CHAPTER 9

Basic Statistical Methods and Control Chart Principles

If your experiment needs a statistician, you need a better experiment.

—Ernest Rutherford

There are complementary methods to measure the impact of a change or innovation on quality and performance—statistically and using control charts. Statistical methods to determine changes in performance rely on the performance of statistical tests to determine if changes in quality, performance, or other metrics are "statistically significant." Graphical approaches, on the other hand, use specialized charts known as statistical process control (SPC) charts (and specific rules to aid the interpretation of those graphs) to determine if a change in quality or performance is in fact occurring. This chapter discusses how both of these methods can be employed for quality and performance improvement.

Statistical Methods for Detecting Changes in Quality or Performance

I chose the epigraph at the start of this chapter rather tongue-in-cheek. My intent with the quotation isn't to say that statistics (and statisticians) should be avoided, but rather that the job of analytics professionals (including statisticians) is to make statistics more *accessible* and easily understood to all users of information through the use of the right tools in addressing the right problems (those of the quality and performance issues of the organization).

Statistics offers a wide range of methods with which to analyze quality and performance data. In this section, it is my intention only to introduce some basic statistical tests and terminology as they pertain to quality and performance improvement projects. I strongly encourage the reader to explore additional resources for more in-depth coverage of additional statistical topics.

Analytics is comprised of the tools, techniques, and systems necessary for obtaining deeper insight into the performance of an organization. Statistics is but one of those tools—an important tool to be sure, but not the only tool. I often say that analytics for healthcare quality improvement (QI) projects does not require *fancy* statistics, but rather *appropriate* statistics. A large part of the discussion in this book is about ensuring that high-quality data is available, that it is compiled into relevant indicators, and that it is made available to QI teams using structured methodologies. I have seen many analysis efforts become derailed because an analyst was overly concerned with applying complex statistical analysis, when this level of analysis was not necessary. Hopefully this section will demonstrate how statistics can be but one valuable tool in the quality and performance improvement toolbox.

Most QI methodologies do not require extensive statistical knowledge and the use of exotic statistical methods. Rather, statistical methods can best be used to identify any unusual variations in performance and to help pinpoint the causes of this variation.[1] This is more common in some methodologies (such as Six Sigma) that require more statistical analysis than others (such as Lean). Most QI projects can benefit from some basic statistical summaries such as average, median, and percentiles to report baseline information and current performance (as discussed in Chapter 6). Six Sigma delves deeper into more statistics, such as analysis of variance (ANOVA) and other tests, to either detect differences in performance or to increase the certainty of a result. In my experience, I have always tried to use statistics where appropriate and necessary to clarify and strengthen conclusions, not to search for a needle in a haystack.

If extreme statistical analyses are necessary to detect a change in performance, then a few scenarios are likely:

- The quality of data being used is poor.
- The wrong indicators were developed or used.
- There is no change in performance, or it is too small to be relevant.

Statistics are typically applied in one of two ways. *Descriptive statistics* are used to describe a large number of values or observations (representing an entire population or a sample thereof). There is a wide variety of descriptive statistics; the most commonly used ones include the mean (for example, the average weight of patients visiting a clinic) and the median (for example, length of stay [LOS] of emergency department patients). *Inferential statistics*,

on the other hand, analyze a sample of data to help evaluate and draw conclusions about a population. See Chapter 6 for a discussion on measures of central tendency and the use of descriptive statistics. This section will focus more on inferential statistics used to confirm the statistical significance of a change in performance.

Hypothesis Testing

Consider a facility that was observing longer than desired lengths of hospital stay and decided to implement a new streamlined patient discharge protocol. Prior to the implementation of the protocol, three months of baseline data showed an average length of hospital stay of 4.54 days. Following the implementation of the new protocol, the results were evaluated and the three-month post-implementation average length of hospital stay was 3.56 days. See Figure 9.1 for a graph illustrating the results. The difference pre- and post-implementation was 0.98 days. QI teams needed to determine whether this difference is the result of the new protocols, or whether the protocols made no difference and the observed difference is entirely by chance.

The process of determining whether this difference in values is due to natural variation and chance or the result of the change in process is known as *hypothesis testing* and typically involves a test of *statistical significance.* Because of natural variations in performance, no two sets of randomly selected data will ever be exactly the same even if the two samples are drawn from the same population of patients. Hypothesis testing and tests of statistical significance will help to determine if any observed differences between two (or more) groups are likely due to *actual* differences in the populations being studied (the result of a process change or other intervention), or if the observed differences are due to random variation and chance.[2]

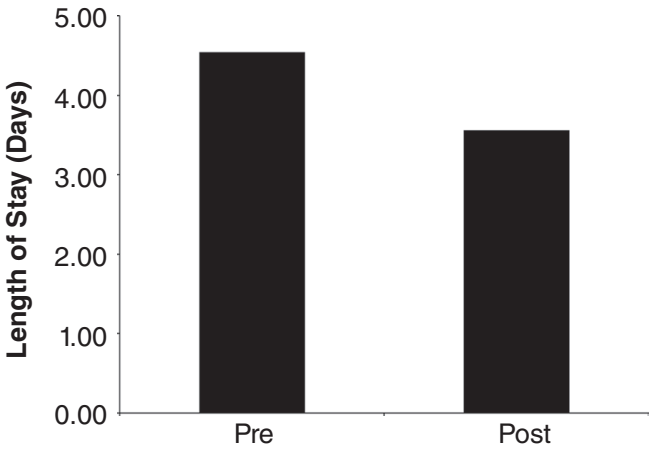

FIGURE 9.1 Sample Length of Hospital Stay before and after Implementation of New Discharge Protocols

