Why Logarithmic Transformations Are a Mistake

Donald J. Wheeler

You have heard it said that the chart for individual values requires normally distributed data. You have been told that when your data are bounded on one side you may need a logarithmic transformation to reduce the false alarm rate on a chart for individual values. And you may even believe that high-level, report-card summaries can be used to determine when a process is being operated predictably. These ideas are respectively wrong, false, and incorrect. The erroneous thinking behind these misconceptions is the topic of this paper.

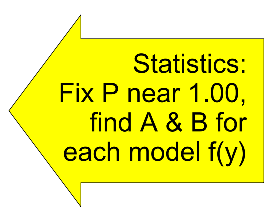
Since this erroneous thinking occurs at a fundamental level we will have to go into some background material in order to explain why these ideas are so completely wrong. While some of this may fall into the category of "more than you ever wanted to know," spending the time to go through this background material can free you from much of the superstitious nonsense that is taught today.

"NORMALLY DISTRIBUTED DATA"

The purpose of data analysis is to filter out the noise within the data so that we can detect any signals contained therein. The myth that we need "normally distributed data" in order for a process behavior chart to work has been around ever since 1935 when E. S. Pearson failed to understand how Walter Shewhart's approach to filtering out the noise differed from the statistical approach pioneered by Pearson and his father, Karl Pearson.

This statistical approach for filtering out the noise used by Pearson can be summarized as follows. Since we can never filter out all of the noise, we have to begin by choosing how much of the noise we wish to filter out. Let P denote this proportion. Common values for P are 0.95 or 0.99. Next we identify some test statistic, Y, which is a function of the original data, and find the probability model, $f(y)$, that characterizes the behavior of Y when certain conditions exist. Next we use the equation for the area under a curve to find the critical values A and B that correspond to our value for P for that particular probability model.

$$\int_A^B f(y)\, dy = P$$



Thus, with the statistical approach we fix the value for P, and then for a specific probability model, we find the exact critical values A and B that filter out the proportion P of the noise. Then any value for our test statistic Y that falls outside the interval from A to B is interpreted as a potential signal. This approach is summarized in Figure 1. There we use a P value of 0.99 to find the critical values A and B for seven probability models.

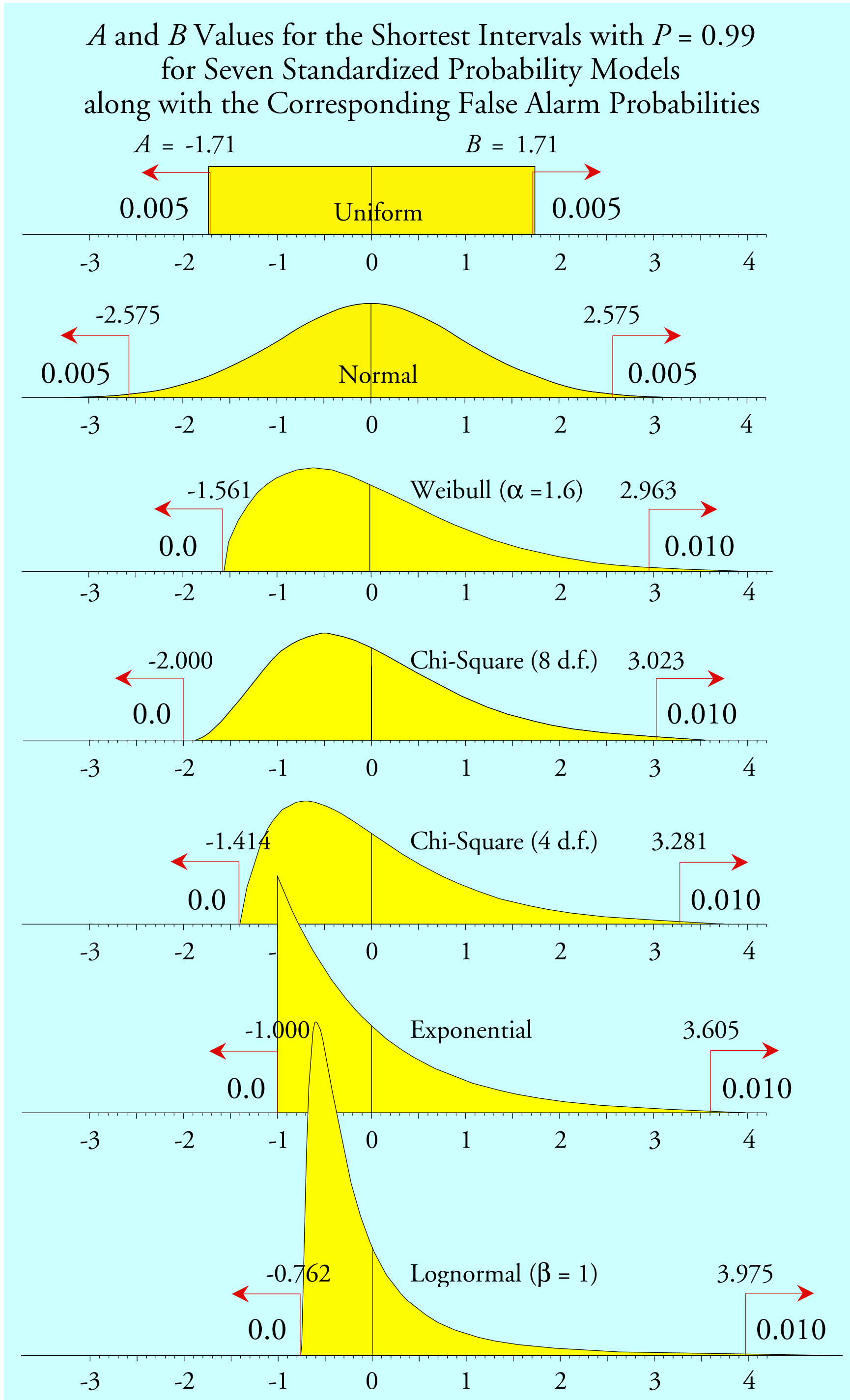
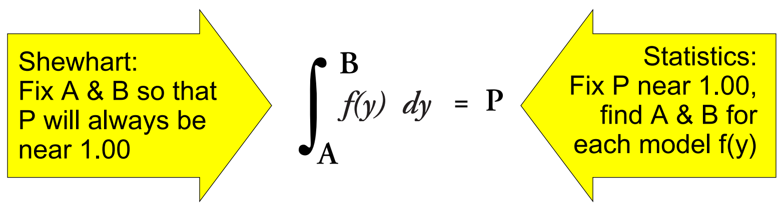


