Wysowl Pty Ltd

ACN 010 677 022

18 Miller Street
CHERMSIDE QLD 4032
Ph: Intl + 61 7 3359 9918
Mob: Intl + 61 407 149 966
john@wysowl.com.au
www.wysowl.com.au

Wysowl Pty Ltd
Newsletter Number 34
January 2017

The Enormous Initial *STATISTICAL* Mistake

or

Why the use of three sigma limits can be the short road to statistical hades.

INTRODUCTION

Every now and again each of us discovers that something we have been doing for many years is, in fact, poor practice. This can be difficult to accept, especially when this poor practice is exactly what we were taught to do during our education. The internal conflict this causes can be gut-wrenching and confusing. To illustrate, nearly all readers were taught to develop control limits for a process by setting these limits two or three standard deviations either side of the average. It is the proposition of this document that this is poor practice that can conceal more than it reveals and which has the potential to cause significant economic loss.

The primary purpose for calculating upper and lower control limits is to determine whether the process under examination is in a state of statistical control, or stable. Such a study should separate random (stable) from non-random (unstable) variation and avoid confusing one for the other.

This document will explain how the practice of setting control limits two or three standard deviations either side of the average was developed for descriptive statistics and distribution theory, and whilst the practice works well in that environment, it regularly fails when transferred to analytical studies and time series data.

TERMINOLOGY and ASSUMPTIONS

Statistical studies may be considered in two broad groupings: descriptive studies and analytical (or inferential) studies.

Descriptive studies. A descriptive study is one where we are concerned with the data to hand only, and wish to describe it in statistical terms. A common way to do this is to present a frequency distribution complete with a measure of location (sometimes called central tendency) such as mean, median or mode, a measure of dispersion (perhaps standard deviation, variance or range) and perhaps an indication of shape, such as a measure of normality, skewness and/or kurtosis. Descriptive studies do not attempt to make predictions or to infer what the data might look like if new data were collected again, perhaps at some time in the future. Data for a descriptive study are not time ordered. They can be likened to a photograph, where even if the data were created in time ordered sequence, this sequence is lost, as is the case when data are presented as a frequency distribution. The data become frozen; a snapshot in time.

Analytical studies. Here we are using a sample to draw conclusions about a population. In most cases, analytical studies (also known as inferential or inductive studies) use data in time ordered sequence. Commonly, but not always, data are presented as a time-ordered plot of points, and they resemble a movie, frame by frame, rather than a photograph. In this case, it is possible to see how the data behave over time. Perhaps they are trending, or following a repeatable pattern. Usually, those conducting the study are trying to understand the process that produced the data, perhaps in an attempt to improve it; or perhaps they wish to make a prediction about what data created in the near future will look like either if nothing changes or if a specified change were to be made. A Shewhart control chart is a common tool for analytical studies.

Underlying assumptions. When we make a frequency distribution for a descriptive study we are assuming that the data to hand are from a single distribution, population or universe. In terms that Dr Shewhart used, we are assuming that all the data come from the same constant cause system. For many descriptive studies, such as stocktakes, quality control checks on batches of incoming material and census studies or similar, this assumption is either generally correct or it makes little to no difference it is not, and a frequency distribution is a good way to display the data. However, if we are using time-ordered data for an analytical study, this assumption, if made, is likely to be wrong more often than it is likely to be correct because as Dr Shewhart observed, *"No universe exists until control is established* $_{(1)}$, and because statistical control does not happen by accident. More often, it is the result of determined efforts. Shewhart also said, *"Formal distribution theory will give valid predictions only in a state of statistical control"* $_{((2)}$. An analytical study makes no such assumptions. Frequently the question to hand at the outset of an analytical study is, "do these data exhibit statistical control"; or "do these data come from a single constant cause system"; or "is anything changing?"

WHAT IS STANDARD DEVIATION?

Standard deviation means the root-mean-square-deviation of the data. In this document "standard deviation" is the response received when one presses the σ button on a calculator or types the excel command =STDEVP. For the sample standard deviation one would press the σ_{n-1} button or type the excel command =STDEV, but for the purpose of this document there is no practical difference between these approaches. There are other ways to calculate standard deviation for binomially or Poisson distributed data, and Dr Shewhart's control charts use the standard error of the mean to develop an estimate of the standard deviation of sub-groups. Many readers are scientists who are quite aware of the varying ways standard deviation can be calculated. However, in the great majority of cases, when managers and scientists use the term "standard deviation" or "sigma" they mean the root-mean-square-deviation. Because standard deviation is a general term with several potential definitions and formulae, this document uses the term RMS Deviation, which is a specific term.