Hypothesis testing starts with the assumption that there actually is no difference between the groups (that is, any observed differences are caused by random variation) unless there is compelling evidence to demonstrate otherwise. This is called the *null hypothesis*, and is expressed as:

$$H_0: \mu_1 = \mu_2 \text{ or } H_0: \mu_1 - \mu_2 = 0$$

The null hypothesis states that the means of data sets 1 and 2 are equivalent (that is, subtracting the mean of one data set from the mean of the other would return zero). In the case of the streamlined discharge protocol, H_0 states that there is no difference in the mean hospital LOS before and after the implementation of the new protocol—or that the new protocol had no effect on patient LOS. In the event that the null hypothesis is demonstrated to be false, the *alternative hypothesis* is then assumed to be true (that is, that the means of data sets 1 and 2 are not equal, and the differences observed between two data sets are likely *not* due to chance).

$$H_a: \mu_1 \neq \mu_1 \text{ or } H_a: \mu_1 - \mu_2 \neq 0$$

In our discharge protocol example, the alternative hypothesis is that there is a true difference in the means of discharge times measured before and after the protocol was introduced, suggesting that the new protocol *did* have an effect on patient LOS in hospital.

Comparing Performance between Two Groups

One common statistical test to evaluate situations like the pre-post evaluation of the protocol implementation is the *t*-test. The *t*-test is a statistical method that can be used to help determine if a statistical parameter (such as the mean, or average) is the same when compared between two groups (the null hypothesis) or different (the alternative hypothesis).[3]

LEARNING MORE ABOUT STATISTICS

If you are interested in learning more about the scientific and statistical basis behind hypothesis testing and statistical significance, there are many good statistical textbooks that cover these topics. I would also encourage you to review the resources listed in this chapter for more information. You can also visit this book's web site, http://HealthcareAnalyticsBook.com, for links to relevant resources.

A *t*-test can be used in two situations, depending on the number of samples. The *one-sample t*-test is used to compare one point of interest to a sample. For example, the one-sample *t*-test can be used to compare a sample's average performance to the target value of an indicator. If an emergency department's average left without being seen (LWBS) is 3.4 percent and the target LWBS rate is 2.5 percent, a one-sample *t*-test could be used to determine if the difference between actual performance and the target value is statistically significant.

A *two-sample t*-test is used to compare the performance of two groups. There are two varieties of this type of *t*-test; the best one to use depends on the two samples being tested. For example, if you are testing the hospital LOS at two different hospitals—Hospital A versus Hospital B—then the test to use is the *independent t*-test. The independent *t*-test assumes that the two populations are indeed independent, and are normally distributed. The independent *t*-test would not be appropriate in our example of the pre-post analysis of the streamlined discharge protocols. A pre-post study evaluation, also known as a *repeated measures* design, requires use of the *dependent t*-test variant.

A *t*-test can be applied to the "pre-change" and "post-change" groups in the example highlighted in Figure 9.1 to see if the difference of 0.98 days is statistically significant. Normally, statistical tests such as the *t*-test would be performed in a statistical software package or a spreadsheet with statistical capabilities. In the case of our discharge protocol example, running a dependent *t*-test on the two groups generates the following output from the statistical software used to run the test, which in this case is R:

```
t = -33.3139, df = 89, p-value < 2.2e-16
alternative hypothesis: true difference in means is not
  equal to 0
95 percent confidence interval:
 -1.0385635 -0.9216485
sample estimates:
mean of the differences
               -0.980106
```

What do these results mean? Consider if we repeated the discharge protocol pre-post test a second time and the results were similar, with a difference of 0.92 days; chances are our confidence in the results would improve, with two repeated tests demonstrating the same trend. Now, if we repeated the pre-post test 100 times and found that 95 out of these 100 times produced similar results, our confidence would be pretty high that the discharge protocols actually did decrease LOS for patients. If we repeated the test 100 times and found that the LOS of the protocol patients

was shorter than nonprotocol patients for only 60 of the trials, we would be less confident in the results. Finally, if each group had the shorter LOS 50 percent of the time over 100 repetitions of the pre-post test, we would likely deem that the protocols did not result in shorter lengths of stay.

Of course, it would be extremely time consuming and expensive to repeat trials such as our discharge protocol evaluation the necessary number of times to fully gauge confidence in results. This is where statistical tests are very useful, to determine how confident we can be that any differences observed are the result of a process change, or whether the observed difference likely occurred by chance.

More formally, the statistical significance is the probability of obtaining the observed (or more extreme) results if the null hypothesis were in fact true.[4] This chance or probability is calculated on the basis that the null hypothesis is correct; the smaller this chance, the stronger the evidence against the null hypothesis. Statistical analyses such as the *t*-test provides a quantitative assessment of this confidence with the *p*-value. A common *p*-value target often used in scientific research and QI is 0.05 or less, which means that there would be less than a 5 percent chance of obtaining the observed results if the null hypothesis was true.

In the previous example, the *p*-value is estimated to be less than $2.2e^{-16}$ by the computer software, suggesting that there is an extremely small chance of observing an LOS difference of 0.98 days if there was in fact no difference between the protocol and nonprotocol groups. This small *p*-value can provide the QI team with confidence that the discharge protocols actually are making a difference in lengths of stay of patients.

