Working Paper

BEYOND VARIANCE ANALYSIS
Part 3: SIGNALS FROM NOISE

Steve Morlidge
August 2016
Acknowledgements

This paper is the preliminary third chapter in Steve Morlidge’s next book provisionally titled ‘Present Sense: A practical guide to the science of measuring performance and communicating it’.

We have agreed to ‘serialise’ the book in the form of Working Papers that will be added to our members’ knowledge base and shared in our network. We would like to offer our appreciation to Steve Morlidge for allowing us to publish his book as it evolves in this manner. This is highly interesting and relevant for our members.

In the previous paper, Steve outlined the case for using dynamic measures of performance and provided examples of how these can be applied in practice. But while presenting history in this manner helps set the right context for interpreting data, trend analyses are of limited value in making sense of recent news. Every period we get a new piece of data which potentially could contain significant information. We cannot afford to wait for several periods until the picture gets clearer. The problem we have in trying to make sense of what is going on ‘in the moment’ is the result of the mixed blessing of having large amounts of granular data. Our ability to capture data on a scale way beyond what was possible in even the very recent past holds the tantalising prospect of new insights and greater agility, but the superabundance and variability of our new found data riches can easily overwhelm our ability to make sense of it. To be able to exploit these data seams and extract actionable insights we need to have a technique for finding the nuggets amongst the rubble - to detect signals amongst all this noise. This is the subject of this paper.

Steve Morlidge has spent most of his professional career in designing and running performance management systems in Unilever. Steve co-authored ‘Future Ready: Mastering business forecasting’ (John Wiley, 2010) and has written many papers on business forecasting. Steve has a PhD in Management Cybernetics.

Steve is the former Chairman of the BBRT and remains a tremendous support of our network for which we are most grateful.

About Beyond Budgeting Institute and BBRT

The Beyond Budgeting Institute is at the heart of a movement that is searching for ways to build lean, adaptive and ethical enterprises that can sustain superior competitive performance. We promote a set of principles that lead to more dynamic processes and front-line accountability. Organizations that follow this approach transform their management model in line with these principles.

Our ideas are spread through the Beyond Budgeting Round Table (BBRT); a shared learning network of member organizations with a common interest in transforming their performance management models to enable sustained, superior performance. We help organizations learn from worldwide best practice studies and encourage them to share information and experiences to move beyond command and control.

For more information about Beyond Budgeting and BBRT membership, please contact:

Beyond Budgeting Institute
One Kingdom Street
Paddington Central
London W2 6BD
United Kingdom
Tel: +44 20 3755 3692
Mail: info@bbrt.org
Web: www.bbrt.org

© Beyond Budgeting Association Ltd.
No part of this publication may be reproduced or transmitted in any form or by any means without permission in writing from the author.
Introduction

This is the third paper in a series of four on the subject of variance analysis.

Finding a way of making sense of performance without using simplistic comparisons between target and actual is important for the BB community because financial people are taught to analyse performance using variance analysis – a technique that demands a budget or a target. Also, although full-blown variance analysis is peculiar to finance, the use of fixed targets to measure performance pervades the traditional approach to management - with many negative consequences.

Over this series of papers I will make the case for using alternative analytical approaches to measuring and making sense of performance and demonstrate how they address the weaknesses of the traditional approach. These techniques can be applied to any kind of data, although most of the examples I cite refer to ‘historic’ financial data, because I am seeking to show how we dispense with traditional financial variance analysis.

I am conscious that focussing on historic data leaves me open to the criticism that my proposals are of limited value because they are backward rather than forward looking; like ‘driving a car using the rear view mirror’ according to critics. I think this kind of criticism is overstated and needs to be addressed before we move to the subject of this paper: how to extract signals from noise.

Why history is important

I co-authored a book about forecasting (Future Ready) and spend much of my time in the forecasting community, but I think that the value of forecasts can be overstated.

The reason for this is simple: even the best forecasts are no more than educated guesses about what might happen in the future, based on assumptions about the environment and how this could impact business performance. Because we can never be sure that our assumptions are correct, the only thing that we can be certain of is that our forecasts will be wrong, sometimes by an insignificant amount, but sometimes in a way that can be positively misleading. In particular it is notoriously difficult to forecast changes in a trend – what economists call structural breaks.

Why, then, do we forecast?

The reason we forecast is that many of the decisions that we make will take some time to have an impact – weeks, months or perhaps even years. As a result we need to anticipate the future to ensure that we take appropriate action in good time, in just the same way that the captain of a super tanker needs radar to navigate safely.

So, when key decision have long lead times it is important to have the ability to anticipate the future. But although we rely upon forecasts to make decisions we need to be aware that we can never predict the future perfectly. We need to have a good understanding of the risk attached to our forecasts and put in place measures to mitigate these risks (or exploit unforecast opportunities).

Actual data, on the other hand, is not based on guesswork. Furthermore, forecasts are not somehow separate from history; they are dependant on it, because:

1. The most commonly used forecasting techniques are based on the extrapolation of historic trends (using time series analysis)
2. Forecasts are compared to actual outcomes to validate the assumptions used to produce them
3. The best way to assess the credibility of forecasts is to compare them to a range of alternative scenarios based on extrapolation from the past, in the way outlined in my previous paper.
Forecasts reflect our assumptions about the world. Real data is the only way we can ever know the world, and so has to be our primary source if we want to understand performance. The issue is not whether history is important, but how we make sense of it, which is the subject of this paper.

In summary, I think the criticisms often made of the use of historic information are misplaced. The inadequacies of traditional methods of performance analysis do not lie with the data source but with the techniques used to analyse it because:

1. They rely on arbitrary targets as a comparator
2. They focus on single data points, which means that any sense of the dynamics of performance is lost
3. They cannot distinguish between the signal carried by the data, which we might need to act upon, and irrelevant noise.

In the second paper in this series, I tackled the second of these weaknesses, demonstrating some simple approaches to help expose performance trends. The same techniques also go some way to dealing with the third problem since any kind of average will suppress noise in the data to some extent (because randomness is cancels itself out in the process of aggregation).

But trend analysis on its own cannot be the answer because the frequency at which we collect financial data does not match the need to take decisions. Most businesses cannot afford to wait for half a dozen periods or more to establish whether historic trends in performance have continued or have changed. They need to act more quickly to exploit opportunities in the market place or to address problems or threats, based on the results for one or two recent periods. The question is, given the way that data can vary from period to period, how can we be sure that we are not over or under reacting to this new information? This problem has become more pressing since only a decade or so ago data was so sparse that we had time to interrogate and analyse it but today the volume, velocity and variety of data makes this impossible.

We need a more structured and scientific approach to help us make sense of the huge data sets that are now available to us. Our challenge is to extract actionable signals from noisy data and to do that we need to be able to answer a very basic question: what is noise?

**What is noise?**

In my experience we accountants struggle to get to grips with the concept of noise. Unlike engineers who encounter it in everything they do, we don’t even see it because we have never been made aware of its existence. Before we can take measures to deal with noise we need to understand what it is, which we will do with the help of an everyday example.