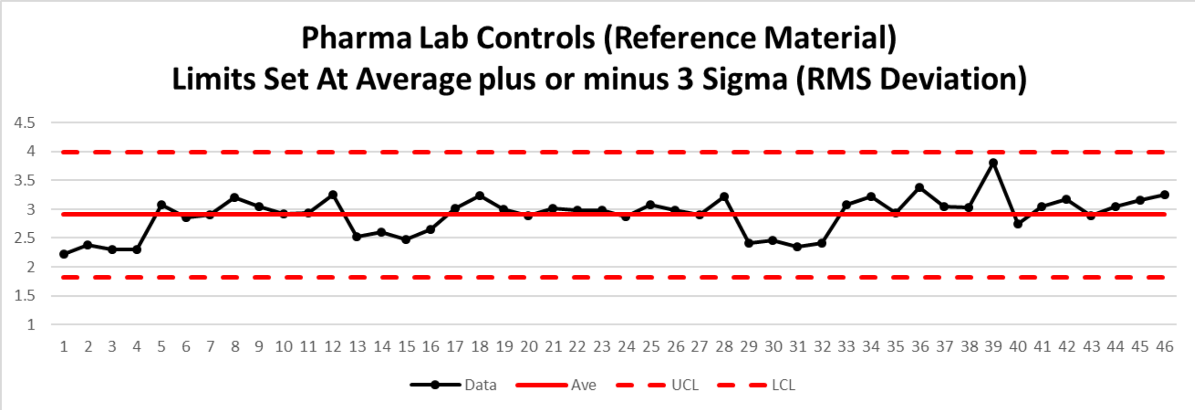
RMS Deviation is calculated in the reverse order of its full name. First, we calculate the difference between each datum and the average. These are the deviations. Next these deviations are squared to give the squared deviations. The third step is to calculate a mean of these squared deviations. This gives the variance. Finally, the square root of the variance yields the RMS Deviation or σ_{RMS}, the root-mean-square-deviation.

ILLUSTRATING THE ISSUES

To illustrate the issue underlying the hypothesis, at Figure 1 is a set of Laboratory Controls (or reference material) data from a pharmaceuticals laboratory in the USA. This is a reproduction of the chart produced by the laboratory manager and his technical staff. The limits are three RMS Deviations either side of the average.

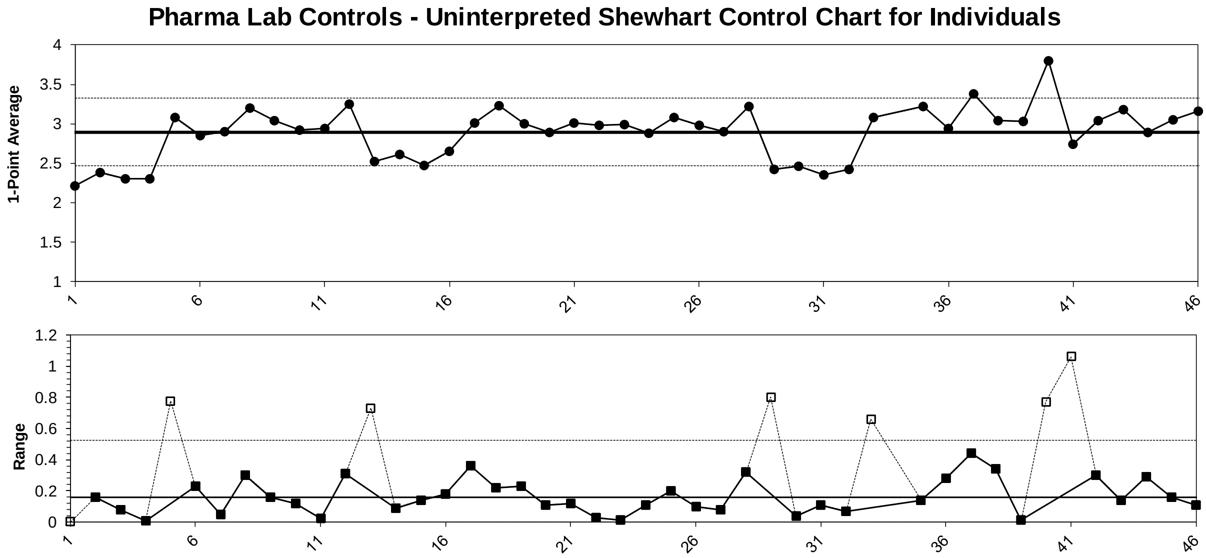
The laboratory manager and the technical staff concluded that there were no points beyond the control limits and that the data exhibited satisfactory statistical control, or were stable. The average is 2.9 and the RMS Deviation is 0.33.