Another value reported by the computer software on the example above is the 95 percent confidence interval (CI). The CI is a computed range of numbers within which the true value is expected to lie.[5] In this case, the *t*-test calculated that the 95 percent CI, or the range of values in which the difference in LOS for protocol and nonprotocol patients can be expected to lie, is most likely between −1.039 and −0.922 (with rounding). In other words, we can say there is only a 1 in 20 chance that the true difference between the groups is *not* within that range.

If the CI included zero (for example, if the CI was between −0.5 and +0.5), it would imply that "no difference" in means, or a difference of zero, was as likely as other values within the CI. Because zero is not within the CI, however, it is likely that there is in fact a true difference.

The description of the *t*-test and test of significance earlier is to provide a flavor of how tests of statistical significance can help determine whether an actual change is occurring in a process. The *t*-test is ideal for comparing two groups, but what if more than two groups need to be tested at once, or you needed to test categorical data, or if other assumptions required of the data to perform a *t*-test are not met?

Comparing Performance of More Than Two Groups

What if an analyst needs to compare the performance of more than two groups? For example, consider the case where an analyst needs to compare the hospital LOS between three different facilities for patients who undergo a coronary artery bypass graft (CABG) procedure. Table 9.1 illustrates sample CABG patient LOS for three different facilities—A, B, and C. What would be the best approach to determine if there is a statistically significant difference between groups, or if the differences observed are simply the result of random variation?

The first instinct might be to perform pairwise comparisons—that is, compare A to B, A to C, and B to C using standard *t*-tests. This approach has two drawbacks. First, as the number of groups to compare grows, the number of pairwise comparisons that are required becomes unwieldy very quickly; after just seven groups, the number of comparisons required would be 21. Technically, we can get computers to run multiple *t*-tests quite simply, so the number of comparisons is not really a concern. However, as more *t*-tests are performed, the risk of obtaining a statistically significant difference *purely by chance* increases. Although there are corrections (such as the Bonferroni correction) that can be made for this when performing multiple tests (such as *t*-tests) on the same set of data, other statistical options are available.

In this case, and other cases when the *t*-test is not appropriate to use, there are other tests that can be used. The ANOVA (*an*alysis *o*f *va*riance) test is helpful when you need to compare more than two samples to each other to determine whether any of the sample means is statistically different from the other sample means.[6] A *one-way* ANOVA is valid if the groups are independent (as three sites would be), the data is normally distributed, and the variance in the populations is similar. Without going into the formulas, ANOVA works by comparing the variance *within* each group with the variance *between* the groups, and comparing the ratio of the within-group and between-group variance; the ratio is known as the *F-statistic*. If the variation between groups is much higher than the variation within groups, and the F-statistic exceeds a critical value, then a difference observed between the groups can be considered statistically significant. (Note that the critical F-statistic value can be looked up on a specially designed table of critical values, but more often than not, this will be performed by computer.)

TABLE 9.1 Sample Hospital LOS for CABG Patients for Three Sites

Facility	Hospital Length of Stay (days) for CABG Patients
Hospital A	8.5
Hospital B	9.8
Hospital C	8.9

Comparing Observations of Normal and Ordinal Values

What if the data that needs to be compared between two (or more) groups is nominal or ordinal, that is, data for which a mean cannot be generated for a test like a *t*-test or an ANOVA? A *chi-square test* is useful for determining if there is in fact a relationship between two categorical variables,[7] and would be appropriate in this situation. Rather than comparing a statistic such as the mean of two or more groups, the chi-square sums the squared differences observed and expected frequency of observations within each category.[8]

The different scenarios for which you may consider using the different types of ANOVA tests include:

- **One-way between groups.** Use the one-way between-groups ANOVA when the performance of three or more groups needs to be compared (as in the above example).
- **One-way repeated measures.** When performance has been measured a few times (for example, prior to a QI project, during the execution of the QI project, and after the QI project), the one-way repeated measures ANOVA can test for a statistically significant change in performance.
- **Two-way between groups.** This is used when looking for more complex interactions. For example, if comparing hospital LOS for CABG procedures, QI teams may be interested in understanding the interaction between whether the site is a teaching hospital or community hospital, and the overall hospital LOS.
- **Two-way repeated measures.** This is similar to the one-way repeated measure, but includes an interaction effect (for example, if you wanted to test whether type of X-ray had any impact on changes in the processing time of diagnostic imaging patients).

Lessons Learned

Many software packages include statistical tests built in, so it is relatively simple to perform a *t*-test, ANOVA, or other statistical test. Keep in mind that although software makes it easy, applying a statistical test to a set of numbers on a spreadsheet may not achieve accurate results. Before proceeding with a statistical test, always ensure that the basic assumptions required of the test are met (for example, does the test require normally distributed data?), and that the test can provide the type of answer being sought.

Graphical Methods for Detecting Changes in Quality or Performance

Hypothesis testing and tests of statistical significance are one method of determining if any change is occurring in quality and performance within a healthcare organization (HCO). The challenge with statistical tests, however, is that most require large samples of data to be accurate, and can be cumbersome to run every time the performance or quality associated with a process needs to be measured. Another issue is that they tend to utilize aggregated data (for example, determining if the mean of two samples is statistically significant). If all analysis is done in aggregate, it is possible to lose sight of variations in the way that processes are performed and in the outcomes of those processes. One danger of solely relying on aggregate data and statistical analysis is that although *average* values of data sets might be meeting a target value, individual performance and quality may vary so widely that the inconsistency poses a risk to patient safety.