Imagine talking to a friend in a noisy bar where you have to shout to make yourself heard. In this context, it is clear what the signal is - it is the content of your conversation. Everything else that is going on – mainly other peoples’ conversations – is noise.

In their own context all these other conversations have meaning – they are the signal passing between two different people – but because they are not relevant to your purpose from your perspective they are just noise. And ironically, when there are lots of other conversations going on, we often find it easier to ignore them because many meaningful conversations going on at the same time eventually generate noise that is more random in nature and so less likely to interfere with our ‘signals’. This is the reason why actors in a movie crowd scene are told to mumble ‘rhubarb, rhubarb, rhubarb’ to mimic the effect of a lot of people all talking at the same time.

It should now be clear, at least in principle, what noise is. It is anything that distorts the signal that you are seeking to uncover, which in our context might be the level of sales, cost or market share. So for an insurance company for instance, the average rate of car accidents within a certain period is the signal – which they use...
to set insurance premiums - and the distribution of actual accidents around this average is the noise. Note that signals and noise are patterns in data; they cannot be traced back to individual data points. So while the pattern of accidents may be random, this doesn’t mean that individual accidents are random; each one can be explained – by a mistake or a failure in a car. We can’t point to an individual crash and say ‘this is the signal’ or ‘this is the noise’ since both properties are a product of a collection of accidents, just as we cannot point to a single value in a data set and say ‘this is the average’.

Insurance companies look at a large population of accidents over a period and use ‘population statistics’ to help them detect the signal but we do not have this luxury when we are assessing business performance. We haven’t got a large static data set we can analyse - we get one new data point at a time. The challenge we have is to determine whether the difference between our new data point and the one that preceded it is the result of chance or whether it is a sign that there has been a change in the signal, which we might need to act upon.

How can we do this?

The challenge

One thing is clear: we cannot rely upon our intuition to differentiate between signal and noise. The same pattern seeking qualities of our brain that makes graphical presentation of trend data such an effective way of communicating can also fool us into seeing patterns that do not exist in data. The fact that our ancestors saw the gods at play in the constellations of stars and we see faces in clouds shows that this is part of our genetic makeup.

But randomness can also fool simple arithmetic.

To illustrate this, take a look at this device:

It is called a ‘Galton Board’ after its Victorian era inventor, Francis Galton, who was an early statistician and cousin of Charles Darwin. He built it to explore the nature of statistical distributions, which – it had recently been discovered – often follow a distinctive pattern with specific mathematical properties.

A Galton Board is made from wood, nails and metal balls, rather like a bagatelle. Instead of the single metal ball in a bagatelle Galton used a large number that were held in a funnel at the top of the device as shown in the
diagrams on the right. When the balls are released, they fall down and bounce off a series of regularly spaced nails before dropping into a number of traps at the bottom of the device.

What Francis Galton was able to demonstrate with his machine was that with a sufficiently large number of balls the distribution across the traps always formed a bell shape that approximated a ‘Gaussian’ or ‘Normal’ distribution. Galton was fascinated that, while every ball obeyed the strict laws of Newtonian physics, the location of any individual ball could not be predicted because of the impact of random factors – the nails. This meant that nothing could be inferred about the location of the bottom of the funnel – the signal in this system - from a single ball; it could only be deduced from a population of balls.

To map this simple ‘game’ to our challenge we can think of the nails being like different conversations in a noisy room or the myriad number of decisions that might result in car crashes. The position of the funnel is like what our friend has just said, or the ‘typical’ level of accidents upon which insurance premiums should be set. What makes it so difficult to spot our signal is we receive data about our business ‘one ball at a time’. This is particularly problematic because if we only have a small sample a random sequence can look like a pattern— even to an arithmetical algorithm.

For example, below are three graphs drawn using the output from a simulated Galton Board where each ‘trap’ is assigned a number from 1 to 15. The results are plotted in the order in which they arrive – as a time series.

The green line is the signal because the funnel was placed above trap number 8. This doesn’t move but as you can see the sequence in which the simulated balls arrive make it look like there is a downward trend in the first ‘year’ made up of the first 12 balls. This impression is ‘confirmed’ by Excel when we fit a trend line to the data. And while the second year is ‘flat’ – in line with what we know the true signal to be - in year three we ‘see’ growth. Not only is our eye fooled into thinking there is a trend in this data; the trend-fitting algorithm is fooled as well.

In real life, our job is even more challenging because we have no way of knowing in advance whether the level of noise in any data set is small or large in comparison with the strength of the signal. We don’t know how many ‘nails’ there are that interfere with the signal. Neither can we assume that the signal is fixed in the way that the funnel is. There is likely to be some sort of a trend, or recurring pattern in the data, and any trend or pattern may change at any time, permanently or temporarily, all of which makes it even more difficult to separate the signal from the noise.

Faced with the difficulty of separating signal from noise in a single data series it is not surprising that most people shy away from the challenge of trying to make sense of the huge data sets that businesses now diligently harvest. With existing techniques it is simply too difficult to extract meaning from many hundreds of data series displaying confusing patterns of behaviour. But ignoring the problem doesn’t make it go away. Not only is it wasteful to ignore this data resource, it also means that we miss out on valuable insights.

We need help!
The solution

So we cannot trust our intuition, and we cannot trust simple arithmetic. What is the solution?

When we are dealing with data heavily infected with noise, we need to work with statistics and probabilities, not simple arithmetic. The mere mention of the word ‘statistics’ is almost guaranteed to scare many of you who have been forced to learn about it at school or as part of their professional qualification, because the way it is traditionally taught is heavily biased towards theory. The good news is that the vast majority of what performance analysts need to know about statistics in practice is easy to understand and use, not least because we only have one type of problem – making sense out of time series data. Also, unlike academics, we are not in search of ‘the truth’ in a strict scientific sense but just enough understanding to be able to make the right decision, quickly, most of the time. The decision may be to take action or it might be to do nothing because we do not have the evidence to justify it. Taking action when we should be sitting on our hands is not just wasteful – it usually makes things worse.

There are two ways to detect signals in a noisy data series. The first involves analysing the sequence in which data arrives, the second the size of the data values. Let’s first look at a simple way to detect signals by looking at data sequences from a probabilistic perspective.

Sequences

What are the chances of flipping a coin and getting eight heads or eight tails in a row? And at what point should you become suspicious that the coin you are using is ‘bent’ – that it has either two heads or two tails rather than one of each?

In this is the problem we face whenever we are analysing a sequence of numbers. The values have gone up, or gone down for the last ‘x’ periods. What are is chance that the sequence of numbers is evidence of a trend in the data rather than the outcome of chance process like the balls falling into the traps of the Galton Board?

We struggle with the problem intuitively because our brains have not evolved to tackle questions like this, but the logic behind the answer is very simple, as shown below.

