charts for the Vane Opening Data

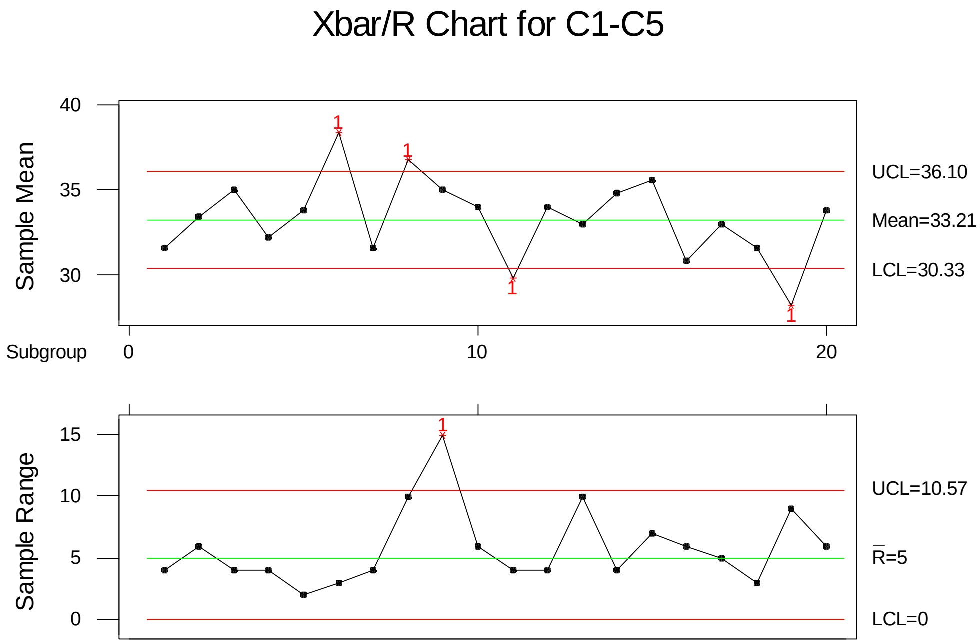

Figure 2. Minitab $\bar{X}$ and R Control Charts for the Vane Opening Data with Revised Control Limits

Notice that in the revised control chart in Figure 2 we have treated the first 20 samples as estimation data with which to establish a set of control limits that we hope will be appropriate for monitoring future production. As each new sample becomes available, the values of $\bar{X}$ and R should be computed and plotted on the control charts.

It is desirable to revise the control limits occasionally, even if the process remains stable. Some organizations mandate that control limits should be revised no less frequently that quarterly, even for a stable process. Control limits should always be revised when either process improvements or significant process changes are made.

Control Charts for Individual Measurements

In many situations, the sample size used for process monitoring and control is $n = 1$; that is, the sample consists of an individual unit. Some examples of these situations are as follows.

1. Automated inspection and measurement technology is used, and every unit manufactured is analyzed.
2. The production rate is very slow, and it is inconvenient to allow sample sizes of $n > 1$ to accumulate before being analyzed.
3. Repeat measurements on the process differ only because of laboratory or analysis error, as in many chemical processes.
4. In process plants, such as papermaking, measurements on some production units will differ very little and produce a standard deviation that is much too small if the objective is to monitor the variability in the measured characteristic between units.

In such situations, the control chart for individuals is potentially useful. The control chart for individuals uses the moving range of two successive observations to estimate the process variability. The moving range is defined as $MR_i = |X_i - X_{i-1}|$. Letting $\overline{MR}$ be the average of the moving ranges, an estimate of σ is

$$\hat{\sigma} = \frac{\overline{MR}}{d_2} = \frac{\overline{MR}}{1.128} \tag{8}$$

because d_2 = 1.128 when two consecutive observations are used to calculate a moving range. It is also possible to establish a control chart on the moving range using D_3 and D_4 for n = 2. The parameters for the control chart for individuals are defined as follows:

$$\begin{aligned} UCL &= \bar{X} + 3\frac{\overline{MR}}{d_2} = \bar{X} + 2.6596\overline{MR} \\ CL &= \bar{X} \\ LCL &= \bar{X} - 3\frac{\overline{MR}}{d_2} = \bar{X} - 2.6596\overline{MR} \end{aligned} \tag{9}$$

