Columbia Business School

Columbia CaseWorks

CU226
ID#080207

PUBLISHED ON
AUGUST 16, 2010

Statistical Process Control

BY NELSON M. FRAIMAN*

ABSTRACT

This note discusses the notion of statistical control and shows how to create and utilize control charts in order to identify and reduce variation in a process. In addition, it discusses the differences between specification limits, process capability, and performance limits.

CONTENTS

* *Professor of Professional Practice and Codirector of the W. Edwards Deming Center, Columbia Business School*

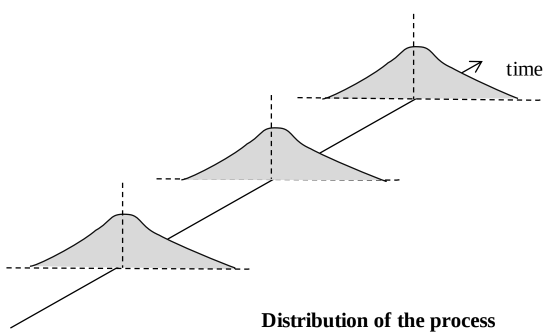
Introduction

Statistical Process (Quality) Control (SPC) is a set of statistical techniques and concepts used to identify and reduce undesired variation in a process. SPC has its origins in manufacturing, and one of its first advocates was Walter Shewhart at Bell Laboratories in the 1920s. While studying process data, Shewhart made the distinction between controlled and uncontrolled variation, due to what we now call *common* and *assignable* causes, respectively. He developed a simple but powerful tool to separate the two—the control chart. Shewhart set the stage for SPC by asking in his seminal work *Economic Control of Quality of Manufactured Product*: “How much may quality vary and yet be controlled? How much variation should we leave to chance?” W. Edwards Deming brought SPC to Japan in the 1950s, and these same concepts lie at the heart of Six Sigma programs that are being implemented both in manufacturing firms such as GE, Honeywell, and Motorola, and in service firms such as J.P. Morgan, American Express and Federated Department Stores. In 1950 Deming stated in Tokyo: “Variation is a rule of nature. Repetition of any procedure will produce variable results.”

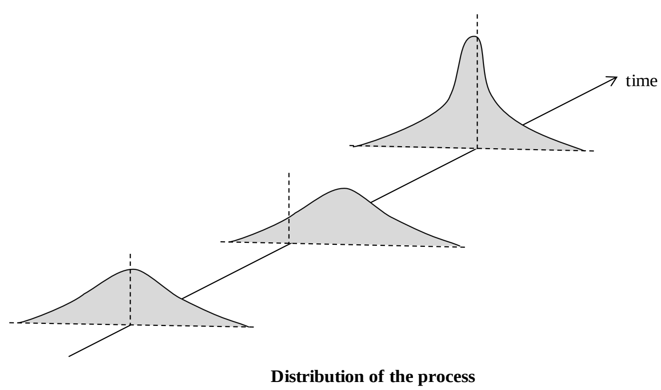
A central principle of SPC is that the measured aspect of a product is always subject to a certain amount of inherent variation, and that this variation is the result of chance. Thus, variation within predictable limits is inevitable. The acidity of a solution that Benetton uses in dying textiles, the weight of the Corn Flakes box that General Mills uses in packaging, the level of cash that Citibank maintains, and the thickness of a wood chip that International Paper uses in its digester are all subject to variation. In these instances, it is certain that variation will occur, but if the pattern of variation is stable, then the process is under statistical control.

Let us look at Figures 1 and 2 below to illustrate the notion of statistical stability:

FIGURE 1
THE CONCEPT OF *STATISTICAL CONTROL*
Definition: A process whose distribution is stable over time.



A PROCESS NOT UNDER *STATISTICAL CONTROL*
Definition: A process whose distribution changes over time.



BY NELSON M. FRAIMAN*

In the process under statistical control (see Figure 1, left), the output is variable but the distribution is stable. The distribution of values we experienced yesterday will be the same distribution we see today, tomorrow, and in the future. Another way of saying this is that all the variability is due to *common causes*. When this is the case, the process is under statistical control. When variability is due to common causes, no single source of variability can be identified. In a process not under statistical control (see Figure 1, right), the distribution changes over time. Thus, we cannot predict what the future distribution will look like. In a process not under statistical control, a substantial part of the variability is due to *assignable* or *special causes*. In addition to the common causes, there are identifiable special causes that substantially influence the variability of the system.

The main statistical tool by which we can determine whether or not a process is under control is the control chart. Its purposes are to:

- identify if a state of statistical control exists, and if not,
- isolate assignable causes of variation (to help eliminate them) and then
- monitor the system as part of a continuous improvement process for quality and productivity.

In short, the control chart helps to bring a process under control.

Managing quality with process control can be thought of as a preventive approach. To quote Columbia Business School Professor Garrett van Ryzin: We want to "keep the patient [process] healthy by regular checkups [process control] and preventive medicine [problem solving] rather than waiting until a heart attack [defect] strikes." This preventative approach requires:

- tracking quality metrics
- identifying variation due to assignable causes
- determining process capability
- instituting formal methods for continuous problem diagnosis and solution.

Controls charts rely on the first requirement. They perform the second requirement and are a part of the fourth. And they are a necessary first step for the third requirement. (Process capability is discussed in section three of this note.)

Control charts are based on statistical principles and, consequently, are vulnerable to what statisticians call type 1 and type 2 errors. The type 1 error occurs when we assume that a variation is a special cause variation, when in fact it is a common cause variation. The type 2 error occurs when we assume that a variation is due to common causes, when in fact it is due to a special cause. The error of the first kind lowers quality because it leads to changing a system with no problems. Errors of the second kind can result in missed opportunities to make significant process improvements. These are standard problems in statistics, and a key issue in developing control charts is to find an appropriate balance between them.