<table>
<thead>
<tr>
<th>Coin Flips</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chances of All Heads or Tails</td>
<td>1:1</td>
<td>1:2</td>
<td>1:4</td>
<td>1:8</td>
<td>1:16</td>
<td>1:32</td>
<td>1:64</td>
<td>1:128</td>
<td>1:256</td>
<td>1:512</td>
</tr>
<tr>
<td>Probability</td>
<td>100.0%</td>
<td>50.0%</td>
<td>25.0%</td>
<td>12.5%</td>
<td>6.3%</td>
<td>3.1%</td>
<td>1.6%</td>
<td>0.8%</td>
<td>0.4%</td>
<td>0.2%</td>
</tr>
</tbody>
</table>

The chances of getting either a head or a tail on one throw are – obviously – 100%. The chance of getting either another head, or another tail, on the second throw is half of this since whenever you flip the coin there is a 50% chance of getting a head and a 50% chance of getting a tail. And on the third throw the chance is reduced by a further 50%, and so on. Proceeding in this way it is easy to calculate the probability of any number of heads or tail in sequence.

There are two obvious consequences we can draw from this.

Firstly, whenever there is any degree of chance involved, it is impossible to make a useful judgement from a single data point by the direction of movement alone.

Our coin will either land on head or tails, just in the same way that any new value will always either be higher or
lower than the last one. The fact that the second value is higher or lower has absolutely no significance at all; it’s a 50:50 call. But too often we believe that any and every change is significant. This is partly because traditional techniques do not provide us with the ability to make probabilistic judgements but also perhaps because we can always discover something that has ‘caused’ a value to move. So if revenue is higher this month than last, we will always be able to point to, say, a customer who has bought more than normal. The mistake we make is that we think that this is an explanation, which it isn’t. The logic is tautological, rather like saying we won a football match because we scored a goal: we wouldn’t have won the match if we hadn’t! Unless we get lucky, all we have done is identified a component of randomness so our commentary on the results, which refers to the upswing in sales for customer x, has no more value than claiming that Elvis has appeared to us in the shape of a cloud.

How does that make you feel about variance analysis now?

The second conclusion we can draw is that we can never be absolutely sure. Whenever there is randomness involved in a process there is always a chance that the result might simply be the result of ‘luck’. Any judgement we make can be wrong. But we cannot avoid making judgements, so these have to be made by explicitly taking into account the consequences of making a mistake.

For example, if a country were on high alert for terrorism, a policeman would be more worried about missing the signs of an imminent attack and so might make an arrest based on suspicion, perhaps accepting that there is a 1 in 10 chance of arresting an innocent person. Before someone is charged though, the chance for error needs to be smaller – perhaps 1 in 20 and a successful conviction requires that there be proof should be ‘beyond all reasonable doubt’ – perhaps the chance of a ‘false positive’ might be 1 in a hundred or even higher.

Let’s now apply this logic to a business scenario. Just as with coin flips, if we have 4 monthly data points in a row moving in the same direction, there is a 1 in 8 chance of us being wrong if we interpret this as a trend. So, if this were our criterion, on average we would have one or two false alarms (false positives) a year. Five data points in a row would reduce this risk to 1 in 16, meaning the chance of us being wrong is about 5%.¹

Ultimately the level at which you commit to a judgement or an action is a choice based upon a balance between the risk of false positives and false negatives and the cost attached to each outcome. The problem that many business now have is that they have large quantities of noisy data that that currently don’t have the time to analyse in detail. Consequently they either ignore this source of insights or jump to conclusions without sufficient evidence. This suggests that the bar for flagging up potential signals should be set very high – perhaps the equivalent of 8 values in a row since this means that the chance of a false alarm is less than 1 in 100. We now have a very simple statistical rule that helps us to identify the small number of significant events in very large data sets that doesn’t require anything more that the ability to count.

Size

Looking at sequence is a good start in helping us distinguish between signals and noise but it has some weaknesses. Specifically one stray data point can break a simple sequence and also no account is taken of the size of the difference from one data point to the next. Obviously the bigger the difference the more confidence we should have that something significant has happened that cannot be attributed to noise. But how do we decide what size of difference is significant? For example, in a large and stable market a change in the level of sales of 1% may be significant but in a very small volatile market the difference every consecutive data points is likely greater than this, purely because the level of noise in the market place is higher – so maybe something closer to 10% would be a better criterion. But how can we set the correct threshold?

¹ When we are making inferences from data, it is critical that we gather evidence from multiple sources. In legal parlance this increases the ‘weight’ of evidence, dramatically reducing the probability of false positives that we would calculate based on a single data source. But this only works if the corroborating evidence comes from a source that is independent of our primary data – which is why variance ‘analyses’ that refer to a component of the variance do not have any explanatory value.
To solve this problem we need a statistical method that helps us to set thresholds so we know what data we can safely ignore – where there is a high risk of false positives. Fortunately there is a simple robust technique that has been in extensive use for nearly 100 years called ‘Control Charting’.

Control charts allow us to calculate control limits for a time series based on sound statistical criteria, thereby enabling us to scientifically identify those data points or patterns of data points that differ significantly before from what has gone before. The control limits identify the normal level of variation in a data series. When the control limits are breached this is a sign that either the signal or the level of noise in the system has changed, or both.

At the last BBRT meeting in London (March 2016) I ran an exercise to illustrate some of the challenges we have in analysing performance data and demonstrate how we can use this technique to distinguish between signals and noise.

The first part of the exercise involved participants using the stopwatch functions on their smartphones. With their eyes closed they were asked to try and stop the clock at exactly 10 seconds, note the results, and then repeat the exercise 9 further times.

This part of the exercise gave participants an understanding of noise in a process and the impact that it has on performance measurement. I deliberately set up a very simple task, one that intuitively we might believe is easy to target and assess performance, and where there should be relatively little noise compared to the real world we normally work with, because of the lack of ‘moving parts’.

Based on your experience what do you think would constitute ‘good performance’? And what level of variance would be worthy of explanation and perhaps coded amber or red on a RAG chart? 5%? 10%?

Here is an example of the results for an individual participant.

**Variance from Target**

![Variance from Target](image)

The vertical scale shows the difference between each timed exercise and the target of 10 seconds, expressed as a percentage. The horizontal scale records the result of each of the 10 trials.

As you can see, despite being a very simple exercise involving nothing more than counting up to 10 in your head, there was quite a lot of variability in the results. If this represented nearly a year’s worth of revenue data rather than the results of simple exercise, what conclusions might you be tempted to draw? In other words, what would you treat as noise and what would you treat as a signal?

Now look at the results from all 21 participants.
Timing Exercise Performance

How would you judge who performed well or performed badly based on this data? Even with this very simple exercise there is clearly a lot of noise, so how do we go about spotting an exceptional data point or a change in the level of performance? Also, when we assess performance, should we take notice only of the average variation or should we factor the range of outcomes into our assessment? In other words, who has performed better – participant 4 who has a low average but a large range or participant 18 whose average was higher but with a lower range?

It is clear that even in a very simple situation like this, there are a number of considerations to make in assessing relative performance and simple arithmetic does not give us a rational basis on which to base judgements.

But there is an even more scary thought. Might these results – even the averages – be meaningless? Might they simply be the outcome of random variation?

This is where Control Charts come in.

The next step in the process we ran in the meeting required participants to create control charts for their own data. While the theory behind Control Charts is sophisticated, the calculations involved are very straightforward, involving two simple steps.

