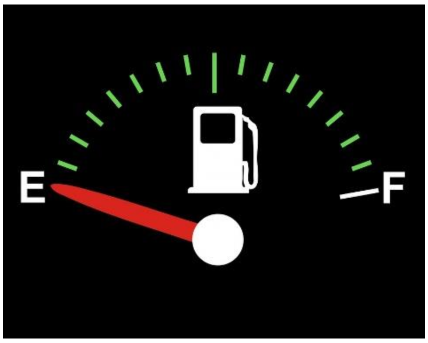
Control Charts: A 'how to' guide

(Author: Steve Schefer, Version: 2.0, Date: July 2022)



This guide is to help people understand the purpose and importance of a **control chart**[1], and gain a foundational level of understanding of how to create and use one.

Note: It doesn't cover the rather large (and important) topic of 'what' to measure. It is concerned with 'how to' use a measure.

An upfront definition: What is a control chart?

A control chart is a graph to study **how a given measure changes over time**. Data are plotted sequentially (i.e. in time order) along the X axes to show how they vary.

The chart has a set of horizontal lines added to it (a centre line and control limits), which are calculated from **a baseline set of historical data**.

We then **compare current data to these lines** and can **draw conclusions** about whether the measure is stable or not, and what this might mean.

Right, so why would we want that?! Here's an explanation…

A very high-level summary:

- We come across, rely upon, and are affected by measurement in all aspects of our lives. In particular: as used 'by us' at work; and as used 'on us' as citizens of the world (by the media, public services, organisations etc.)

- Much of conventional measurement (often involving tables of numbers) leads to incorrect insights being drawn and poor decision making

- We need to see our data visually, over time in appropriate charts if we are to see, and understand what is going on

- There is variation in everything, and our charts should uncover this

[1] **The name 'Control chart'**: Donald Wheeler refers to control charts as 'process behaviour charts' as he believes that this naming makes their purpose clear. John Seddon refers to control charts as 'capability charts' as I don't think he likes what the word 'control' suggests. I use the traditional term of 'control charts' throughout this guide but I see both Wheeler's and Seddon's points.

- If we want to draw the right conclusions as to what is going on, the variation must be separated out into its two component parts:
    a. routine variation (noise); and
    b. exceptional variation (signals that something has changed)

- If we do not do this, then we risk:
    a. mistaking noise as signals and taking inappropriate actions (tampering); or, conversely,
    b. missing important signals

- There is a method for separating out signals from the noise, using control limits and detection tests; and finally

- There is a right way to perform this method (and many a wrong way).

The following will unpack the above and add some 'meat onto the bones':

'Seeing the wood for the trees'

A great deal of conventional measurement that we use consists of a bunch numbers, often presented in tabular form, and sometimes discussed within a paragraph of text.

However, whatever form these numbers take, we humans find them very hard to digest.

> *"The fact is that the human brain is singularly incompetent in the matter of handling figures."* (Stafford Beer)

To be able to properly consider what is happening, we need to **visualise our data**, usually over time**.** We need to see it 'come to life' graphically, to 'see the picture'.

A bunch of numbers is not enough.

A red herring – the addition of 'useful analysis'

Often, rather than simply visualising the data over time, yet more numbers are added to tables to provide some (supposedly) useful analysis.

Common examples of additional analysis are:

- Calculations of variances (e.g. this week vs. last week; this week vs. average)
- Turning these variances into 'percentage change'
- Adding some sort of up/down arrows or 'traffic lighting' symbols to (supposedly) assist anyone taking a cursory glance; and then
- The addition of some comment that 'explains what we are seeing'

All of these techniques are about adding a narrative to the existing numbers, rather than considering whether we have enough of the right information, presented in the right way to derive valid insights.

We should be mindful of **the Narrative Fallacy**:

> *'A human vulnerability for over-interpreting the information we have available to us and **deriving a compact story that neatly fits**. The less information we have, then the easier it is to come up with such a story.'*

If we want to improve the performance of a system, and processes within, then **we should be about understanding and prediction, rather than about narration.**

Turning to an example: Throughout this guide I use a familiar process to most of us – the refuelling of your car at a fuel station. This example has been deliberately selected as being simple and non-work related so that you can focus on the points being made.

Once you 'get it', you can then apply the 'fuel station' learnings to any system of interest and its associated measures.

Here's a typical looking table showing how long it took me to refuel my car today, as compared to the last time I refuelled and compared to the average time it takes me to refuel. Additional analysis has been added as per the 'red herring' list above:

Table 1: Time to refuel 'management dashboard report'

Measure	This fill	Last fill	Variance		Average fill	Variance	Comment
Time to refuel (Minutes)	7.81	8.01	0.2 (-3%)		7.45	0.36 (5%)	Blah blah blah

You can see that:

- I was faster today by 0.2 minutes (i.e. 12 seconds) from last time, which is a 3% 'improvement' and gets me a green traffic light; but
- I was slower by 0.36 minutes (i.e. 22 seconds) as compared to my average, which is a 5% 'deficit' and gets me a red traffic light; and
- ….I could make up a story to (supposedly) explain.

Not just any old chart

So perhaps you agree that visualising data is a good thing. But how? There are lots of charting types at our disposal, and lots of ways of presenting each.

The statistician Donald Wheeler[2] cautions against what he refers to as *"graphical purgatory".*

> *"Along with the proliferation of graphing options has come the abundance of chart junk, decoration and inappropriate graphics.*
>
> *Good graphics should amplify and clarify a table of numbers, not compete with it."*

Some points of guidance:

- Don't attempt to present too much information in a single graph
- Resist superimposing curves of 'fit'. Don't confuse a mathematical model with reality
- Think hard before using a pie chart: people find it more difficult to distinguish differences in angles than heights of bars
- Don't use radar plots: they misleadingly connect things that are not comparable.

In short, **keep it simple and make it clear**.

Variation everywhere

'Variation is the spice of life'. There is variation (of differing degrees) within everything. Further, if we want to understand performance (i.e. what's actually happening), **we need to see this variation**.

To explain variation, we'll return to the 'refuelling your car at a fuel station' process.

Let's say that the last time you performed this process you timed yourself[2] and it took you 7 minutes and 27 seconds.

- How long do you think that it will take you when you fill up next?

- Would you think that it was a bit weird if it took you exactly 7 minutes and 27 seconds again? I would!

You recognise that there are lots of different reasons that might speed you up or slow you down.

If we have a quick think about what some of these factors might be, the list may include:

- What other cars were already in the fuel station, and at which pumps? (did you have to wait a bit?)
- Which pump did you drive up to today? (from the closest to the furthest)
- How easy was it for you to get out of the car? (did you park too close again so that you had to squeeze out?!)
- How easy did you find it to open the fuel cap today? (did you remember to pull the internal lever that opens the fuel flap?)
- How twisted was the fuel line coming from the pump? (It can get very tangled - did you have to unwind it?)
- How empty was your tank, and how much fuel did you put in?
- How smooth were you in putting the fuel cap back? (did you misplace it?)
- Was there a queue when you went to pay? (and how many pay desks were open?)
- Who was the cashier? (and what mood were they in? Chatty or not?)
- Did you pay by cash or card?

[2] **Timing yourself refuel:** We'll assume that you have devised a consistent way of measuring e.g. using a stop watch and timing from the point that your front wheels enter the fuel station forecourt to the point at which your back wheels leave.

- If card, which one did you use? (Did you enter the wrong pin number for that card and have to start again?)
- Did you lock your car? Where did you put your keys?
- Were other cars entering/ leaving the fuel station when you wanted to drive off?
- Could you pull out into the traffic?
- …

You could carry on (and on and on) with such a list, and probably never complete it because there are so many possible reasons that would cause variation.

You can see that there would be a great deal of variation even within each factor above. Further, much (most) of this variation is outside of your control.

Given the above, you’d expect that the refuelling of your car might normally take between a range of times. You can also clearly see why it would be weird if it always took exactly 7 minutes and 27 seconds.

Charts that show variation

Whilst there are lots of different charting options available, two types of charts allow us to see the variation within a set of data:

- A tally chart (**histogram**); and
- A time-series chart (which we will later use as a **control chart**)

Both types of chart show us the variation present within a set of data, but a time-series chart has the additional quality of **showing variation over time**. This time-ordered sequence will contain information about the process being charted.