Figure 1: Statistical Approach with P = 0.99

Thus, when we hold P fixed we end up with different critical values for each probability model. While the statistical approach works quite nicely when we know the probability model to use for some test statistic, Y, Shewhart observed that we will never have enough data to fully specify a probability model for the original data. In the absence of a probability model the statistical approach breaks down, so Shewhart chose to approach the problem from the other end. He decided to use fixed, generic critical values A and B. Specifically, Shewhart chose to use symmetric critical values defined by [Average ± 3 Sigma]. These generic critical values will *always* result in a value for P that is reasonably close to 1.000 *regardless of what probability model f(y) applies.*



Thus Shewhart's approach to the problem of how to filter out the noise is 180 degrees opposite from the traditional statistical approach. By choosing to use generic, symmetric, fixed-width critical values Shewhart freed his technique from the tyranny of determining which probability model to use. Figure 2 summarizes Shewhart's approach using the seven probability models from Figure 1.

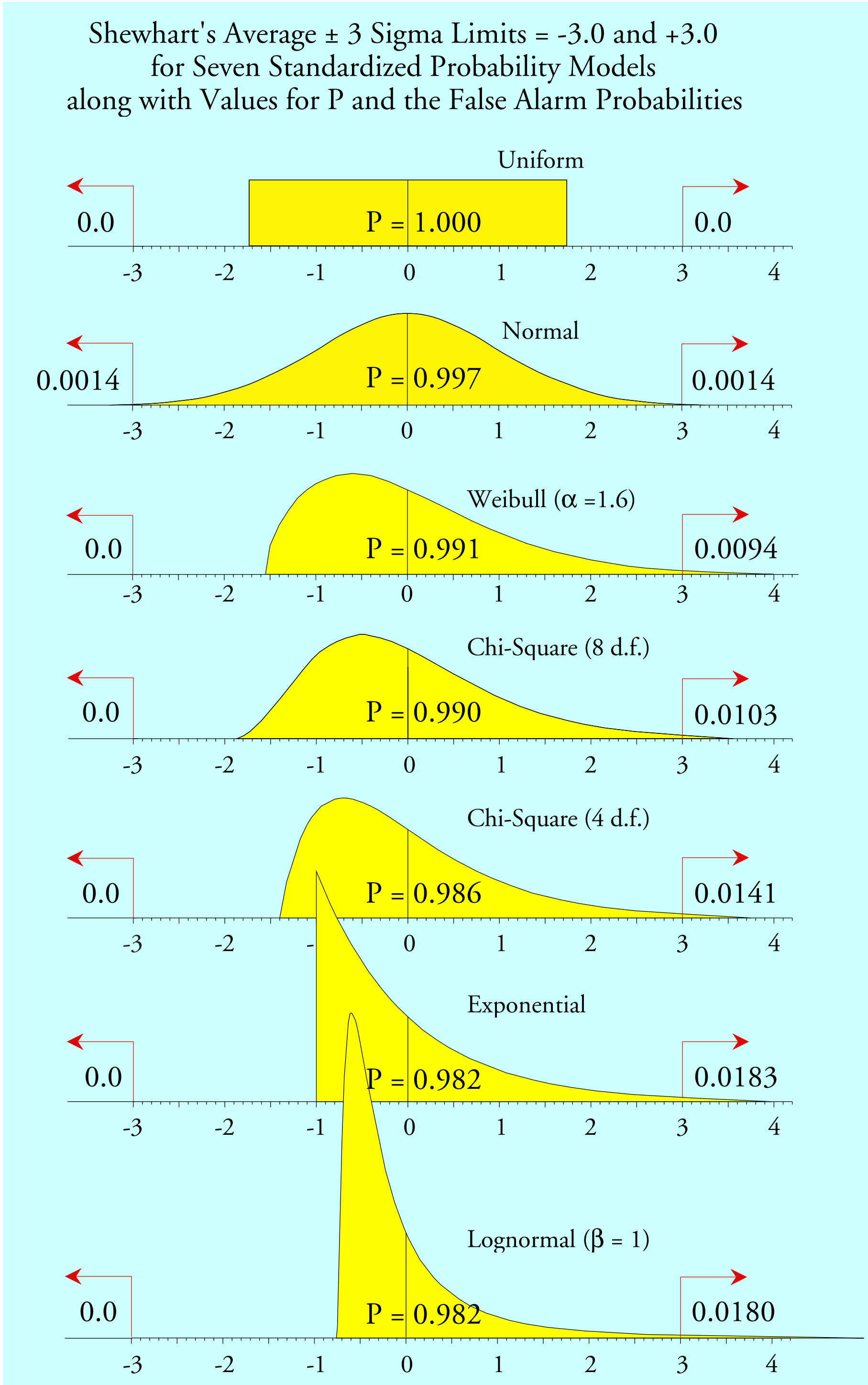


Figure 2: Shewhart's Approach with Symmetrical Fixed-Width Limits

Symmetric, three-sigma limits will always result in a *P* value that is reasonably close to 1.000 for any predictable process regardless of the shape of the histogram. They will filter out virtually all of the noise so that anything left over may be interpreted as a potential signal.

Three-sigma limits allow us to make the right decision almost every time without having to obsess about having the right distribution in order to obtain some predetermined value for P.