FIGURE 1



The same data were then used to make a Shewhart control chart for individuals, seen at Figure 2 (often called a Single Point and Moving Range control chart). This type of control chart extrapolates from the average of the Moving Ranges (lower) chart to develop control limits, and has been shown to be more robust and better able to detect disturbances in the data.

FIGURE 2



In a Shewhart control chart the Moving Ranges (lower) chart displays the dispersion or variation in the data and the top chart (the individual data) shows the location. This is an important distinction because variable data can change in location, but not in variability, or the opposite, or in any possible combination. Knowing whether a change or disturbance has impacted the location, the variation, or both, can be important. This separate information about location and variation is missing from charts like the one at Figure 1, which has only one plot of points.

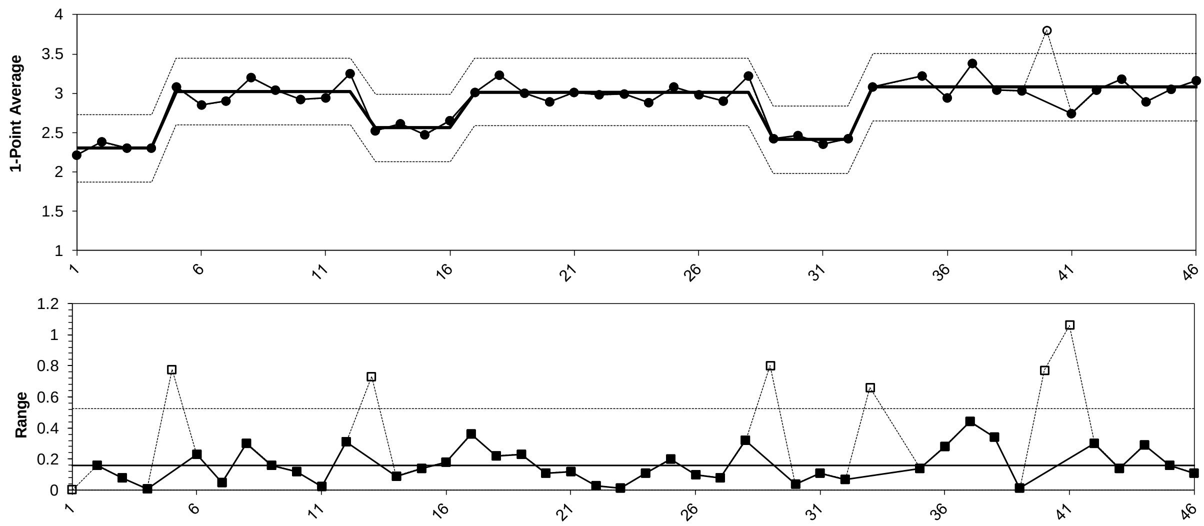
When reading a Shewhart control chart it is good practice to read the Moving Ranges chart first. In this case four individual "spikes" that extend beyond the Upper Control Limit can be seen. For this type of chart this is strong evidence that the location or average of the individual data has changed at each of these “spikes”. In addition, the two consecutive moving Ranges points that correspond with data point 40 lie beyond the UCL of the Moving Ranges chart, indication a disturbance/upset, or assignable/special cause.

The broken lines and open points indicate that whilst these signals were received, and remain visible, they have been removed from the calculations to provide an uncorrupted or more reliable estimate for R-bar, the centre line of the Moving Ranges chart. Because the formulae extrapolate the control limits for both charts from R-bar it can be said that if R-bar is a reliable estimator, the spread of the control limits will be similarly reliable. If R-bar is a corrupted or unreliable estimate, the control limits will be similarly unreliable or corrupted. When using control charts, one is always attempting to separate random from non-random variation. The Moving Ranges chart at Figure 2 shows five instances of non-randomness. Four are changes to the mean of the process, and one is an assignable or special cause. It is good practice to treat the non-random variation separately from the random or systematic variation. Only the random or systematic variation in the Moving Ranges chart at Figure 2 (solid lines and filled points) has been used to develop control limits, in order to separate random from non-random variation.