First the average moving range for a data series is calculated using the average change in the times recorded from one period (or trial) to the next. This is then multiplied by 2.66 to give the position of control limits: approximately three standard deviations away from the mean. Any value that is in excess (on the upside or downside) of these limits can be considered to be statistically highly significant because it is extremely unlikely to be the product of chance (less than 0.5%).

This is what the results look like when plotted on a graph, using the data from our earlier individual example (See Control Chart page 11).

As we can see, none of the data points breach either the upper or lower control limits, so we can say with confidence that there is no evidence for anything of any significance in this data series. We can safely assume that all the variation we see is just noise – the results of this individual sometimes doing better, sometimes worse for no reason that we can identify. Had we instead intuitively set the limits of acceptable variation at say 5% (0.5 seconds) we would have treated the values in periods 2 and 3 as being significant, and perhaps everything that happened subsequently as evidence of improvement or as part of an improving trend. But we would have been wrong.
What does this picture look like for the individual participants?

Timing Exercise Performance

In the chart above, the calculated control limits are shown as drop bars, and as you can see there is a wide variation in the position of the control lines. This is because the use of the moving range to measure variation results in the data ‘setting its own control limits’ which are in all cases wider than the actual range in the data, sometimes significantly so. This means that this technique automatically reflects differences or changes in the level of volatility in a data series, which is one of its great strengths. Every participant gets control limits that reflect the nature of their own process rather than having an arbitrary target imposed on them, and the target chances to reflect changes in their performance.

But the true test is: does this technique help us make sense of the data we generated, which was designed to be a small-scale simulation of a real life business scenario?

Had we taken +/- 5% as our guide to determining a significant variance, we would have been driven to investigate 119 of the 210 of these data points. In other words, over 50% of the trials would have required explanation! Even using +/- 10% gives a total of 50 alarms. On the other hand, only 3 out of 210 data point were outside the
statistically calculated control limits and all of these came at the beginning or the end of the exercise and so can probably be attributed to experimental design rather than evidence of any significant event.

In summary, the level of ‘false alarms’ we would avoid by using control charts methodology is somewhere between 25% and 50% of this entire data set. The vast majority of the data can be safely ignored since there is no evidence of any ‘performance problems’ or any deterioration or improvement.

The potential saving in time and the consequences of taking action when we should have done nothing are potentially enormous. Instead, we can focus our attention on the real issues – the signal that would otherwise be obscured by noise.

In this case, the real issues include gaining an understanding of what was behind the three statistically significant ‘alarms’. Is it a process (performance issue) or is it a data problem? For example, in real life this kind of analysis could provide management with an audit capability helping them detect unusual transactions or patterns of behaviour.

Control Charts can be used to track performance over time and alert management to positive or negative changes. For example, it could be that management is concerned about the average level of performance or the high level of variation in the processes, since this makes analysing and forecasting performance more difficult and make the performance of operational processes more variable and costly to run. If this were the case action might be taken to reduce the average values or the level of variation and control charts used to determine whether this had the desired effect.

Control charts work best where data is plentiful and there are a limited number of causal factors at work: if the system being measured is complex the control limits tend to be set too wide. This means that they may not be the best way to analyse aggregated data, like revenue for a business unit, for example. But we have other means of analysing high-level data, such as the trend analysis methods described in the last paper, whereas we have no weapons to help extract insights from large and noisy data sets, and this is where control charts excel.

To illustrate how Control Charts are best used in practice I have attached to this paper an excerpt from the best introductory book on the topic: ‘Understanding Variation’ by Donald J. Wheeler, Ph.D.². The author has very kindly allowed me to reproduce this for you so I would urge you to respect his copyright by not copying it or distributing it without permission. In it he describes two real life examples of the use of control charts for analysing inventory and on time shipments – exactly the kind of granular data that we struggle to interpret using traditional techniques.

In summary, Control Charts provide a simple, powerful and versatile tool that enables managers to filter large amounts of data to identify the relatively small number of data points that warrant attention. They also enable us to measure and track a quality of a data series that is an important dimension of performance – its variability. In our world, the word we use to describe variability is ‘risk’.

Summary

All real world data is infected by noise; random perturbations that obscure the signal upon which we may need to act.

Traditional performance analysis methods are based on pairwise comparisons of target to actual or actual to actual. This is based on the assumption that the difference between two numbers is significant, which in a world full of noise is unlikely to be the case. Without understanding the impact of noise, most of our attempts to analyse periodic data may be little more than stories we spin in attempt to explain the impact of randomness.

² Understanding Variation: The Key to Managing Chaos, 2nd Edition by Donald J. Wheeler, Ph.D. Copyright © 2000 SPC Press, Knoxville, Tennessee, USA. All Rights Reserved
The only way of determining the level of noise is to look at a data series rather than individual data points using statistical rather than arithmetic logic. In this paper, I have described two ways in which it is possible to detect a signal with a defined level of statistical confidence:

1. By examining the sequence of data points
2. By analysing the size of variation around the mean value using control limits that capture the normal level of noise.

Neither of these two techniques uses arbitrary targets to make a judgement about performance. This, along with the fact that targets often distort patterns of performance and drive dysfunctional behaviour, has led some people to argue that targets are damaging and completely unnecessary.

While members of the Beyond Budgeting community sympathise with this point of view, many of our colleagues have difficulty in imagining how it is possible to manage without any kind of targets, so a ‘no targets’ message is a hard sell. In addition, in my view, our position on targets is subtler than a crude binary choice between ‘targets/no targets’.

Even if we abandon budgets and traditional fixed targets, we still need comparators to be able to assess performance, whether they are the performance of internal or external peers or performance trends. While it is clear that most businesses have too many targets, many of which are abused and misused, the question remains. Given that comparators cannot be dispensed with altogether what form should they take? To reflect the dynamic nature of performance should they be moving targets rather than fixed at a point in time, and if so how can we track performance against them. And how do we assess performance against them if our data is infected with noise?

The last paper in this series describes a technique that enables us to track performance in a dynamic and noisy environment, thereby providing a control mechanism that does not rely on fixed, time bound budgets. We will also discuss alternative ways to think about target setting that avoid the problems associated with the traditional budget based approach.

The aim of this series of working papers is to share insights that might be of interest to the Beyond Budgeting community; particularly those that might help make the journey out of the land of budgeting easier. The other purpose is to expose material and arguments that I plan to use in a forthcoming book to friends who I hope will help me do the best job I can of communicating it to a more sceptical audience. I would therefore very much appreciate receiving your comments, criticisms and suggestions and any contributions that you think might enhance the work. You can contact me at: steve.morlidge@satoripartners.co.uk
Appendix 1 – How to set up a control chart

Control Charts were invented by Walter Shewhart in the 1920’s to analyse and improve the quality of manufacturing on a manufacturing line for an early telephone. Due to the rigour of the thinking behind them and the simplicity of their use, they have been in continuous use ever since. Today they are increasingly being used in service industries, like those supported by Vanguard, the Beyond Budgeting partner in the UK.