Those who fail to understand the difference in these two approaches will be caught up in the gyrations involved in determining which probability model to use and finding the critical values for some predetermined value for P. And when they cannot determine which probability model to use they may be tempted to transform the data to make them appear to be more "normally distributed."

Hence, those who teach that you have to have a normal distribution to use a chart for individual values are approaching the problem from the wrong direction. Anytime anyone begins by talking about finding a probability model or transforming the data before using a process behavior chart you can be sure they are making the same mistake that E. S. Pearson made. They either have not learned about Shewhart's distribution-free approach, or else they know about Shewhart and are deliberately choosing to complicate matters out of a fear that their ideas may not seem profound.

DO HIGH LEVEL SUMMARIES NEED TO BE TRANSFORMED?

In spite of the above, you will still hear people who use the statistical approach telling you that your high-level, report-card summaries will need to be transformed before they can be placed on a chart for individual values. It turns out that this is wrong even within the context of the statistical approach. The central limit theorem tells us why.

In the Spring of 1810 Pierre Simon Laplace published a theorem that today is called the central limit theorem. This theorem tells us something about how high-level summary measures will behave. Assume that if you have a summary measure X that consists of the sum or average of several components, A, B, C, etc., and assume you are going to repeatedly observe the value for X over time. The central limit theorem tells us that, as the number of components in the sum or average increases, the histogram for the measure X will become increasingly normal. And in most cases X can be said to be approximately normally distributed when there are as few as 5 to 10 components in the sum.

