PRACTICAL MEASUREMENT SYSTEMS ANALYSIS

#1
LAYING THE FOUNDATIONS

PREFACE

This eBook is the first volume of a series of future publications on Practical Measurement Systems Analysis and is a collection of thoughts and concepts I have learned, applied and developed over the years.

Measurement Systems Analysis (MSA) is a subject I have always gravitated towards, and having had the opportunity to learn both from great teachers and develop my skills through self-study and exploration throughout my career, I want to pay it forward; I have set out to educate fellow professionals and provide them a practical approach to Measurement Systems Analysis and valuable tools to work with.

This particular eBook focuses on explaining basic concepts of MSA, sheds some light on traditional MSA methodologies and some of their drawbacks, introduces the reader to a practical approach to MSA and explains in great detail a few practical tools. As a gift, the practical tools described in this eBook are free to download and use.

I hope you will enjoy this eBook!

Gabor Szabo
Practical MSA

BASIC MSA PRINCIPLES

MEASUREMENT ERROR

Measurement error is present everywhere. There is not a single measurement that does not contain some amount of variation in it.

The typical scenario goes like this: You measure something, and you get a value. You measure it a second time, and you get a different value. Sound confusing? It happens all the time.

How is one supposed to trust their measurements if they can't even get the same measurement twice?

Why does one need to trust the measurements in the first place? Because important decisions – decisions that can have an impact on safety, quality, efficiency, effectiveness or other important aspects – could rely on measurements.

And that is exactly why understanding measurement variation is of great importance.

THE GOAL OF MEASUREMENT SYSTEMS ANALYSIS

Measurement Systems Analysis activities aim to *identify*, *assess* and *quantify* measurement error, and ultimately *unconfound* it from the variation of what is being measured (whether it is a product or process characteristic or any other characteristic of interest).

These are the type of questions Measurement Systems Analysis deals with:

How much measurement error is there?

How much measurement error matters?

How much is too much?

What kinds of measurement error are there?

What happens when you measure the same thing multiple times?

By different people or different labs?

Anything that is measured will have some amount of measurement error associated with it. This measurement error can come from different sources, which I describe in more detail later in this eBook.

As I mentioned earlier, one of the goals of MSA is to characterize and quantify this measurement error and help practitioners learn and make an informed decision on the suitability of the measurement system.

How is all this achieved, you ask? Before we get down to the nitty-gritty, let's talk basic principles of Measurement Systems Analysis

ELEMENTS OF A MEASUREMENT SYSTEM

The elements of a measurement system go well beyond the gage itself. In reality, a measurement system is a set of physical and non-physical components, as follows:

GAGE	This is the apparatus (device) that takes the measurements.
THE THINGS BEING MEASURED	These are the objects/units of interest. This includes the specific characteristic(s) we are interested in measuring
SPECIFIC MEASUREMENT TECHNIQUE	This is the specific measurement method/technique, which includes description of steps and their sequence
FIXTURES	Any type of locating or holding fixtures that come in contact with the units being measured
OPERATORS	The personnel that perform the measurement
ENVIRONMENT	Temperature, pressure, brightness and other environmental conditions

Each of these elements is an important part of the system and can be a source of measurement error.

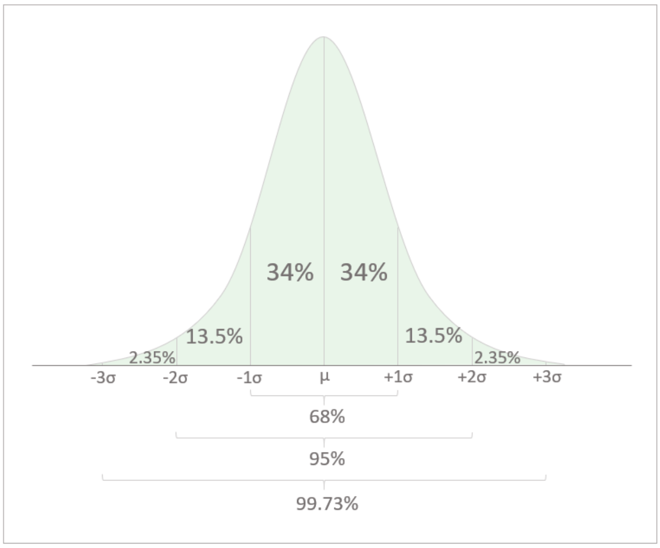
SOURCES OF MEASUREMENT ERROR

There are multiple sources from which measurement error can originate. One should understand what these sources are.

