

But the Limits Are Too Wide!

When the XmR Chart Doesn't Seem to Work

Donald J. Wheeler

Last month I described what makes the XmR Chart work. This month I will describe some common failure modes for the XmR chart and show how they come from a failure to follow the two fundamental principles behind the XmR chart.

When administrative and managerial data are placed on an *XmR* Chart the first reaction will frequently be that the limits are far too wide—"we have to react before we get to that limit." So what are we to do when this happens? Are the limits really too wide? There are three cases to consider: the data are full of noise; the data are full of signals; and the data, in turn, represent different processes.

THE DATA ARE FULL OF NOISE

Administrative and managerial data tend to be report card data. The measures are accumulated across departments, across plants, across regions, and even across countries before being presented for consumption by the executives. At each stage, as the data are aggregated, the noise of each data stream is also being aggregated into the total, so that, by the time the finished value is presented, it has a lot of noise in the background. For example consider the quarterly sales values for one company shown in Figure 1.

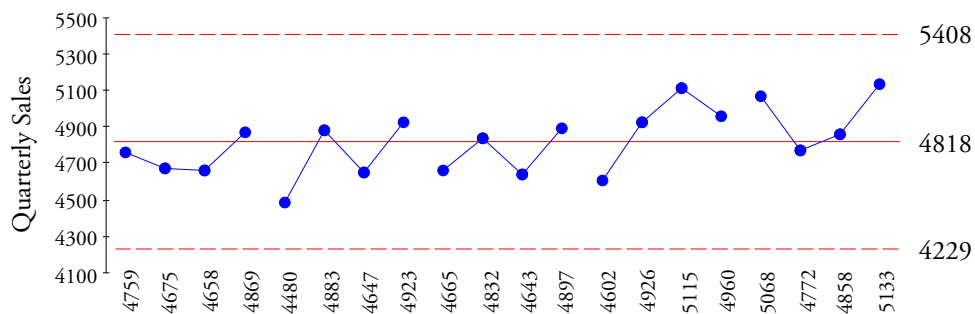


Figure 1: *X* Chart for Quarterly Sales for All Regions for Years 01 to 05

If they ever reached the upper limit on the *X* Chart, they would be the darling of Wall Street. Conversely, if the sales dropped to equal 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 have gone up, or why they have gone down. While you may believe that the advertising campaign helped, these data contain far too much noise to allow you to demonstrate that the advertising campaign had an effect upon the sales! In short, report card data result in report card charts, and the report card does not identify what needs to be done to change things.

Consider the regional sales that were combined to give the graph in Figure 1. These six time-

series are shown on their own X-Charts in Figure 2. There we see many signals of changes that were occurring at the regional level. However, when these six time-series were combined into the report card measure in Figure 1, all of the signals of change were lost in the noise. Thus, one way to deal with data that have very wide limits is to disaggregate the time-series into its constituent components and consider each component separately.