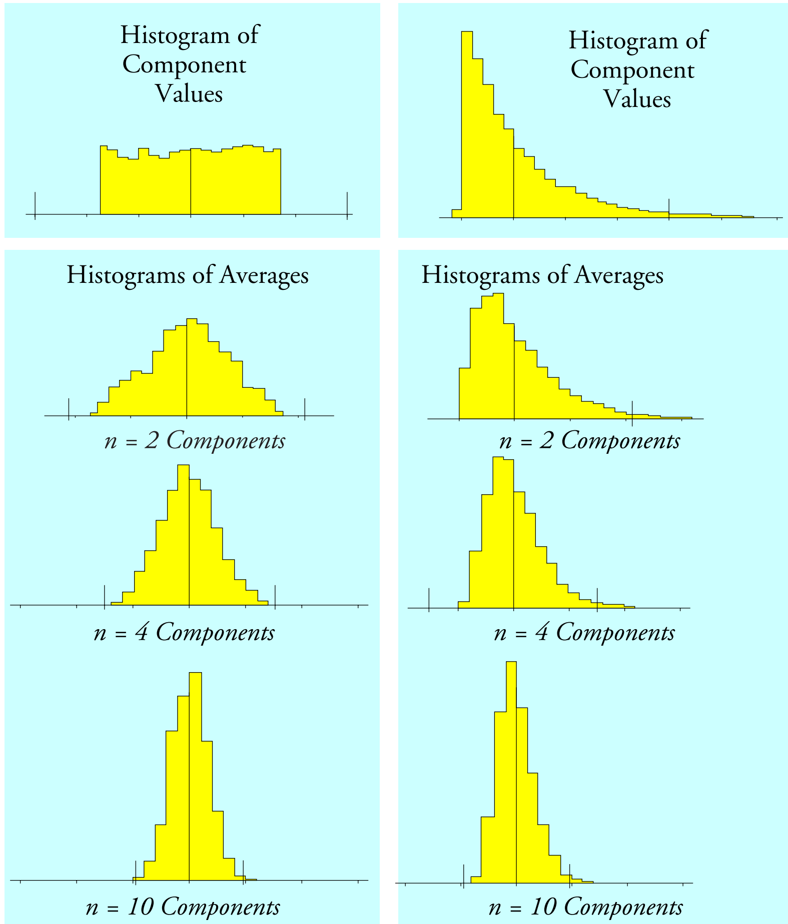


Figure 3: The Central Limit Theorem Illustrated

Thus, the central limit theorem undercuts those who argue that we need to logarithmically transform high-level, report-card measures to make them "more normal." *The central limit theorem guarantees that the aggregation involved in every high-level summary is itself an inherently normalizing transformation.*

Furthermore, the central limit theorem also tells us that it is illogical to assume that a high-level report-card measure might ever be lognormally distributed. Assume that X is a high-level, report-card measure and that it is lognormally distributed. Then $\log X$ will be normally distributed. If we think of $\log X$, as being the sum of several components, $\log A$, $\log B$, and $\log C$, then the original measure, X, *would have to be the product of components A, B, and C!*

$$\log X = \log A + \log B + \log C + \ldots$$
implies that
$$X = A * B * C * \ldots$$

The mathematics will not allow any other interpretation. Since virtually no high-level, report-card measure is obtained by *multiplying* all of the component measures together, it is a mathematical absurdity to assume that any report-card measure will ever be lognormally distributed. This means that even when we use the statistical approach, it will never make sense to take the logarithm of our high-level summary measures as part of our analysis.

TRANSFORMING SKEWED DATA

So, while high-level summaries are unlikely to ever be highly skewed, can individual measures have a skewed distribution? Yes, indeed. There are two ways we end up with skewed data. Occasionally we will get a skewed histogram when the data from a predictable process piles up close to a barrier or boundary condition. However, most of the time a skewed histogram results when a process is operated unpredictably and the changes in the process cause one of the tails to grow more than the other. Figure 4 shows four histograms of data from Shewhart's first book. These data came from the same process over four sequential time periods. The process was operated unpredictably during the first three time periods. After they found the assignable causes and made them part controlled process inputs the process was operated predictably during the fourth time period. Thus, as can be seen from the fourth histogram, the skewness in the first three histograms is not a property of the process, but merely the result of the random walk taken by the process when it was operated unpredictably.