CONSISTENCY

Consistency is a vital aspect of a measurement system. A consistent measurement system means that measurements made with the system are steady over time. This is an important aspect because only consistent measurements are *predictable*.

The underlying principle for consistency is the normal distribution and its associated pillars, such as the *68-95-99.7 rule* and Laplace's *Limit Theorem*, which are not only important considerations for the consistency of the measurement system itself, but also play an important role in specific measurement systems analysis methods.



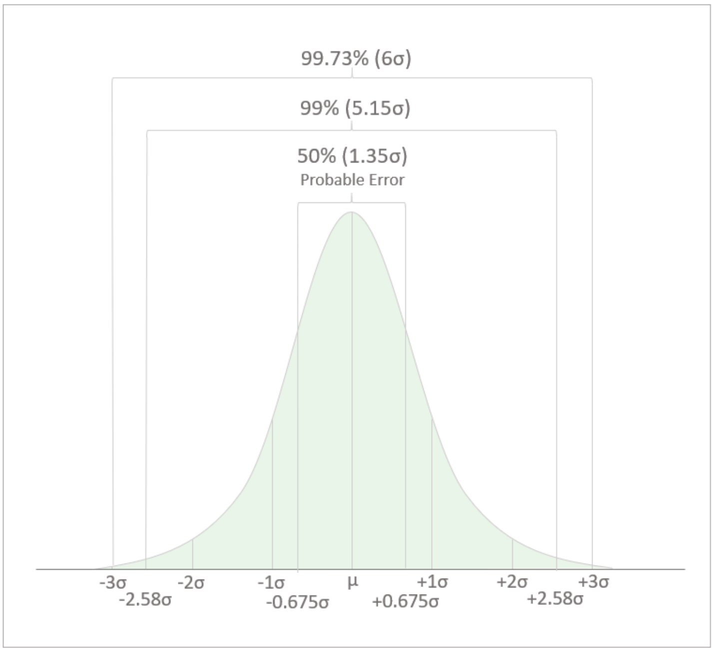
The 69-95-99.73 rule

Elements of consistency:

REPEATABILITY

Random measurement error. Assessed through measuring the same thing(s) multiple times under the same conditions (the same gage, the same operator etc.).

Repeatability error is random error, and as such, it is best modeled by the normal distribution (since Laplace's Limit Theorem dating all the way back to 1810). Repeatability error is quantified as a *standard deviation*, and can be characterized by a confidence interval of a specific width around the average of the repeated measurements.



Confidence intervals for repeatability error

REPRODUCIBILITY

Measurement error related to how well measurements from the same thing under different conditions agree with each other, and as such, assessed through measuring the same thing(s) under different conditions. Now, different conditions can be different operators, different gages, different fixtures etc. The most common interpretation of reproducibility error is *different operators;* that is what most folks think of when they hear the term reproducibility error.

The one thing that is important to note here is that reproducibility error, in the context of consistency (random error), is driven by repeatability error. Unless there is a significant amount of bias present (operator bias, method bias etc.), the best way to reduce reproducibility error is to reduce repeatability error. I discuss this more in detail in later sections of this eBook.

STABILITY

Consistency over time. Assessed through measuring the same thing(s) over time under the same conditions (the same gage, the same operator etc.).

CONSISTENCY OF REPEATABILITY ERROR

AMONG OPERATORS

Since different operators can measure the same objects differently, that can result in the repeatability error varying operator to operator.

OVER THE RANGE OF MEASUREMENTS

Also called **CV EFFECT** or **LINEARITY OF REPEATABILITY ERROR**

Repeatability error can vary over the range of the units being measured, and this is assessed by comparing repeatability error between across the expected range of measurements. Although this is not something that happens very often, there are certain scenarios where non-linear repeatability error could pose an issue.

BIAS

Bias is also an important consideration. The definition of bias, as it pertains to measurement systems analysis, is the systemic difference between the expected (average) value of results and the true value, or the systemic difference between the expected (average) values of certain elements of a measurement system, such as operators or gages.