In practice, there are two types of control charts that are commonly used:

- **Control charts for variables.** These are used when it is possible to measure the actual degree of variation of a quality characteristic. Examples of quality measures of this kind include weight and size.
- **Control charts for attributes**. When it is not possible to measure the degree of variation, one can usually measure whether or not it conforms. In other words, the product (or service) either succeeds or fails, fits or doesn't fit, satisfies or doesn't satisfy, etc.

Control Charts for Variables

Two commonly used controls charts for variables are the X-bar chart and the R chart. X-bar and R charts are constructed from a series of sample observations. The X-bar chart plots the mean of each sample, and the R chart plots the range of each sample. The central idea of control charts is based on the size and composition of the sample. In control chart terminology, each sample is called a rational subgroup.

Perhaps surprisingly, the subgroups are gathered in a relatively small size so that:

- Within the subgroup, the only variation is due to common cause variation
- Variation due to special causes between subgroups is detectable. A subgroup should contain only data produced under homogeneous conditions. (For instance, subgroups should share the same equipment, environmental conditions, workers, etc.)

Typically, the number of observations (*n*) in a subgroup (sample) ranges from 2 to 10. The goal with rational subgroups is to minimize variation within groups and maximize it between groups. Once we have determined the size of the subgroups, we need to take measurements. We usually use 20 or more subgroups (each containing *n* observations), but one should not have less than 10 subgroups of observations. (One rule of thumb is to use 100 or more observations subdivided into 20-25 subgroups, each containing 4-5 observations.) When we first build a control chart, we assume that the process is somewhat stable; if not, we keep taking samples until we see some stability. Then we calculate the overall mean values and build the control limits. Control charts are best introduced with an example.

CONSTRUCTION OF X-BAR AND R CHARTS: THE SOFTWOOD CHIP EXAMPLE

Wood chips are the basic raw material for a pulp mill. Data were gathered from a vendor supplying wood chips to a paper mill in Moss Point, Mississippi. Truckloads of chips were delivered each day from the vendor to the mill. The paper mill randomly sampled truckloads to determine whether the shipments were acceptable. Acceptability was based on the weight of the truckload. (Heavier truckloads were preferred to lighter truckloads.)

In practice, there are subtleties in collecting the data. For example, where do we weigh the truckload? (Initially, it was weighed at the mill, after the truck arrived from the supplier.) Or, for another example, how should the truckload weight actually be determined? (Again, initially, the weight was determined from a bucket of chips selected from the top of the truckload.) These are important questions, but for simplicity, let us assume that there is no measurement error in the data. Table 1 presents the data collected.

TABLE 1
SOFTWOOD CHIP TRUCKLOAD WEIGHTS

Day	1	2	3	4
1	76.3	74.2	82.1	77.7
2	77.9	79.1	82.2	75.0
3	80.4	76.3	82.9	77.7
4	79.7	69.7	79.6	77.7
5	77.5	78.1	75.4	71.3
6	76.4	82.1	88.5	80.6
7	80.3	76.7	78.9	80.5
8	84.8	86.3	80.9	81.9
9	79.9	81.0	81.9	78.8
10	78.1	79.2	81.9	81.0
11	76.1	75.1	82.1	76.5
12	75.1	78.8	74.8	80.6
13	77.5	79.8	73.7	77.4
14	77.7	79.6	79.6	71.3
15	77.3	77.7	72.4	76.3
16	75.8	76.8	74.1	74.2
17	79.1	73.3	77.4	78.0
18	78.9	78.8	79.4	81.7
19	79.2	79.3	81.4	79.9
20	78.3	77.2	78.3	81.5
21	80.9	82.6	76.8	82.8
22	80.4	86.7	81.3	79.4
23	82.3	76.9	83.1	85.0
24	78.7	83.9	86.7	87.2
25	88.2	83.8	83.0	86.8
26	83.8	88.7	80.1	82.6
27	86.3	78.5	83.8	78.2
28	80.5	81.8	78.5	80.8
29	82.5	84.1	81.1	80.1
30	78.2	83.5	84.8	84.0

Each day, four truckloads were randomly selected and weighed. The chart shows the measurements taken each day for a period of 30 days. A total of 120 truckloads were sampled. In order to construct the X-bar and R charts, we proceed as follows:

1. Take samples of data. We have already done this above. In Table 1 above, we have 30 samples of size $n = 4$.
2. For each sample, calculate the *sample average*:

$$\bar{X}_j = \frac{X_{j1} + X_{j2} + \cdots + X_{jn}}{n}$$

where j stands for the day (a row in Table 1) and $X_{j1}, X_{j2}, \ldots, X_{jn}$ are the observations on that day. For example, for day 4 we have:

$$\bar{X}_4 = \frac{79.7 + 69.7 + 79.6 + 77.7}{4} = 76.7$$

We also compute the *range* for each sample:

$$R_j = \max\{X_{j1}, X_{j2}, \ldots, X_{jn}\} - \min\{X_{j1}, X_{j2}, \ldots, X_{jn}\}$$

For example, for day 4 we have:

$$R_4 = \max\{79.7, 69.7, 79.6, 77.7\} - \min\{79.7, 69.7, 79.6, 77.7\} = 10.0$$

Table 2 below provides the computations using the data from Table 1.