There are many types of control charts, but the most common is called the XmR chart because it requires only two things to be calculated and plotted – the mean of a data series (X) and the moving range (mR). The steps in the process are as follows:

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Capture a data series A data series is a set of consecutive values. A control chart needs a minimum of 5 typical data points (one where there have been no deliberate attempt to change it) but the longer the series the better.</td>
</tr>
<tr>
<td>2.</td>
<td>Calculate the mean of the data series Average of results over time.</td>
</tr>
<tr>
<td>3.</td>
<td>Calculate the average moving range (AMR) The moving range is the difference between two consecutive data points. So if a value of 100 were recorded in period 1 and 120 in period 2, the moving range would be 20.</td>
</tr>
<tr>
<td>4.</td>
<td>Calculate the control limits for the X chart The control limits are set at 2.66 x AMR above and below mean for the data series. So if the mean were 80 and the AMR 15, the upper control limit (UCL) would be set at 119.9 (80 + (15 x 2.66 = 39.9)) and the lower control limit (LCL) at 40.1 (80 - 39.9). The UCL and LCL are taken to measure of ‘common cause’ variation that is associated with the normal operation of the process.</td>
</tr>
<tr>
<td>5.</td>
<td>Calculate the control limit for the mR chart The control limit for the mR chart is 3.27 x AMR. For example 15 x 3.27 = 49</td>
</tr>
<tr>
<td>6.</td>
<td>Plot the control limits and individual values on the X and mR charts Actual data is plotted on the X chart along with the mean and the UCL and LCL. Moving range data is plotted on the mR chart along with the control limit. The two charts are usually presented together with the X chart above the mR. In the main text above, only the X chart was shown for the sake of simplicity.</td>
</tr>
<tr>
<td>7.</td>
<td>Determine whether any alarms are triggered A simple set of rules are used to determine whether there is any evidence for ‘special cause variation’ - that which is indicative of an abnormal event or a change in the system.</td>
</tr>
</tbody>
</table>
The detection rules all indicate the presence of special cause variation at a significance level of 3 standard deviations, which means it is highly unlikely (<0.05%) to be the result of chance. The rules are:

<table>
<thead>
<tr>
<th>Points outside the limits</th>
<th>A single point outside the limits of the X or mR chart. This should be interpreted as an indication of a special cause with a dominant effect.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Runs about the centre line</td>
<td>8 or more successive data points on either side of the mean on the X chart. This is evidence of a change in the mean value of the process.</td>
</tr>
<tr>
<td>Runs near the limits</td>
<td>3 out of 4 successive data points outside the inner 75% of the upper or lower range on the X chart. This is a manifestation of a moderate but sustained effect.</td>
</tr>
</tbody>
</table>

In addition to these detection rules, we can use the control limits to assess risk, since we would expect 80% of all future values to fall within the inner 50% of the control limits.

Unsurprisingly, given the technique has been in continuous use for near 100 years, a large body of knowledge has grown up around control charts, but for our purposes, we can ignore most of this and concentrate on one very simple question: is there any evidence of special cause variation? If there is, we should investigate it to determine whether it is the result of bad data or a change in the process. If there isn’t, we should ignore the data since there is no evidence of any significant change.

There are a few technical issues we need to be aware of when using control charts, which are otherwise remarkably simple and robust. First, they work best when they are used with ‘low level’ data since when data is aggregated the range of different forces at play tend to distort the position of the control limits. However, since our aim is to find ways of filtering large volumes of disaggregated data, in practice this ‘weakness’ needn’t concern us. Second, control charts are also less sensitive when they are dealing with data that is cyclical or where there are strong trends. For these reasons and the practical challenge of dealing with many data series, serious implementers of control charts should consider buying specialist software. For an example, go to www.lightfootsolutions.com. Also, go to www.vanguard-method.net to learn more about how these methods are used in business and their links to Beyond Budgeting.

To learn more about control charts and the analysis of time series data explained in simple terms, I strongly recommend “Understanding Variation: The Key to Managing Chaos” by Donald J. Wheeler, Ph.D.
Appendix 2 – Extract from “Understanding Variation: The Key to Managing Chaos” by Donald J. Wheeler, Ph.D..

THREE

THE PURPOSE OF ANALYSIS IS INSIGHT

Data are generally collected as a basis for action. However, unless potential signals are separated from probable noise, the actions taken may be totally inconsistent with the data. Thus, the proper use of data requires that you have simple and effective methods of analysis which will properly separate potential signals from probable noise.

The following examples are intended to illustrate how you can use process behavior charts to gain insight into data. For purposes of comparison, the traditional approaches to management data will be shown alongside the process behavior chart. Each of the situations described below actually happened. Since these stories involve real people in real companies, it was necessary to disguise the data in order to preserve confidentiality. In every other respect these examples are true to the original.

Let us return to the monthly report for July which was first shown in Figure 1.1, and is reproduced in Figure 3.1. When faced with a table of
Understanding Variation / Managing Chaos

numbers such as that in Figure 3.1 most people begin to scan the percent difference columns to see which numbers have changed the most. The idea is to single out those values which have changed most dramatically and question why they have changed.

<table>
<thead>
<tr>
<th>Monthly Report for July</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>On-Time Shipments (%)</td>
<td>20</td>
<td>91.0</td>
<td>91.3</td>
<td>0.3</td>
<td>0.9</td>
<td>0.6</td>
</tr>
<tr>
<td>First Time Approval (%)</td>
<td>12</td>
<td>54</td>
<td>70</td>
<td>23.0</td>
<td>10.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Pounds Scrapped (per 1000 lbs production)</td>
<td>3</td>
<td>75</td>
<td>100</td>
<td>12.5</td>
<td>10.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Production:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Production Volume (1000's lbs)</td>
<td>3</td>
<td>34.5</td>
<td>36</td>
<td>4.2</td>
<td>2.0</td>
<td>0.2</td>
</tr>
<tr>
<td>Material Costs ($/100 lbs)</td>
<td>13</td>
<td>196.29</td>
<td>201.22</td>
<td>1.5</td>
<td>1.9</td>
<td>1.4</td>
</tr>
<tr>
<td>Manhours per 100 lbs</td>
<td>13</td>
<td>4.45</td>
<td>4.16</td>
<td>7.0</td>
<td>4.5</td>
<td>7.2</td>
</tr>
<tr>
<td>Energy &amp; Fixed Costs / 100 lbs</td>
<td>13</td>
<td>11.34</td>
<td>11.27</td>
<td>0.6</td>
<td>11.3</td>
<td>2.2</td>
</tr>
<tr>
<td>Total Production Costs/100 lbs</td>
<td>13</td>
<td>280.83</td>
<td>278.82</td>
<td>0.7</td>
<td>0.9</td>
<td>0.7</td>
</tr>
<tr>
<td>In-Process Inventory (100's lbs)</td>
<td>17</td>
<td>28</td>
<td>19.7</td>
<td>42.0</td>
<td>12.0</td>
<td>9.6</td>
</tr>
<tr>
<td>Operations:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>On-Time Closings of Accounts (%)</td>
<td>3</td>
<td>74.3</td>
<td>95</td>
<td>21.8</td>
<td>23.5</td>
<td>7.6</td>
</tr>
</tbody>
</table>

Figure 3.1: A Typical Management Report

Of course there are three problems with the use of percent differences as a basis for interpreting the values in a monthly report.