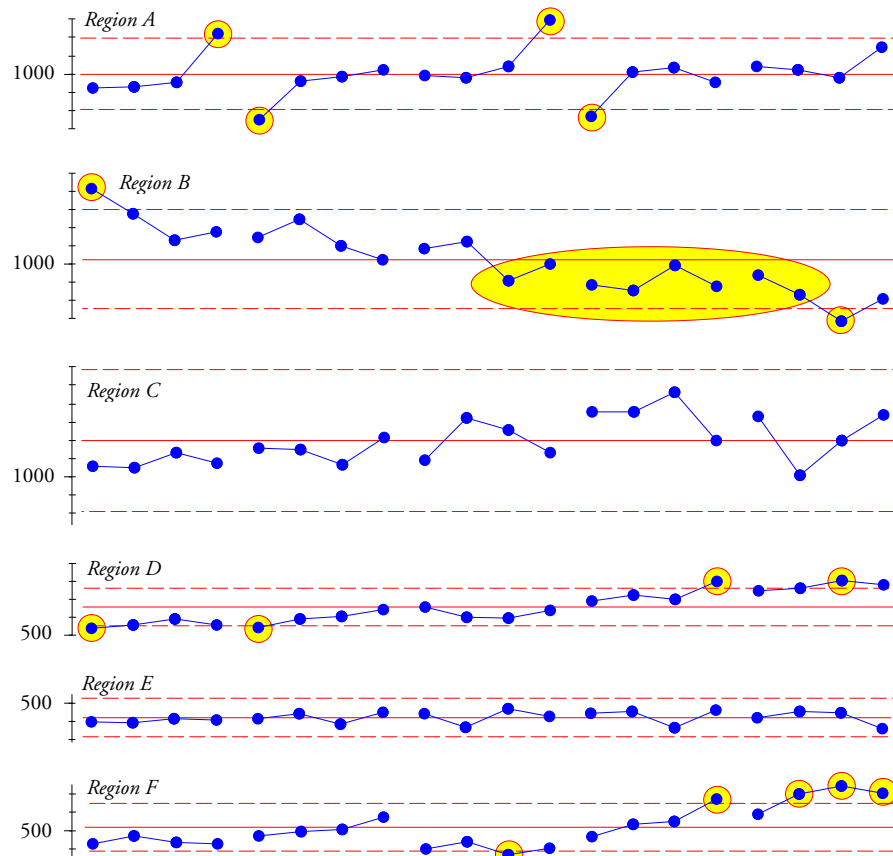


Figure 2: X Charts for Quarterly Sales by Region, Years 01 to 05

So while report card charts are valid, they will often have very wide limits because the data are full of noise. The wide limits serve as a warning that it is virtually impossible to assign an explanation to why a point goes up or why it goes down—the routine variation inherent in the data stream itself is sufficient to be the cause of any particular increase or decrease. While movement in one direction may be good and movement in the other direction may be bad, *the metric itself is too full of noise to be used to run the business!*

Any failure to appreciate this point will result in the interpretation of noise as if it amounted to signals, which is the first type of mistake that can be made in interpreting data.

THE DATA ARE FULL OF SIGNALS

Because of the way the limits are computed they can become inflated whenever changes become commonplace.

Moving ranges are used in order to separate the exceptional variation from the routine variation—by computing limits using either the average or the median of the moving ranges we hope to dilute the impact of any exceptional variation that is present. *Thus, in the end, the computations implicitly assume that exceptional variation will be an occasional thing.* As more and more of the moving ranges are affected by some Assignable Cause it will become harder to distinguish between exceptional variation and routine variation. Finally, at some point, the exceptional variation may become so commonplace that it will look like routine variation to the computations.

So how can we tell if the data are full of signals? In some cases the context will provide the key, in others the only way will be to have a reference period where the process is operated without the influence of the Assignable Cause.

As an example of the first case, where the context will provide a clue that the data are full of signals, we use the data for neonatal autopsies for one hospital over a ten year period. The data and X chart are shown in Figure 3.