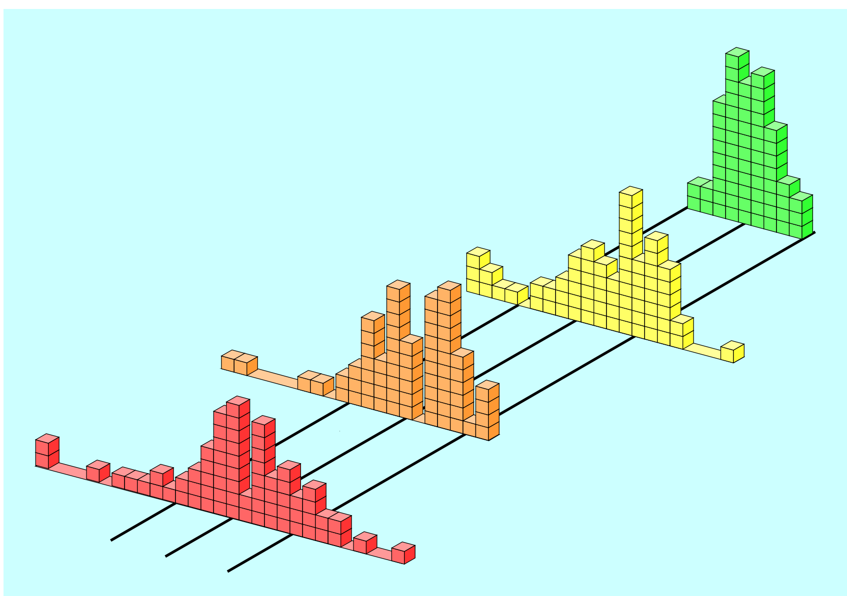


Figure 4: Four Histograms from the Same Process

Nevertheless, those who take the statistical approach to filtering out noise may react to a skewed histogram by suggesting a transformation of the data "in order to make the data look more normal."

To illustrate the effect of a logarithmic transformation we will use the Hot Metal Transit Times (shown written in rows in Figure 5). These values are the times (to the nearest five minutes) between the phone call alerting the steel furnace that a load of hot metal was on the way and the actual arrival of that load at the steel furnace ladle house.

40	45	125	100	40	40	100	65	55	40	125	65	40	45	95
105	45	110	40	50	120	45	65	105	35	70	55	25	50	55
50	40	40	45	55	50	45	125	55	100	40	70	40	40	110
55	50	30	50	105	45	45	55	50	25	65	60	60	55	70
55	45	100	60	45	145	45	50	65	180	60	45	35	35	55
55	55	50	120	35	45	35	45	55	50	70	45	75	60	45
60	40	60	40	50	60	65	95	65	60	50	25	25	100	50
60	45	35	40	30	180	50	30	30	30	65	130	80	20	45
65	65	45	40	50	25	120	30	115	50	85	40	35	40	40
55	50	25	75	55	50									

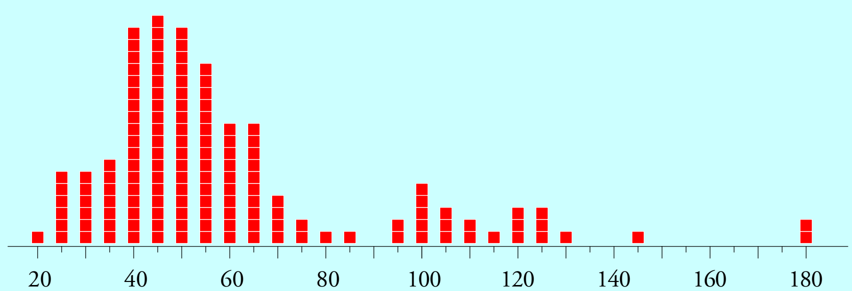


Figure 5: Hot Metal Transit Times

Given the skewed nature of these data some authors would suggest using a logarithmic transformation. When we take the natural logarithm of each of these transit times we get the results in the histogram in Figure 6. (The horizontal scales show both the original and transformed values.) Notice how the values on the left of Figure 6 are spaced out while those on the right are crowded together. After the transformation the distance from 20 to 25 minutes is about the same size as the distance from 140 to 180 minutes. (How would you begin to explain to the superintendent of the steel furnace that your "statistical analysis" requires that you treat a five minute delay as being equivalent to a 40 minute delay?)