TABLE 2
X-BAR AND R VALUES FOR SOFTWOOD CHIPS

Day	X-bar ($\bar{X}$)	Range (R)	Day	X-bar ($\bar{X}$)	Range (R)
1	77.6	7.9	16	75.2	2.7
2	78.6	7.2	17	77.0	5.8
3	79.3	6.6	18	79.7	2.9
4	76.7	10.0	19	80.0	2.2
5	75.6	6.8	20	78.8	4.3
6	81.9	12.1	21	80.8	6.0
7	79.1	3.8	22	82.0	7.3
8	83.5	5.4	23	81.8	8.1
9	80.4	3.1	24	84.1	8.5
10	80.1	3.8	25	85.5	5.2
11	77.5	7.0	26	83.8	8.6
12	77.3	5.8	27	81.7	8.1
13	77.1	6.1	28	80.4	3.3
14	77.1	8.3	29	82.0	4.0
15	75.9	5.3	30	82.6	6.6

3. Compute the overall average, which is called $\bar{\bar{X}}$ (X-double bar). This is simply the average of the X-bars of all the samples, namely,

$$\bar{\bar{X}} = \frac{\bar{X}_1+\bar{X}_2+\cdots+\bar{X}_m}{m},$$

which in our example, with m=30, gives $\bar{\bar{X}} = 79.8$.

4. Compute the average of the R values, $\bar{R}$ (R bar):

$$\bar{R} = \frac{\bar{R}_1 + \bar{R}_2 + \cdots + \bar{R}_m}{m}$$

In our example, $\bar{R}$ is 6.1.

5. We can now compute what are called control limits. For the X-bar chart, these are the formulas:

$$Upper\ Control\ Limit\ (UCL) = \bar{\bar{X}} + 3 \cdot \frac{\bar{R}}{d_2 \cdot \sqrt{n}}$$

$$Center\ Line = \bar{\bar{X}}$$

$$Lower\ Control\ Limit\ (LCL) = \bar{\bar{X}} - 3 \cdot \frac{\bar{R}}{d_2 \cdot \sqrt{n}}$$

where the factor d_2 depends on the sample size and is tabulated in Table 3.

To understand these formulas, think of X-double bar as a process mean and think of $\bar{R}/d_2$ as the process standard deviation. Therefore, $\bar{R}/(d_2\sqrt{n})$ is the standard deviation of the sample average, also referred to as the standard error. The control limits look like a confidence interval centered at the sample mean plus/minus three standard errors. The d_2 factor makes $\bar{R}$ act like a standard deviation of the process. (Formally, X-double bar is the *estimate* of the process mean, and $\bar{R}/d_2$ is the *estimate* of the process standard deviation.) In the example, with n=4, the factor d_2 is equal to 2.059.

TABLE 3
CONTROL CHART PARAMETERS FOR 99.7 PERCENT CONFIDENCE

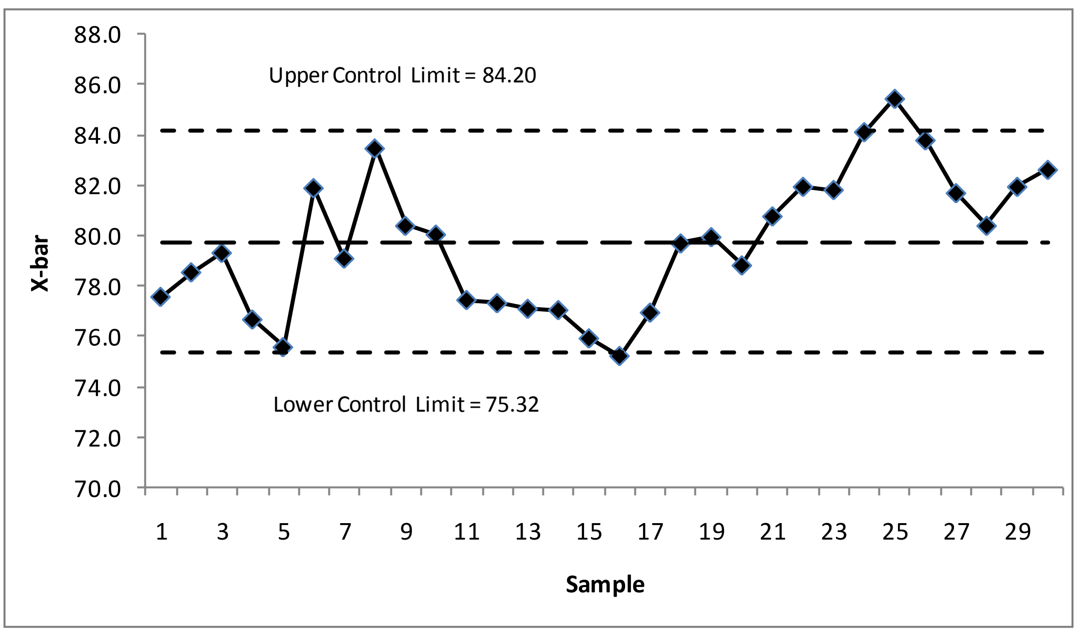
Number of observations in subgroup (n)	Factor to Estimate Standard Deviation based on the Range (d2)	Factor for Lower Control Limit in R chart (D3)	Factor for Upper Control Limit in R chart (D4)
2	1.128	0	3.27
3	1.693	0	2.57
4	2.059	0	2.28
5	2.326	0	2.11
6	2.534	0	2.00
7	2.704	0.08	1.92
8	2.847	0.14	1.86
9	2.970	0.18	1.82
10	3.078	0.22	1.78

The D3 and D4 factors in Table 3 are used to compute the control limits for the R-bar chart as follows:

$$Upper\ Control\ Limit = D_4 \cdot \bar{R}$$
$$Center\ Line = \bar{R}$$
$$Lower\ Control\ Limit = D_3 \cdot \bar{R}$$