Elements of bias:

BIAS - GENERIC The systemic difference between the expected (average) value of multiple measurements of an object and the true value of the object.

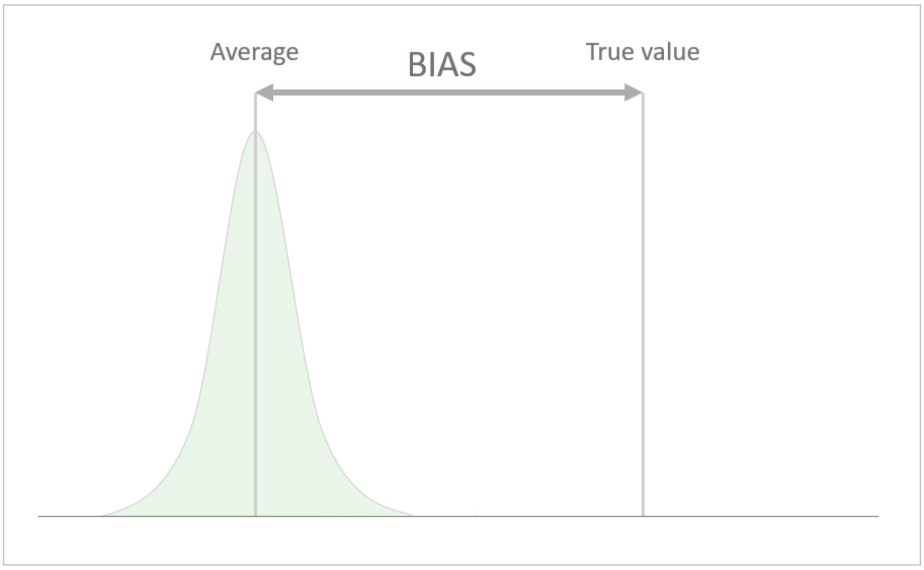


Illustration of bias

OPERATOR BIAS The systemic difference between average measurements of different operators.

METHOD BIAS The systemic difference between average measurements of different measurement methods, or the same measurement method in different physical locations, pertaining to a specific measurement application.

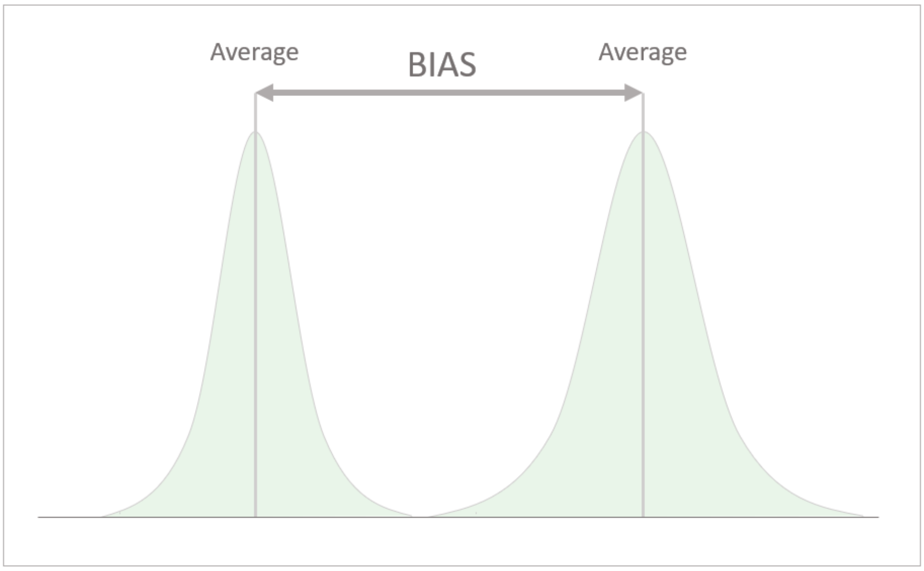


Illustration of operator bias or method bias

MEASUREMENT RESOLUTION

Measurement resolution is yet another important aspect of measurement systems analysis and a potential source of measurement error. The basic concept is that certain sensitivity is required for measurement data to appear continuous and be of actual utility.