First, the size of the percent difference will partially depend upon the magnitude of the base number—a ten unit change from 100 to 110 is a 10% change, yet a ten unit change from 300 to 310 is a 3.3% change. Percentages show the relative size of a change rather than the actual amount of change. Therefore, comparing one percentage change with another is not a reliable way to find the interesting parts of the data because it does not take into account the difference in the base numbers.
Second, the practice of comparing lines in a monthly report by comparing the size of the percent differences assumes that all lines should show the same amount of relative variation month to month. Yet, in any collection of time series, each time series will have its own inherent amount of month-to-month variation. Some lines will show large percent differences month to month, and others will show small percent differences from month to month. Therefore, comparing percent differences will guarantee that some lines receive more attention than they deserve while others receive less attention than they deserve.

Third, when considering the percent differences based upon a comparison of the current value with last year’s value, a large percent difference may be due to an unusual value in the past rather than an unusual value in the present.

All of these problems make the comparison of percent differences a weak and unreliable guide to finding the potential signals within the data. Nevertheless, it is a common practice, so we shall use it to select certain lines out of the monthly report for discussion.

As we scan the percent difference column under the July portion of the monthly report three lines stand out. First-time approvals are down 23 percent, in-process inventory is up 42 percent, and on-time closings of accounts is down 21.8 percent.

PROCESS BEHAVIOR CHARTS
FOR INDIVIDUAL VALUES AND MOVING RANGES

Since the monthly meeting would probably focus on the 42 percent increase in the in-process inventory value for Department 17, we will also begin with this line from the monthly report. First we will consider the usual approach to interpreting management data, and then we will use these values to illustrate the construction of a process behavior chart.
Understanding Variation / Managing Chaos

<table>
<thead>
<tr>
<th>Production:</th>
<th>July Monthly</th>
<th>% Diff from July</th>
<th>Year-to-Date Values</th>
<th>This YTD as % Diff of Last YTD</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-Process Inventory (100's lbs)</td>
<td>Dept Value</td>
<td>Monthly Plan Value</td>
<td>% from July</td>
<td>Average Plan</td>
</tr>
<tr>
<td>In-Process Inventory</td>
<td>17</td>
<td>28</td>
<td>19.7</td>
<td>21.6</td>
</tr>
</tbody>
</table>

**Figure 3.2: In-Process Inventory for Dept. 17**

In July the in-process inventory value was 28. This was the largest value for in-process inventory ever recorded in Department 17. This value was 12% above the value for last July, which is bad, and it was 42% above the plan value, which is very bad. For this year as a whole, the year-to-date in-process inventory is running 5.9% above last year and 9.6% above plan—two more bad values.

If you had responsibility for Department 17 and were shown these data, what would you do? When you have four bad numbers in a row you usually get to prepare a report.

The main problem with the “write a report” approach is the creativity required in its preparation. These reports often become works of fiction whose only purpose is to allow a manager to pretend that something is being done about a perceived problem.

In this case the “write a report” approach has a problem with the data themselves. In-process inventory is a result, not a cause, and it cannot be directly managed. It can only be changed by actions which affect the causes. However, many of these causes will also affect other outputs and parts of the process. So, if you are not careful, pressure to reduce the in-process inventory may have a detrimental effect on some other characteristic of the process such as production volumes. (Remember the three ways of meeting a goal?)

Of course the whole “write a report” approach is based upon the assumption that the in-process inventory value for July is a signal. But is it a signal—or is it just noise? How can you know?
Three / The Purpose of Analysis is Insight

<table>
<thead>
<tr>
<th>In-process Inventory for Department 17 (Hundreds of Pounds)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Year One</td>
</tr>
<tr>
<td>Year Two</td>
</tr>
<tr>
<td>Year Three</td>
</tr>
</tbody>
</table>

**Figure 3.3: In-Process Inventory Values**

Before you can detect a potential signal within the data you must first filter out the probable noise. And to filter out noise you must start with the past data. The table in Figure 3.3 shows the in-process inventory values for Department 17 for the past 31 months.

The average of the 24 values for Years One and Two is 20.04. This value was used as the central line in the time-series graph of the in-process inventory values seen in Figure 3.4.

![Time Series for Monthly In-Process Inventory](image)

**Figure 3.4: Time Series for Monthly In-Process Inventory**

A glance at Figure 3.4 shows no long-term trends, nor any other systematic patterns\(^6\). So while the time series graph adds to our understanding of the data as a whole, it still does not answer the question of whether or not the July value is exceptional. To answer this question we will have to filter out the routine month-to-month variation, which means that we shall have to measure the month-to-month variation.

---

\(^6\) The gaps in the running record are visual breaks used to delineate the data for each year. They have no other meaning in this, or in subsequent, graphs.
Understanding Variation / Managing Chaos

FINDING THE MOVING RANGES

To measure the month-to-month variation we compute the differences between the successive monthly values. These values are called moving ranges. They are computed in the following manner.

<table>
<thead>
<tr>
<th>Year One</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>mR Values</td>
<td>8</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3.5: Moving Ranges for In-Process Inventory for Year One**

The difference between the January value of 19 and the February value of 27 is 8, thus the first moving range is 8.

The next moving range is 7. It is the difference between the February value of 27 and the March value of 20.

The third moving range is 4. It is the difference between the March value of 20 and the April value of 16.

Continuing in this manner, using all 31 values from Figure 3.3, we obtain the 30 moving ranges shown in Figure 3.6.

<table>
<thead>
<tr>
<th>Year One</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>mR Values</td>
<td>8</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3.6: Moving Ranges for In-Process Inventory**

The time series graph of these moving ranges is shown in Figure 3.7. These moving ranges directly measure the month-to-month variation. Their average will be called the Average Moving Range. The central line for the time series graph of the moving ranges is commonly taken to be the Average Moving Range.
Three / The Purpose of Analysis is Insight

Figure 3.7: Graph of the Moving Ranges

THE TWO GRAPHS IN AN XmR CHART

To construct a process behavior chart for Individual Values and a Moving Range (an XmR chart) we begin with the two time series graphs shown in Figures 3.4 and 3.7. First these two graphs are shown together, as in Figure 3.8.

Figure 3.8: Combined Time Series Graphs
Understanding Variation / Managing Chaos

The time series for individual values is sometimes referred to as "the X-chart." The time series for the moving ranges is sometimes referred to as "the range chart," or "the moving range chart." Once the time series graphs for the individual values and the moving ranges have been constructed, the central lines for each graph are computed. The average of the individual values is the usual central line for the X-chart. The Average Moving Range is the usual central line for the moving range chart. The average for Years One and Two is 20.04, and the Average Moving Range is 4.35. These lines are shown in Figure 3.8.