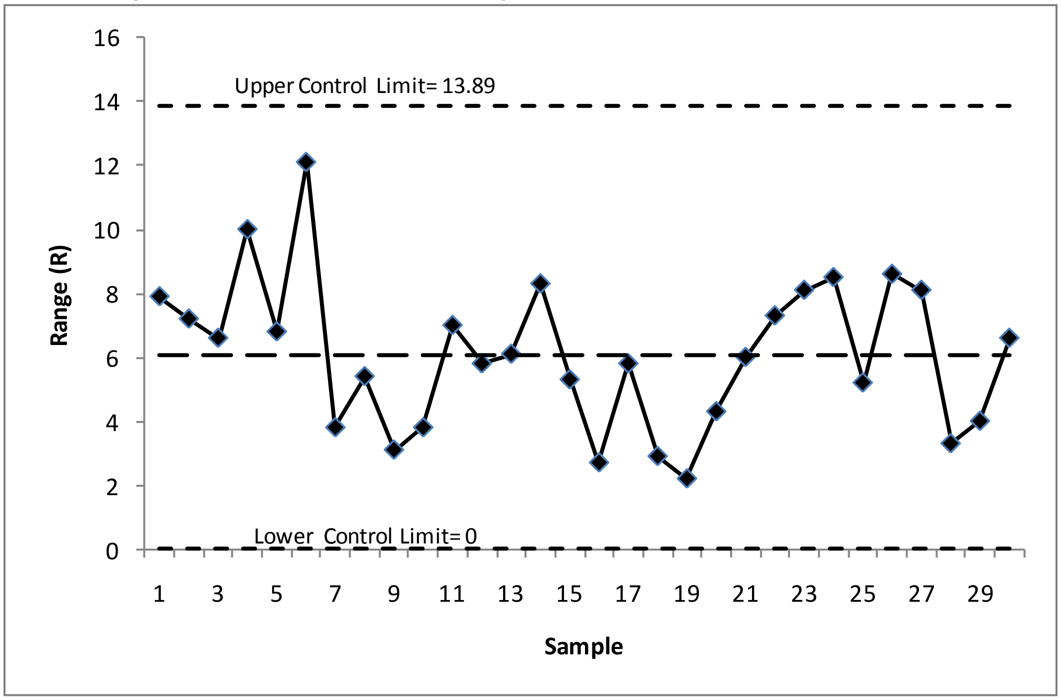
For the example, D_3=0 and D_4= 2.28. Figures 2 and 3 below depict the X-bar and R Charts.

FIGURE 2
X-BAR CHART (SOFTWOOD CHIP EXAMPLE)



Columbia CaseWorks

FIGURE 3
R CHART (SOFTWOOD CHIP EXAMPLE)



EXAMPLE COMMENTARY

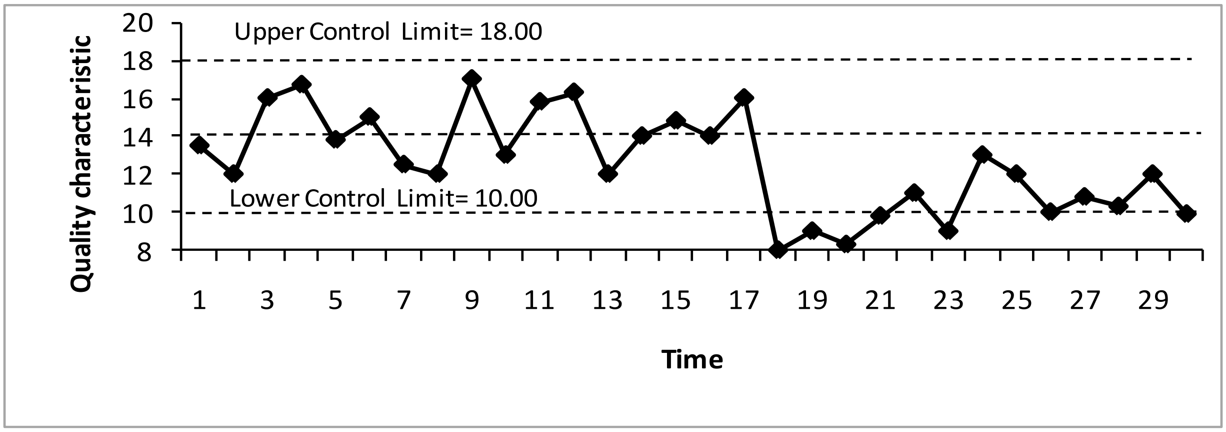
On the X-bar chart, two points are outside the control limits: one on day 16 and one on day 25. Points outside the limits indicate that assignable causes were probably present. In other words, something unusual happened on these days. On day 16, we were out on the low side, which, in the case of Softwood Chips, is undesirable; and on day 25, we were out on the high side, which, in this particular case, is desirable. Both situations indicate a lack of control. In the low situation, we have to find out what the problem is and eliminate it. In the high situation, we should find out what caused it so we can exploit it in the future.

With respect to the R chart, the upper and lower control limits were 13.9 and 0 respectively. All points are within the control limits.