Control charts are a very common visual approach to evaluate performance and quality with associated rules to determine if a process is in control and improving (or getting worse). Graphical analysis is a highly regarded approach in healthcare QI. It has been recommended that "methods for the analysis of data should be almost exclusively graphical (always including a run order plot), with minimum of aggregation of the data before the initial graphical display."[9]

Graphical analysis of performance data provides visual evidence of the variability inherent in a process. Measuring and understanding the variation in a process is merited because it is "important to eliminate extraneous process variation wherever possible, while moving well-defined metrics toward their target values."[10]

Variation in Performance

There are many different causes of variation in performance. Causes can range from differences in the way individuals perform tasks to calibration differences in equipment. All the different causes of variation, however, can be divided into two categories:

1. **Common (or random).** These are causes of variation that are inherent in the work being performed, affect everyone who performs the work, and affects all outcomes of the process.[11] Common cause variation is generally predictable and expected and can be caused by myriad reasons ranging from complexity of patient needs to materials available. An example is the natural variations in the time it takes to triage an emergency patient; although every triage is different because each

patient presentation is unique, there is a typical range in the time it normally takes to complete a patient triage.

2. **Special (or assignable).** These are causes of variation that are *external* to the system or the work being performed, and do not occur all the time; they arise due to special circumstances.[12] An example of special-cause variation would be a nurse who takes significantly longer to triage patients than is typical. This may be caused, for example, by a nurse who is improperly trained on the use of the triage system.

Quality not only means that a process is able to meet target performance on average, but it must accomplish this within certain tolerances and consistency; that is, it must be considered *stable.* A stable process refers to one that is free of special-cause variation. The term "in control" is also used when variations in data are present and exhibit a pattern that is random.[13] (Note that "in control" does not mean an *absence* of variation, since even the best processes will demonstrate some variability.) In addition, a statistically "in control" process may still not be acceptable if the variation falls outside a range that is deemed safe or otherwise acceptable by the HCO, clinical experts, or governing bodies.

One of the tenets of process improvement is that a process must be stable before it can be improved. Strictly speaking, even the act of changing a process from one that is out of statistical control to one that is within statistical control (i.e., with reduced variability in the output of the process) can be considered an improvement.

Almost every report showing any metric will display some variability in the performance of a process. No process in healthcare is so stable that it is able to produce the same results every single time. The question is how to determine how much variability in a process is too much, and how much is acceptable. Statistical process control (SPC) is a technique that QI teams use to improve, evaluate, predict, and control process through control charts.[14] In essence, an SPC chart is the chronological time series plot of an indicator, metric, or other important variable and is used for, among other things, analyzing the occurrence of variations within a process. Many statistics can be plotted on an SPC chart, including averages, proportions, rates, or other quantities of interest.[15]

Rather than simply plotting values on a graph, one of the unique components of SPC charts is the addition of upper and lower reference thresholds, which are called *control limits*. The control limits are calculated based on the process data itself; the plotted points of data must almost always fall within the control limit boundaries, as the control limits specify the natural range of variation within the data. Points falling outside of the control limit boundaries "may indicate that all data were not produced by the same process, either because of a lack of standardization or because a change in the

process may have occurred."[16] When looking for changes in performance, then, a reduction in variation and/or a deliberate and consistent shift to values near (or outside of) the control limits may signal that changes in a process are occurring.

Statistical Process Control Chart Basics

Many analytics tools with even basic visualization capabilities can be used to generate SPC and run charts. There are some stand-alone software tools (as well as plug-ins for Microsoft Excel) that can generate excellent SPC and run charts (and provide other visualization tools for quality and performance improvement). Even without dedicated SPC generation capabilities, very useful charts similar to SPC charts can be generated with the basic graphics capabilities of most analytics and business intelligence software provided that the basics of SPC charts are understood (and a little creativity is applied).

Tip

For a list of software tools that can be used to generate SPC and run charts, and examples on how to build them, please visit this book's web site at http://HealthcareAnalyticsBook.com.

See Figure 9.2 for a sample control chart. The important features of control charts are:

- **Data points** that represent a quality or performance indicator associated with a process (and may be a statistic such as mean or proportion).
- A **centerline (CL)** that is drawn at the mean value of the statistic.
- An **upper control limit (UCL)** and **lower control limit (LCL)**, which represent the values outside which performance of the process is considered statistically unlikely.

The centerline of a control chart is drawn at the mean ($\bar{x}$) or average value of the observations being plotted. Upper and lower control limits are typically drawn at $+3\sigma$ and -3σ (where σ is one standard deviation) from the centerline. The sample SPC chart in Figure 9.2 demonstrates a process that would generally be considered to be in control. All the data points are randomly scattered around the mean ($\bar{x} = 9.20$) and all fall within the upper control limit (UCL = CL$+3\sigma$ = 9.47) and the lower control limit (LCL = CL -3σ = 8.94).