Table 2 shows 20 observations on concentration for the output of a chemical process. The observations are taken at one-hour intervals. If several observations are taken at the same time, the observed concentration reading will differ only because of measurement error. Since the measurement error is small, only one observation is taken each hour. To set up the control chart for individuals, note that the sample average of the 20 concentration readings is $\bar{X} =$ 99.1 and that the average of the moving ranges of two observations shown in the last column of Table 2 is 2.59. To set up the moving range chart, we note that D_3 = 0 and D_4 = 3.267 for n = 2. Therefore, the moving-range chart has center line 2.59, LCL = 0, and UCL = (3.267)(2.59) = 8.46. The control chart is shown as the lower control chart in Fig. 3. This control chart was constructed by Minitab. Because no points exceed the upper control limit, we may now set up the control chart for individual concentration measurements. Since the moving range of n = 2 observations is used, then for the data in Table 2 we have

$$\begin{aligned} UCL &= 99.1 + 2.6596(2.59) = 105.99 \\ LCL &= 99.1 - 2.6596(2.59) = 92.21 \end{aligned}$$

The control chart for individual concentration measurements is shown as the upper control chart in Fig. 3. There is no indication of an out-of-control condition.

The chart for individuals can be interpreted much like an ordinary $\bar{X}$ control chart. A shift in the process average will result in one or more points outside the control limits, or a pattern consisting of a run on one side of the center line. Some care should be exercised in interpreting patterns on the moving-range chart. The moving ranges are correlated, and this correlation may often induce a pattern of runs or cycles on the chart. The individual measurements are assumed to be uncorrelated, however, and any apparent pattern on the individuals' control chart should be carefully investigated.

Table 2 Concentration Data for Individuals Control Chart

Observation	Concentration	Observation	Concentration
1	102.0	11	101.3
2	94.8	12	98.7
3	98.3	13	101.1
4	98.4	14	98.4
5	102.0	15	97.0
6	98.5	16	96.7
7	99.0	17	100.3
8	97.7	18	101.4
9	100.0	19	97.2
10	98.1	20	101.0

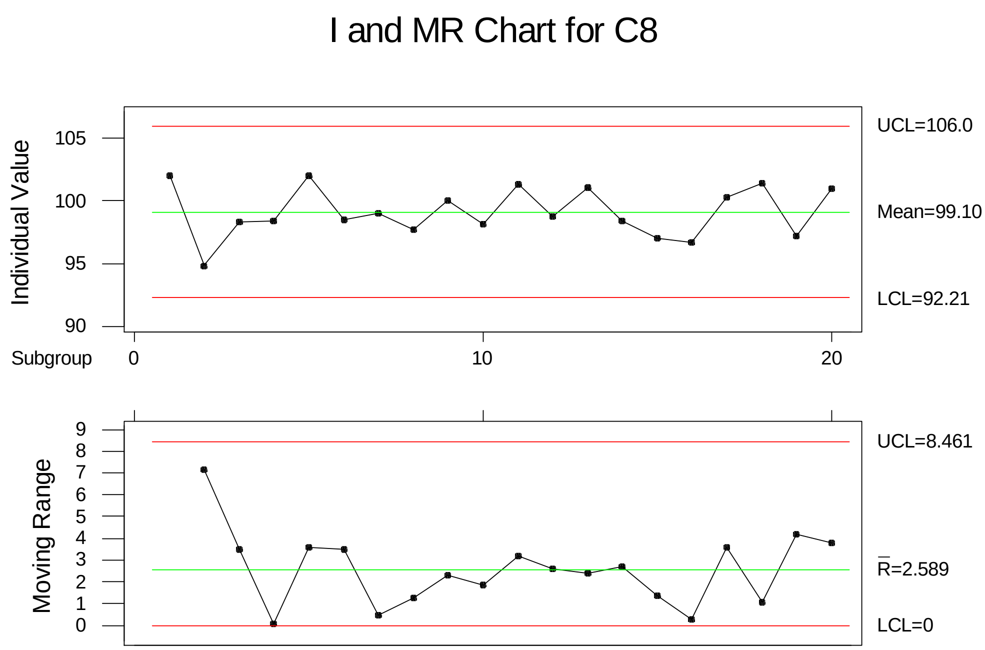

Figure 3 Individuals and Moving Range Control Charts for Chemical Concentrations

The control chart for individuals is very insensitive to small shifts in the process mean. For example, if the size of the shift in the mean is one standard deviation, the average number of sample points that must be plotted to detect this shift is 43.9. While the performance of the control chart for individuals is much better for large shifts, in many situations the shift of interest is not large and more rapid shift detection is desirable. In these cases, we recommend the cumulative sum (CUSUM) control chart or the exponentially weighted moving-average (EWMA) chart. These charts are briefly discussed in a subsequent section. See Montgomery (2001) for more details.