INTERPRETING CHARTS—OTHER EXAMPLES

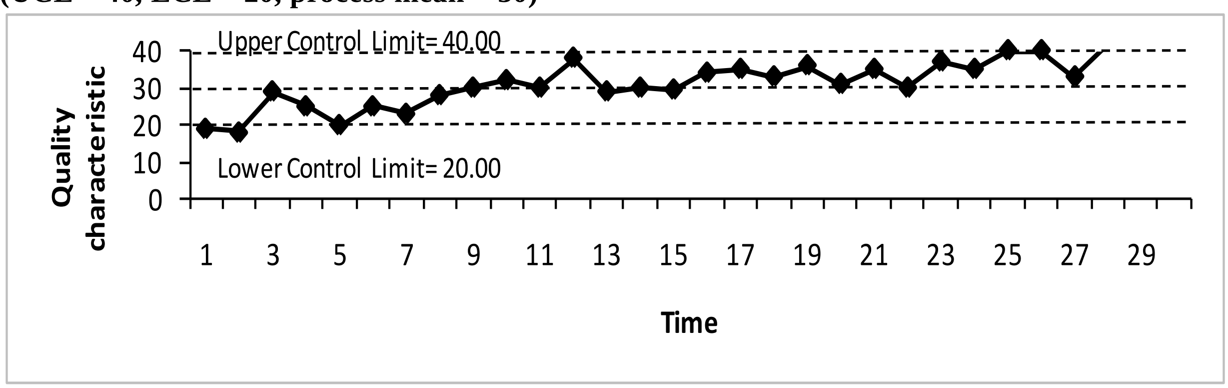
Even if all the data are within control limits, the pattern of the data can provide information. The position of points on the control charts, their concentration, and/or the various patterns they display provide us with signals as to whether the process is in control, going out of control, or already out of control. If a process is in control, all points will be within the control limits. Further, their pattern within the limits will look random. We will expect close to 64% of the points to be within a standard error, which we are estimating by $\bar{R}/(d_2\sqrt{n})$. We should also expect about the same number of points above the center line as below it. In short, no particular concentration, sequence, or direction should be discernable. If a pattern is discernible, then the process is out of control. To help you recognize discernible patterns, we consider some examples.

FIGURE 4
SUDDEN SHIFT IN LEVEL
Change in a Pattern with a Large Number of Points Outside
(UCL = 18, LCL = 10, process mean = 14)



In Figure 4, we observe a definite pattern. The first group of points are almost all above the center line (value of center line is 14), whereas the second group is entirely below the center line. This is definitely nonrandom. This situation is common during the beginning of control chart activity before a state of control is achieved. It may indicate that we took observations from two different populations—for instance, populations with different raw materials, different machine settings, etc. Even if there are no points outside the control limits, we would still regard the process as being out of control. Let us take a look at some other examples.

FIGURE 5
TRENDS
A Long Series of Points without a Change in Direction (Up or Down)
(UCL = 40, LCL = 20, process mean = 30)

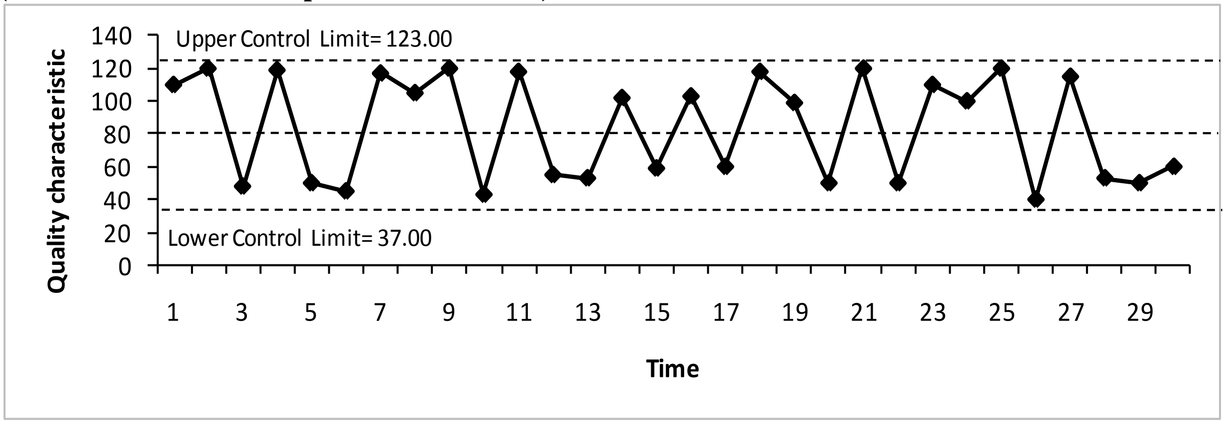


In Figure 5, the nonrandom pattern in the chart is a trend. Even though almost all the points are within the control limits, we suspect that something is happening to the process. Typically, trends are associated with some gradual change such as gradual deterioration of equipment, decreasing operator attention, etc. One way to identify trends is to count consecutive points in either an ascending or descending direction. A process is probably out of control if you have 8 consecutive points either

BY NELSON M. FRAIMAN*
Columbia CaseWorks

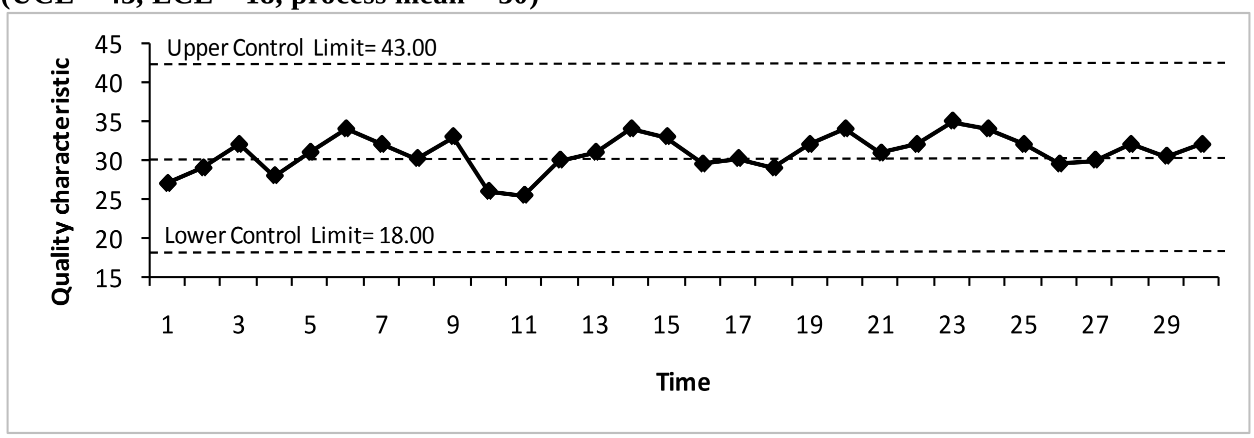
increasing or decreasing. For similar reasons, if you have 8 consecutive points either above or below the center line, then the process is probably out of control as well.