If the measurement units to which the measurements are recorded are inadequate, you have a measurement resolution issue at hand. Inadequate measurement resolution can cause excess variation or no variation at all (when there is actually variation present). There are methods for detecting inadequate measurement resolution, some of which I discuss in later sections of this eBook.

CONVENTIONAL APPROACHES

AIAG GAGE R&R - THE STANDARD METHOD

The most common method associated with quantifying repeatability and reproducibility error is Gage R&R, or Gage Repeatability and Reproducibility studies. Gage R&R is an approach that most everyone in engineering and scientific fields has either heard of or applied themselves.

In simple terms, Gage R&R studies are *designed experiments*, in which multiple units (typically 5-10) are measured multiple times (typically 2-3) by multiple appraisers (typically 2-3) with the goal to quantify the amount of repeatability error and reproducibility error for a specific measurement application.

Gage R&R as a specific method was first developed and used in the automotive industry[1] decades ago and eventually became the standard, “the one and only” method for evaluating measurement error.

Most everyone knows and uses it…so why try and come up with a “practical approach” when there is already an approach to do just that, and it is widely used and accepted in industry?

The answer is manifold, but it boils down to these two things:

- The widely accepted conventional Gage R&R method[1] has many fallacies associated with it, which makes a lot of its elements not very practical.

[1] AIAG Measurement Systems Analysis Reference Manual, 4th Edition (2010)

- We, humans, tend to look for easy yes/no answers. We expect metrics and summary statistics to tell us how we are doing - "are we winning or losing?"

I strongly believe a practical approach to measurement systems analysis is very important, and I explain aspects of a not-so-practical approach and eventually more practical approaches in the following sections.

METRICS-BASED THINKING VS. CRITICAL THINKING

Don't let metrics get in the way of critical thinking

We all love a good metric, don't we? They tell us if we are "winning" or "losing", or if we have achieved a pre-determined goal or target. Granted, metrics can be useful for planning and determining where we are or where we need to be, but the real world is more complex than solely relying on pre-determined metrics. Metrics are simple to understand, follow and communicate, but without context as to how they came to exist and what they really mean, they can be dangerous when followed blindly, and one could jump to conclusions without understanding what is really happening.

Ask yourself these questions before you decide to blindly try and follow or meet a metric:

How was the metric determined?
If it is a mathematical calculation of some sort, what is the formula?
How is it calculated?
Does the metric even make sense or apply to the situation at hand?
What happens if you fail said metric? What happens if you meet it?

So, what are conventional Gage R&R metrics? When is a measurement system deemed "acceptable" by those conventional metrics?

Below are the conventional Gage R&R metrics, or summary statistics, widely used and accepted in industry (AIAG MSA Reference Manual), for continuous characteristics:

%TOLERANCE (OR P/T RATIO)

This metric is the ratio of the total observed measurement error (both repeatability and reproducibility) and the tolerance band (Upper Specification Limit minus Lower Specification Limit).