The control chart for individuals is very sensitive to the normality assumption. Figure 4 shows the normal probability plot of the chemical process concentration data in table 2. There are no obvious problems with the normality assumption. However, note that even small departures from normality such as a true distribution that is slightly skewed or has slightly heavier tails than the normal will seriously affect the performance of the control chart, resulting in a higher than advertised rate of false alarms. The exponentially weighted moving average control chart is a better alternative in these situations. More details on this aspect of the individuals control chart are in Montgomery (2001).

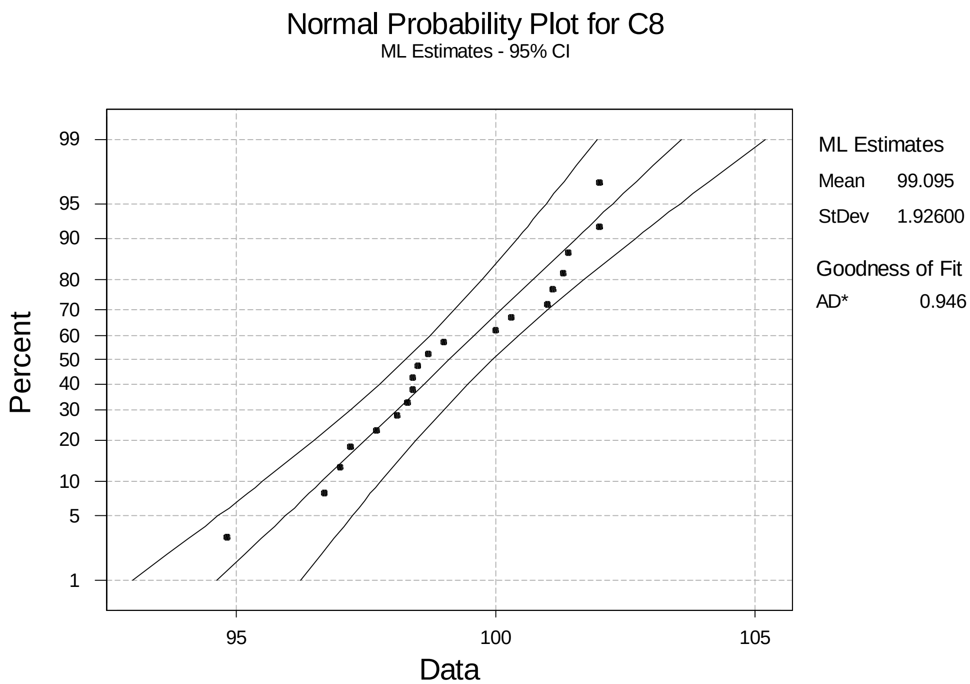

Figure 4. Normal probability Plot of the Chemical Process Concentration Data

Process Capability

It is usually necessary to obtain some information about the capability of the process; that is, the performance of the process when it is operating in control. The histogram is helpful in assessing process capability. The histogram for all 20 samples from the vane-manufacturing process is shown in Fig. 5. The upper and lower specifications on vane opening are 0.5020 and 0.5040 inches. In terms of the coded data, the upper specification limit is USL = 40 and the lower specification limit is LSL = 20. The general impression from examining this histogram is that the process is barely capable of meeting the specifications and that it is operating slightly off-center.

