

No Data Have Meaning Apart from Their Context

How should you organize and plot your data?

Donald J. Wheeler

As Davis Balestracci frequently emphasized in his column, it is fundamental to understand the context of the data before you begin to do any computations. It is the background for your data that determines how you should organize the data, how you should analyze the data, and how you should interpret the results of your analysis. Once you ignore the context you are like a train that has gone off the track, with the inevitable result being a train wreck.

One day a company sent me some data that they had spent over a month collecting. These data represented the results of an experimental study carried out using production batches. For each of 30 batches they recorded all sorts of production information, along with the experimental conditions that applied to that batch. At the end of the production process they had taken 40 items from each batch and measured the property of interest. Thus, they had a total of 1200 values: 40 values for each of 30 batches.

With 40 values per batch, they had gotten busy and drawn a histogram for the values for each batch. They had also computed the average, the standard deviation statistic, and even the skewness and kurtosis statistics. In their attempt to make sense of this mountain of data they had created a summary sheet for each batch. These 30 summary sheets each contained the production and experimental information for each batch, along with the histogram and the set of descriptive statistics. Having gotten this far, they had stalled out. There was still too much information to assimilate, hence they had decided to call on my services.

As they were preparing to send me the data, someone realized that I was likely to ask for a control chart, so they quickly placed the data on an Average and Standard Deviation Chart using the 30 subgroups of size 40 defined by the 30 batches. This chart was stapled on to the stack of 30 summary sheets and the whole package was delivered to my office.

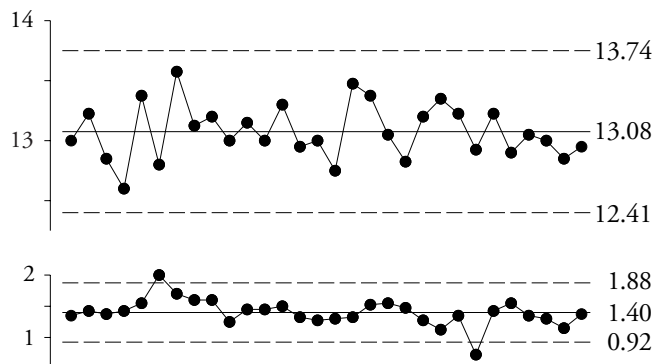


Figure 1. Original Average and Standard Deviation Chart

While the summary sheets contained all sorts of contextual information, none of that information had been used in creating the Average and Standard Deviation Chart. In fact, the order of the points on the chart made it clear that the chart had been done as an afterthought—the order of the points on the chart matched the order of the summary sheets in the stack, and the stack had been rather thoroughly shuffled prior to being sent to me. This chart is shown in Figure 1.

Upon recognizing that the Average and Standard Deviation Chart had not been organized in any manner that respected the context for the data, I immediately rearranged the chart using the dates when each batch of rubber was compounded.

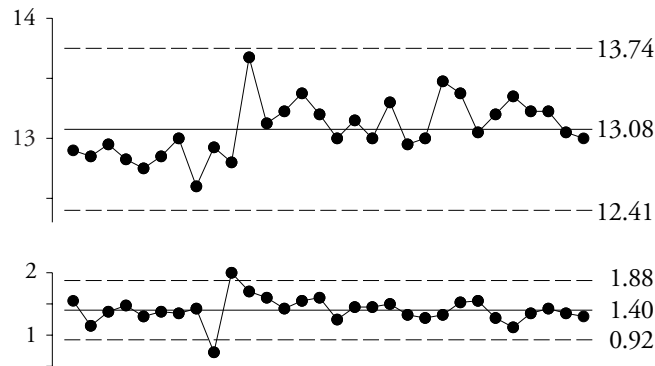


Figure 2. Average and Standard Deviation Chart in Time Order

With the new ordering I hit the jackpot, as may be seen in Figure 2. All of the lower values for the property of interest occurred before a certain date. All of the elevated values for the property of interest occurred after that date. This one change is the largest signal within these data. It did not line up with any of the experimental factors. Hence, regardless of whether or not the experimental factors have any influence upon the property of interest, there is at least one dominant factor that they have overlooked. If they continue to ignore this dominant factor, they will continue to have an unpredictable production process. Tweaking the factors they studied will not make this unknown-but-dominant factor go away.

February's column was entitled "First, Look at the Data." This is important even when the data come from an experimental study. Failure to do so can result in a train wreck of an analysis. After discovering the presence of this dominant factor in their data, I went to their plant, showed them my Average Chart for their data, and watched their jaws drop open. Then I flew home and sent them a bill for simply plotting their data in time order sequence—something which they should have done themselves.