At Figure 3 the chart for the individual data (Figure 2) has been “split” where the moving Ranges chart indicated a change in the average, to better interpret the data. The chart shows six “systems” of data, and an upset, or assignable/special cause at point 40.

FIGURE 3

Pharma Lab Controls - Interpreted Shewhart Control Chart for Individuals



Once the changes in the control chart were noted an investigation soon revealed that the changes in the average of the process were occurring at shift change. Both day and night shift had similar variability, but the average result was lower for night shift than it was for day shift. The process average changed at each shift change. The cause of the disturbance at point 40 was not discovered.

The RMS Deviation of all the data, as noted in Figure 1 was 0.33. The Shewhart control chart, using Shewhart’s methods, calculated a RMS Deviation for the systematic data of 0.142. The average plus or minus three RMS Deviations approach used in Figure 1 combines all variation, random and non-random. The Shewhart control chart separates random from non-random variation and the estimate for RMS Deviation of 0.142 includes only the random or systematic variation in this calculation. The Shewhart chart shows that the underlying, systematic variability is less than half that indicated by the chart at Figure 1. Put another way, the disturbances more than doubled the estimate for RMS Deviation, and gave far wider limits when the approach of average plus or minus three RMS Deviations was used to develop limits.

DISCUSSION

Interpreting the data. The charts at Figures 1 and 3 are very different, and are likely to generate quite different interpretations from most of the people who view them. Most people see only stable variation in Figure 1; not all, but most. Every person to whom the chart at Figure 3 has been shown has concluded that there were six "systems" of data, and one assignable/special cause or upset. If the assumption that an important part of the reason for developing control limits is to separate random from random variation, and to avoid confusing the two is correct, then twenty years of showing both charts to thousands of people suggests the chart at Figure 3 is a superior interpretation of the data.

Is the data stable? Both the statistical evidence (the control chart at Figure 3) and the technical evidence (records that demonstrated that each shift in the mean in the control chart was caused by a shift change) suggest that the average of the process was changing. It was not stable, despite being appearing that way in Figure 1. Readers are invited to draw their own conclusions about the correctness and importance of this conclusion.

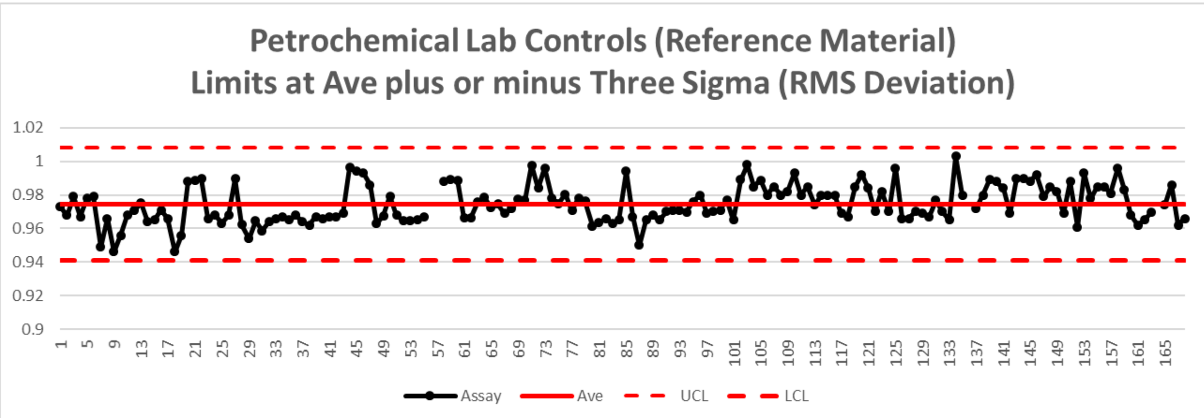
Why the difference in estimates for RMS Deviation? The RMS Deviation used in Figure 1 is more than twice that generated by Shewhart's methods. This is because when the process average changed as night shift came on duty, the actual deviation values were in fact quite small, given that the data were close to the average **for that shift**. However, when the data are rolled into one set, the overall average is used to calculate the deviations and therefore RMS Deviation. When night shift data are presented, the deviations of each of these data from the **overall average** are used, and are large. These large deviations are then squared, making the error orders of magnitude bigger. Shifts in the process average and upsets to the process tend to produce much bigger deviation values. In turn, these bigger deviation values generate a larger value for RMS Deviation that would be created if the data were in fact stable.