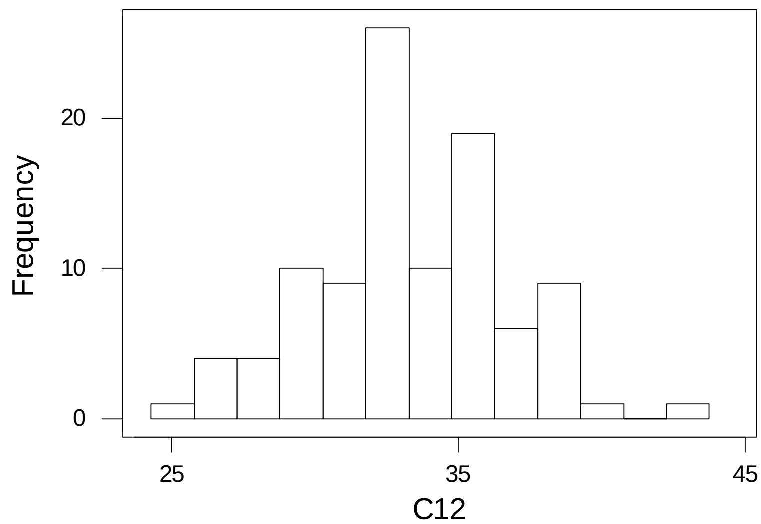

Figure 5. Histogram of the Vane Opening Measurements

Figure 6 presents a normal probability plot of the vane opening measurements. The plot indicates that assuming normality for the distribution of vane opening measurements is not unreasonable.

Notice that the observations from the samples that were likely taken when the process is out of control have been included in both Figures 5 and 6. Inferences about process capability for an out of control process are potentially problematic, because the very nature of an out of control is that the process is not predictable. Inferences about process capability are essentially predictions about future process performance, or about the quality level of material that has been shipped to customers. If the process is not in control, then these predictions may be very unreliable.

Many organizations express process capability in terms of the process capability ratio

$$C_p = \frac{USL - LSL}{6\sigma} \tag{10}$$

Notice that the numerator of this ratio is the width of the specifications and that the denominator can be thought of as the natural spread in the process. Ideally, this ratio should exceed unity. For the vane opening manufacturing process,

$$C_p = \frac{USL - LSL}{6\sigma} = \frac{40 - 20}{6(2.15)} = 1.55$$

where we have used the estimate of process standard deviation obtained from the R chart in Figure 2, $\hat{\sigma} = \bar{R} / d_2 = 5.0 / 2.326 = 2.15$. The process capability ratio has a simple interpretation, because $1/C_p$ is the fraction of the tolerance band that the process uses.

Notice that if the vane opening manufacturing process was in statistical control, we estimate that it uses about (1/1.55)100 = 64.5% of the available tolerance.

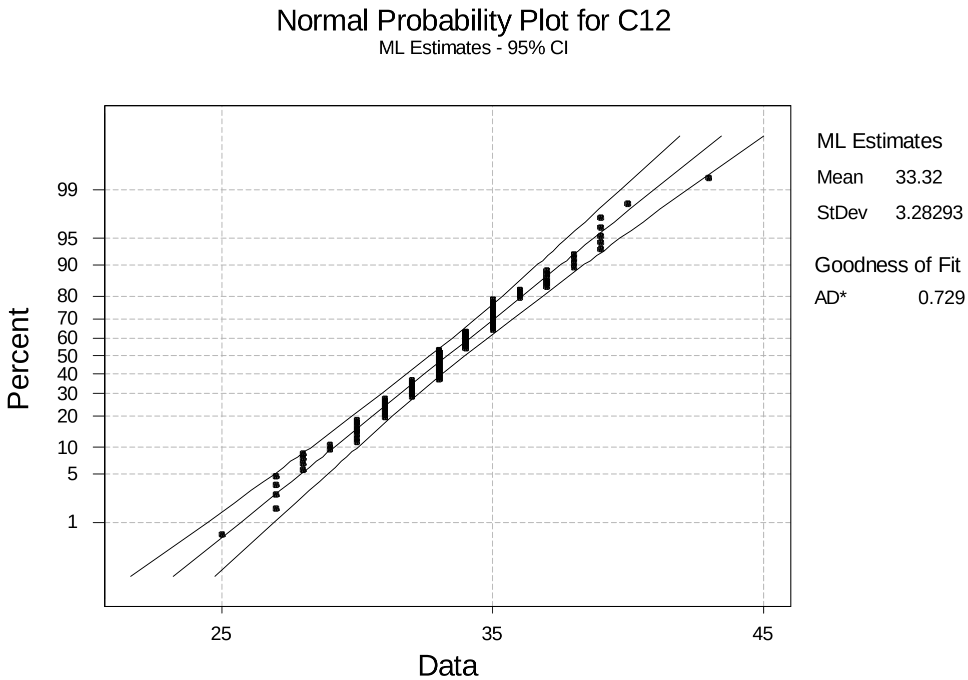

Figure 6. Normal Probability Plot for the Vane Opening Measurements

The process capability ratio C_p does not take process centering into account, so it really reflects only the potential capability that might be achieved by a perfectly centered process. If the process is not perfectly centered, then a measure of actual process capability is often used. This ratio, called C_{pk}, is defined as follows:

$$C_{pk} = \min\left[\frac{USL-\mu}{3\sigma}, \frac{\mu-LSL}{3\sigma}\right] \tag{11}$$