To illustrate, let’s continue with the ‘refuelling a car’ example and assume that I’ve timed the last 30 visits that I’ve made to my local fuel station. Here’s the data:

<u>Table 2:</u> Time to refuel – 30 data points

Refuelling instance (X)	1	2	3	4	5	6	7	8	9	10
Time taken (minutes)	8.01	7.81	7.25	7.62	8.70	8.20	7.30	8.17	7.40	8.43

Refuelling instance (X)	11	12	13	14	15	16	17	18	19	20
Time taken (minutes)	7.96	8.08	7.67	8.53	8.53	12.51	7.46	8.01	7.62	7.09

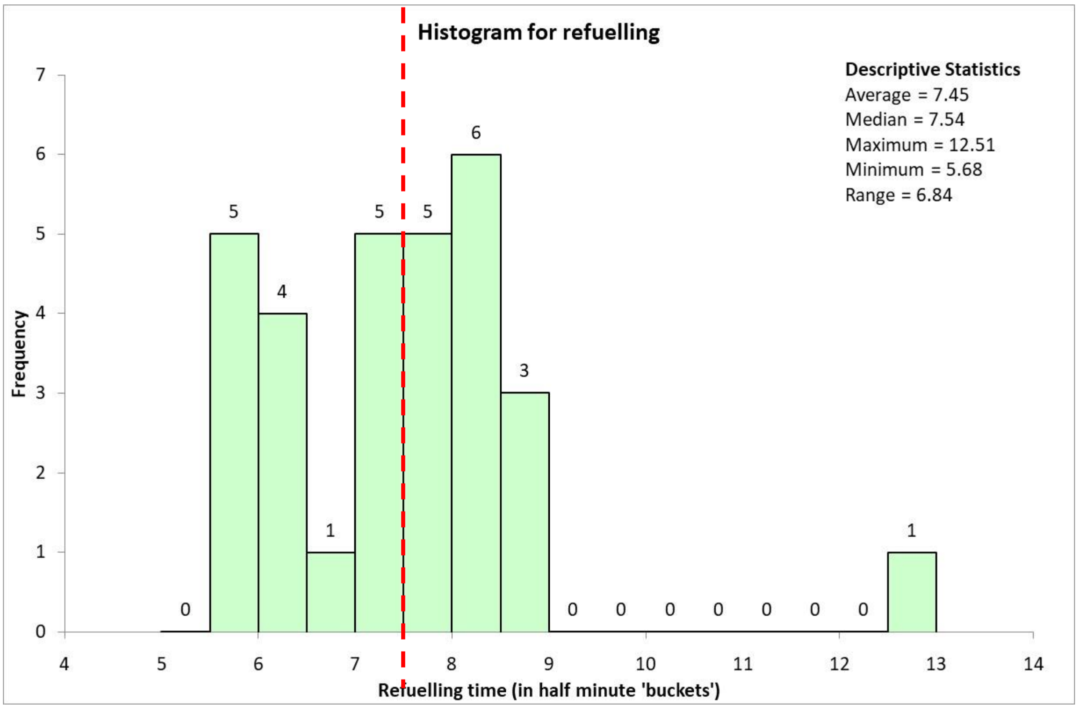
Refuelling instance (X)	21	22	23	24	25	26	27	28	29	30
Time taken (minutes)	6.69	5.80	6.18	6.45	5.91	5.99	6.47	6.15	5.68	5.96

The first point to reiterate is that we find it incredibly hard to digest and perform accurate yet quick analysis using a table of numbers.

If you got out your calculator (or spread sheet) you would find that the average of these 30 refuelling instances is 7.45 minutes (or 7 minutes, 27 seconds). But how useful is that average? What does it hide?

Here are the 30 data points presented as a histogram[3], with the red dotted line denoting the average:

<u>Figure 1:</u> Histogram of time to refuel – 30 data points



We can see that:

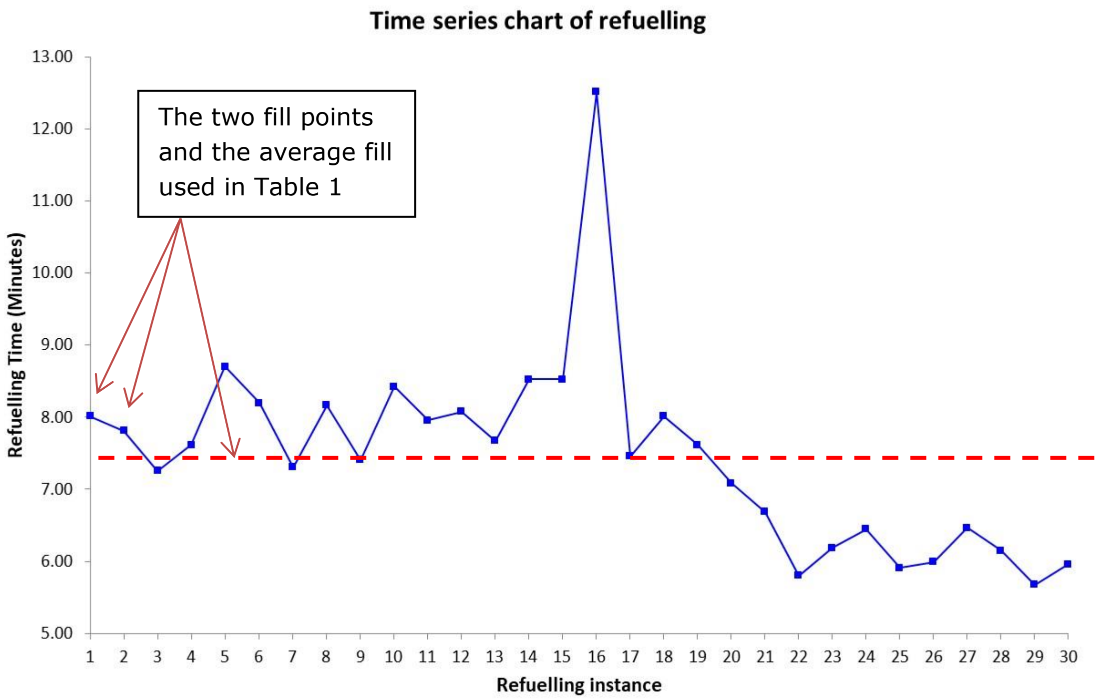
- there is plenty of variation, as we should expect given our list of potential factors above; and
- the average is not particularly representative/ useful. It masks how wide the range of values is and hides any outliers.

Fig. 1 shows a great deal more than Table 2, but we are still missing the dimension of time.

[3] **Histogram:** Some care needs to be taken when creating a histogram, particularly with regards to selecting the 'buckets' to place the data into (the intervals).

Here are the same 30 data points presented as a time-series chart, with the red dotted line denoting the average. I’ve also pointed out the three data points used in the Table 1 ‘Management dashboard’ report:

<u>Figure 2:</u> Time series chart of time to refuel – 30 data points



That’s far more illuminating - some interesting things appear to be going on. More on this later.

Averages hide what is happening

There can be a whole host of different sets of data that will have **the same average value but vastly different variation** within.

- With regards to the histogram, we could have:
    - A tight distribution around the average; or
    - A wide distribution; or
    - Skewed distributions; or
    - Multiple peaks; or
    - ….and so on

- With regards to the time series chart, in addition to the potential histogram differences, we could also have:
    - Consistent variation over time; or
    - Changing variation over time (going up or going down or widening or narrowing or….and so on)

In short, **when we use an average we are virtually blind as to what is actually going on**.

A nice quote to consider in this regard is:

> *"Beware drowning in a river of average depth one metre"* (Bicheno)

If you've ever crossed a river, you'll know that it's often shallow in some parts and (really) deep in others. Using the average depth to decide whether to cross could kill you!

Is there a difference?

The big question that we constantly ask ourselves when we look at data is **'has a process changed?'** (whether for better or worse). This is because we aim to make real on-going improvements and, conversely, to notice, halt and reverse any decline.

Sticking with the car refuelling example:

- A reminder that it took you 7.81 minutes (7 minutes and 49 seconds) to refuel today. That's 12 seconds quicker than the 8 minutes and 1 second it took last time! Should you be pleased about this 'improvement in performance'? Does it suggest that the refuelling process has changed? I doubt it.

- Conversely, what if it had taken you 12 minutes and 31 seconds to refuel today? (a difference of nearly 5 minutes). Well, yes, that could be frustrating, and probably abnormal. It looks like something different happened – a change to the process?

So, if a difference of 12 seconds is nothing to bother about, yet a difference of 5 minutes is, then presumably there's a divide somewhere between as to what's routine and what's exceptional? If so, where is this divide? Or more specifically, what range of times would you consider as being routine, and why?

If you look back at the time series chart at Fig. 2 above, you can see lots of ups and downs. The question that we want to answer is 'which ones should we be interested in and which should we ignore?'

Introducing the work of Walter Shewhart

This problem of understanding variation and its significance was solved by Walter Shewhart (1891 – 1967) whilst working as a statistician at the Hawthorne Works (Illinois, USA) of the Western Electric Company.

He came up with a method[4] of visualising variation and of **effectively separating out variation into two components** for the purpose of quality improvement.

W. Edwards Deming (1900 – 1993) came across Shewhart's work and began collaborating with him. Deming then took Shewhart's ideas to Japan from 1950 onwards.

Deming is credited with being a significant factor in Japan's post-war economic recovery.

[4] **Shewhart:** Today we recognise Shewhart's method as Statistical Process Control (SPC), using control charts.

Components of variation

All data collected about a given process display variation. Shewhart devised how we can split this variation into two components:

- **Routine variation** (also known as 'assignable' or 'common cause' variation); and
- **Exceptional variation** (also known as 'chance' or 'special cause' variation).

Routine variation (as the name suggests) is to be expected within a process and is often simply referred to as **noise**. It is in-built **within the current design of the process** and, as such, is unavoidable. This routineness makes it predictable.

> *Clarification: Just because such routine variation is (currently[5]) unavoidable doesn't make it desirable.*

Exceptional variation is caused by something 'out of the ordinary'. Such exceptions are **signals of a process change**, which could be **temporary or permanent**. Such variation is unpredictable.