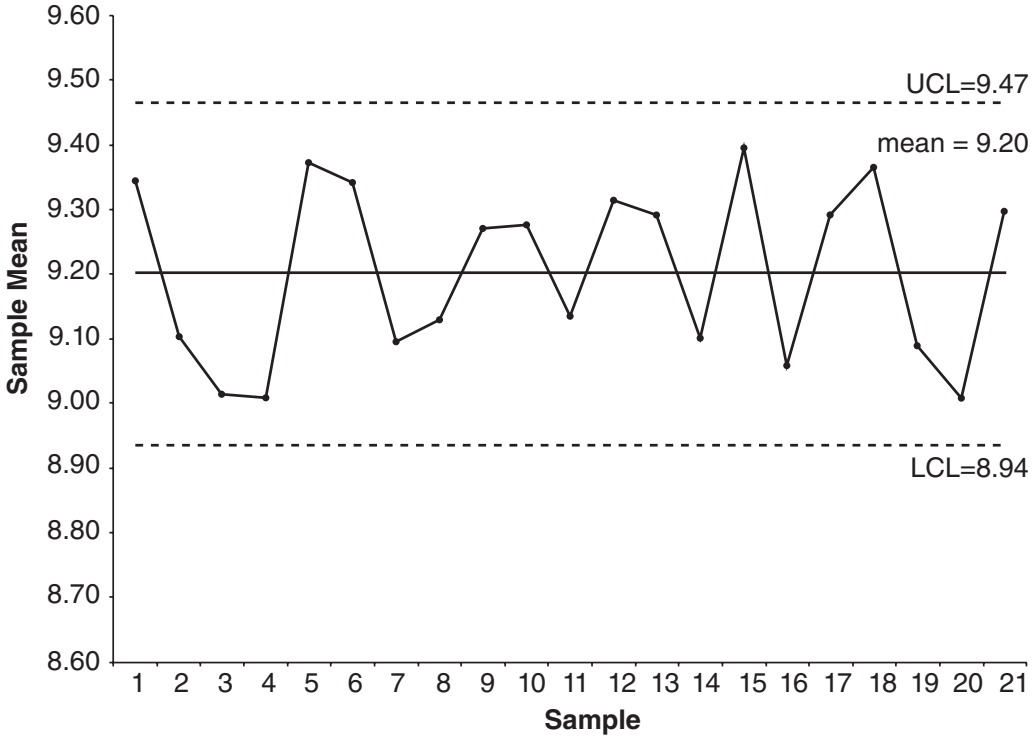

FIGURE 9.2 Sample Control Chart

When a change in process occurs as the result of a QI activity (or some other cause of change), the SPC chart can be used to monitor if a change in performance and/or outcomes has occurred. The limits of an SPC chart should be revised when the existing limits are no longer relevant or useful. When a shift in process occurs, it is helpful to reset the mean and control limit lines to better isolate the new process from the old process in the chart. If new mean and control limits are not reset, the existing mean and control limits will expand (or otherwise adjust) as new data is added to the calculations. This may make it more difficult to identify any actual change in outcomes or performance.

Figure 9.3 illustrates an updated version of the Figure 9.2 chart, this time with new data points added after a process change, and with the new mean and control limits added. With this chart, the baseline performance is shown, the time at which the new process was introduced is clearly evident, and the performance of the new process stands out from the baseline data. When the SPC chart is drawn in this way, the new performance can be evaluated not only to see if the desired target performance is being met but also to investigate the stability of the new process and whether it is in control.

Data Considerations for Statistical Process Control Charts

As mentioned earlier, when developing analytics for quality and performance improvement, it is important to use the right data, and that the

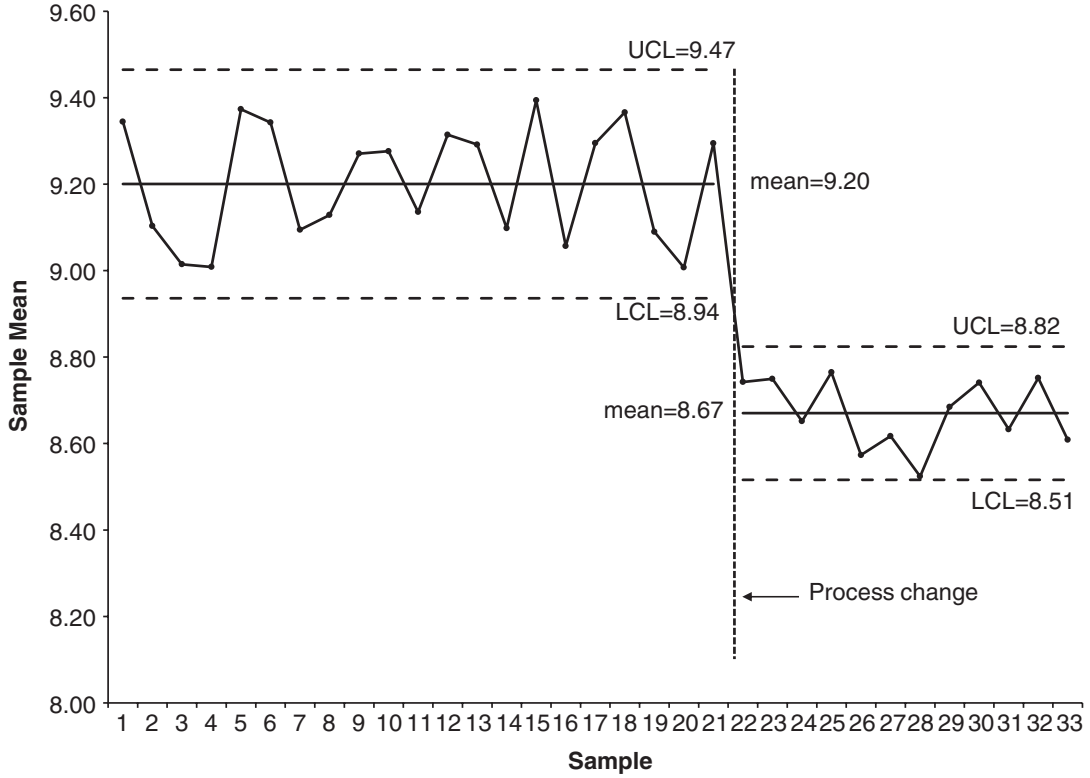

FIGURE 9.3 Sample Control Chart Highlighting Performance before and after a Change in Process

underlying assumptions of any statistical test or other tool are being met, otherwise inaccurate results are possible. SPC charts are no different; there are a few data considerations to ensure that SPC charts are accurate and that any conclusions drawn from them are valid.