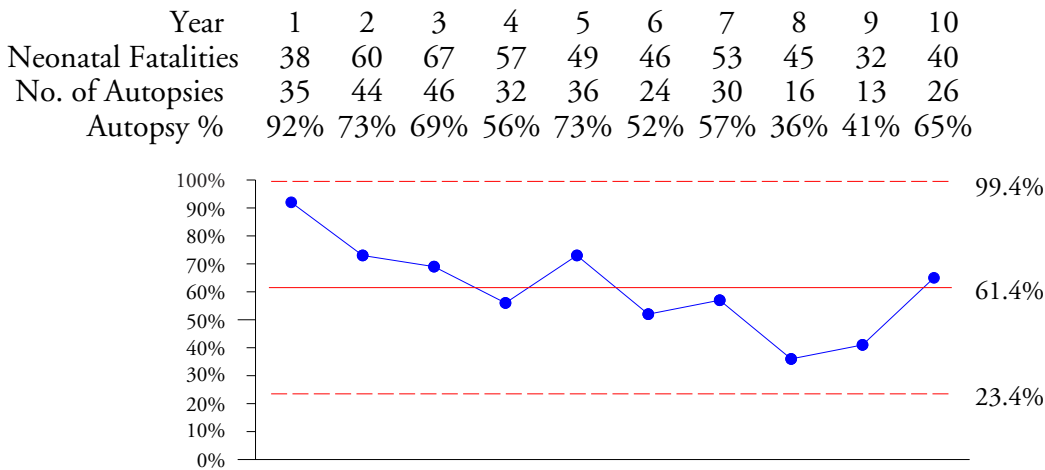


Figure 3: X Chart for Neonatal Autopsies

The limits of 99.4% and 23.4% provide no discrimination here. They are far too wide for any practical purpose. However, any process that changes from 92% in one year to 56% in another year, and then to 36% in a subsequent year is clearly not the same from year-to-year. The limits do not show these changes because these data are full of signals. Fortunately, limits are not needed here since common sense is sufficient. The first principle for understanding data is that no data have any meaning apart from their context. This means that context must always be the starting point for any analysis. Here the context tells us that this process is changing from year-to-year.

(One author suggested placing these data on a *p*-chart as a way of fixing these wide limits. However, since the likelihood of a neonatal fatality being autopsied is not likely to be the same for all fatalities in a given year, these data can not be modeled with a binomial distribution. For more on this problem see my column “What About *p*-Charts?” from October of 2011.)

So how does this differ from the situation where the data are full of noise? When we are working with a highly-aggregated metric, wide limits are most likely to be due to excessive noise.

When we are working with a simple, localized metric, wide limits may be due to changing conditions from period to period. Judgment, experience, and contextual knowledge are required.

With annual values it is often helpful to break the data down into shorter time periods. Given the small counts in Figure 3, quarterly summaries would be about as far as we should go (monthly values would involve very small counts indeed). Figure 4 shows what the quarterly data for Neonatal Autopsies might look like.

| Neonatal Autopsies By Quarter | | | | | | | | | | | | | | | | |
|-------------------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| Year | 1 | | | | 2 | | | | 3 | | | | 4 | | | |
| Quarter | I | II | III | IV | I | II | III | IV | I | II | III | IV | I | II | III | IV |
| Fatalities | 17 | 6 | 8 | 7 | 9 | 17 | 11 | 23 | 17 | 15 | 17 | 18 | 13 | 16 | 18 | 10 |
| Autopsy | 16 | 5 | 8 | 6 | 6 | 12 | 7 | 19 | 9 | 12 | 12 | 13 | 7 | 8 | 13 | 4 |
| Proportion | .941 | .833 | 1.00 | .857 | .667 | .706 | .636 | .826 | .529 | .800 | .706 | .722 | .539 | .500 | .722 | .400 |
| Year | 5 | | | | 6 | | | | 7 | | | | 8 | | | |
| Quarter | I | II | III | IV | I | II | III | IV | I | II | III | IV | I | II | III | IV |
| Fatalities | 6 | 10 | 13 | 10 | 13 | 3 | 17 | 13 | 10 | 13 | 13 | 17 | 15 | 8 | 7 | 15 |
| Autopsy | 4 | 8 | 10 | 4 | 7 | 2 | 10 | 5 | 3 | 7 | 8 | 12 | 6 | 1 | 3 | 6 |
| Proportion | .75 | .800 | .769 | .400 | .538 | .667 | .588 | .385 | .300 | .538 | .615 | .706 | .400 | .125 | .429 | .400 |
| Year | 9 | | | | 10 | | | | | | | | | | | |
| Quarter | I | II | III | IV | I | II | III | IV | | | | | | | | |
| Fatalities | 8 | 9 | 9 | 6 | 10 | 10 | 10 | 10 | | | | | | | | |
| Autopsy | 3 | 4 | 5 | 1 | 6 | 7 | 6 | 7 | | | | | | | | |
| Proportion | .375 | .444 | .556 | .167 | .600 | .700 | .600 | .700 | | | | | | | | |