**COMPUTING LIMITS FOR AN XMR CHART**

To obtain the Upper Range Limit for the moving range chart you must multiply the Average Moving Range by a scaling factor of 3.27. This value of 3.27 is the number required to convert the average range into an appropriate upper bound for ranges. The value of 3.27 is a constant for this type of process behavior chart.

\[
Upper \text{ Range Limit} = URL = 3.27 \times \overline{mR} = 3.27 \times 4.35 = 14.2
\]

This Upper Range Limit is plotted as a horizontal line on the moving range portion of the combined graph. This line is shown in Figure 3.9.

The limits for the Chart for Individual Values (the X-chart) are commonly called the Natural Process Limits. They are centered on the central line. The distance from the central line to either of these limits is computed by multiplying the Average Moving Range by a second scaling factor: 2.66. The value of 2.66 is constant for this type of process behavior chart—it is the value required to convert the Average Moving Range into the appropriate amount of spread for the running record of individual values.
The **Upper Natural Process Limit** is found by first multiplying the Average Moving Range by 2.66 and then adding the product to the central line of the $X$-chart.

$$Upper \ Natural \ Process \ Limit = \bar{X} + (2.66 \times \overline{mR})$$

$$= 20.04 + (2.66 \times 4.35) = 31.6$$

The **Lower Natural Process Limit** is found by first multiplying the Average Moving Range by 2.66 and then subtracting the product from the central line of the $X$-chart.

$$Lower \ Natural \ Process \ Limit = \bar{X} - (2.66 \times \overline{mR})$$

$$= 20.04 - (2.66 \times 4.35) = 8.5$$

These Natural Process Limits are plotted on the individual values portion of the combined graph. The complete $XmR$ chart is shown in Figure 3.10.
INTERPRETING THE XmR CHART

The interpretation of the process behavior chart in Figure 3.10 is as follows. The month-to-month variation is seen on the moving range portion of the chart. The Upper Range Limit of 14.2 means that if the amount of in-process inventory changes (up or down) by more than 1420 pounds from one month to the next, then you should look for an explanation. A change of this amount from one month to the next is excessive, and it is likely to be the direct result of some assignable cause.

The actual monthly values are seen on the individual values portion of the chart. The limits on this part of the chart define how large or how small a single monthly value must be before it represents a definite departure from the historic average. Here, a monthly value in excess of 31.6
would be a signal that the amount of in-process inventory had shifted upward. Likewise, a monthly value below 8.5 would be taken as a signal of a downward shift. In either case, you would be justified in looking for an assignable cause of such a shift.

Thus, the July value of 28 is not, by itself, a signal. There is no evidence of any real change in the in-process inventory. This means that asking for an explanation for the July value would be an exercise in futility. The man-in-the-moon would be as valid an explanation as any found in a written report.

Now some may feel disconcerted when they see limits for individual values which go from 8.5 to 31.6. Surely we can hold the in-process inventory more steady than that! But that is precisely what cannot be done. At least it cannot be done unless fundamental changes are made in the underlying process. The Natural Process Limits are the Voice of the Process. They define what the process will deliver as long as it continues to operate as consistently as possible.

When a process is operated predictably it is also operating as consistently as possible. The process doesn’t really care whether or not you like the Natural Process Limits, and it certainly does not know what the specifications may be (specifications should be thought of as the Voice of the Customer, which is distinctly different from the Voice of the Process).

Therefore, if the manager of Department 17 issued a decree that the in-process inventory should not vary by more than ± 20 percent from its average value, what would happen? A ± 20 percent variation from a value of 20.0 is 16.0 to 24.0. Is this process going to operate within these limits? The process behavior chart says that it has not done so in the past, and it should not be expected to do so in the future—at least not without some fundamental change in the underlying process. Thus, such a decree will simply encourage the workers in Department 17 to distort the system or to distort the data. Such decrees, by themselves, do nothing to change or improve the system.

Likewise, dissatisfaction with the Natural Process Limits cannot be cured by finding some alternative method for computing the Natural
Understanding Variation / Managing Chaos

Process Limits. Any method that results in appreciably different limits is simply incorrect. The Voice of the Process will still be defined by the limits computed in the manner given above.

Therefore, if you are not pleased with the amount of variation shown by the Natural Process Limits, then you must go to work on the system, to change the underlying process, rather than setting arbitrary goals, jawboning the workers, or looking for alternative ways of computing the limits.

7 A valid alternative for computing limits for the XmR chart is to use the Median Moving Range instead of the Average Moving Range. When this is done the scaling factors have to be changed. The value of 2.66 is changed to 3.14, and the value of 3.27 is changed to 3.87.

Using all 31 individual values for the in-process inventory you would get an Average of 20.39, while the 30 moving ranges would have a Median Moving Range of 3.5. With the formulas given on page 136 you would get the limits shown in the following XmR chart.

![XmR Chart for In-Process Inventory (Using the Median mR)](chart)

Figure 3.10 (a): XmR Chart for In-Process Inventory (Using the Median mR)

While the limits above are slightly different from those in Figure 3.10 they are not appreciably different, and the story they tell remains the same.
A CHART FOR ON-TIME SHIPMENTS

Now return to the monthly report for July (Figure 3.1) and look for the line with the smallest percent differences. The line for “on-time shipments” is the winner in this category. In most discussions of the monthly report this line would get little more than a cursory glance.

<table>
<thead>
<tr>
<th>Quality: On-Time Shipments (%)</th>
<th>July Actual Value</th>
<th>Monthly Average Value</th>
<th>% Diff from July</th>
<th>Year-to-Date Values Actual Value</th>
<th>Year-to-Date Values Plan or Average</th>
<th>% Diff from Last YR</th>
<th>This YTD as % Diff of Last YTD</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>91.0</td>
<td>91.3</td>
<td>-0.3</td>
<td>90.8</td>
<td>91.3</td>
<td>-0.6</td>
<td>-0.3</td>
</tr>
</tbody>
</table>

Figure 3.11: On-Time Shipments for July

In July, 91.0% of the shipments were shipped on time. This is 0.3% below the historic average, and 0.9% below the value for last July. The year-to-date value is 90.8% shipped on time, which is 0.6% below the average and 0.3% below last year. Thus, by these traditional comparisons, the on-time shipments performance is slightly below, but essentially unchanged from last year. Nothing to get excited about here—or is there? How can you know?

We shall put these data on an XmR chart. The raw data for the past 31 months are shown in Figure 3.12. The percentage of on-time shipments for each month is computed by dividing the total number of shipments into the total number of shipments which were made on or before the customer request date.

