Avoiding Statistical Jabberwocky

Donald J. Wheeler

In my August column, "**Do You Have Leptokurtophobia?**" I carefully explained how the process behavior chart does not require that you have "normally distributed data." I explained how three-sigma limits effectively strike a balance between the economic consequences of getting false alarms and those of missing signals of economic importance. I also demonstrated this balance in two ways: first by showing how the chart for individual values avoided false alarms, even when used with synthetic data generated by an exponential distribution; and then by showing how the individuals chart also detected signals within a real data set, even when those data possessed a very lopsided histogram.

On August 24, Forrest Breyfogle responded with an article that claimed that the control charts require normally distributed data and that we should transform our data to make them "more normal" in certain situations. He presented no support for this claim that the charts require normal data. He simply stated it as a article of faith. He then used a very extreme probability model to generate some synthetic data and observed that the individuals chart had 3.3 percent false alarms. Of course, Breyfogle interpreted this as evidence that the chart did not work because he did not get the normal theory false alarm rate of 3 per thousand. To understand Breyfogle's example it is important to know that virtually all practical models for original data will have a skewness of 3 or less and a kurtosis of 12 or less. Breyfogle's model had a skewness of 6 and a kurtosis of 114. Yet in spite of this unusually extreme probability model, his model had a theoretical value of only 1.8% beyond the upper three-sigma limit. Both this theoretical result and Breyfogle's observed result are consistent with the definition of what will happen with a robust procedure. (Robust procedures are those where conservative theoretical critical values will continue to deliver less than 5% false alarms when you change the probability model.) Therefore, in light of the extreme nature of Breyfogle's probability model, the 3.3% false alarm rate is actually evidence of the robustness of the individuals chart, rather than being evidence of a failure of the technique. Breyfogle simply misinterpreted his own example!

In my September column, "**Transforming the Data Can be Fatal to Your Analysis**," I explained the shortcomings of Breyfogle's paper and presented Shewhart's conceptual foundation for the use of process behavior charts. There I showed that *Shewhart was not concerned with achieving a particular risk of a false alarm*, but rather was looking for a general, distribution-free approach that would have a reasonably small risk of a false alarm for all types of data. I also showed why a process behavior chart for the original data should always be the first step in any analysis of process related data.

On September 16, Breyfogle responded to my September column with a 17 page broadside arguing that data have to be transformed prior to using them on a process behavior chart. In this paper he painted my response as being too narrow in scope, conveniently overlooking the fact that I have often made exactly the same points he was making about the kinds of action possible in practice. (For example, see "**Four Possibilities for Any Process**," *Quality Digest*, December

1997.) Before we consider the rest of Breyfogle's paper we will need some background material. Most of what we will need may be found in my articles in this column during the past nine months, but first we shall begin with the **Four Foundations of Shewhart's Charts**.

The first foundation of Shewhart's charts is the generic use of *three-sigma limits*. This frees us from having to specify a probability model and determine specific critical values in order to achieve a particular alpha-level. As Shewhart noted, in practice, we really do not care what the alpha-level is as long as it is reasonably small, and three-sigma limits have been proven to provide reasonably small alpha-levels with all sorts of real world data.

The second foundation of Shewhart's charts is the use of *within-subgroup variation* to compute the three-sigma limits. Global measures of dispersion are descriptive, but the foundation of all modern statistical analyses, from ANOVA and the Analysis of Means to process behavior charts, is the use of the within-subgroup variation as the filter for separating probable noise from potential signals.

The third foundation of Shewhart's charts is the use of *rational sampling and rational subgrouping*. This is simply the flip side of using the within-subgroup variation. In order to filter out the noise we need to estimate the routine variation. This means that we will have to select and organize the data in such a way that the subgroups will be logically homogeneous. As Shewhart said, "Specify how the original data are to be taken [collected] and how they are to be broken up into subsamples [subgroups] upon the basis of human judgments about whether the conditions under which the data were taken were essentially the same or not." Two numbers belong in the same subgroup only when they can be said to have been obtained under essentially the same conditions. Numbers that might have come from different conditions belong in different subgroups. Rational sampling and rational subgrouping are *rational* simply because they are the result of careful thought about the context for the data. When we ignore the principles of rational sampling and rational subgrouping we undermine the computations and are likely to end up with nonsense charts.

And the fourth foundation of Shewhart's charts is the ability of the organization to use the knowledge gained from the charts. This topic is so broad that Deming developed his Fourteen Points to address the many issues that occur here. However, one thing has been repeatedly demonstrated. *Organizations that refuse to listen to the data will fail.*

In my February column, "**First, Look at the Data**," I demonstrated the importance of plotting the original data in time order on an XmR Chart. There we considered the number of major hurricanes in the North Atlantic from 1940 to the present. When we plotted these original data on an individuals chart we found a distinct oscillation between quiet periods and active periods. These data consisted of small counts, were bounded on one side by zero, and had a histogram that looked like a ski slope. By simply listening to the story told by original data themselves we learned something. Had we transformed these data we would have missed this important aspect of these data.