When obtaining data for SPC charts, it is recommended that there be a minimum of 20 to 30 consecutive subgroups,[17] which are comprised of at least 100 consecutive observations.[18] For example, if the sample SPC chart in Figure 9.2 was evaluating the emergency department LOS for patients to be admitted, it would be ideal to plot at least 100 admissions over a period of 20 days for optimal validity of the control chart. In this example, assuming that there are at least five admissions from the emergency department every day, each subgroup would be the average LOS for admitted patients over a 20-day period. The mean for each subgroup would be plotted on the *y* axis, and each day would be plotted *in chronological order* along the *x* axis. This number of observations is necessary because, as in most evaluation methods, insufficient data may lead to inaccurate results.

Graphically Displaying the Stability of a Process

As long as the basic data requirements are met, a change in process can be quite clearly identified on an SPC chart. It takes more than simply

"eyeballing" it, however, to determine a change in performance or to detect undesirable variations and trends in the data. Figure 9.4 outlines a set of rules that can be used to determine the stability of a process based on data plotted on a control chart. The rules help quality teams to interpret the process patterns on the charts, specifically to special causes of variation. Figure 9.5 is a visual representation of the rules specified in Figure 9.4.

Different patterns that manifest on control charts may signal different issues or different causes of variation. In manufacturing and other industries, many of the rules help detect problems with machinery and other manufacturing issues. Healthcare is in many ways much more complex than manufacturing, so changes in control charts may be caused by any number of reasons. For example, a single point above the UCL (or below the LCL) may indicate that a single abnormality occurred that day (with possible causes ranging from a multicasualty incident causing a surge of patients at the emergency department to a lab equipment glitch requiring all blood work to be redone).

Some of the indications in Figure 9.4 and Figure 9.5 that a significant change has occurred or a process may not be in control include:

- Eight (or more) points in a row above or below the centerline
- Four of five points between $+1\sigma$ and $+2\sigma$ (or -1σ and -2σ)
- Two of three points between $+2\sigma$ and the UCL (or -2σ and the LCL)
- Any one point above the UCL or below the LCL

When reviewing SPC charts, the important point to remember is that any time the chart stops exhibiting random variation and patterns begin to manifest in one or more of the ways described, it is an indication that *something* is causing a process to change, whether as the result of deliberate intention or due to inconsistent practices, performance, or other causes.