8 For those who expected a p-chart with these data, see note on page 140.
Understanding Variation / Managing Chaos

Percentage On-Time Shipments, Department 20

<table>
<thead>
<tr>
<th>Month Year</th>
<th>Total No. Shipments</th>
<th>No. Shipped On-Time</th>
<th>Percentage Shipped On-Time</th>
<th>Moving Ranges</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 01</td>
<td>191</td>
<td>176</td>
<td>92.1</td>
<td></td>
</tr>
<tr>
<td>February 01</td>
<td>203</td>
<td>186</td>
<td>91.6</td>
<td>0.5</td>
</tr>
<tr>
<td>March 01</td>
<td>220</td>
<td>202</td>
<td>91.8</td>
<td>0.2</td>
</tr>
<tr>
<td>April 01</td>
<td>200</td>
<td>183</td>
<td>91.5</td>
<td>0.3</td>
</tr>
<tr>
<td>May 01</td>
<td>236</td>
<td>215</td>
<td>91.1</td>
<td>0.4</td>
</tr>
<tr>
<td>June 01</td>
<td>213</td>
<td>194</td>
<td>91.1</td>
<td>0.0</td>
</tr>
<tr>
<td>July 01</td>
<td>212</td>
<td>191</td>
<td>90.1</td>
<td>1.0</td>
</tr>
<tr>
<td>August 01</td>
<td>241</td>
<td>215</td>
<td>89.2</td>
<td>0.9</td>
</tr>
<tr>
<td>September 01</td>
<td>159</td>
<td>143</td>
<td>89.9</td>
<td>0.7</td>
</tr>
<tr>
<td>October 01</td>
<td>217</td>
<td>197</td>
<td>90.8</td>
<td>0.9</td>
</tr>
<tr>
<td>November 01</td>
<td>181</td>
<td>165</td>
<td>91.2</td>
<td>0.4</td>
</tr>
<tr>
<td>December 01</td>
<td>113</td>
<td>103</td>
<td>91.2</td>
<td>0.0</td>
</tr>
<tr>
<td>January 02</td>
<td>170</td>
<td>155</td>
<td>91.2</td>
<td>0.0</td>
</tr>
<tr>
<td>February 02</td>
<td>270</td>
<td>246</td>
<td>91.1</td>
<td>0.1</td>
</tr>
<tr>
<td>March 02</td>
<td>167</td>
<td>151</td>
<td>90.4</td>
<td>0.7</td>
</tr>
<tr>
<td>April 02</td>
<td>216</td>
<td>196</td>
<td>90.7</td>
<td>0.3</td>
</tr>
<tr>
<td>May 02</td>
<td>227</td>
<td>206</td>
<td>90.7</td>
<td>0.0</td>
</tr>
<tr>
<td>June 02</td>
<td>149</td>
<td>136</td>
<td>91.3</td>
<td>0.6</td>
</tr>
<tr>
<td>July 02</td>
<td>182</td>
<td>167</td>
<td>91.8</td>
<td>0.5</td>
</tr>
<tr>
<td>August 02</td>
<td>224</td>
<td>206</td>
<td>92.0</td>
<td>0.2</td>
</tr>
<tr>
<td>September 02</td>
<td>246</td>
<td>225</td>
<td>91.5</td>
<td>0.5</td>
</tr>
<tr>
<td>October 02</td>
<td>185</td>
<td>170</td>
<td>91.9</td>
<td>0.4</td>
</tr>
<tr>
<td>November 02</td>
<td>261</td>
<td>239</td>
<td>91.6</td>
<td>0.3</td>
</tr>
<tr>
<td>December 02</td>
<td>140</td>
<td>128</td>
<td>91.4</td>
<td>0.2</td>
</tr>
<tr>
<td>January 03</td>
<td>216</td>
<td>198</td>
<td>91.7</td>
<td>0.3</td>
</tr>
<tr>
<td>February 03</td>
<td>247</td>
<td>225</td>
<td>91.1</td>
<td>0.6</td>
</tr>
<tr>
<td>March 03</td>
<td>230</td>
<td>209</td>
<td>90.9</td>
<td>0.2</td>
</tr>
<tr>
<td>April 03</td>
<td>265</td>
<td>239</td>
<td>90.2</td>
<td>0.7</td>
</tr>
<tr>
<td>May 03</td>
<td>184</td>
<td>165</td>
<td>89.7</td>
<td>0.5</td>
</tr>
<tr>
<td>June 03</td>
<td>207</td>
<td>188</td>
<td>90.8</td>
<td>1.1</td>
</tr>
<tr>
<td>July 03</td>
<td>178</td>
<td>162</td>
<td>91.0</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Figure 3.12: The On-Time Shipments Data

The limits will be based, somewhat arbitrarily, upon the data for Year Two. Columns two and three of Figure 3.12 show that 2225 of the 2437 shipments in Year Two were shipped on time, for an annual percentage of 91.30%. This value will be used as the central line for the individual values. Using the 12 moving ranges for Year Two, the Average Moving Range is computed to be 0.317. This value will be used as the central line for the moving ranges.
Three / The Purpose of Analysis is Insight

Using these values, the Natural Process Limits are computed as:

\[ \text{UNPL} = 91.30 + (2.66 \times 0.317) = 92.14\% \]
\[ \text{LNPL} = 91.30 - (2.66 \times 0.317) = 90.46\% \]

And the Upper Range Limit for the moving range chart will be:

\[ \text{URL} = 3.27 \times 0.317 = 1.037\% \]

The \( XmR \) chart for the on-time shipments is shown in Figure 3.13. It would be interpreted as follows. Allowing for the month-to-month variation, if the shipping process was operating as consistently as possible, with an average of 91.30 percent on-time shipments, then the monthly values should fall between 90.46 percent and 92.14 percent. Moreover, the values for successive months should differ by no more than 1.037 percentage points.

![XmR Chart for On-Time Shipments](image)

**Figure 3.13: XmR Chart for On-Time Shipments**

Six of the individual values and one of the moving ranges fall outside the limits shown in Figure 3.13. Thus, once the routine month-to-month variation has been taken into account, there is still too much variation in this time series to be due to chance alone. The six values below the Natural Process Limit should be treated as signals. You should have
Understanding Variation / Managing Chaos

looked for an explanation of why the percentage of on-time shipments dropped during these months.

The six values outside the Natural Process Limits are exceptional. The process is trying to tell you that it has a problem. Something was happening in the Summer of Year One, in March of Year Two, and in April and May of Year Three. This something could reoccur, and it could be worse next time.

- You have already missed three opportunities to improve this process.
- The process has already done all that it can do to alert you to the presence of this problem.
- How many more signals are you going to miss before this problem causes you to lose a large account?
- If you concentrate on the percent differences in the monthly report (Figure 3.11) you are not likely to ever be aware of this problem until it is already too late.

Process behavior charts are the way to listen to your process. When you listen to the Voice of the Process as revealed by process behavior charts, you can often detect signals that you would otherwise miss.
SUMMARY

- Large percent differences do not necessarily indicate a signal.
- Small percent differences do not necessarily indicate a lack of a signal.
- Points outside the limits are signals—they are opportunities to discover how to improve a process. (A large moving range signifies a break in the original time series.) Shewhart argued, and experience has shown, that it is economically worthwhile to investigate all such signals of exceptional variation.
- The process behavior chart focuses data so that the user will ask the interesting and important questions.
- A single value beyond the limits of a process behavior chart is a signal.
- Another pattern which is taken to be a signal consists of at least three out of four consecutive values which are closer to one of the limits than they are to the central line.
- The process behavior chart filters out the probable noise in order to detect the potential signals in any data set.
- By filtering out the noise, the process behavior chart minimizes the number of times that you will interpret a bit of noise as if it were a signal.
- By causing the potential signals to stand out, the process behavior chart will also minimize the number of times that you miss a signal.
- Process behavior charts are the beginning of knowledge because they help you to ask the right questions.