In my March column, "**The Wrong Place to Start Your Analysis,**" I used the same hurricane data to demonstrate that even though you might find a probability model that provides a reasonable fit to your histogram, this does not mean that your data actually came from a single system. The erroneous idea that you can infer things from how well your data fit a particular probability model is known as the Quetelet Fallacy, after the Belgian astronomer who had this mistaken idea in 1840. Quetelet's Fallacy was exposed by Sir Francis Galton in an 1875 paper that proved to be the foundation of modern statistical analysis. In this paper Galton demonstrated that a collection of completely different processes, having different outcomes, could still yield a histogram that looked like a histogram produced by a single process having consistent outcomes. For the past 134 years statisticians have known better than to read too much into the shape of the histogram. Unfortunately, each generation of students of statistics has some individuals who follow in the footsteps of Quetelet. Some of them even write articles about their profoundly erroneous insights.

In my April column, "**No Data Have Meaning Apart From Their Context**," I demonstrated that you cannot create meaning with arithmetic. The meaning of any data set is derived from the context for those data. I showed that if you do not respect that context when you analyze your data then you are likely to end up with complete nonsense. Shuffling the data, or transforming the data, in ways that distort or destroy the context for those data may give you a pretty histogram, or a pretty running record, but it will not provide any insight into the process represented by the data.

In my May column, "**All Outliers Are Evidence**," I demonstrated the fallacy of "removing the outliers" from your data. Once again, simply plotting the original data on an *XmR* chart provided the insight needed to understand the data. Moreover, this approach worked even though the original data had a kurtosis statistic of 12.3! No transformation was needed. No subjective decision had to be made about which nonlinear transformation to use. We simply plotted the data and got on with the interpretation of what was happening.

In my June and July columns, "**Don't the Outliers Distort the Limits?**" and "**Good Limits from Bad Data**," I demonstrated that the correct way of computing the limits is robust to outliers. These correct computations will always use the within-subgroup variation and are a direct consequence of Sir Francis Galton's 1875 paper. But this approach imposes a requirement of rational subgrouping upon the use of process behavior charts. Where, you might ask, are the subgroups on an *XmR* chart? They are the pairings of successive values used to compute the moving ranges. *The requirement of rational subgrouping means that these successive values must be logically comparable.* In other words, we cannot mix apples and oranges together on an *XmR* chart and expect the chart to work the same as it would if we charted all apples. While the definition of what is logically comparable may depend upon the context of what we are trying to accomplish with the chart, we must still work to avoid irrational subgroupings.

In my August column, "**Do You Have Leptokurtophobia?**" I once again showed how the use of a nonlinear transformation will inevitably distort the original data and obscure the signals contained within those data. *These are facts of life, not matters of opinion.* However, in order to be as clear as possible about this, consider Figure 1 which reproduces part of a graph from an article in a scientific journal. There the distance from the zero point to the first arrow represents one million years, but the distance between the first arrow and the second arrow represents 13,699 units of one million years each!

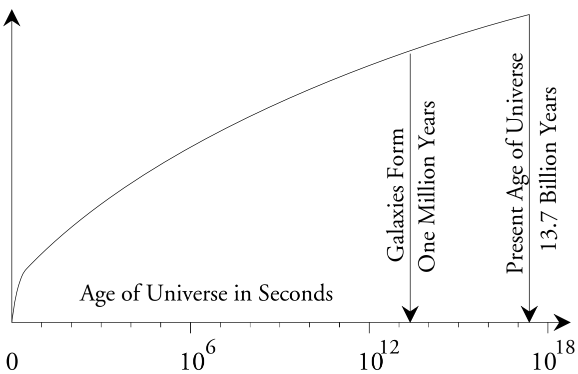

Figure 1: One Million Years versus 13.7 Billion Years on a Logarithmic Scale

When I saw this graph I was sure that there must be some mistake on the horizontal scale. One million years after the Big Bang should be a lot closer to the beginning than it looks on this scale. However, calculation of the times involved show the scale of Figure 1 to be correct, *even though it defies all logic and preception*.

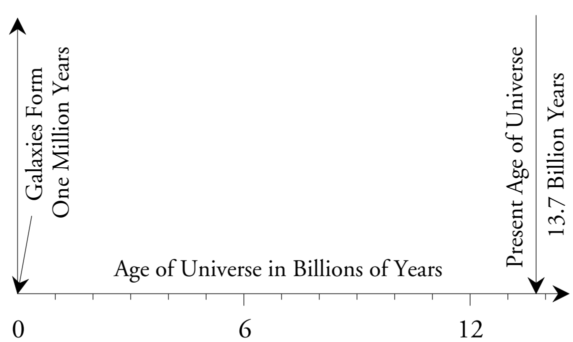

Figure 2: One Million Years versus 13.7 Billion Years on a Linear Scale

To understand the distortion created by the logarithmic scale of Figure 1 consider the same points plotted on a linear scale in Figure 2. Exponential and logarithmic scales are simply not intuitive. *We do not think in these terms.* As a result, when we use non-linear transformations we cannot even begin to understand the profound distortions that we are creating. Yet, as we will see, Breyfogle wants to use non-linear transformations of the original data simply for cosmetic purposes.

