The Chart For Individual Values

By Don Wheeler

On page 314 of Walter A. Shewhart's *Economic Control of Quality of Manufactured Product* (Van Nostrand, 1931 original publication; American Society for Quality Control, 1980 commemorative reissue) wrote:

"Obviously, if the cause system is changing, the sample [subgroup] size should be as small as possible so that the averages of samples do not mask the changes. In fact single observations would be the most sensitive to such changes. Why then do we not use a sample size of unity? The answer is that if we do, we are faced with the difficulty of choosing the standard deviation to be used in the control charts. ... In fact, the sensitivity of the test [the process behavior chart] will increase, in general, with the decrease in subgroup size until the size of the subgroup is such that the data in any given subgroup come from a constant system of chance [common] causes. In the absence of any a priori information making it possible to divide the data into rational subgroups, there would be some advantage therefore in reducing the subsample [subgroup] size to unity."

Thus, from the very beginning, the philosophy behind the process behavior chart has been pushing us in the direction of the chart for individual values. Initially, a technical problem prevented us from reaching this goal. While Shewhart found a way around this problem for certain types of count data (p-charts and c-charts are charts for individual values), his solution would not generalize to other types of data.

In 1942, W. J. Jennett provided an answer to the problem posed by Shewhart. In his work with the M O Valve Co. Ltd., Jennett proposed using a technique that had been used in studying the ballistics of artillery shells in the 1890s. This technique was known as the method of successive differences. By using the differences between successive values as a measure of dispersion, Jennett effectively created the chart for individual values and a moving range, or XmR chart.

The attribution to Jennett comes from a 1953 article by Joan Keen and Denys J. Page in the journal *Applied Statistics,* which claims that the XmR chart was used extensively at General Electric Ltd. in Wembley, England, from 1943 to 1953. During this period, it also appears to have found its way into the textile industry, since physicist and statistician Leonard H.C. Tippett included a description and example of the XmR chart in his 1950 text. In this same time period, theoretical work by John von Neumann of the Institute for Advanced Study, in Princeton; and H.O. Hartley and A.R. Kamat of University College, in London; established that the average two-point moving range is 100-percent efficient, and that it is sufficiently similar to the ordinary range for n = 2 that the usual scaling factors can be used. In short, the two-point moving range has the highbrow theoretical ancestry that we like to have in our statistical techniques. Moreover, its use in the XmR chart is completely consistent with this mathematical background.

However, for the next 40 years, in book after book, the XmR chart would get no more than a brief mention as a miscellaneous technique. Then, one night in 1985 while we were at dinner, Deming asked me to tell him about this strange chart he had seen at a company where we were both consulting. It turned out that he had never seen an XmR chart before. While XmR charts had been around since 1942, they were barely known and rarely used.

Fortunately, in working with several different chemical companies during the 1980s, I had seen the power and utility of the XmR chart over and over again. With logical subgroups of size one, the XmR chart was ready-made for their single-point data. But it's not only the process industries whose data occur one value at a time. All sorts of service sector, administrative, and managerial data, fall into this category. So, in 1993, to open up new areas to the power and utility of process behavior charts, I produced a book on how to use XmR charts to transform the job of interpreting management data. This book, *Understanding Variation, the Key to Managing Chaos*, was excerpted in *Quality Digest* in August, 1993, and has proven to be one of the most popular books on data analysis ever written.

Today, the XmR chart is an integral part of almost every package of SPC software. Unfortunately, just as with the average chart, most of these software packages provide options that compute the limits incorrectly.

The chart for individual values done right

The first hurdle to using a chart for individual values is the hurdle of rational subgrouping. Even though the subgroup size is unity, the fact that we are going to compute the limits using the successive differences means that successive values must be logically comparable. That is, under the conditions expected to prevail when the process is operated predictably, successive values should differ only by what amounts to routine variation. In this way, the moving ranges will capture the routine variation and the limits will filter out the routine variation. In other words, you cannot *knowingly* place apples and oranges together on an XmR chart. Of course, as always, rational subgrouping is a matter of judgment that requires process knowledge. Then, once the data have been organized in a rational manner, there remains the question of how to compute the limits.

The first, and most common, way to compute limits for an XmR chart is to compute the average and the average moving range, and then to combine them using:

Approach 1
Limits for *X* = *average* ± 2.66 * *average moving range*
and upper limit for *mR* = 3.268 * *average moving range*
This approach should always be the default option for computing limits for individual values.