Essentially, C_{pk} is a one-sided ratio that measures how close the process mean is to the nearest specification limit. For the vane opening manufacturing process, if we consider the only the in control samples, an estimate of the process mean is $\bar{\bar{X}} = 33.19$, so the process is indeed off-center. Therefore C_{pk} is

$$C_{pk} = \min\left[\frac{40-33.19}{3(2.15)} = 1.06, \frac{33.19-20}{3(2.15)} = 2.04\right] = 1.06$$

Notice that the actual capability is much less than the potential capability. To achieve the potential capability, the process will have to be in statistical control and the mean will have to be adjusted so that it is closer to the target or nominal dimension.

Montgomery (2001) presents a table reporting the parts per million defective achieved by a normally distributed process in statistical control as a function of the process capability ratio. The assumptions of normality and stability are essential; indeed it is well known that even minor deviation from normality renders predictions of process defective highly inaccurate.

A six-sigma process is defined as one that has the process mean located no closer that six standard deviations from the nearest specification limit. Consequently, a six-sigma process has a process capability ratio C_{pk} that exceeds 2.

Attribute Control Charts

Often it is desirable to classify a product as either defective or nondefective on the basis of comparison with a standard. This classification is usually done to achieve economy and simplicity in the inspection operation. For example, the diameter of a ball bearing may be checked by determining whether it will pass through a gauge consisting of circular holes cut in a template. This kind of measurement would be much simpler than directly measuring the diameter with a device such as a micrometer. Control charts for attributes are used in these situations.

Attribute control charts often require a considerably larger sample size than do their variable measurements counterparts. In this section, we will present the fraction defective control chart, or *P* chart. Sometimes the *P* chart is called the control chart for fraction nonconforming. We will also discuss control charts for defects.

Suppose that *D* is the number of defective units in a random sample of size *n*. We assume that *D* is a binomial random variable with unknown parameter *p*. The fraction defective control chart is defined as follows:

$$\begin{aligned} UCL &= p + 3\sqrt{\frac{p(1-p)}{n}} \\ CL &= p \\ LCL &= p - 3\sqrt{\frac{p(1-p)}{n}} \end{aligned} \tag{12}$$

Values of the sample fraction defective $\hat{p} = D/n$ are plotted on the chart. When the true process fraction defective *p* is unknown, it must be estimated from preliminary data. If *m* preliminary samples are available and if p_i is the sample fraction defective for each sample, then $\overline{p} = (1/m)\sum_{i=1}^{m} p_i$ replaces *p* in equation (12). Note that this control chart is based on the normal approximation to the binomial distribution.

It is sometimes necessary to monitor the number of defects in a unit of product. Let the sample consist of *n* units and suppose that there are a total of *C* defects in the sample. Then $U = C/n$ is the average number of defects per unit. The control chart for defects per unit is defined as follows:

$$
\begin{aligned}
UCL &= U + 3\sqrt{\frac{U}{n}} \\
CL &= U \\
LCL &= U - 3\sqrt{\frac{U}{n}}
\end{aligned} \tag{13}
$$

When U is unknown, it may be estimated from preliminary samples. If there are m of these preliminary samples and U_i, $i = 1,2, \ldots, m$ are the observed number of defects per unit in these samples, then $\bar{U} = (1/m)\sum_{i=1}^{m} U_i$ replaces U in equation 13.

Many control chart computer programs produce a control chart of C, the total number of observed defects in the sampled units. This variation of the U chart is called the C chart.

Both the U and the C chart assume that the number of defects observed in a unit follows the Poisson distribution. These charts can produce erroneous results if the distribution of the number of defects is not Poisson. Situations where defects occur in clusters or groups often are symptoms of violation of the Poisson assumption. In the Poisson distribution the mean and the variance are the same, and this information can sometimes be used to provide a rough check of the validity of the Poisson assumption. When the sample variance greatly exceeds the sample average in count data, the Poisson assumption may be inappropriate. In practice, we often find that the variance is considerable greater than the mean. This overdispersion can lead to control limits that are much wider than they should be and greatly impacts the usefulness of the control chart. Consult Montgomery (2001) for more details.