Using these two definitions with our car refuelling example, you can see that:

- All of the factors causing variation in the list of factors above are routine i.e. within the current design of the process;
- Examples of exceptional variation (i.e. a change to the process) might be that:
    - You met a long-lost friend on the forecourt and had a lengthy chat; or
    - The fuel station 'upgraded' their computer system (hopefully for the better!)

The first example is a temporary change to the process, the second is not.

Why does this matter?

My deliberately simple 'refuelling the car' example can (and should) be extended to <u>all</u> systems, and processes within.

When people (particularly those in positions of management) look at numbers, they usually look at each of the ups and downs, search for an explanation, come to conclusions, and take actions accordingly.

However, given the two components of variation, we can see that some (usually most) ups and downs in data **require no explanation** (because nothing's actually changed). Any such explanation[6] would be incorrect and would likely lead to unjustifiable actions.

---

[5] **Routine variation:** If you want to reduce routine variation then you would have to (successfully) redesign your process. Looking at specific data points (e.g. refuelling instances) will not help.

[6] **Explanations:** Such explanations often focus on people (he/ she must have done better or worse) and the resultant actions commonly involve compliments or reprimands. Most of the time, the ups or downs are not about individuals – they are because of the system that they work within.

Wheeler writes that:

> *"The numerically naïve approach of …'two numbers that are not the same are different!'…will turn everything into a signal – two points will always define a trend – and explanations will be required for all of the (unfavourable) noise in the monthly report."*

This leads on to the two mistakes possible when attempting to use data:

> ***Mistake 1:*** *we interpret noise (routine variation) as a signal and sound a false alarm and [thus] chase after explanations that do not exist;*
>
> ***Mistake 2:*** *we interpret a signal as merely a bit of noise (routine variation), thereby miss the signal and fail to learn that which could be learned.*

If we are to avoid making either mistake, then we must use Shewhart's method to separate out the noise (routine variation) from any signals of exceptions.

Tampering

Making Mistake 1 - taking actions because of ups and downs when nothing's actually changed - is referred to as **tampering**. You might think that this is harmless. However, tampering makes things worse!

If you are curious to understand this point more fully then there is a useful exercise called the Funnel experiment that nicely demonstrates this fact (See 'Tampering').

As a summary of this pint, Wheeler writes that:

> *"Questions about signals lead to improvements, questions about noise merely create chaos."*

Introducing the control chart

So, we want to:

- see the variation within our process over time; and
- separate out the noise (routine variation) to uncover any signals of exceptions that may be present.

This is where the **control chart** (as first devised by Shewhart) comes into play.

A control chart helps us perform what Wheeler terms an **Observational Study:**

> *"[We] are not looking for a difference that we think is there, but* ***asking if an unknown change has occurred*** *within a process that we are trying to operate in some steady state.*
>
> *Since a change can occur at any time,* ***this question needs to be continually asked of the process****.*
>
> *As each value is obtained you can use it to see if a change has occurred. This makes every point added to a chart a separate act of (sequential) analysis."*

We will see whether a change has (or hasn't) occurred by constructing a control chart through performing a two-step process of:

- calculating and adding some summary statistics (a centre line and limits) to a time-series chart of our data (creating **a baseline**); and then
- applying some tests against the resultant chart **as each new data point is added**.

How to carry out these steps is explained in detail below.

The XmR control chart

There are several different types of control charts, but the simplest (and perhaps most useful) is the chart of individual values (X) and their moving ranges (mR). This is referred to as an **XmR chart**.

Here's the first 10 instances of our refuelling example again, but also with their moving ranges calculated:

<u>Table 3:</u> Time to refuel – First 10 data points and their moving ranges:

Refuelling instance (X)	1	2	3	4	5	6	7	8	9	10
Time taken (minutes)	8.01	7.81	7.25	7.62	8.70	8.20	7.30	8.17	7.40	8.43
Moving range (mR)		0.20	0.56	0.37	1.08	0.50	0.90	0.87	0.77	1.03

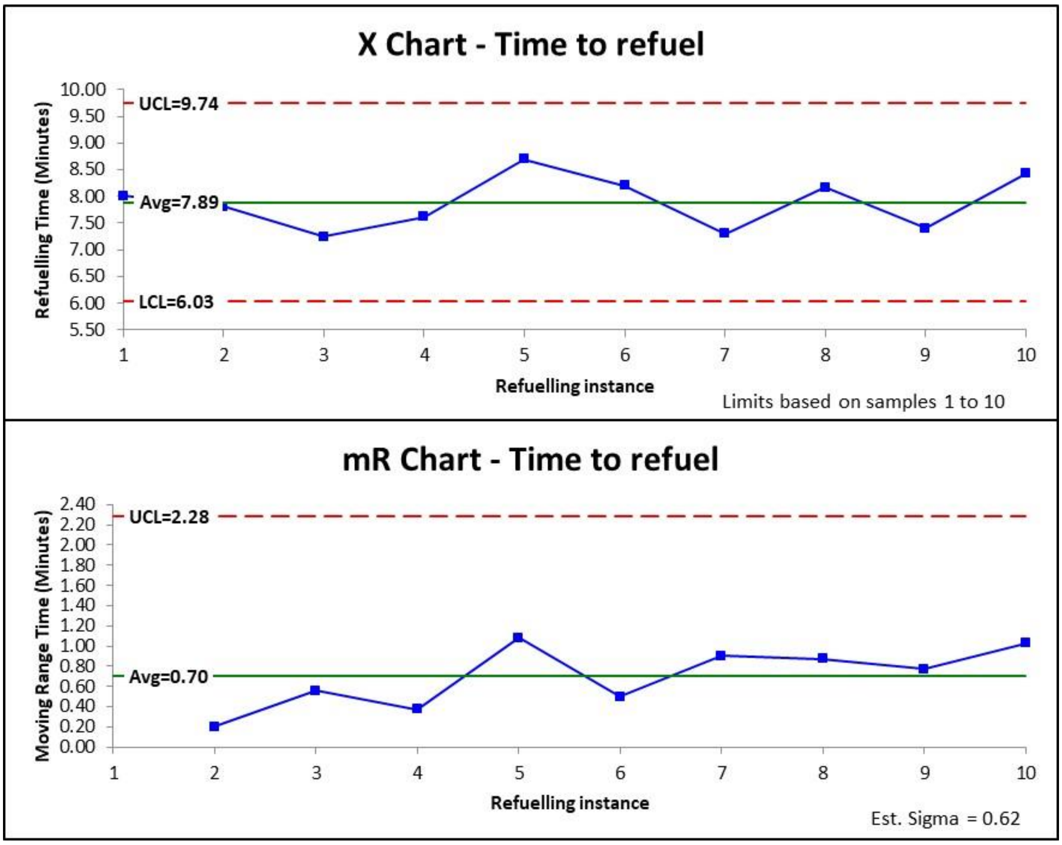
The **moving range** (mR) reveals the point-to-point variation and is calculated as **the difference between each sequential individual value** (X).

So, the moving range between refuelling instance 1 and 2 is the difference between 8.01 and 7.81 (which = 0.20). We are only interested in the magnitude of the movement and so we ignore the sign[7].

[7] i.e. all moving ranges are shown with a positive sign.

Here is an XmR chart visualising the Table 3 data:

<u>Figure 3:</u> XmR chart of time to refuel – Baseline of first 10 data points



Note that there are two parts to the chart – a plot of the individual values, X, and (below this) a plot of their moving ranges, mR.

The red and green lines applied to the chart will be explained below.

'Apples with apples'

Before going further, we must consider whether we are comparing 'apples with apples'.

> *"You can pick the best graphical format, and compute the best numerical summaries, and still end up with nonsense because you are comparing apples with oranges."* (Wheeler)

We want to be comparing **data points that have come from the same cause system**. If not, then what are we doing?

Some examples to consider:

- If you are monitoring your blood pressure, then you should be taking readings when your body is in a comparable state. Taking some readings when you wake up and some after rigorous exercise and then mixing these all together will provide misleading information. Sure, you'll get a graph over time…but what would it mean?

- There's no point putting the times to refuel an engine-driven car on the same chart as the times to recharge an electric car. To truly get 'apples with apples' you'd want to compare refuelling times for the same car type, using the same fuel station design.
- …and so on

Checking for patterns

Many systems will contain repeating patterns in their time-series data (e.g. there might be a daily or weekly cycle), which suggests that there are different cause-systems within.

Once you've charted your data as a time-series, and **before you start to use software to interrogate it, stand back and take a look**. Do the movements look random or are there any patterns that would cause anyone to ask *"I wonder what lies behind that?"* If there are, then go and investigate.

Wheeler writes that *"you [with your human brain] are the best pattern recognition device that we know about."*

The XmR chart should not be used to compare apples with oranges and so, if you have a repeating pattern, and once you've investigated why, you will need to separate out the different elements of the pattern so that you are comparing 'like with like':

- You might need to separate out 'days of the week', to compare Mondays with Mondays
- You might need to separate out daytime operations from 'after hours'
- …and so on.

Calculating limits for the XmR chart

So, assuming that we have ensured that we are only dealing with metaphorical 'apples'…

Starting somewhere - setting a base line:

Whilst we will be testing each new point added to the control chart with the behaviour of those previously collected, we must start somewhere. We do this by creating a baseline to calculate limits from.