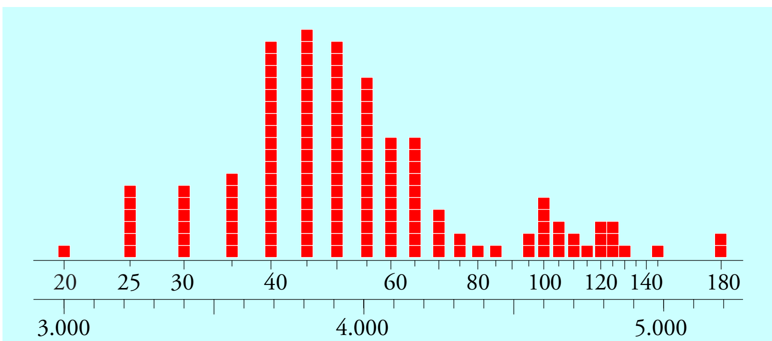


Figure 6: Logarithms of the Hot Metal Transit Times

By itself, this distortion of the data is sufficient to call into question the practice of transforming the data to achieve statistical properties. However, the impact of non-linear transformations is not confined to the histograms. The logarithm changes the scale for the data. Since the scale of the data is the basis for filtering out the noise, we end up with a completely different analysis.

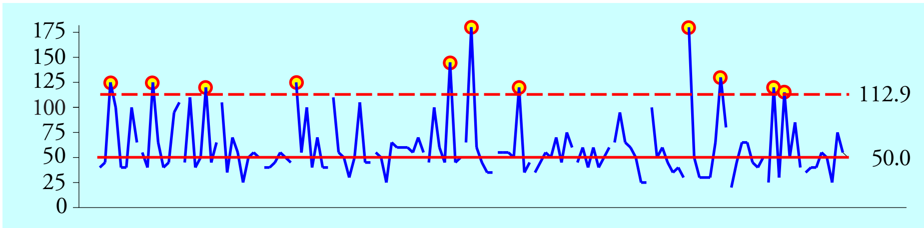


Figure 7: *X* Chart for the Hot Metal Transit Times

Figure 7 shows the *X* Chart for the original, untransformed data. The central line shown is the median of the *X* values, and the limits are based on the median moving range of 20 minutes. Eleven of the 141 transit times are above the upper limit, confirming the impression given by the histogram that these data come from a mixture of at least two different processes. Even after the steel furnace gets the phone call, they still have no idea when the hot metal will arrive at the ladle house.

However, if we transform the data before we put them on a process behavior chart we end up with Figure 8. There we find no points outside the limits!

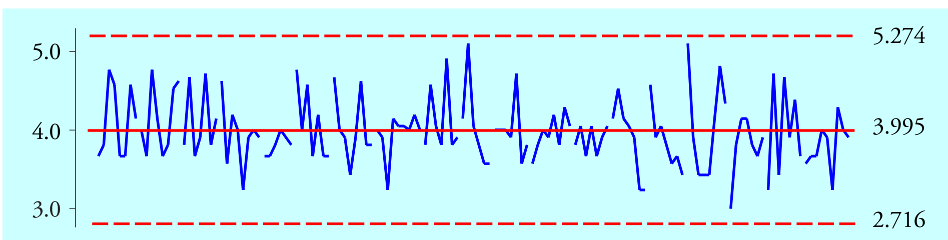


Figure 8: *X* Chart for the Logarithms of the Hot Metal Transit Times

Clearly the logarithmic transformation has obliterated the signals. What good is a transformation that changes the message contained within the data? Transforming the data to make the histogram more bell-shaped is simply a complex way of distorting both the data and the truth.

HOW CAN SYMMETRIC LIMITS WORK WITH SKEWED DATA?

Rather than getting lost in the statistical approach, we can use Shewhart's symmetric, three-sigma limits with skewed data. As may be seen back in Figure 2, when a process is operating close to some boundary condition the histogram can become skewed. When these boundary conditions fall within the computed limits the boundary condition will take precedence over the computed limit and we will essentially have a one-sided chart. Notice that regardless of how skewed the distribution may be, the upper three-sigma limit continues to cover the bulk of the long tail. More extensive studies in references [1], [2], and [3] show that three-sigma limits will always cover more than 97.6% of the routine variation even when used with extremely skewed probability models. Since filtering out more than 97.6% of the noise meets the traditional criterion for a conservative analysis, we can conclude that the process behavior chart will provide a conservative analysis regardless of what probability model you might think is appropriate.