Control Chart Performance

Specifying the control limits is a critical decision in designing a control chart. By moving the control limits further from the center line, we decrease the risk of a type I error; that is, the risk of a point falling beyond the control limits, indicating an out-of-control condition when no assignable cause is present. However, widening the control limits will also increase the risk of a type II error; that is, the risk of a point falling between the control limits when the process is really out of control. If we move the control limits closer to the center line, the opposite effect is obtained: The risk of type I error is increased, while the risk of type II error is decreased. The control limits on a Shewhart control chart are customarily located a distance of plus or minus three standard deviations of the variable plotted on the chart from the center line. These limits are called 3-sigma control limits.

We may evaluate decisions regarding sample size and sampling frequency through the average run length (ARL) of the control chart. Essentially, the ARL is the average number of points that must be plotted before a point indicates an out-of-control condition. For any Shewhart control chart, the ARL can be calculated from the mean of a geometric random variable (for details, see Montgomery 2001). Suppose that p is the probability that any point exceeds the control limits. Then for an $\bar{X}$ control chart with three-sigma limits we have $p = 0.0027$ as the probability that a single point exceeds the limits when the process is in control, so the in control ARL is

$$\text{ARL}=\frac{1}{p}=\frac{1}{0.0027}=370$$

So on the average, a false alarm will be generated every 370 points. To find the out of control ARL, or the number of points plotted on the average to signal when a shift has occurred, we need to find the probability *p* of a point plotting outside the control limits. For example, if the process mean shifts by three standard deviations of the variable plotted on the control chart then it is straightforward to show that $p = 0.5$. Therefore, the out of control ARL is ARL = 1/0.5 = 2.

Studies of the ARL performance of Shewhart control charts reveal that they are relatively insensitive to small to moderate shifts of magnitude of up to about 2 sigma. That is, their out of control ARLs are relatively large. For larger shifts, they are quite effective; note that for the three-sigma shift above it only requires two periods on the average to detect the shift. A brief table of ARL values follows.

Shift size, $\sigma\sqrt{n}$	0	0.5	1	1.5	2	3
ARL	370	155.2	43.9	15	6.3	2

Because Shewhart control charts are relative insensitive to small to moderate size shifts, interest often focuses on control charts that perform better in the detection of these smaller shifts. Two such procedures are the cumulative sum control chart and the exponentially weighted moving average control chart. These procedures are often very effective alternatives to using a Shewhart chart with additional sensitizing rules because they can detect small shifts just as quickly yet they do not suffer increased rates of false alarms.

Cumulative Sum and Exponentially Weighted Moving Average Control Charts

Two very effective alternatives to the Shewhart control chart are the cumulative sum (or CUSUM) control chart and the Exponentially Weighted Moving Average (or EWMA) control chart. These charts have much better performance (in terms of ARL) for detecting small shifts than the Shewhart chart, but do not cause the in-control ARL to drop significantly. This section will outline the basic procedures for using these control charts.

The CUSUM

The CUSUM chart plots the cumulative sums of the deviations of the sample values from a target value. For example, suppose that samples of size $n \geq 1$ are collected, and $\bar{X}_j$ is the average of the *j*th sample. Then if μ_0 is the target for the process mean, the CUSUM control chart is formed by plotting the quantity

$$C_i = \sum_{j=1}^{i}(\bar{X}_j - \mu_0) \qquad (14)$$

against the sample number.. Now, C_i is called the cumulative sum up to and including the *i*th sample. Because they combine information from several samples, cumulative sum charts are more effective than Shewhart charts for detecting small process shifts. Furthermore, they are particularly effective with samples of $n = 1$. This makes the cumulative sum control chart a good candidate for use in the chemical and process industries where rational subgroups are frequently of size 1, as well as in discrete parts manufacturing with automatic measurement of each part.

If the process remains in control at the target value, the cumulative sum defined in equation 14 should fluctuate around zero. However, if the mean shifts upward then an upward or positive drift will develop in the cumulative sum statistic. Conversely, if the mean shifts downward, then a downward or negative drift in the CUSUM will develop. Therefore, if a trend develops in the plotted points either upward or downward, we should consider this as evidence that the process mean has shifted.