There are a few things to be aware of when doing this:

- Choose a baseline period in which the process is **free from known changes**

- The more data points used in the baseline, then the more solid (i.e. useful) the limits will be. However:
    - You don't need a lot of data points to reach such solidity. **Around 20 data points should do it**
    - You can calculate useful limits for an XmR chart with **as few as 6 data points** *if* you take good care to avoid known or obvious changes within the baseline period. So, whilst using 20 or more points is desirable, this is not necessary

- Once you've calculated your baseline summary statistics then you **don't change them** again…unless there is good reason to[8]. Key point: You don't recalculate each time you add a new data point

The above advice doesn't give you an exact answer regarding setting your baseline and this is deliberate - **you must think about the data you have and their context**.

Wheeler makes clear that:

> *"The objective is not to compute the 'right' limits, but to use limits to characterise the behaviour of the process."*
>
> *"The priority for choosing a baseline is not the number of data points within that baseline, but rather the rationality of that baseline given the context of the data."*

Note: Your first go at setting a baseline may necessarily be arbitrary (because you don't know much about the process when you begin to analyse at it) but, as you learn about your process, you can usually define a baseline that makes sense given the context of the data.

Finally, Wheeler provides a caution regarding using statistical software:

> *"The software does not know whether a given baseline is rational or not. Do not let the software choose the points to be included in your baseline."*

Which calculation?

Before we get to the calculation for creating limits to plot on an XmR chart, we should understand two statistics within – measures of:

- Location; and
- Dispersion.

A **measure of location** is a statistic about a set of data that characterises its 'mid-point'. There are two different methods for doing this:

- the average (the 'balance point' for the data); and
- the median (which divides the data into two halves).

When using XmR charts we **generally use the average** in our calculations. This is because it is more efficient than the median in its use of the data – you need less data points before achieving solid limits. However, the average is susceptible to extreme values.

Wheeler writes that:

> *"Medians are less efficient than averages, but they are less susceptible to extreme values. So, if your average moving range might have been unduly inflated by some exceptionally large moving ranges, you might wish to consider using the median moving range instead."*

So, don't use the median by default. Use the average unless your data clearly suggests otherwise.

---

[8] These 'good reasons' are covered later in this guide.

A **measure of dispersion** tells us how the data are spread out around the measure of location.

There are a few different methods for calculating a measure of dispersion (which include the range and standard deviation). However, the only relevant one for using with an XmR chart is **calculated from the set of moving ranges**. The resultant measure of dispersion is called **sigma**[9].

Clarification: We do not use the standard deviation as the measure of dispersion

The calculations:

Looking at the baseline XmR chart (Fig. 3 above), you will see some (green and red) lines drawn across the charts. They are calculated as follows:

Table 4 XmR chart calculations:

	Method 1: If using Average mR (for general use)	**Method 2:** If using Median mR (only if applicable)
For the X chart:		
Centre line (CL_x)	Average of X ($\bar{X}$)	Median of X ($\tilde{X}$)
Upper Limit (UL_x)	$\bar{X}$ + 2.660 $\overline{mR}$	$\tilde{X}$ + 3.145 $\widetilde{mR}$
Lower Limit (LL_x)	$\bar{X}$ - 2.660 $\overline{mR}$	$\tilde{X}$ - 3.145 $\widetilde{mR}$
For the mR chart:		
Centre line (CL_{mR})	Average of mR ($\overline{mR}$)	Median of mR ($\widetilde{mR}$)
Upper Limit (UL_{mR})	3.268 $\overline{mR}$	3.865 $\widetilde{mR}$

Some notes on the above calculations:

- The bar notation (line) across the X and mR in Method 1 denotes the average
- The tilde notation (squiggle), across the X and mR in Method 2 denotes the median
- The **scaling factors** within the calculations (example: the 2.66) are constants[10], but please note that they are different between the X chart and mR chart, and between the two methods.
- The upper and lower limits calculate **'three-sigma limits'**, where sigma is our measure of dispersion
- If the calculated lower limit for the X chart is less than zero and this doesn't make sense (i.e. you can't have a negative time to refuel your car) then there is no lower limit[11].

[9] The term sigma denotes a standard unit of dispersion.

[10] **Origin of the scaling factors:** See footnote b) at the end of the guide for further information.

[11] This would be a one-sided X chart with a boundary condition of zero on the bottom and a control limit on the top.

- There is no lower limit for the mR chart since the values are always positive.

Warning: Many statistical software packages can (and will) calculate, and add, limits to control charts for you. However, many can (and do) perform this calculation incorrectly. You will need to make sure that the calculation is set up, and working, correctly.

Using Method 1 (i.e. average mR) calculations on our baseline refuelling data (Table 3) we get the following:

Table 5: Calculations for baseline refuelling data (Table 3)

For the X chart:	Calculation:	Result[12]:
Centre line (CL_x)	Average of X ($\bar{X}$)	7.89
Upper Limit (UL_x)	$\bar{X}$ + 2.660 $\overline{mR}$	7.89 + (2.660 * 0.70) = 9.75
Lower Limit (LL_x)	$\bar{X}$ - 2.660 $\overline{mR}$	7.89 - (2.660 * 0.70) = 6.03
For the mR chart:		
Centre line (CL_{mR})	Average of mR ($\overline{mR}$)	0.70
Upper Limit (UL_{mR})	3.268 $\overline{mR}$	3.268 * 0.70 = 2.29

Why might (should) we show the mR chart as well as the X chart?

Many people only show the top half (the X chart) of the XmR chart. They choose not to show the mR chart. Whilst this is usually fine, there are some reasons as to why you might show both:

- The mR chart can detect sudden shifts even when the individual values sit within the X chart limits
- By including the mR chart you are showing to those reading it that you know **the correct way to compute the limits using moving ranges** (and, for instance, haven't used the standard deviation in error)
- The mR chart allows you to check for the problem of 'chunky data' (a more involved subject outside the scope of this foundational guide).

Has a change occurred? Detection tests to apply

So, we've got a baseline (Fig. 3 above), with control limits applied. We now apply a set of four detection tests (as derived by Shewhart) against the data and these limits to detect signals, and we reapply these tests as each new data point is added.

[12] **Rounding:** You might notice that there is a very small 'rounding' difference between my calculations and what is shown for some of the lines in the charts. This is because my calculations are using the averages to 2 decimal places whereas the charting software is being more precise.

What are the detection tests?

The tests are as follows:

Detection rule one:

A single point falls outside the three-sigma limits on either the X chart OR the mR chart.

- The further a point is outside the limits, then the stronger the evidence that a process change has occurred.

Detection rule two (a run beyond two-sigma):

Two out of three successive values on the X chart are:

- On the same side of the central line AND
- More than two-sigma units away from the central line

Detection rule three (a run beyond one-sigma):

Four out of five successive values on the X chart are:

- On the same side of the central line AND
- More than one-sigma units away from the central line

Detection rule four (a run about the central line):

Whenever eight successive values on the X chart all fall on the same side of the central line

Detection rule one looks for **large process change**, whilst Detection rules two, three and four look for **successively smaller sustained changes**[13].

But what about other detection tests?

Wheeler writes that:

> *"The most commonly used detection rules are rule's one and four. The complimentary nature of these two rules allows them to work well together…you will seldom need more than these two rules.*
>
> *The **Western Electric [Shewhart's four] tests form a coherent set**…which have been proven to provide the maximum boost in sensitivity without an undue increase in the risk of false alarms."*

[13] **Detection rules:** Rules two, three and four can't be applied to the mR chart. Further, they can only be used with the X chart if the data is shown sequentially.

Many statistical software packages allow you to apply lots more detection tests[14], and to change how each is calculated (i.e. the number of data points required to trigger them). Wheeler goes on to write that:

> *"Using a battery of such run tests is like refusing to give up. Use enough detection rules, and even a table of random numbers will show you a signal.*
>
> *You should, in general, avoid the use of additional detection rules…remember, you are the best pattern recognition device that we know about…let the software create the charts, but do not let it interpret the charts."*

Understanding what action to take

The distinction between two components of variation leads to **two different routes to process improvement**.

a) If we have exceptional variation present, then:
   - We have some special causes as well as many common causes of variation (i.e. signals buried within the noise)
   - The process is said to be **unstable**. We can't predict how it will operate going forwards in this state
   - **Improvement will come from removing the effect of the special cause**
   - If this special cause is:
     - Internal to the process (i.e. a cause in our control) then we should identify and remove the effect of the special cause (i.e. investigate and take action), thus improving the process and its outcomes;
     - External to the process (i.e. an upset beyond our control) then there's nothing to do. If the cause is temporary, then we can expect the process to revert to its previous condition.

b) If we have only routine variation present, then:
   - There will be many common causes of variation creating this noise, with no special causes present
   - The process is said to be **stable**. If we make no change to our process, then we can predict how it will continue to operate going forwards (i.e. within the range of times between our limits)
   - **Improvement will require fundamental change to the process**
   - Whether we choose to focus our efforts on this stable process will depend on whether:
     - Its current performance is satisfactory/ desirable or not; and
     - We have other areas that deserve our attention

[14] **Other detection tests:** You might see further tests referred to as Juran, Montgomery, AIAG (and others).

What does our baseline tell us about the refuelling process?