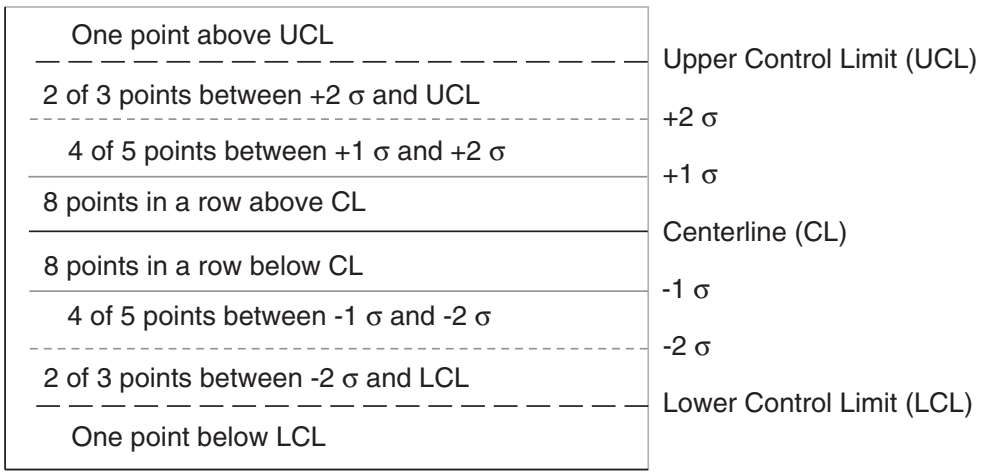

FIGURE 9.4 Detecting Stability in a Process Using Control Charts

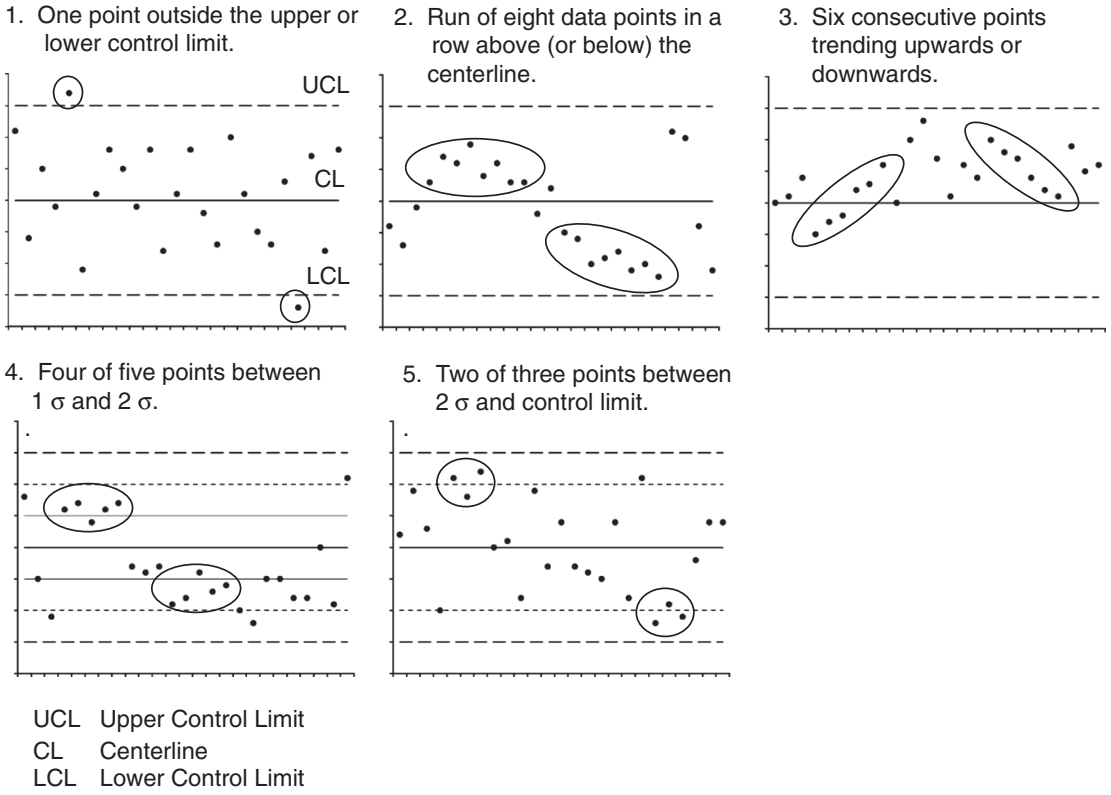

FIGURE 9.5 Sample SPC Chart Rule Violations

Keep in mind that even though a value is within the UCL and the LCL, it might not be acceptable from a clinical perspective. In addition to the σ and control limit values on an SPC, often a *specification limit* will be added. The specification limit is the range of values that is acceptable to the customer (or, in this case, to the patient and/or best clinical practice guidelines).

Consider, for example, an HCO that is improving its care of patients who experience a stroke. To achieve acceptable clinical standards, the HCO might identify a target duration of three hours from the time a patient experiences a stroke to the time rt-PA is administered. If rt-PA administration times for patients are plotted on an SPC, a specification limit of three hours would be added as a visual indicator. In this case, rt-PA administration times that were within the UCL (meaning within statistical control) but outside of the specification limit would still be a cause for concern. As the assessment and treatment of stroke patients was improved and variation in performance decreased, it is likely that UCL and LCL would tighten to the point where they were inside (or very close to) the specification limit.

Types of Statistical Control Charts

This chapter discusses the fundamentals of SPC charts, and there are actually several different kinds of SPC charts that can be used.

TABLE 9.2 Examples of Common Control Chart Types and How They Are Used

Data Type	Chart Type	Usage
Discrete	P-chart	Percentages
	C-chart	Counts
	U-chart	Rates
	T-chart	Time between rare events
Continuous	I-Chart (Sometimes called X-MR, where MR = moving range)	Individually measured data points
	X-Bar	Subgroups of data at the same point in time

Source: www.qihub.scot.nhs.uk/knowledge-centre/quality-improvement-tools/shewhart-control-charts.aspx.

The selection of a type of control chart depends on several factors,[19] including:

- The type of data being used (continuous versus discrete).
- Sample size available.
- What is being plotted (such as percentages, counts, rates, or time between rare events).

Table 9.2 shows a collection of common control chart types, what type of data they are appropriate for, and how they are used.

A summary of the different types of SPC charts, and a guide to selecting the best one for your particular needs, is downloadable from the book's web site at http://HealthcareAnalyticsBook.com.

Putting It Together

Critical to the development of analytics is the knowledge of who needs to use information, and how they can best make use of it. For example, QI experts working with Six Sigma and other methodologies often use SPC charts and statistical analysis in raw form to study the performance of a process and its resultant quality. In doing so, they are using SPC charts and statistics to analyze one (or a few) process changes in depth to uncover opportunities for further performance improvements and changes in quality. Healthcare executives, managers, and other healthcare leaders, who are usually concerned with the operations of an entire unit, hospital, or system,

on the other hand, are not likely to benefit from SPC charts and *t*-tests outlining every performance indicator relevant to a QI activity.

Knowing who needs what kind of information is important to developing effective analytics. Analytics is able to provide deeper insight into the performance of an HCO, and is designed to make decision making easier for QI teams and healthcare leaders. Statistical analysis and SPC charts were invented long before modern analytics. The power of analytics is in the synthesis of information and insight from statistical and graphical analysis into more meaningful and easier-to-interpret formats where appropriate, and presentation of more detailed information when necessary.