Most practical applications of the CUSUM employ a tabular procedure in which the CUSUM statistic in equation 14 is accumulated as two one-sided statistics defined as

$$\begin{aligned} C_i^+ &= \max[0, \overline{X}_i - (\mu_0 + K) + C_{i-1}^+] \\ C_i^- &= \max[0, (\mu_0 - K) - \overline{X}_i + C_{i-1}^-] \end{aligned} \qquad (15)$$

where we usually take $C_0^+ = C_0^- = 0$. The constant K in equation 15 is called the reference value for the CUSUM, and it is usually selected to be about one-half of the magnitude of the shift that we wish to detect. If either of the one-sided CUSUM statistics in equation 15 exceeds a decision interval H the process is considered to be out of control.

The CUSUM is much more responsive to shifts that is the Shewhart control chart. The table below shows the ARL values for a CUSUM with $K = 0.5\sigma_{\overline{X}}$ and $H = 5\sigma_{\overline{X}}$, values that are widely used in practice.

Shift, in multiples of $\sigma_{\overline{X}}$	0	0.25	0.50	0.75	1.00	1.50	2.0	2.50	3.00	4.00
ARL	465	139	38	17	10.4	5.75	4.01	3.11	2.57	2.01

Notice that for detecting a one-sigma shift in the mean (with $n = 1$) the CUSUM ARL is 10.4, whereas for the individuals control chart it is 43.9. Generally, the CUSUM is much more effective than the Shewhart chart for shifts up to about two standard deviations, and roughly comparable for larger shifts. On the other hand, the CUSUM is more difficult to use in bringing an out of control process into a state of statistical control because patterns on the CUSUM are not interpretable (because successive CUSUM values are correlated), and interpretation or analysis of patterns is often very useful in identifying assignable cause when control charts are first applied.

Figure 7 presents the Minitab results of applying a CUSUM to the chemical concentration data in Table 2. We have assumed that the process target is $\mu_0 = 99$. This type of display is usually called a CUSUM status chart. Notice that the process is in control.

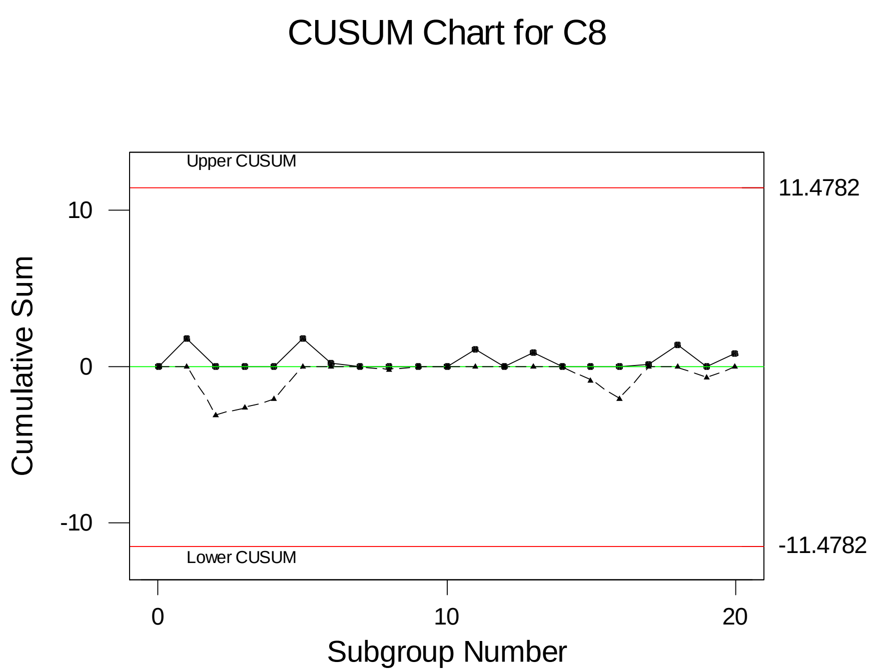

Figure 7. A CUSUM Status Chart for the Chemical Process Concentration Data

Because the CUSUM is very effective at detecting small process shifts and in detecting small departures from a desired target, and because it works very well with individual measurements, it has been widely deployed in the chemical and process industries. It's ise in discret parts manufacturing is growing rapidly.