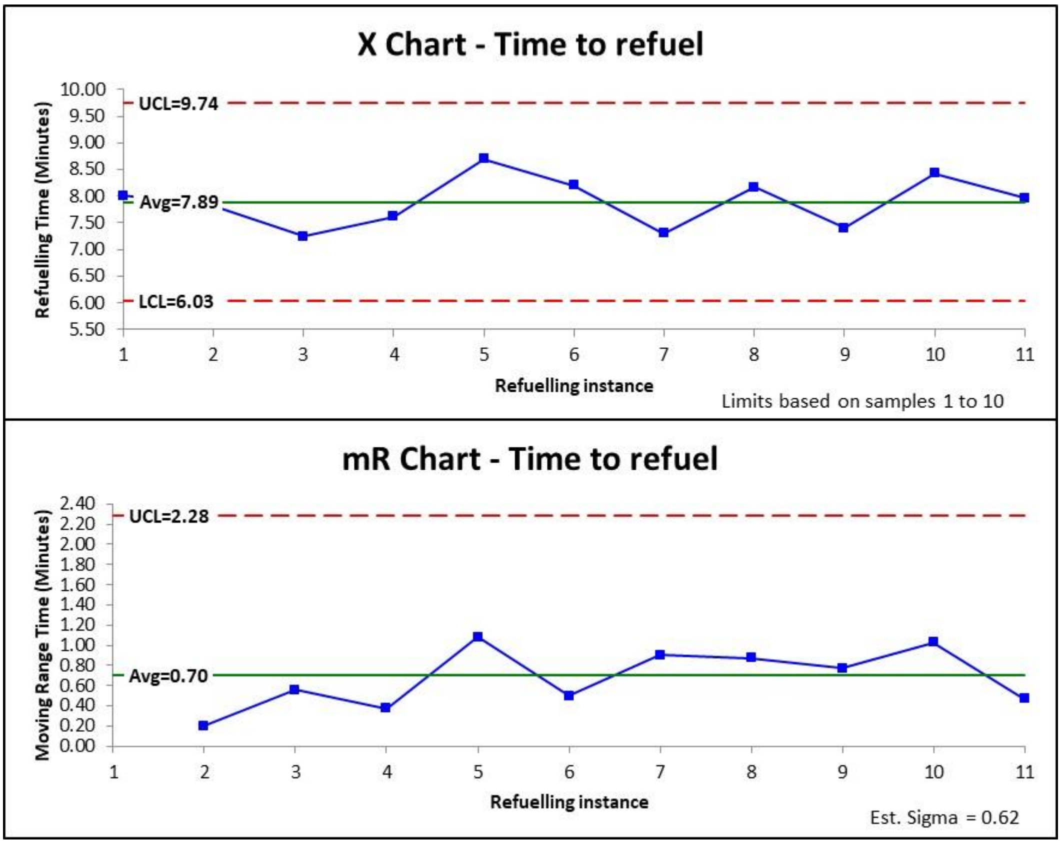
Looking back at our baseline (Fig. 3 above), we can see that all 10 data points fall within our three-sigma upper and lower limits (i.e. pass detection rule one for both X and mR charts). Further, they do not trigger any of detection rules two, three or four for the X chart.

This means that we can say that all the variation is routine (noise) and the process is stable. As such, there is nothing further to explain and, assuming the process is not changed, we can predict that future refuelling instances will take between 6.03 and 9.75 minutes (i.e. between the lower and upper limits).

Adding the next point to the data

Remember that we are testing each new piece of data against the behaviour of our baseline – an observational study. The next chart shows the 11th data point added and, as expected, it doesn't trigger our detection rules – **there is nothing to explain** (even though it is 'different' than the last one).

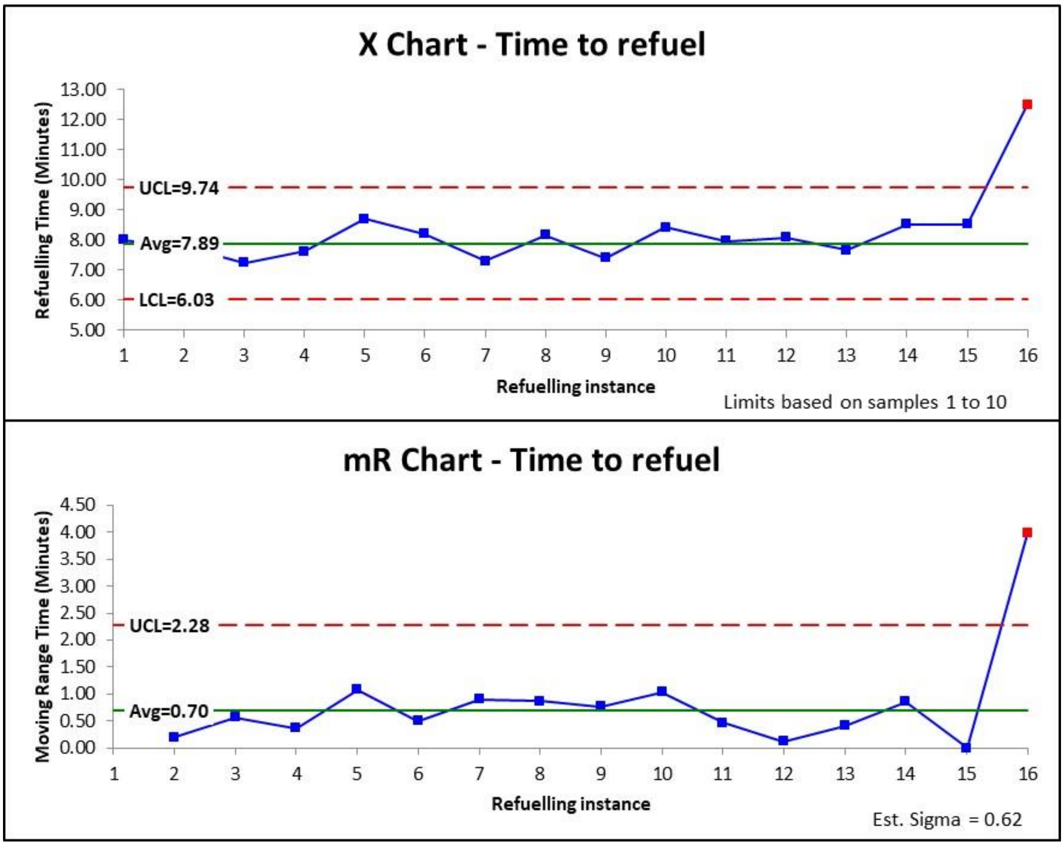
<u>Figure 4:</u> XmR chart of time to refuel – 11th point added.



Seeing and reacting to a signal

We carry on collecting our data, adding them to our XmR chart and running our detection tests (against our baseline). Our process is stable until we reach refuelling instance 16:

<u>Figure 5:</u> XmR chart of time to refuel – 16th point added.



You can see that refuelling instance 16 (taking 12.51 mins) has triggered detection rule 1 on both charts – it falls a long way outside the upper limits and so is **a strong signal that something exceptional happened.**

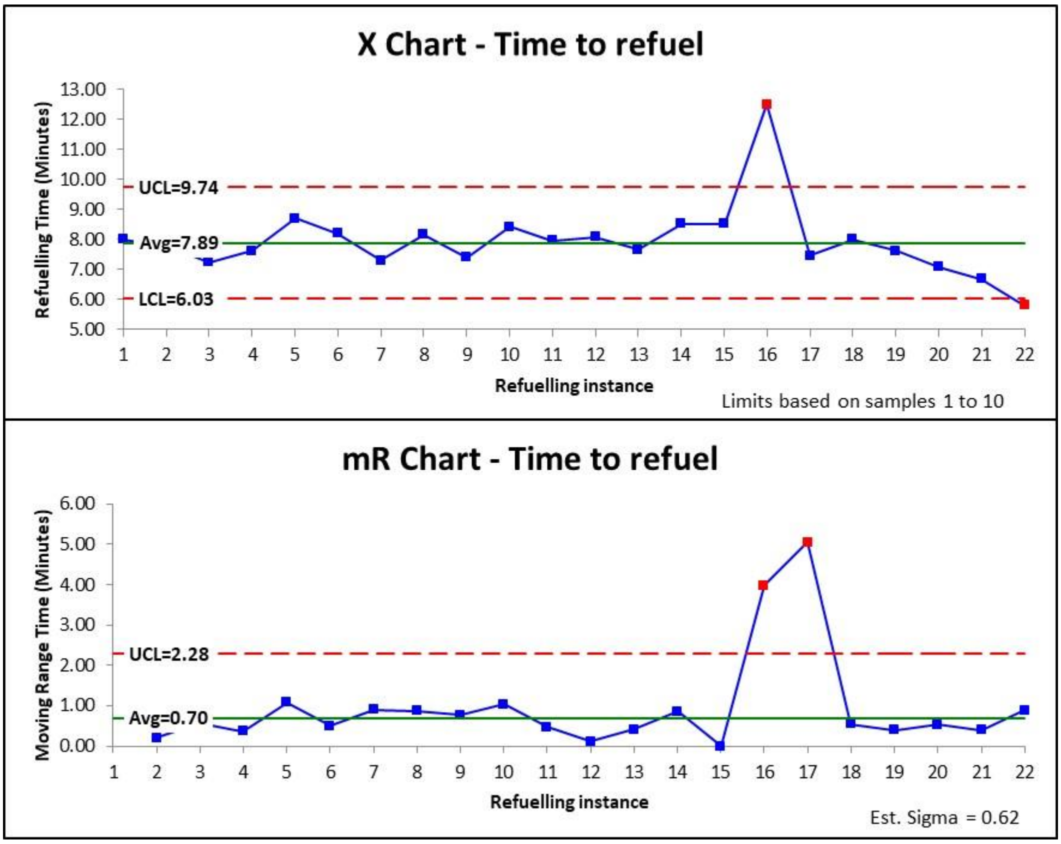
Having detected the exception, we should **investigate what happened and why**, to determine whether we can and should do something about it.