For our example consider the first 16 values in the data set presented on page 20 of Shewhart's *Economic Control of Quality of Manufactured Product*. The values are the resistances, in megohms, for an insulation material. The values and moving ranges are shown in figure1. The average is 4135.5, and the average moving range is 355.93. Using these values we find the limits shown in figure 2.

X	mR	X	mR	X	mR	X	mR
5045	—	4290	315	3980	305	3300	460
4350	695	4430	140	3925	55	3685	385
4350	0	4485	55	3645	280	3463	222

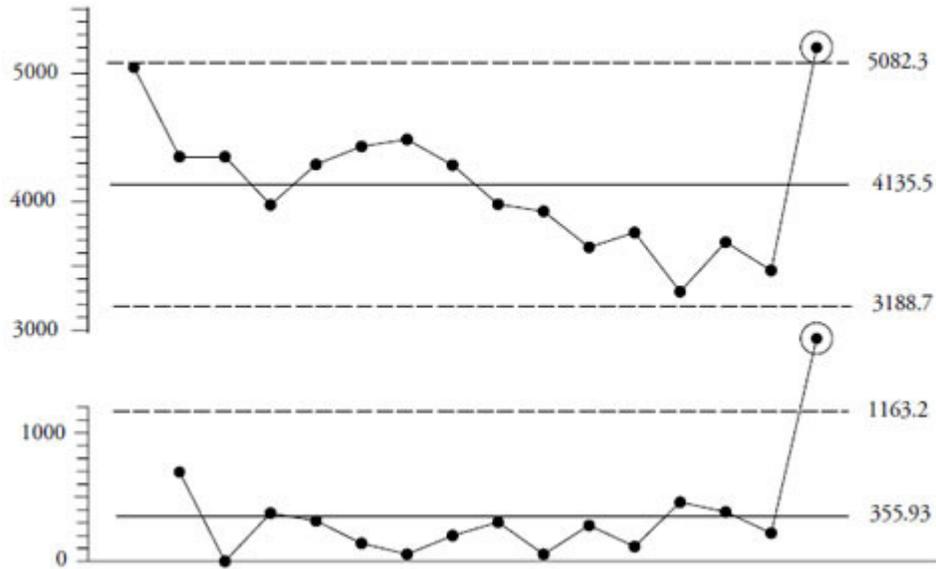|	3975	375		4285	200		3760	115		5200	1737

Figure 1: Sixteen resistance values

Figure 2: XmR chart for resistances with limits based on average moving range

Here, one observation and one moving range fall outside their limits. The large moving range confirms the impression from the X chart that a sudden change occurred between the 15th and 16th values. This is sufficient evidence to say that this process is changing. When this happens the important questions are no longer questions about the process location or the process dispersion, but rather questions about what is causing this process to change. Estimation is moot. Discovery of the assignable cause is paramount.

An alternate way to compute limits for an XmR chart is to use the median moving range:

Approach 2
Limits for X = average ± 3.145 * median moving range
and upper limit for mR = 3.865 * median moving range

Because of the lower efficiency of the median moving range this approach should not be used as the default approach. However, when some large moving ranges may have inflated the average moving range, this approach will allow the computation of limits that will be less severely inflated by those large ranges. For the data of Figure 1 the median moving range is 280.0, resulting in the XmR chart shown in figure 3.
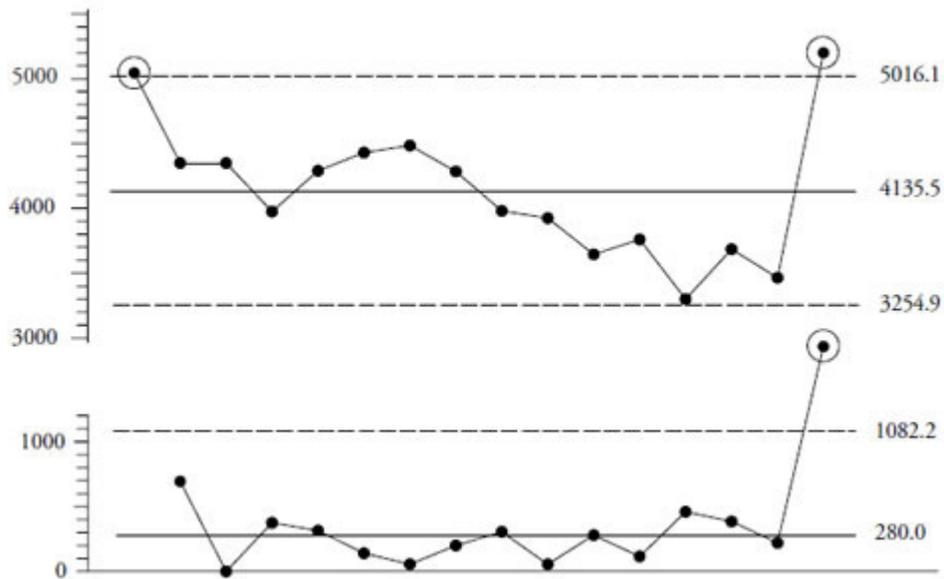
Figure 3: An alternate XmR chart for resistances using the median moving range