There are many important variations of the CUSUM. Some of these include a fast initial response or headstart feature that allows more rapid detection of a process that is still off-target following an adjustment, combined CUSUM-Shewhart schemes that improve detection of large shifts, and CUSUMs for a variety of sample statistics, including counts or time between events. For discussion of some of these procedures, see Montgomery (2001).

The EWMA

The Exponentially Weighted Moving Average (EWMA) defined as

$$Z_i = \lambda \bar{X}_i + (1-\lambda) Z_{i-1} \text{ with } 0 \le \lambda < 1, Z_0 = \mu_0 \tag{16}$$

can be used as the basis of a control chart. The procedure consists of plotting the EWMA statistic Z_i versus the sample number on a control chart with center line $CL = \mu_0$ and upper and lower control limits at

$$\begin{aligned} UCL &= \mu_0 + k\sigma_{\bar{X}}\sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2i}]} \\ LCL &= \mu_0 - k\sigma_{\bar{X}}\sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2i}]} \end{aligned} \tag{17}$$

The EWMA control chart is very flexible. For large values of the parameter λ it closely mimics the performance of a Shewhart chart, while for smaller values of λ it performs like the CUSUM. It is usually applied in situations where CUSUMs are appropriate; namely, the detection of small shifts is of interest. The design parameters are the width of the control limits k and the EWMA parameter λ. Montgomery gives a table of recommended values for these parameters to achieve certain average run length performance.

Figure 8 is the Minitab EWMA control chart applied to the chemical process concentration data in Table 2. In constructing this chart we have used a target value of 99 for the process mean and λ =0.1 and k = 2.8, as these values result in ARL performance that closely matches the CUSUM with $K = 0.5\sigma_{\bar{X}}$ and $H = 5\sigma_{\bar{X}}$. The process is in control.

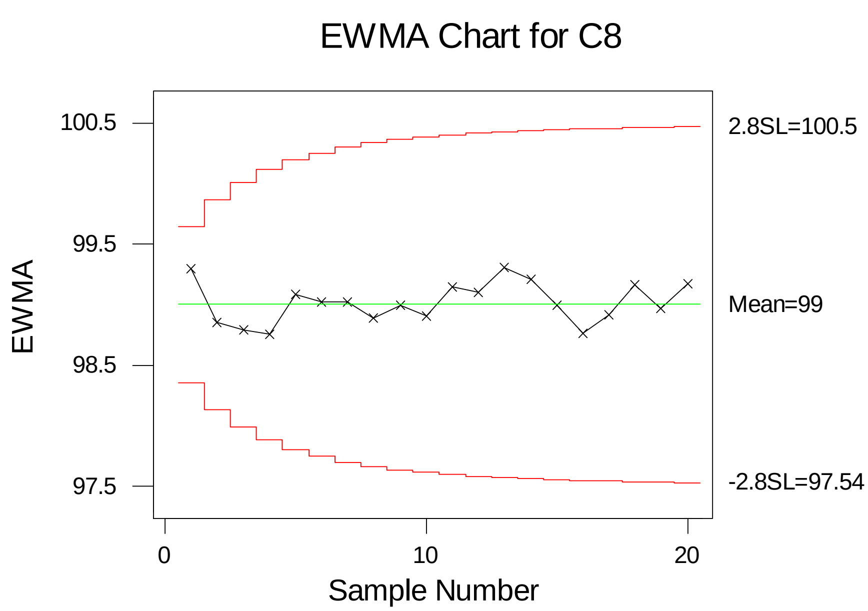

Figure 8 An EWMA Control Chart for the Chemical Process Concentration Data

The EWMA control chart is an excellent alternative to the CUSUM, as it has comparable ARL performance and many practitioners feel that it is easier to implement and use. It can also incorporate a headstart or fast initial response feature, it can be adapted for use with other sample statistics (such as variances), and it can be applied to Poisson counts. The EWMA has also been adapted to monitor processes with autocorrelated data, and because of the near-optimal one-step-ahead predictor features of the EWMA, it is often the basis of many types of feedback control schemes that are widely used for process adjustment. For more details on these aspects of the EWMA, see Montgomery (2001).

References

Montgomery, D. C. (2001). *Introduction to Statistical Quality Control*, 4th edition. John Wiley & Sons, New York.

Western Electric (1956). *Statistical Quality Control Handbook*. Western Electric Corporation, Indianapolis, IN.