Figure 4: Neonatal Autopsies By Quarter

If we use the first year as our baseline, the *XmR* Chart will have an average of 0.908 and an average moving range of 0.139. This will give a lower limit for the X Chart of 0.538, and an upper limit that exceeds 1.00. With limits of 54% to 100% this chart will hardly provide the most precise analysis, but it is sufficient to begin to tell the story contained in the data, as may be seen in Figure 5.

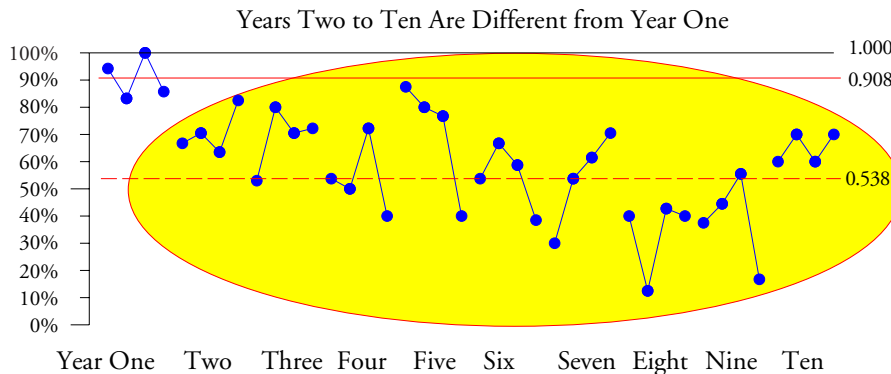


Figure 5: X Chart for Quarterly Neonatal Autopsies

With three out of four points closer to the lower limit than the central line, Year Two can be said to be different from Year One. This impression is further confirmed when the first point of

Year Three falls below the lower limit.

But wait, the limits in Figure 5 are only based on 4 points! Yes, that is true, but this analysis is sufficient to show that a change has occurred. The objective is to gain insight and to share it with others, and to this end, the best analysis is the simplest analysis that provides the needed insight. It is not a matter of using the right amount of data, or computing the best estimates of the limits, but rather using the data in context to tell the story of what is happening. The chart in Figure 3 failed to do this because it used annual values and the year-to-year changes inflated the limits. The chart in Figure 5 succeeds in doing this because, in spite of the small counts involved and the small number of points used in computing the limits, the limits were not inflated by the year-to-year differences.

In fact, the chart in Figure 5 can be improved by using multiple sets of limits to tell the story. Since Years Two and Three look different from Year One, use them to compute new limits. Now Years Four, Five, Six and Seven may be seen to be detectably different from Years Two and Three. So compute new limits using Years Four through Seven. Now Years Eight and Nine are seen to be detectably different from Years Four, Five, Six and Seven. So compute new limits using Years Eight and Nine. Now we see that Year Ten is detectably different from Years Eight and Nine, and so we compute new limits using Year Ten. In this way we end up with Figure 6.