In this instance, the reason for the long refuel was that an old friend pulled up at the fuel pump next to me and we had a conversation that slowed me down considerably. Knowing the cause, I can say that it was likely a one-off (temporary change) and that the process should return to its previous condition.

Continuing to sequentially add data points

We continue to add data points and run detection tests against our baseline.

<u>Figure 6:</u> XmR chart of time to refuel – points 17 through to 22 added.



Refuelling instance 17 confirmed that instance 16 was a temporary exception and the process returned to its previous stable condition. Note that instance 17 triggered an exceptional moving range (detection rule one) showing it swinging back to normal.

Instances 18, 19, 20 and 21 are within our limits and don't trigger any detection tests…but appear to be trending downwards. Has the process changed? We can't be sure yet. It could be noise.

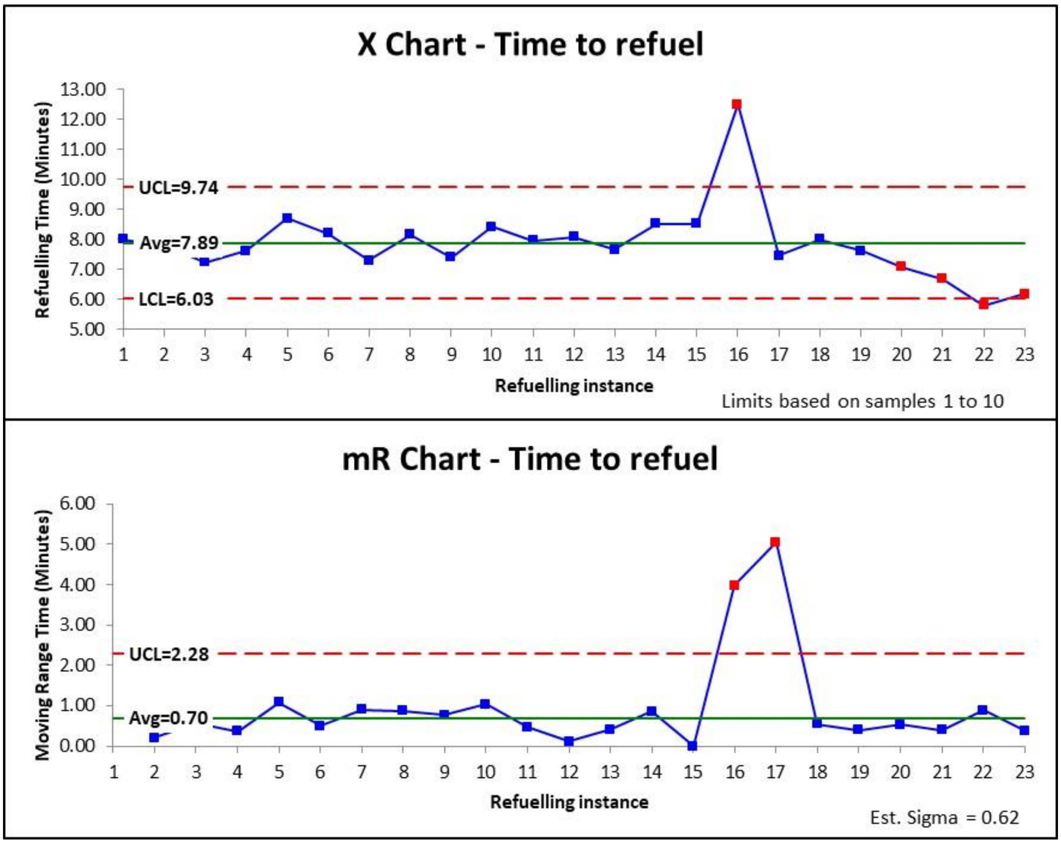
However, when we add point 22 we trigger detection rule one. We are back to investigating. In particular, we want to investigate whether this is a permanent change.

In this instance, your investigation finds out that the refuelling station has changed its computer system, to make serving customers much quicker. That makes sense…and should be a permanent improvement…let's see if it is.

Has the intended improvement worked?

We now add the next refuelling instance (23):

<u>Figure 7:</u> XmR chart of time to refuel – 23rd point added.

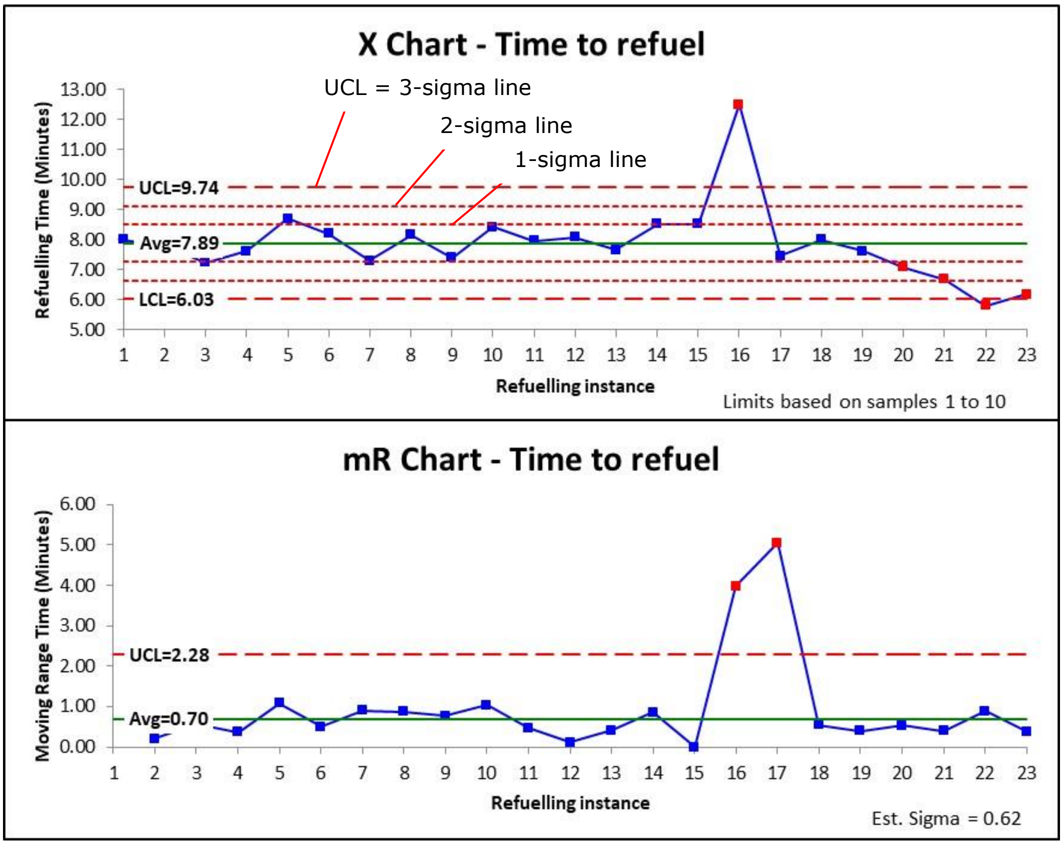


Yes, it looks like refuelling instance 23 has triggered Detection Rules 2 and 3 (a run beyond one and two sigma[15]).

To help you see this, see Fig. 8 below, which is the same as Fig. 7 but has the one-sigma and two-sigma lines added within the normal 3-sigma limit lines:

[15] Note: Points 20 and 21, which were previously blue (probably noise), have now 'turned red' (potential signals) because the addition of points 22 and 23 have uncovered them as part of a run.

Figure 8: XmR chart of time to refuel – 23rd point added, with one, two and three sigma lines showing.



You can see that:

- Four out of five sequential points fall outside the same one-sigma line (Detection Rule three), with these five points being instances 19 – 23; and
- Two out of three sequential points fall outside the same two-sigma line (Detection Rule two), with these three points being instances 21 - 23

This shows, as expected, that this process change is sustained. It now looks like the refuelling process is operating at a different level of performance and our baseline is no longer relevant – we need to consider whether to calculate a new baseline to take forwards

Recalculating limits

We should only recalculate limits when it makes sense to do so. Wheeler sets out four questions to determine when you need to revisit limits[16]. These are:

1. *Do the data display a distinctly different kind of behaviour than in the past?*
2. *Do you know the reason for this change in behaviour?*
3. *Is the new process behaviour desirable?*
4. *Is it intended and expected that the new behaviour will continue?*

[16] **Revising limits:** Wheeler credits Perry Regier for devising the four questions.