FIGURE 6
MIXTURES
Points Too Close to the Control Limit
(UCL = 123, LCL = 37, process mean = 80)



In the above graph, the points are too close to the control limits. This can happen from mixing output from two machines, two operators, or two of anything. The pattern is not random, even though all the points are within the control limits.

FIGURE 7
STRATIFICATION
Points Too Close to Average
(UCL = 43, LCL = 18, process mean = 30)



When all the points are too close to the center line, as in the graph above, we say that there is stratification. There are a number of reasons why this might happen:

- Control limits were not calculated properly
- The rational subgroups (samples) may be incorrect (e.g., the variability within a group is very high, consisting of only high and low values)
- The process has improved, and it may be time to recalculate limits

- The data were fabricated by the operator (not an uncommon phenomenon).

Control Charts for Attributes

Many quality measures are not numerical—they are good/bad (binary) measures. A customer complaint, a lost luggage incident, a crime incident, a medical relapse, a product failure, a warranty repair incident, a bad meal, and a dirty room are all examples of binary measures. They can be described with only two values; for example, conforming/nonconforming, pass/fail, go/no-go, present/absent. In other words, they either have an attribute of interest or they don't. The tool for analyzing such data is called the P-chart. It measures the proportion of nonconforming items in a group of items being inspected. The conforming/nonconforming assessment may be based on evaluating one characteristic (was a particular meal delivered?) or many characteristics (was anything found wrong at the Ritz-Carlton last night?).

Control charts for attributes are useful for several reasons:

- The data are already available in many situations, so no additional expense is involved
- Even if data are not already available, attribute information is quick and inexpensive to obtain
- Attribute-type situations exist in most technical and administrative processes, so attribute analysis can be used in many service operations.

CONSTRUCTING A P-CHART

In discussing X-bar and R charts, we talked about common cause variation. With the P-chart, the common cause variation is estimated by the total proportion of nonconforming items. For a given sample, the question is whether the sample's nonconforming proportion is close enough to the total nonconforming proportion. If a sample is not close enough, we assume that there is an assignable cause. As before, let us proceed with an example.

1. Gather Data

In contrast to X-bar and R charts, charts for attributes generally require large subgroup (sample) sizes in order to detect moderate shifts in performance. Samples could have different numbers of observations, so we denote the size of sample j by n_j. (One rule of thumb is to have the subgroup size be greater than three divided by the expected proportion of nonconforming items. Using the notation we introduce below, we would want $n_j \geq 3/\bar{p}$ for each n_j.) As is the case for X-bar and R charts, it is preferable to have 20-25 subgroups.

2. Calculate the fraction of defects on each sample:

$$Fraction\ in\ sample\ j = \frac{\text{Number of defects in sample } j}{n_j}$$

3. Compute the central line and control limits. In contrast to X-bar and R charts, the control limits can vary for each sample as the formulas below show.

$$\text{Center Line} = \bar{p} = \frac{\text{total number of defects in all samples}}{\text{total number of observations}}$$

$$UCL_j = \bar{p} + 3\sqrt{\frac{\bar{p}\,(1-\bar{p})}{n_j}}$$

$$LCL_j = \bar{p} - 3\sqrt{\frac{\bar{p}\,(1-\bar{p})}{n_j}}$$

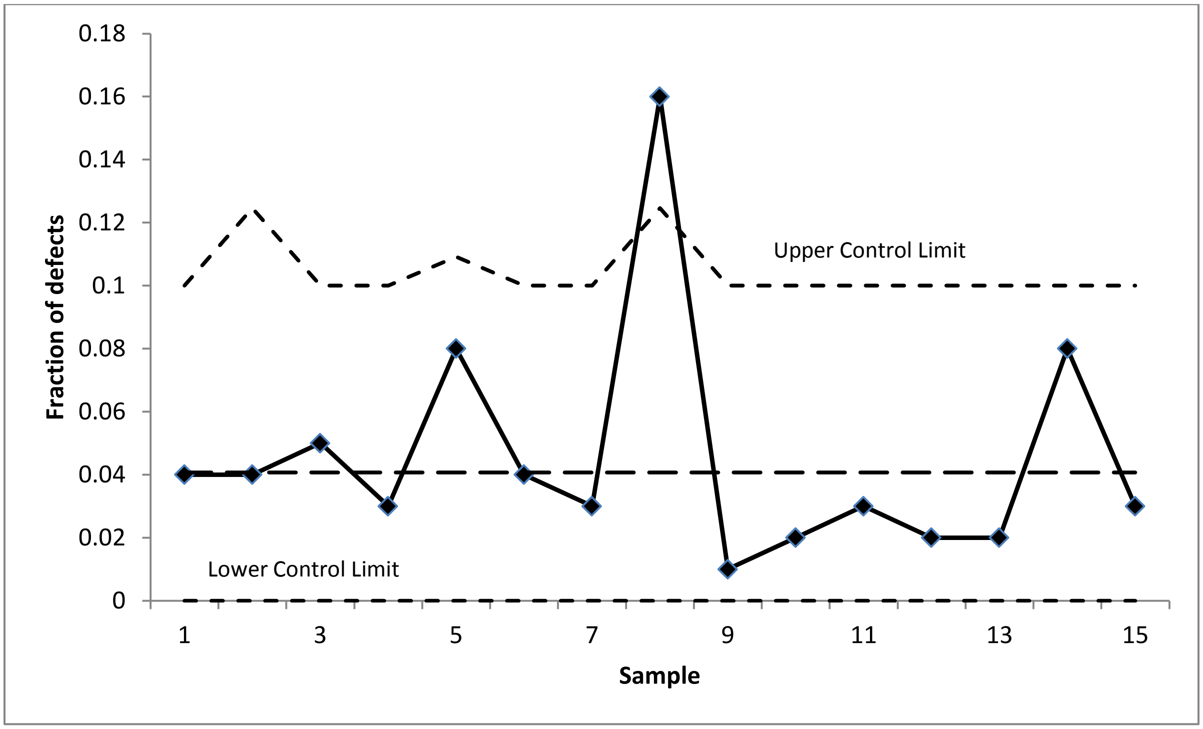
To understand these formulas, think of $\bar{p}$ as the inherent failure rate. As with X-bar charts, each control limit describes a confidence interval with plus or minus three standard errors. The chart below provides an example with calculations.