When is the distinction important? When the data are stable Shewhart's methods and the approach of average plus or minus three RMS Deviations develop similar control limits. However, at the outset of an analytical study one is unlikely to know whether the process is stable. That is one of the first questions to be answered. If one knew, in advance, that the process was stable, one could calculate an average, place control limits three RMS Deviations either side of this average, and in most cases get a reliable, useable result; but this is almost never the case. If it were the case, there would be little or no need to do any further analysis or to draw a chart.

ANOTHER EXAMPLE

The next example comes from the petrochemical industry. The chart at Figure 4 shows a chart developed during a course of training for laboratory managers and senior technical people in a large petrochemical company, using laboratory controls data brought to the course by a senior scientist. Control limits were placed at the average plus or minus three RMS Deviations. The senior scientist concluded that whilst there was evidence of some minor drifts in the data, there were no significant issues with the assay.

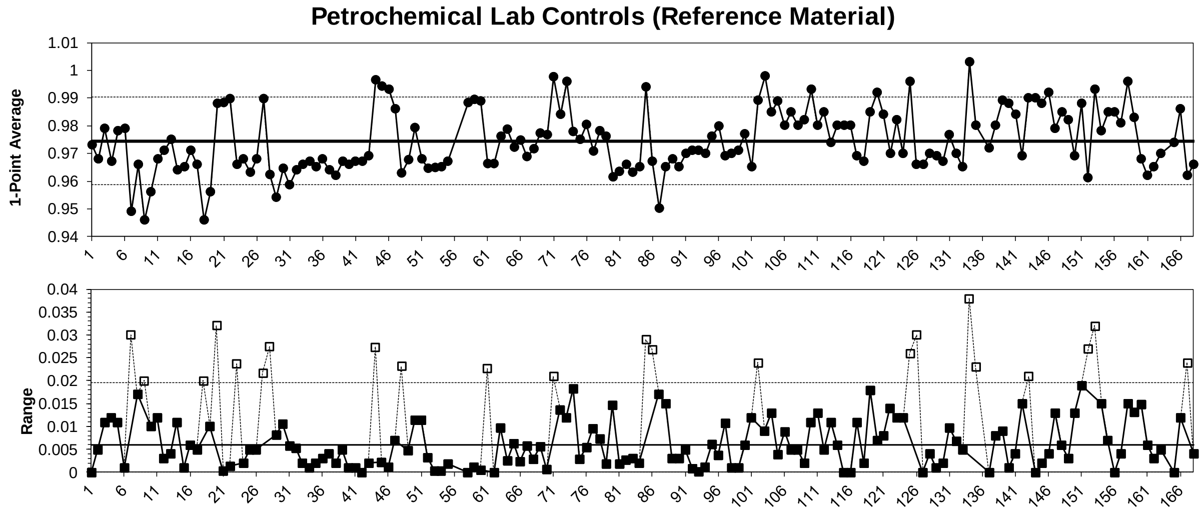
FIGURE 4



In only a few minutes the same data were then used to make a Shewhart control chart (a Single Point and Moving Range control chart) which is at Figure 5.