*"If the answer to **all four questions** is yes…then revise the limits based on data collected since the time when the process changed."*

Commenting on each of the questions in turn:

- Re. Qn. 1: It should be obvious that, if there's been no change in the performance of the process (i.e. variation is still accounted for by routine variation), there's no need to change the limits
- Re. Qn. 2: You should be looking for the reason for any special causes, rather than jumping to change the limits
- Re. Qn. 3: If the new performance isn't desirable then the thing to do is to work on removing the detrimental special cause, rather than playing with limits; and
- Re. Qn. 4: If its temporary, there's no need to change the limits

Going back to our refuelling example at Fig. 7, I believe that we can answer 'Yes' to all four of Wheeler's questions, and so should re-compute our limits.

We use the same calculation (back at Table 4), but on a new baseline set of data, starting at the point that we believe the process changed.

We need to gather enough data to create the baseline. Here's the ten data points from point 20 onwards (i.e. from when the change occurred):

Table 6: Time to refuel – 10 data points from point 20, & their moving ranges:

Refuelling instance (X)	20	21	22	23	24	25	26	27	28	29
Time taken (minutes)	7.09	6.69	5.80	6.18	6.45	5.91	5.99	6.47	6.15	5.68
Moving range (mR)		0.40	0.89	0.38	0.27	0.54	0.08	0.48	0.32	0.47

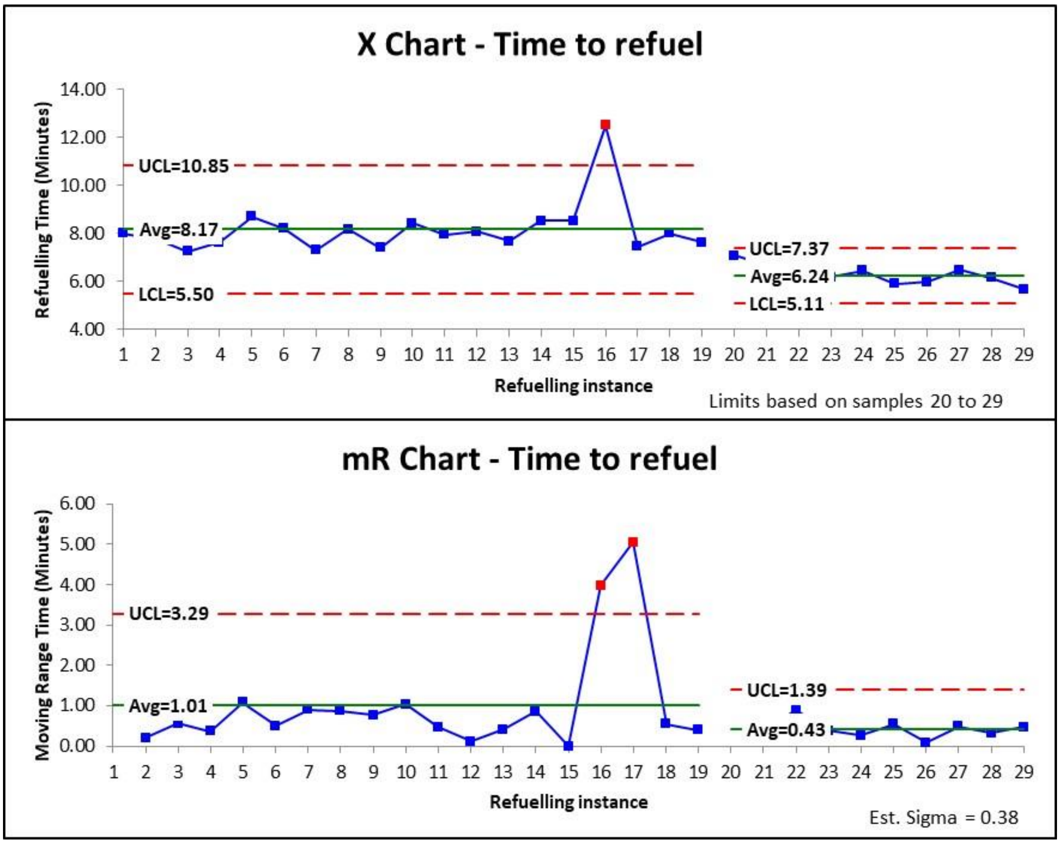
Here's the calculation for the new baseline:

Table 7: Calculations for new baseline refuelling data

	Calculation:	Result:
For the X chart:		
Centre line (CL_x)	Average of X ($\bar{X}$)	6.24
Upper Limit (UL_x)	$\bar{X}$ + 2.660 $\overline{mR}$	6.24 + (2.660 * 0.43) = 7.38
Lower Limit (LL_x)	$\bar{X}$ - 2.660 $\overline{mR}$	6.24 - (2.660 * 0.43) = 5.10
For the mR chart:		
Centre line (CL_{mR})	Average of mR ($\overline{mR}$)	0.43
Upper Limit (UL_{mR})	3.268 $\overline{mR}$	3.268 * 0.43 = 1.41

And here's the XmR Chart showing the move from the old to the new baseline:

<u>Figure 9:</u> XmR chart of time to refuel – New Baseline from point 20



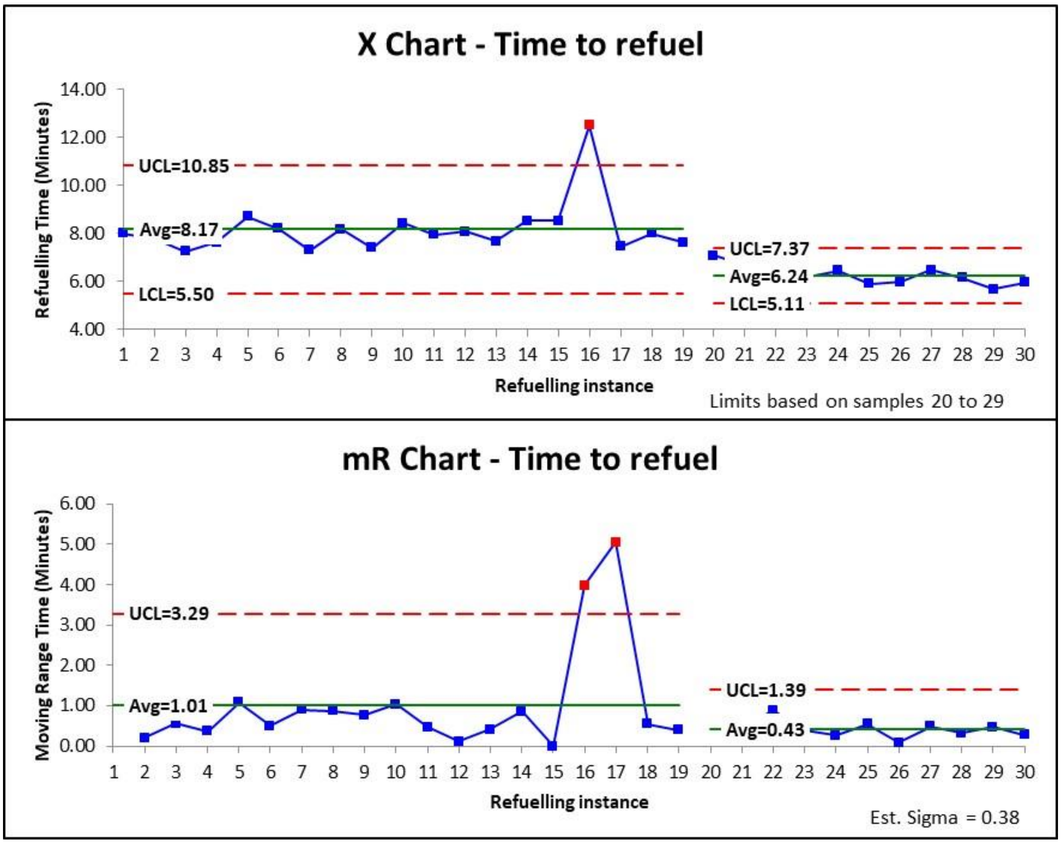
You can see that the new baseline shows the process to be stable, and therefore predictable.

The chart very clearly shows what has happened over time[17]. It has clearly separated out signals from the noise.

We can now continue exactly as before by adding the next refuelling instance (whilst keeping the new baseline limits constant) and running our four detection tests:

[17] You may note that the statistical software I'm using for my charting has recalculated the old baseline to include points 1 – 19, because it is charting how the process was operating for all the data points before the process change at point 20.