TABLE 4
LOWER AND UPPER CONTROL LIMITS FOR P-CHART

Sample	n	Defectives	Fraction	LCL	UCL
1	100	4	0.04	-0.02	0.10
2	50	2	0.04	-0.04	0.12
3	100	5	0.05	-0.02	0.10
4	100	3	0.03	-0.02	0.10
5	75	6	0.08	-0.03	0.11
6	100	4	0.04	-0.02	0.10
7	100	3	0.03	-0.02	0.10
8	50	8	0.16	-0.04	0.12
9	100	1	0.01	-0.02	0.10
10	100	2	0.02	-0.02	0.10
11	100	3	0.03	-0.02	0.10
12	100	2	0.02	-0.02	0.10
13	100	2	0.02	-0.02	0.10
14	100	8	0.08	-0.02	0.10
15	100	3	0.03	-0.02	0.10
Total	1375	56			
p-bar =	**0.04**				

When constructing a P-chart, if the calculation of the lower control limit yields a negative number, the lower control limit is set to zero. Similarly, if the calculation of the upper control limit yields a number greater than one, the upper control limit is set to one. (A proportion must be between zero and one.) For example, compare Table 4 with Figure 8 below. In the chart, the lower control limits have been set to zero.

FIGURE 8
P-CHART



As the control chart shows, there is one point outside the control limits (observation 8). Thus, there is an assignable cause present, and the process is out of control.

Process Capability

Even though a process may be in control, the product or service may not meet customer specifications. To answer the question of whether or not we can meet customer specifications, we need to know what our process is capable of. And to answer the question of what our process is capable of, we need our process to be in control. (For if our process is out of control, how can we predict what our process will produce?)

As a review, remember that *control limits* are used to determine whether variation is due to common causes or special causes. Since control limits are based on actual process data, control charts represent a strictly operational view. For this reason, they can be said to represent the *voice of process*.

If a process is in control, we can compute its *performance limits*. For a process in control, the performance limits are:

$$\text{Upper Performance Limit} = \bar{\bar{X}} + 3\bar{R}/d_2$$

$$\text{Lower Performance Limit} = \bar{\bar{X}} - 3\bar{R}/d_2$$

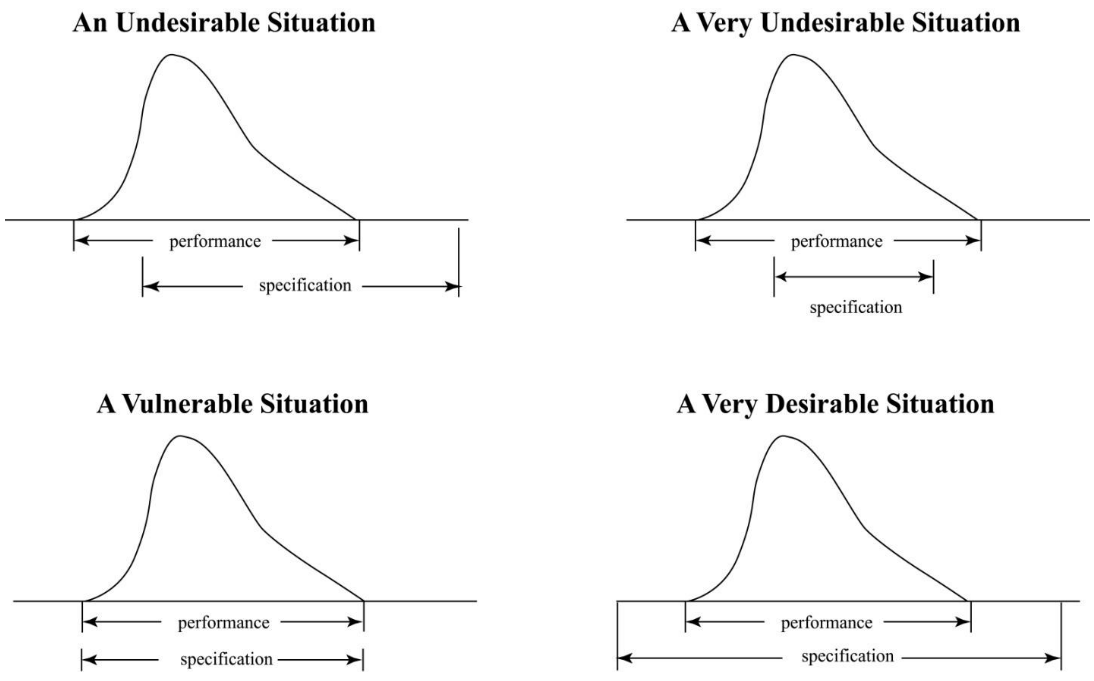
For a process in control, we usually assume that a unit of output from the process will be within these limits 99.7% of the time. (Recall from the section on X-bar charts that R/d_2 is our estimate of the process standard deviation.) Performance limits represent another *voice of process*. We emphasize again that performance limits are meaningless if the process is out of control.