Here we find two individual values and one moving range outside the limits. Compared with the default approach, the limits here are 3.5-percent narrower. While the details change, both figure 2 and figure 3 tell the same story about this process. This process is changing without warning, and we need to find out why.

A variation on method 2 would be to use the median of the X values as the central line for the X chart.

Approach 2.1.

Limits for X = median ± 3.145 * median moving range

and upper limit for mR = 3.865 * median moving range

It is unlikely that you will need to use the median of the X values except in conjunction with the median moving range. This is because the average moving range is more prone to inflation than the average of the X values, so that any time the median of X is appropriate, the median moving range will also be appropriate. For figure 1 the median of the X values is 4132.5, which leaves figure 3 essentially unchanged.

The chart for individual values done almost right

There is another computation that is technically right but practically wrong. This is the use of the square root of the mean square successive difference. To find this measure of dispersion we begin with the moving ranges and then (a) square each moving range, (b) average them, (c) divide by 2.0, and (d) find the square root. This root mean square successive difference (RMSSD) is then used with the average to compute limits for the X chart. This is usually done using the formula:

Approach 3

Limits for X = average ± 3 * RMSSD

Typically this approach doesn't include a chart for dispersion. For the data of figure 1, the root mean square successive difference is *RMSSD* = 382.95. This approach will result in the limits shown in figure 4.
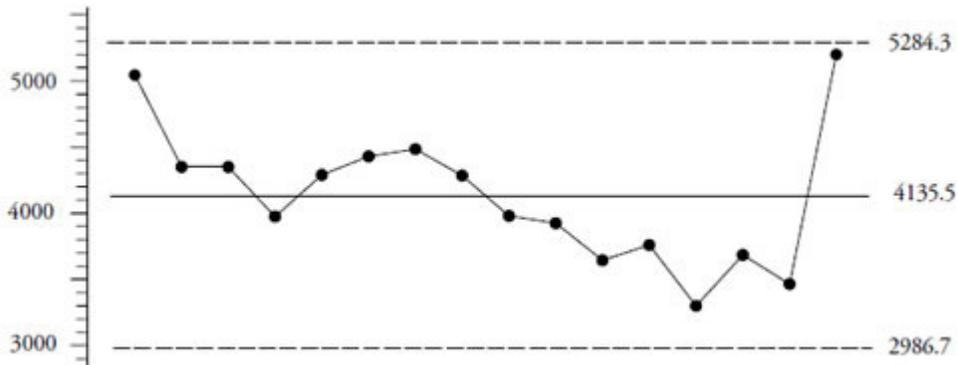

Figure 4: An X chart for resistances done almost right using RMSSD

The RMSSD has exactly the same degrees of freedom as the average moving range. This means that, as an estimator of the process dispersion, it has the same efficiency as the average moving range. However, as we saw earlier, the question here is not one of estimation, but rather one of finding what is causing the changes in the process. In this regard the RMSSD is *not* equivalent to the average moving range, for it lacks the required robustness.

The RMSSD lacks robustness because it averages the squared ranges, which makes it more prone to being inflated by large range values. This will inflate the limits and desensitize the chart. Compared to the default approach in figure 2, the RMSSD limits in figure 4 are inflated by 11 percent. As a result, figure 4 has no points outside the limits. Since this approach is no more efficient, yet is more prone to being inflated, we must conclude that the RMSSD has nothing to offer except complexity for the sake of complexity. There is simply no situation where it will ever prove to be advantageous, and therefore should be avoided in practice.

The chart for individual values done wrong

There are only two measures of dispersion to use when creating a chart for individual values: The average moving range and the median moving range (approaches 1 and 2 above). The use of any other measure of dispersion is wrong.