Figure 10: **XmR chart of time to refuel – point 30 added.**

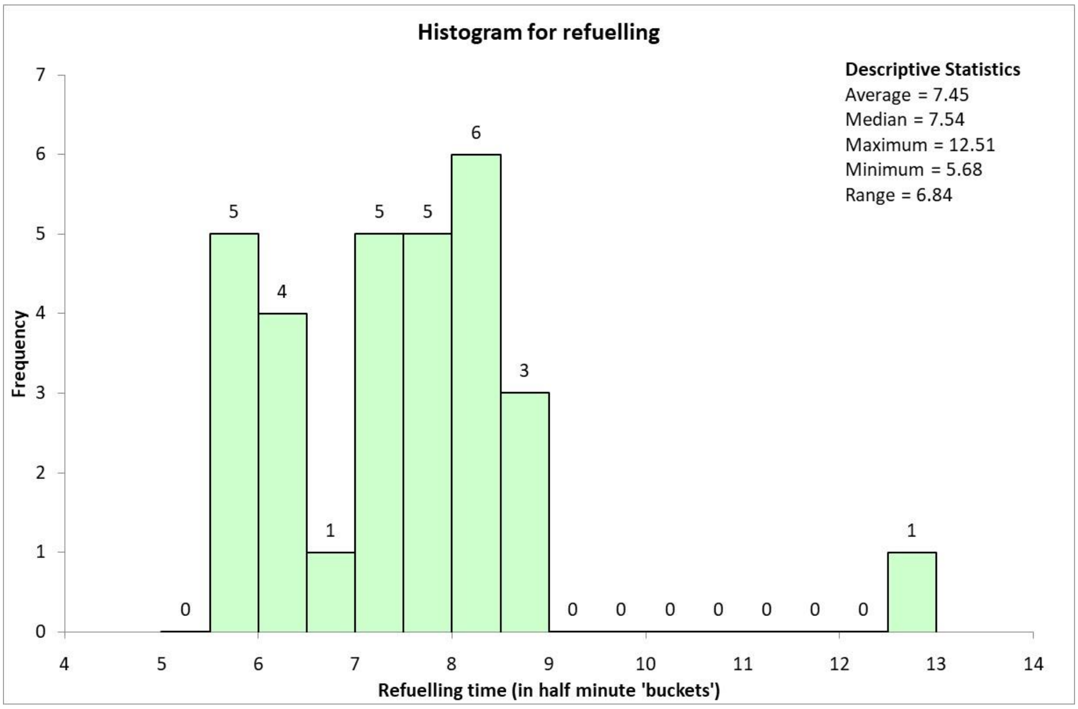


Yes, point 30 is as would be predicted and the new process appears to be stable.

That completes our look at 30 data points (from Table 2). But we aren't done there – it's a never-ending observational study. You would carry on in the same way, repeating each of the steps above as new data is added.

Looking back at the histogram with knowledge

If you look back at the histogram at Fig. 1 (reproduced below) you can now see what was hidden within.

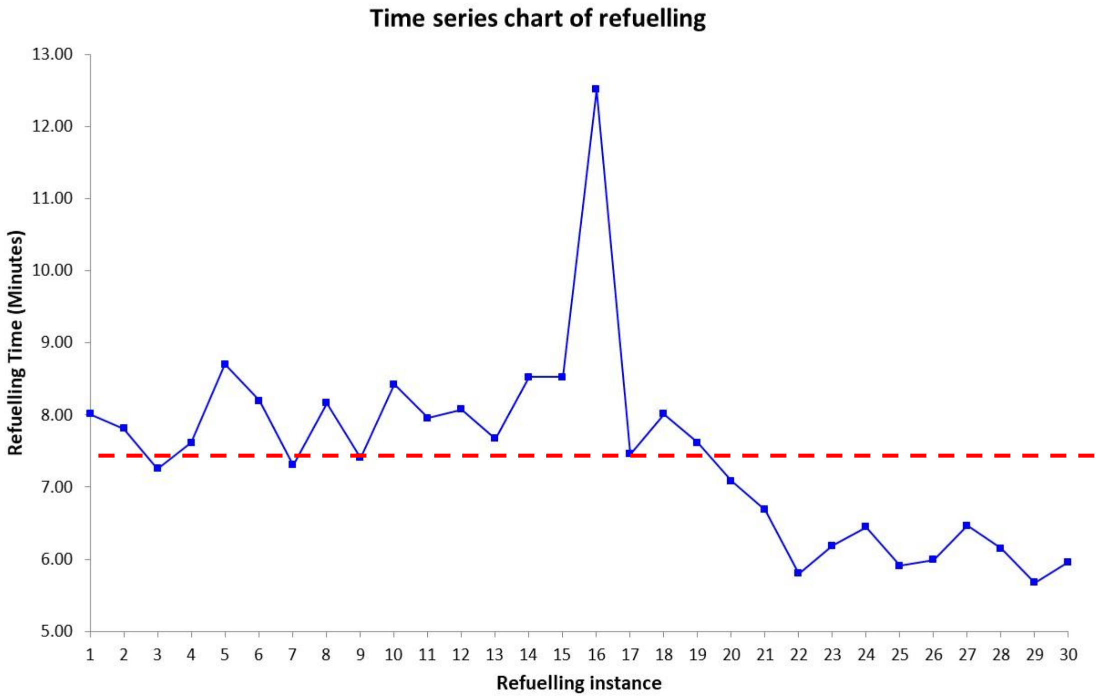


- There was a temporary exception at instance 16 regarding 'a chat with an old friend'. This was the reason for the outlier; and

- There was a permanent process change from around instance 20 because of a new computer system. This is the reason for the two 'humps' in the distribution of the data, because there were two different levels of process operating within.

Looking back at the time series chart with knowledge

If you look back at the time series chart at Fig. 2 (reproduced below) you will see the exact same shape as in the X chart at Figure 10 (which should be no surprise because they are drawn using the same data from Table 2).



Operating the control charts as described above explains <u>why</u> the time series shows what it does. Further, **this knowledge was uncovered as the process was operated** rather than having to wait to see all 30 points on a time series chart.

You can also see that the average line at Fig. 2 doesn't help, and probably hinders.

Back to the start: Binary comparisons

A binary comparison is to compare two numbers with each other and form some opinion about the difference. This is the most common way of using measures – comparing this week with last week, or with this week last year, or with budget or….

However, binary comparisons are 'worse than useless'. Looking back at our data points above, which two points are you picking to compare? Why? What can you conclude from them given that they will all contain a degree of noise?

You might like to reflect on our original Table 1 (reproduced below) and note that this contains two binary comparisons ('This fill' with the 'Last fill' and 'This fill' with the 'Average fill').

Measure	This fill	Last fill	Variance		Average fill	Variance	Comment
Time to refuel (Minutes)	7.81	8.01	0.2 (-3%)		7.45	0.36 (5%)	Blah blah blah

You can see that:

- There was nothing to explain for either of these binary comparisons as the process was stable at that point; and therefore
- Any narrative created from this binary comparison cannot have been correct.

Which takes us full circle to **the problem with conventional measurement.**

A few final words

'We learn by doing': You might have found (some of) the above interesting, even persuasive but nothing will change unless you 'give it a go'. Wheeler writes:

> *"Those who do not use [control] charts have no advantage over those who can't."*

The simple act of playing with your data, using the above, will take you on an interesting journey, posing all sorts of questions to you.

"How did you create those charts?": You will notice that my XmR charts above look rather nicely drawn. That's because I've used some relatively simple software to do this[18].

"How can I add my own lines to the chart?": Some people want to add lines onto their charts to show the level of performance that they desire from a process (often referred to as 'the voice of the customer').

However, the lines created by a control chart have got nothing to do with what you desire – they show **how the process is performing** ('the voice of the process'), whether you like what you see or not.

> *"The voice of the customer is what you want. The voice of the process is what you will get."*

More detailed topics

There's a whole bunch of topics that are beyond the scope of this foundational guide.

They include (but are not limited to):

- Why you shouldn't use standard deviation to calculate control limits
- Why you <u>don't</u> need to transform your data if it doesn't fit a normal distribution
- The problem of chunky data, and how to handle

[18] **Charting tool:** I use an Excel add-in called 'SPC for Excel'. I don't know how it compares to others, it's just the one that I have and use (and quite like). I have previously used Minitab, but this is a more sophisticated (and costly) tool.

- The power of disaggregating data (which is an extension of comparing 'apples with apples')
- Seasonality, and techniques for adjusting for this
- Different types of controls charts
- …and more

You could use Donald Wheeler's excellent books[19] and blog articles to assist.

Footnotes:

a) Source of knowledge: Much of the information within this guide comes from reading the works of Donald J. Wheeler, Walter Shewhart and W. Edwards Deming. With respect to Wheeler, he has written some excellent books on measurement. Three to comment on are:

- **'Understanding Variation – the key to managing chaos':** this book is cited by many other authors and blog writers. It's relatively short, easy to read and gets the points across very well. An excellent starting place.

- **'Twenty Things you need to know'** is written as a technical sister-book to partner alongside 'Understanding Variation' i.e. if you want to 'get the point' then the first book is excellent, but if you want to actually 'do it' (i.e. control charting) and do so properly then this technical book is excellent, though harder going.

- **'Making Sense of Data – SPC for the Service Sector'** is a bigger, more involved book that could be bought instead of the first two. It has been written from a 'service' perspective, which is rather useful.

All Wheeler quotes used in this guide come from these three books.

b) Origin of the scaling factors: Wheeler writes that these scaling factors:

> *"were created for the convenience of the user at a time when all calculations were done by hand [i.e. before calculators] – and are practical constants of great generality which have a fine ancestry of highbrow mathematical theorems (and very complicated formulas)."*
>
> (Wheeler refers to a 1967 piece of work by the statistician Irving Burr)

The scaling factors are *"computational shortcuts"* that combine two steps:

- Calculating sigma (i.e. a measure of dispersion) using 'bias correction factors' for two-point moving ranges; and then
- Multiplying by 3 to obtain three-sigma

[19] See footnote a) for a discussion of these.