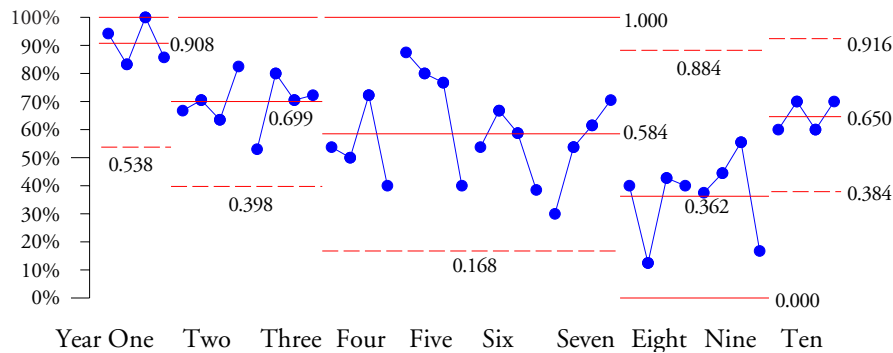


Figure 6: Revised X Chart for Quarterly Neonatal Autopsies

In Year One the autopsy rate was over 90%. In Years Two and Three it dropped to 70%. In Years Four through Seven it dropped to 58%. In Years Eight and Nine it dropped to 36%. In Year Ten it went back up to 65%.

None of the limits obtained here are very tight. This is because these data still contain a substantial amount of noise. However, this does not stop us from telling the story in these data when we disaggregate the annual summaries and use the limits intelligently. Here the interesting question is what happened at the end of Years One, Three, Seven, and Nine. It turns out that each of these points corresponds to a change in personnel. In this hospital it was the job of the chaplain to obtain permission for a neonatal autopsy, and the different chaplains did this job differently.

AN IMPORTANT POINT

The primary question of data analysis is the question of homogeneity. If the data are

reasonably homogeneous, then we are justified in assuming that the underlying process is being operated predictably, and in using the central line, limits, and other statistics computed from the data to characterize that underlying process.

However, if the data show evidence that the process is changing, the focus shifts from the using the computed values to characterize a single process to one of using the computed values to detect when and how the process is changing. Until we understand the story told by the data, we will not know how to operate the process up to its full potential. Here there is no such thing as computing the right limits or having enough data.

THE DATA ARE FULL OF SIGNALS: PART TWO

An example of using a reference period to determine if the data are full of signals is provided by the data on how one plant ran under two different managers. Manager One looked at the Productivity Reports each morning to see how the plant performed. If the output for the final, bottleneck step fell short of the scheduled amount, then he would make adjustments in manpower and operations for the current day. On the whole he ended up making adjustments over 80 percent of the time. Figure 7 shows the X Chart for the plant output for 100 days under Manager One.

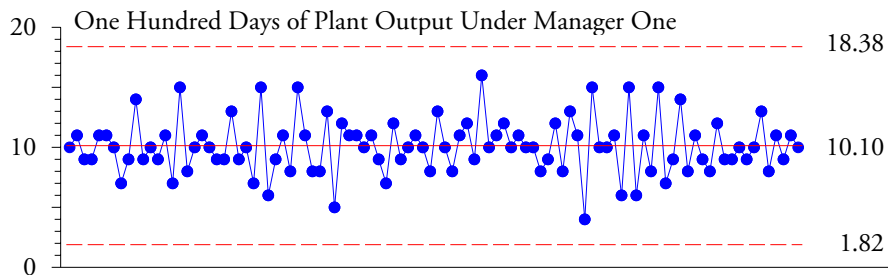


Figure 7: X Chart for Plant Output Under Manager One

Manager Two succeeded Manager One. He also looked at the Productivity Report each morning. However, he plotted the output values on an X Chart and only made adjustments to the operation of the plant when a point went outside the limits. Figure 8 shows the X Chart for the plant output under Manager Two. Manager Two had one-third less variation than Manager One along with a higher average output, using the same work force in the same plant.

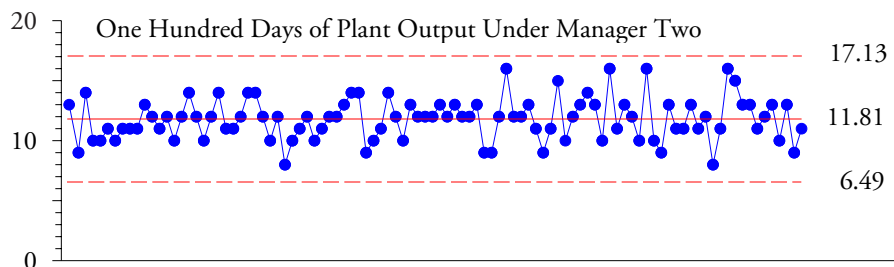


Figure 8: X Chart for Plant Output Under Manager Two

Manager One's adjustments were so frequent (over 80 percent of the time) that the impact of those changes looked exactly like routine variation to the computations. Manager One had 150% of the variation displayed by Manager Two, along with a lower average daily output, yet his data appear to come from a predictable process.