We can now consider the *voice of the customer.* The customer's needs are described by *specification limits*. These are tolerances that represent the customer's definition of quality, and they often reflect fitness for use.

If we have performance limits (implying that our process is in control) and if we have specification limits, we can discuss *process capability*. Quite simply, a process is capable if its performance limits lie within the specification limits. In other words, a process is capable if it reliably (usually defined to be 99.7% of the time) meets the customer specifications.

It is important to note that a process can be in control, but not capable. To illustrate this, consider the examples in Figure 9. In all four situations, we are assuming that the process is in control. In the Very Desirable Situation, note that the performance limits are well within the specification limits. In the Vulnerable Situation, the process is capable, but just barely. And, in the two undesirable situations, note that the performance interval does not lie within the specifications.

FIGURE 9
SPECIFICATION LIMITS VS. PERFORMANCE LIMITS

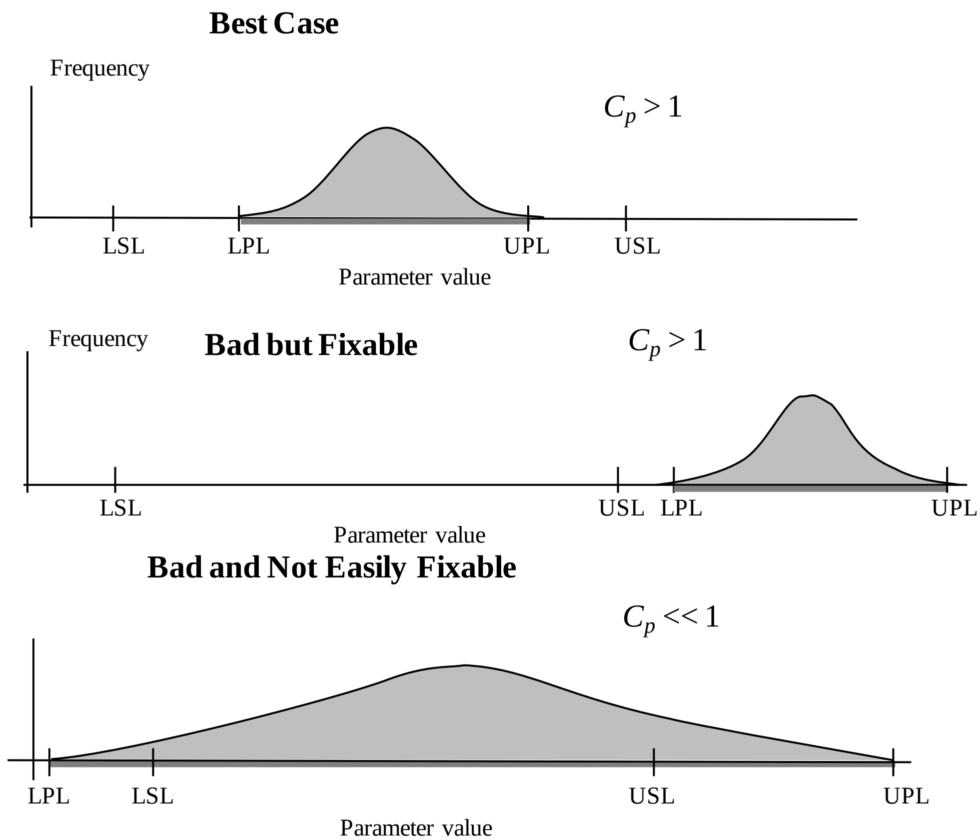


The notion of process capability is sometimes captured with what is called a process capability index, which is defined as:

$$C_p = \frac{\text{allowable process spread}}{\text{actual process spread}} = \frac{\text{upper spec limit} - \text{lower spec limit}}{6\sigma}$$

A process is said to be capable if $Cp \geq 1$. Note that the case of $Cp = 1$ corresponds to the Vulnerable Situation above. Some companies use a higher standard of reliability; i.e., they want to meet customer specifications more than 99.7% of the time. For example, Ford Motor Company requires that $Cp \geq 1.33$, and Motorola's Six Sigma Quality program requires that $Cp \geq 2$. The charts below depict some typical scenarios.

FIGURE 10
PROCESS CAPABILITY SCENARIOS



There are two implicit assumptions in the above charts that should be identified:

- The mean of a process in control can be in general shifted relatively easily
- The variation of a process in control cannot be in general shifted relatively easily.

This is why the middle case is considered bad but fixable, and the bottom case is considered bad and not easily fixable.

Finally, we emphasize once again that having a process under statistical control does not ensure that it is capable. The pictures below describe the four possibilities that can arise when comparing *in control* and *out of control* with *capable* and *not capable*. (A quality purist might object to the *out of control, capable* situation. The purist would maintain that if you are out of control, you should not be talking about capability at all.)

FIGURE 11
POSSIBLE PROCESS STATES