Thus, anyone who starts talking about the skewness of the data as a barrier to placing those data on a process behavior chart simply does not understood the robustness of the process behavior chart approach. Moreover, anyone who wants to use a logarithmic, trigonometric, or square-root transformation on your data either is hopelessly confused about how the process behavior chart works or else is seeking to sweep signals of process change under the rug to make them go away.

USING HIGH-LEVEL REPORT CARD CHARTS

Finally, you may have been told that you can tell if a process is predictable by using a long history of some high-level, report-card measure. This also is wrong.

When we combine multiple measures into a high-level summary we are combining the background noise from each of the components. In consequence, the more components we combine into our summary, the greater the noise level within our summary, and the more "predictable" our summary will look. Consider the *X* chart for the quarterly sales values for one company over the past five years shown in Figure 9.

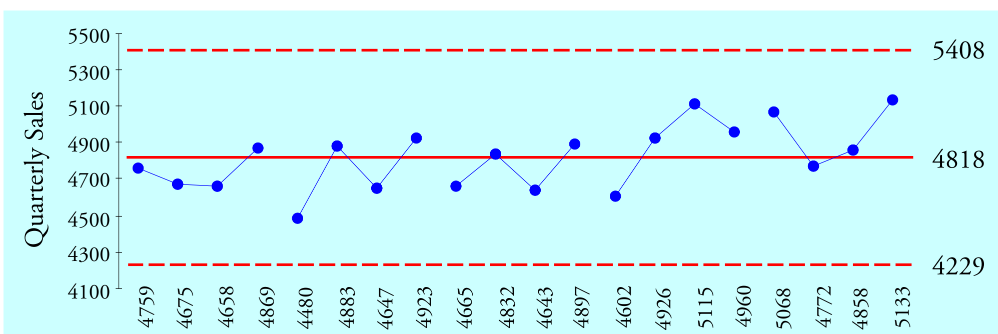


Figure 9: X Chart for the Quarterly Sales for Five Years

If they ever reached the upper limit they would be the darling of Wall Street. Conversely, before sales dropped to the lower limit they would be in serious trouble. So are these limits too wide? Not really. These wide limits are simply warning the reader that it would be a mistake to react to changes in these values. They are so full of noise that you will never be able to pinpoint an explanation for why the sales went up or why they went down in any given quarter.

But if we look at the regional sales that were combined to give the totals in Figure 9 we have the charts in Figure 10. There we see many signals of changes that were occurring at the regional level.