Nevertheless, many software packages compute limits for the X chart using:

Approach 4
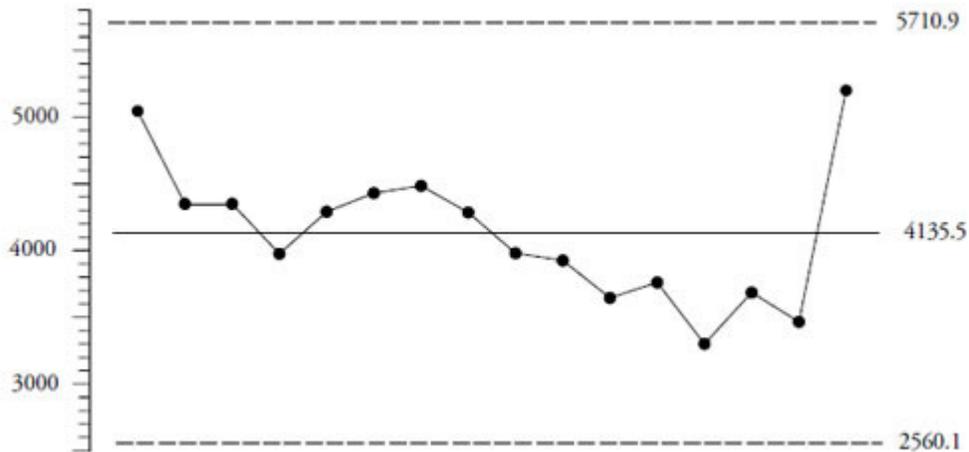Limits for *X* = a*verage* ± 3 * g*lobal standard deviation statistic*

Figure 5: An X chart for resistances done wrong using the standard deviation statistic

While the formula above may be said to give you symmetric, three-standard-deviation limits, this computation is wrong simply because it will be substantially inflated whenever the data are not homogeneous. Because the purpose of the chart is to examine the data for possible evidence of a lack of homogeneity, any approach that implicitly assumes the data to be homogeneous is inherently and philosophically flawed. For the data of figure 1 the global standard deviation statistic is 525.14, resulting in the erroneous limits shown in figure 5.

Specifically, the global standard deviation statistic computed using all of the individual values is wrong for the same reason that it was incorrect for use with an average chart. While you might be able to compute reasonably correct limits when the data happen to be homogeneous, this approach will severely inflate the limits in the presence of signals of process changes. Compared with the default approach, the limits in figure 5 have been inflated by 23 percent. While this may paint a pretty picture of your process, it is simply nothing more than another way to lie with statistics.

Why do we need the moving range chart?

It has been suggested that the moving range (mR) chart adds so little to the X chart that you should not bother to show it. Certainly, approaches 3 and 4 ignore the mR chart. Sometimes even software using approaches 1 and 2 will drop the mR chart. The basis for this recommendation seems to be the documented fact that—in theory at least—the combined XmR chart does not have any appreciably greater ability to detect signals than does the X chart alone. However, this mathematical analysis overlooks the interpretative benefits to be gained by including the mR chart.

For example, inspection of figure 2 will show that the last point could have fallen well within the limits on the X chart (say 4,700 or above) and still would have resulted in an mR value that was above the upper limit on the mR chart. Thus, the mR chart can detect sudden shifts even when the individual values are still within the limits on the X chart. This ability to complement and reinforce the impressions gained from the X chart is one of the practical reasons that you should not ignore the mR chart.

A second practical reason to keep the mR chart is the fact that many software options compute the limits incorrectly. If you are shown a naked X chart you will have no way of knowing if the limits have been computed correctly. However, if you are shown an XmR chart, you will immediately have a higher level of confidence that the limits have been computed correctly. Moreover, by using the central line of the mR chart, you can quickly check to see if the limits are indeed correctly computed. Thus, the mR chart is the secret handshake of those who know the correct way of computing limits for an X chart. Omit it and your readers cannot be sure that you are a member of the club.

A third practical reason for keeping the mR chart is that it will allow you and your audience to check for the problem of chunky data. This problem occurs when the measurement increment is too large for the application at hand. When your data are chunky you will find three or fewer possible values for the moving ranges (including zero) below the upper range limit. When this happens the round off in the measurements will obscure the process variation and the excessive number of zero ranges that result will deflate the average moving range and tighten the limits. In consequence the process behavior chart will have many false alarms and so will become unusable. The only remedy for this problem is to use a smaller measurement increment.

So, while it may be difficult to find a mathematical justification for showing the mR chart, there are three practical reasons to do so, any one of which is sufficient to justify the inclusion of the mR chart with your X chart.

Thus, just as with average charts, there are right and wrong ways to compute limits for a chart for individual values. Because most software will provide you with the option of computing your limits incorrectly, it is your responsibility to determine how to get the right limits using your software. There are many novel ways of computing the limits incorrectly, but only two measures of dispersion that will yield correct limits for an XmR chart.