Rather than simply displaying a collection of graphs and charts, dashboards and reports can be made more analytical by embedding insights gained from statistical analyses and graphical analysis. Figure 9.6 illustrates a sample dashboard for a diagnostic imaging department displaying several key performance indicators for that unit, baseline performance for the previous six months, the indicators' respective targets, the current month's performance data, the performance for the previous month, and a trendline of performance over the last eight weeks. (See Chapter 10 for a discussion on dashboard and data visualization design.)

The dashboard is a simple representation of these indicators, provides an overview of performance, and also includes embedded insight from both statistical analysis and SPC rules. Using superscript values next to the current month's data and descriptive text in the "Notes" section, the dashboard in Figure 9.6 indicates to the user that there was: (1) a statistically significant difference in performance on indicator A between last month and the current month, and (2) an SPC rule violation for indicator D. The statistically significant ($p < 0.05$) decrease in X-ray order to patient pickup times (indicator A) from the last month to the current month may suggest to QI teams and DI managers that efforts to improve processes associated with indicator A might be having a positive effect.

Diagnostic Imaging Department - Performance Dashboard **December**

Indicator	Target	Six-Month Baseline	Last Month	Current Month	8-Week Trend
A. X-Ray order to patient pickup	10.0	33.2	22.5	13.4 [1]	
B. Pickup to X-Ray start	10.0	12.2	11.9	11.8	
C. X-Ray duration	15.0	14.5	15.2	15.1	
D. X-Ray stop to patient return	10.0	12.1	11.9	17.5 [2]	

(All values are in minutes)

Notes:
1) p<0.05 - month over month performance
2) SPC rule violation: two of three points between +2σ and UCL (**Link** to SPC chart)

FIGURE 9.6 Sample Dashboard Including Embedded Analytics

Although dashboards are often designed to be printed out, they are most useful when designed to be viewed on a screen (such as a computer display, smartphone, or tablet) and interacted with. Such interaction allows users to drill down into more detail (such as to view a control chart for performance indicators), select additional or other indicators to view, and even manipulate date ranges and other dashboard parameters. The example in Figure 9.6 shows that the SPC rule violation alert also includes a link to the actual SPC chart that triggered the violation, so an interested user of this dashboard would be able to launch an additional view that contains the desired additional material. The design objective is that the most important features and insight of this dashboard are immediately visible, with additional detail (links to an SPC chart) available via a simple click.

Statistical and nonstatistical approaches to evaluating quality and performance are not at odds but are entirely complementary. When used in concert, and synthesized on interactive information displays such as dashboards and other analytical tools, users of the information can quickly identify where and when performance needs to be improved, and perhaps even what actions need to be taken.

Developing effective analytical tools does require effort and expertise. Analytical teams must understand the context of the data, know control charts (and their associated performance variation rules are used), and be comfortable in enabling the basic statistical analyses required. Although the mechanics of performing the required statistical analyses and building the appropriate charts, graphs, and other data displays are possible in many analytical tools, it still requires a knowledgeable analytics team working closely in concert with quality and performance improvement experts to design concise and effective analytical information displays that provide insight about quality and performance issues, help suggest appropriate mitigation steps, and monitor ongoing results.

Notes

1. Thomas P. Ryan, *Statistical Methods for Quality Improvement, 3rd ed.* (Hoboken, NJ: John Wiley & Sons, 2011), 9.
2. Michael L. George et al., *The Lean Six Sigma Pocket Toolbook: A Quick Reference Guide to 100 Tools for Improving Speed and Quality* (New York: McGraw-Hill, 2005), 156.
3. Ibid, 161.
4. David Freedman et al., *Statistics*, 2nd ed. (New York: W.W. Norton & Company, 1991), 433–435.
5. Ibid., 348–350.

6. Thomas H. Wonnacott and Ronald J. Wonnacott, *Introductory Statistics*, 5th ed. (New York: John Wiley & Sons, 1990), 325.
7. Glenn J. Myatt, *Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining* (Hoboken, NJ: John Wiley & Sons, 2007), 67.
8. George et al., *The Lean Six Sigma Pocket Toolbook*, 182.
9. Sandra Murray and Lloyd P. Provost, *The Health Care Data Guide: Learning from Data for Improvement*, Kindle ed. (Hoboken, NJ: John Wiley & Sons, 2011), location 2134.
10. Frederick W. Faltin, Ron S. Kenett, and Fabrizio Ruggeri, eds., *Statistical Methods in Healthcare*, 1st ed. (Hoboken, NJ: John Wiley & Sons, 2011), 253.
11. Gerald .J Langley, Ronald D. Moen, Kevin M. Nolan, Thomas W. Nolan, Clifford L Norman, and Lloyd P. Provost, *The Improvement Guide: A Practical Approach to Enhancing Organizational Performance*, Kindle ed. (Hoboken, NJ: John Wiley & Sons, 2009), locations 1694–1697.
12. Ibid.
13. Issa Bass, *Six Sigma Statistics with Excel and Minitab*, Kindle ed. (New York: McGraw-Hill, 2007), location 2253.
14. ASQ, "Statistical Process Control," http://asq.org/learn-about-quality/statistical-process-control/overview/overview.html.
15. Fatlin, Kenett, and Ruggeri, *Statistical Methods in Healthcare*, 254.
16. Ibid.
17. Murray and Provost, *The Health Care Data Guide,* location 3633.
18. George et al., *The Lean Six Sigma Pocket Toolbook*, 122.
19. NHSScotland Quality Improvement Hub, "Shewhart Control Charts," www.qihub.scot.nhs.uk/knowledge-centre/quality-improvement-tools/shewhart-control-charts.aspx.