In order for an XmR Chart to work it is important that successive values be logically comparable. *This means among other things that the conditions under which the successive values are obtained will need to have remained the same from period to period.*

If the time periods have been made so large that the system is bound to have changed from one period to the next, or if the data were collected while the process was deliberately being adjusted or changed, then the resulting data are likely to be full of signals and the limits may be inflated by those signals.

THE DATA REPRESENT DIFFERENT PROCESSES

This is simply a special case of the problem where the data are full of signals. Here the data occur in a natural time order, yet they represent two or more conditions, resulting in an apples-to-oranges time series.

The example for this case comes from an allergist who had his patients track their lung congestion using a Peak Expiratory Flow Rate (PEFR) Gauge. The patient would exhale as hard as possible through the gauge and it would record the flow rate in liters per minute. The protocol consisted of the patient getting one flow rate value in the morning and one flow rate value in the evening. The morning reading was to be obtained prior to taking any medication. The evening reading was to be taken 15 minutes after using the bronchodilator inhaler. Thus, this sequence of values represented two states: a.m. pre-med. and p.m. post-med. As a time series, the physician tried to place these values on an XmR Chart. The resulting X Chart is shown in Figure 9.

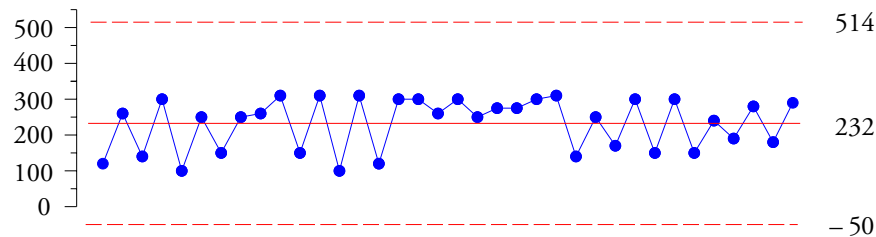


Figure 9: X Chart for PEFR Scores for One Patient

The physician could not make sense of these wide limits. The upper limit was unreasonable for this patient, and the lower limit was nonsense. Of course, these wide limits are a result of the large daily swings. This chart violates the guideline given above. No progress was made in using these charts as an adjunct to clinical practice until the physician started charting the a.m. pre-med. values alone. When he did this, the limits began to make clinical sense.

SUMMARY

When the limits seem to be too wide to be practical, it is important to determine whether this

is because the data are full of noise, or because they are full of signals.

If the data are full of signals, then the wide limits are incorrect and the organization of the chart is at fault. Here it is up to the user to organize the data in such a way that the charts can be useful.

If the data are full of noise, then 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 make use of the data in spite of the noise they contain.* The failure to understand this point will inevitably result in Manager One Syndrome.

It is interesting to note that Manager One Syndrome is encouraged by the emphasis on looking at all of the current values together on a monthly “dashboard.” While this does encourage the use of multiple measures rather than reacting to each value separately, it still lacks the filter that is needed to separate the signals (where there is something to be learned) from the noise (where there is no change from previous months). With the monthly dashboard any two numbers that are not the same are thought to be different. Unfortunately, while this is true when it comes to arithmetic, it is not true when it comes to the interpretation of data. In this world two different numbers may well represent the same thing.

All data contain noise.

Some data also contain signals.

Until you can differentiate between the noise of routine variation and the signals of exceptional variation you are likely to be misled by the noise.