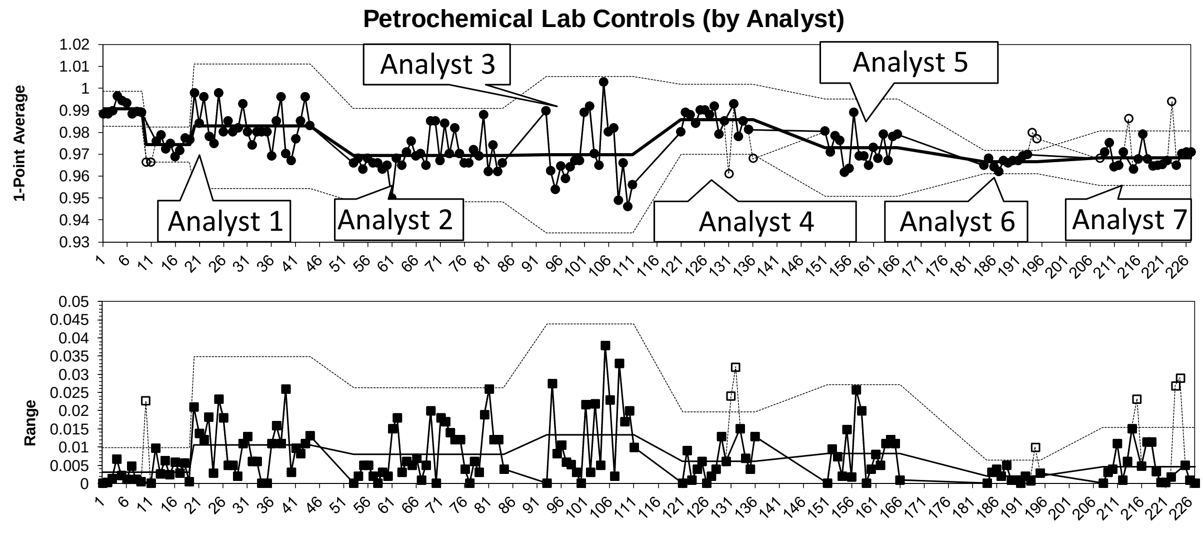
The Shewhart control chart leads towards a very different conclusion; that the data are unstable, which suggests that the assay is not producing a predictable and reliable result. This led the senior scientist to ponder on what the likely causes of this instability were.

FIGURE 5



The group of laboratory managers and senior technical people attending the course of training had already expressed the collective view that the biggest source of error in their laboratories was differences between analysts. To test this notion, in five minutes the data was sorted by analyst and a new chart, showing each analyst separately, was developed. This chart is at Figure 6.

FIGURE 6



In fifteen minutes this senior scientist's opinion about the state and reliability of the assay was profoundly changed. Initially he saw it as being reasonably (but not perfectly) stable, and thought that he had no significant issues. A short time later he saw that the assay was unstable, and that this meant that it was neither predictable nor reliable. He understood that stabilising the assay would greatly reduce analytical error, improving process control in the factory. He knew also why the assay was unstable; the most significant driver of variation in this assay was differences between analysts. He publicly noted his surprise that Analyst 6 was easily the best (most precise and repeatable) analyst.

This analyst was new to the laboratory, having only recently completed the training for laboratory technicians. Analysts 1 and 3 were his most senior and experienced analysts.

CONCLUSIONS

Setting control limits at the average plus or minus three (or two) RMS Deviations is an approach that is largely a self-fulfilling prophecy. For any data set that is approximately bell-shaped and reasonably continuous (that is, there are no significant gaps in the tails of the distribution) limits set at plus or minus three RMS Deviations are likely to capture 99% or more of the data. This is true even if the data are in a state of chaos, as was the case with the petrochemical example.

The approach that sets control limits three RMS Deviations either side of the mean was developed for distribution theory. It assumes that the data under examination come from a single population. In turn, the data must be at least reasonably stable before there can be any surety that a single population or universe exists. In some studies, this is likely to be correct. Census studies are one such example. However, for commercial and industrial data this is rarely correct because good stability is not common. Shewhart wrote, *"For our present purpose, however, ... we are not concerned with the functional form of the universe but merely with the assumption that a universe exists".* *(3)* Such an assumption is dangerous, and may be expensive. One may calculate an overall average, but that average means very little for an analytical study if the mean of the process is drifting around, a common occurrence. One can calculate a RMS Deviation, but if the data are unstable the RMS Deviation calculated will be equally meaningless because it has been inflated by an unknown amount by drifts in the process average and by disturbances to the process. One can calculate skewness or kurtosis, but frequently either skewness or significant kurtosis is the result of special causes or upsets being combined with what would have otherwise been a close to normal distribution, if stable.