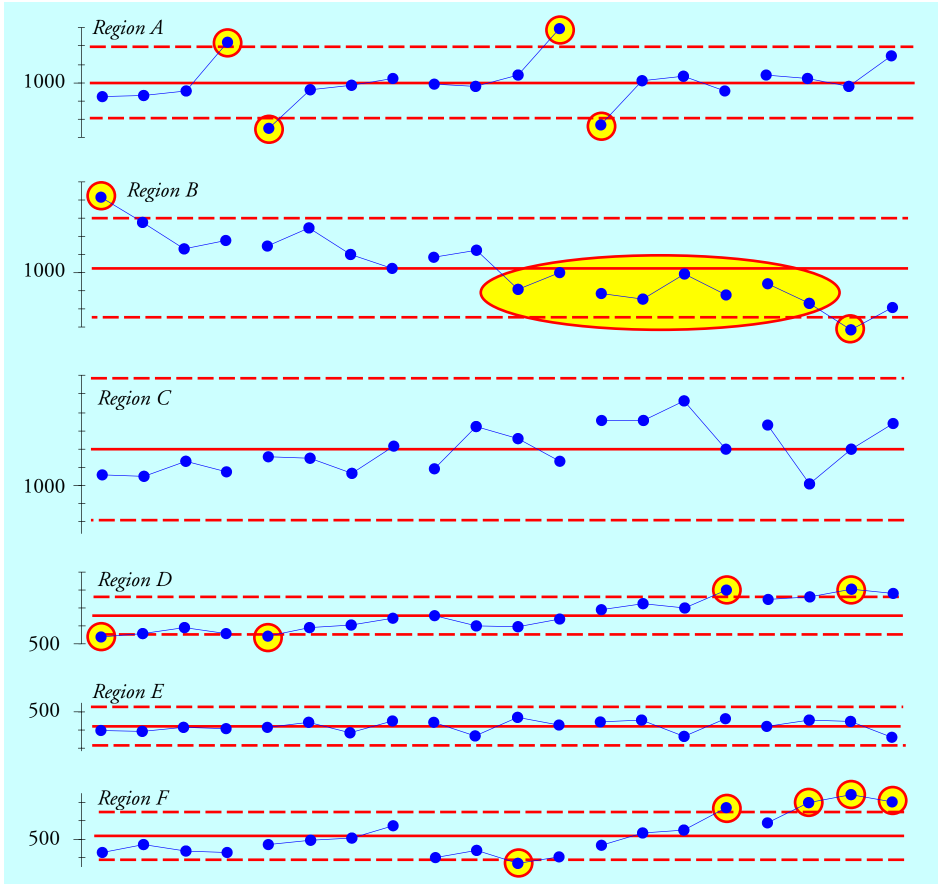


Figure 10: *X* Charts for the Quarterly Sales by Region for Five Years

Region A is playing with the sales figures to get good year-end values, Region B has steadily declining sales, while Region D has increasing sales, and Region F has come back from a disastrous Year Three, yet none of this shows up in Figure 9. If we mark the signals from Figure 10 on the graph in Figure 9 we get Figure 11.

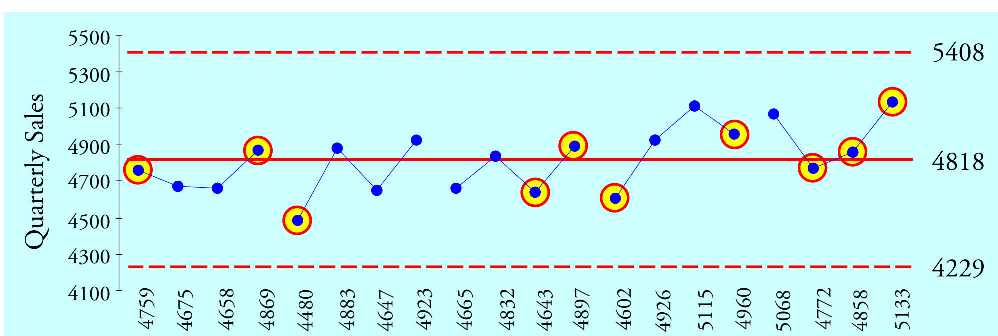


Figure 11: *X* Charts for the Quarterly Sales for Five Years with Signals from Figure 10

Ten of the 20 quarters had one or more points outside the limits at the regional level. Yet none of these points even begin to come close to the limits in Figure 11. So while report-card charts are valid, they will often have very wide limits because the data are full of noise. Any signals that show up on a report-card chart will be real, but *the absence of signals on a report-card chart is not evidence of a predictable process.*

When the data are full of noise the measure will be of little use in running the business. While the chart may serve as a report-card, it cannot be used to identify what caused the values to change. *Here it is not the chart that is at fault, but rather the idea that you can interpret the changes in the data as anything other than noise.* While movement in one direction may be good and movement in the other direction may be bad, interpreting the ups and downs as if they amount to signals of changes in the process is always a mistake.

On the other hand, using a report-card chart to say a process is predictable is simply sweeping the signals under the rug. *Disaggregation is always required to discover what is happening.*

SUMMARY

We have examined the erroneous thinking behind four misconceptions about process behavior charts. Hopefully at this point you understand why the chart for individual values does not require normally distributed data; why you do not need to use a logarithmic transformation before placing skewed data on a process behavior chart; why high-level summaries are never going to be highly skewed; and why the absence of signals on a report-card chart does not mean the processes represented are being operated predictably.

While some these misconceptions have their origin in the use of the statistical approach to filtering out noise, some of the others are just simple confusion arising out of the complexity of the statistical approach. While the statistical approach works nicely when analyzing experimental data, it runs into problems when it is used with observational data arising out of day-to-day operations. Here the inability to ever fully define an appropriate probability model for the original data simply pulls the foundation out from under the statistical approach and leaves the user wandering around in statistical hell. But you can escape and cross back over the river Styx by using Shewhart's approach to filtering out the noise when working with process behavior charts.

References

1. "Properties of Probability Models: Part One, What They Forgot To Tell You About the Weibull Distributions," ***Quality Digest***, August 3, 2015.

2. "Properties of Probability Models: Part Two, What They Forgot To Tell You About the Gammas," ***Quality Digest***, September 1, 2015.

3. "Properties of Probability Models: Part Three, What They Forgot To Tell You About the Lognormals," ***Quality Digest***, October 5, 2015.