So what about Breyfogle's paper of September 16? After four pages of statistical jabberwocky, he finally gets down to an example on page five. The data used are said to be the "time to change from one product to another on a process line." For this example he combines "six months of data from 14 process lines that involved three different types of changeouts, all from a single factory." Clearly, such data are what I have called report-card data, and while we may place report-card data on a process behavior chart, the limitations of such data will impose certain limitations upon any report-card chart.

The first salient feature of a report-card chart is that the aggregation which is inherent in all report-card data will inevitably create a lot of noise. This noise will inflate the routine variation detected by the chart, which will inflate the limits and make report-card charts insensitive. This phenomenon is one reason that some authors, such as Lloyd Provost, do not like report-card charts. They claim that they can never detect any signals with such charts. While such report-card data are inherently a violation of the rational subgrouping requirement, they will sometimes work when the aggregated values are, indeed, logically comparable. Unfortunately, the noise that is part of the aggregation will also tend to make these charts appear to be reasonably predictable, even when the individual components of the aggregated time series are not predictable at all.

But this is not Breyfogle's problem. His chart shows 9 out of 639 points, or 1.4%, above the upper limit. So what does this mean? Given that these data represent three different types of changeouts on 14 different lines, it would be a miracle if there were no signals! The grouping together of the different types of changeouts, which in all likelihood have different averages, is almost certainly an example of an irrational aggregation, and the most plausible interpretation of the 9 points above the upper limit is that they are trying to tell us about this problem with the organization of the data!

The second salient feature of report card charts is that the aggregation will tend to obscure the context for each point. This aggregation, along with the time delay inherent in report-card data, will tend to make it very difficult to use report-card charts to identify specific assignable causes of exceptional variation when a point falls outside the limits. While improvements can be tracked on a report-card chart, it is a rare thing when the report-card chart can be used as the catalyst for process improvement. It is always much easier to identify assignable causes as the data are disaggregated. As the data become more narrowly focused and are plotted in real time they become more useful as a tool for process improvement. (I suspect that Breyfogle will take issue with that statement, but then, by his own admission, his clients have not been able to use SPC successfully. I have many clients that have been using SPC with great success for over 20 years, and it is their experience that I am reporting here, not my opinion.)

So, the inherent noise of a report-card chart, along with the difficulty of identifying specific assignable causes that correspond to the few signals that do occur, will tend to limit their usefulness to merely tracking the business at a fairly high level, which is what Breyfogle seems to be suggesting. However, there is a twist to what Breyfogle is doing.

In the two sentences following the description of the conglomeration of data that he is using, Breyfogle states that the factory is "consistently managed" and that the "corporate leadership considers this process to be predictable enough." These two statements are puzzling. According to the data given on the chart, these 639 changeouts averaged 11.54 hours each. If this plant is operating 24/7, then these changeouts consumed 12 percent of the operating time for the 14 lines in this plant. If this plant is operating 16/5, then these changeouts consumed over 25 percent of the operating time for the plant. It seems to me that anything that reduces productivity by 12 to 25 percent deserves more than a glance from the "30,000 foot" level!

But the real problem with Breyfogle's approach is not that he has an irrational aggregation (a mistake), and it is not that he has used a non-linear transformation in an effort to correct his irrational aggregation (another mistake), or that he has missed the message of this report-card chart about an opportunity for improving productivity (a third mistake). The real problem is seen in the last sentence in the following quote.

> "The corporate leadership considers this process to be predictable enough, as it is run today, to manage a relatively small finished goods inventory. With this process knowledge, what is the optimal method to report the process behavior with a chart?"

Following this sentence, Breyfogle begins to compare and contrast the Individuals Chart of the original data with a chart for the transformed data. But before we get bogged down in the argument about transforming the data, notice the sequence of events in the quote above. Having just said that the executives do not want to find any signals, Breyfogle sets to work to get rid of all of the signals. He is shaping the data to fit a preconceived message. This is not data analysis, but rather data *manipulation*. When accountants do this we put them in jail.

No matter how you might try to dress it up, any attempt to transform the data as the first step in an analysis is an act that does not treat the data with integrity. That which is lost with a non-linear transformation of the data can never be recovered by any subsequent analysis, no matter how elegant, profound, or complex that subsequent analysis may be. Even when the objectives and motivation of the subsequent analysis may be correct, appropriate, and well motivated, the distortion of the initial non-linear transformation makes everything else moot. And when the motivation is to change the message, the use of a non-linear transformation becomes just one more way to lie with statistics.

I have not tried to address many of the points raised by Breyfogle simply because, no matter what he says, no matter what he does, no matter how he justifies his actions, once he has used a non-linear transformation to massage his data into saying what he wants it to say, nothing else that follows can ever make sense.

I have carefully and repeatedly explained the problems of using non-linear transformations. Those who wish to study this topic have sufficient resources in the columns cited herein. Those who wish to continue to transform their data may do so, secure in the knowledge that they will rarely, if ever, be disturbed by the facts contained within their data. Either way, it is now time to move on.