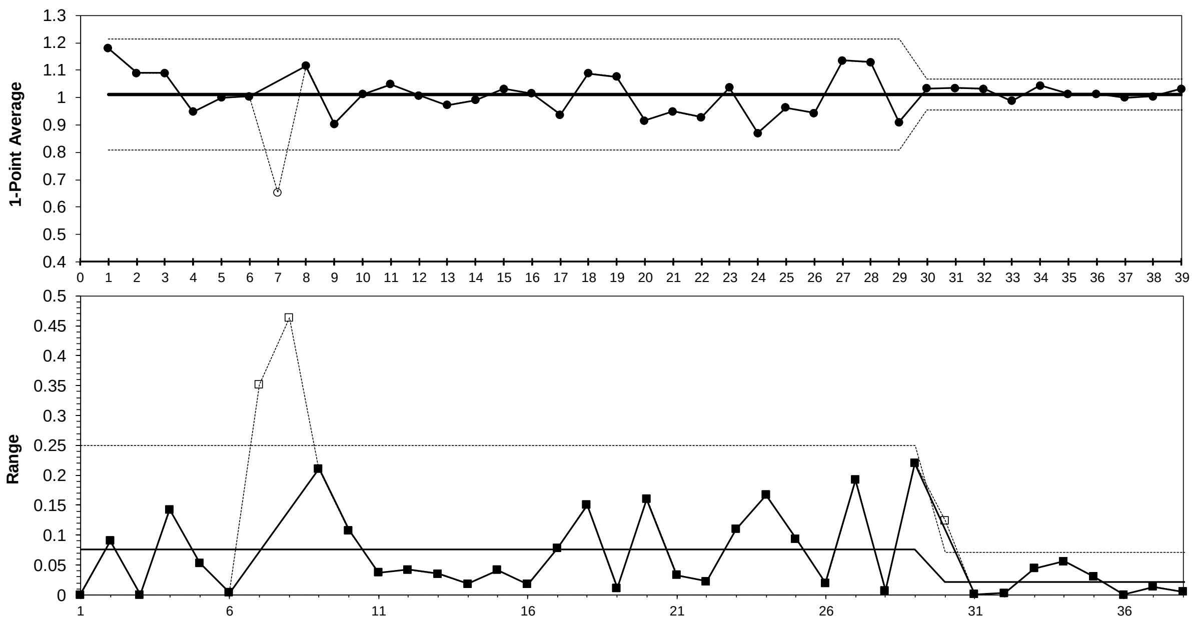
The very characteristic of RMS Deviation that makes it a valuable statistic in distribution theory results in it being a poor basis of determining the limits of controlled variation for an analytical study. As the shape of a distribution changes, so too does RMS Deviation because it responds to a change in dispersion of the data. This is perfect for distribution theory. However, for most analytical studies, often the first questions that are asked are, "is the process stable", and, "what are the limits of controlled variation?" Those who place control limits three RMS Deviations either side of the centre line are trying to answer an analytical question with a descriptive tool ill-suited to the purpose. This approach will habitually set the limits further apart than they should be and make data that are chaotic look much more stable than is actually the case.

Dr Shewhart not only saw these problems as far back as the 1920s, but also, by and large, he solved them. The Shewhart control charts are not perfect; nothing can be absolute or perfect in a probabilistic science such as statistics. Nevertheless, for analytical studies the control charts he developed are much more robust than the methods inherited from descriptive studies and distribution theory. They are more effective at separating random from non-random variation; give a good estimate of the underlying systematic variation independent of drifts and disturbances, are more robust generally and help managers and technical people better understand the state of affairs that currently exists in their businesses as well as providing important clues on how to improve matters.

Dr Brian Nunnally has addressed the issues presented here on many occasions. The chart at Figure 7 comes from operations he managed. He arranged for one of his statisticians to receive training in Shewhart's methods. She was then asked to address the variability in an assay that was in the final stages of development, but which was too variable to take to the validation stage. The results noted were achieved in a week with no new equipment, no fundamental changes in technique; only a focus on doing things exactly the same way, analyst to analyst. Matters do not always proceed so rapidly, but they are unlikely to proceed at all if managers and technical people are not aware of the problems present, or of approaches that might help to solve them.

FIGURE 7

Lab Controls for a Biologic



For analytical studies, setting control limits at plus or minus three RMS Deviations either side of the average frequently is the enormous initial statistical mistake. Managers and technical people who are well versed in Shewhart's methods are in a far superior position to avoid this mistake; to improve productivity, quality, profitability and competitive position.

REFERENCES:

1. *STATISTICAL METHOD, From the Viewpoint of Quality Control*, W. A Shewhart, The Graduate School, The Department of Agriculture, Washington, 1939, page 113
2. *STATISTICAL METHOD, From the Viewpoint of Quality Control*, W. A Shewhart, The Graduate School, The Department of Agriculture, Washington, 1939, page 44
3. *STATISTICAL METHOD, From the Viewpoint of Quality Control*, W. A Shewhart, The Graduate School, The Department of Agriculture, Washington, 1939, page 54