%STUDY VARIATION

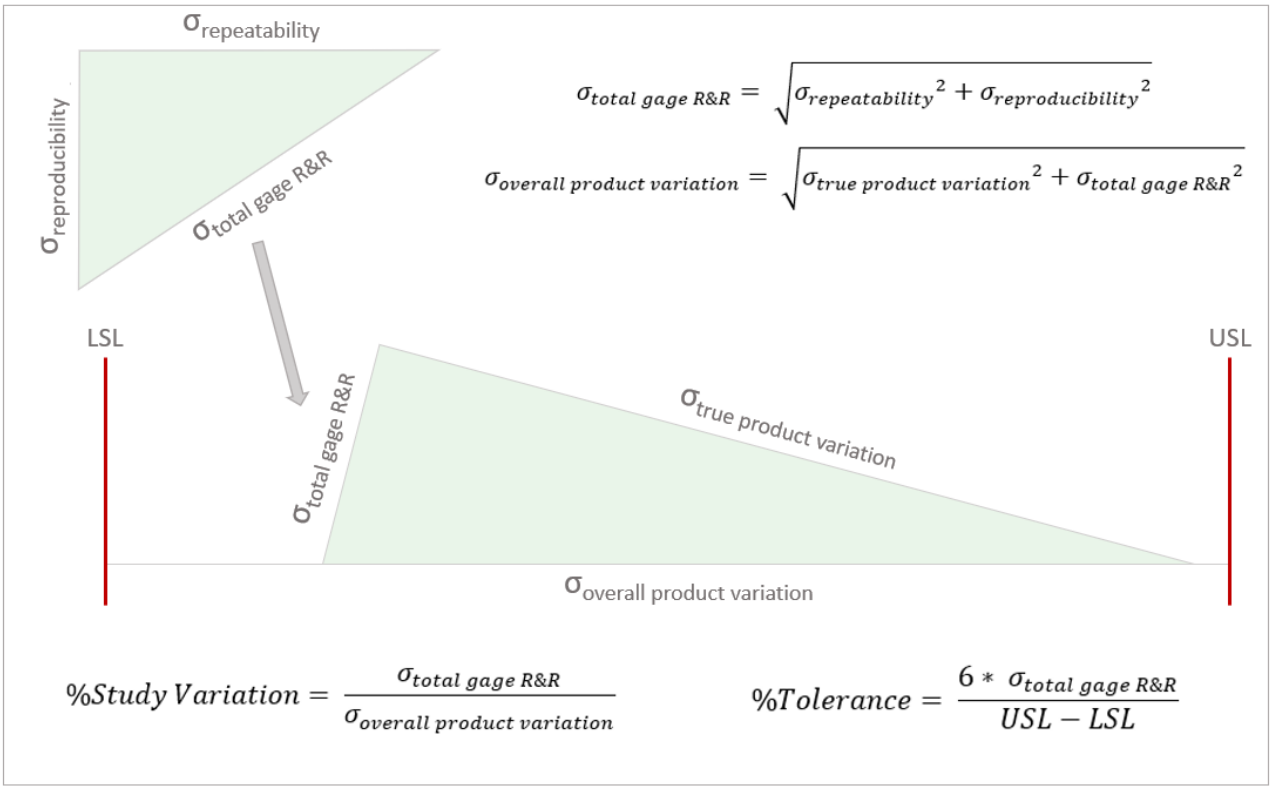
This metric is also a ratio, that of the total observed measurement error (both repeatability and reproducibility) and the observed product variation. The idea is to compare measurement error to the variation of whatever is being measured. While this makes perfect sense, there are some serious fallacies surrounding this metric, which makes its credibility and utility questionable.

NUMBER OF DISTINCT CATEGORIES (NDC)

This metric represents the *number of distinct product categories that can be reliably distinguished by the measurement system.*[1] The NDC metric essentially tells us the same thing as the %Study Variation metric in a different way, in that both the total observed measurement error and the observed product variation are included in both formulas.

Now let's take a look at how the %Study Variation and %Tolerance metrics are calculated. The guiding principle here is the Pythagorean theorem

[1] AIAG Measurement Systems Analysis Reference Manual, 4th Edition (2010)



Calculation of %Study Variation and %Tolerance metrics per Pythagorean theorem

And below are the general acceptability guidelines[1] for both the %Tolerance and %Study Variation metrics:

Acceptability guidelines for %Tolerance and %Study Variation metrics	*Definition of acceptability*	*Details of acceptability*
Under 10%	Generally considered to be an acceptable measurement system.	Recommended, especially useful when trying to sort or classify parts or when tightened process control is required.
Between 10% and 30%	May be acceptable for some applications	Decision should be based upon, for example, importance of application measurement, cost of measurement device, cost of rework or repair. Should be approved by the customer.
Over 30%	Considered to be unacceptable	Every effort should be made to improve the measurement system. This condition may be addressed by the use of an appropriate measurement strategy; for example, using the average result of several readings of the same part characteristic in order to reduce final measurement variation.

What is the problem with these guidelines? There are many, but let's start with the most obvious one: They are not actionable. In other words, where you land on these scales doesn't necessarily determine the acceptability or utility of your measurement system.

[1] AIAG Measurement Systems Analysis Reference Manual, 4th Edition (2010)

Let me elaborate on this and other fallacies of the conventional Gage R&R metrics:

- Since these guideline acceptance criteria were arbitrarily derived, there isn't any empirical, practical evidence to support them. Sure, one could say "the less measurement error you have, the better", and that statement generally holds true, but blindly following arbitrary guidelines and basing the suitability of a measurement system on those guidelines is very far from practical.

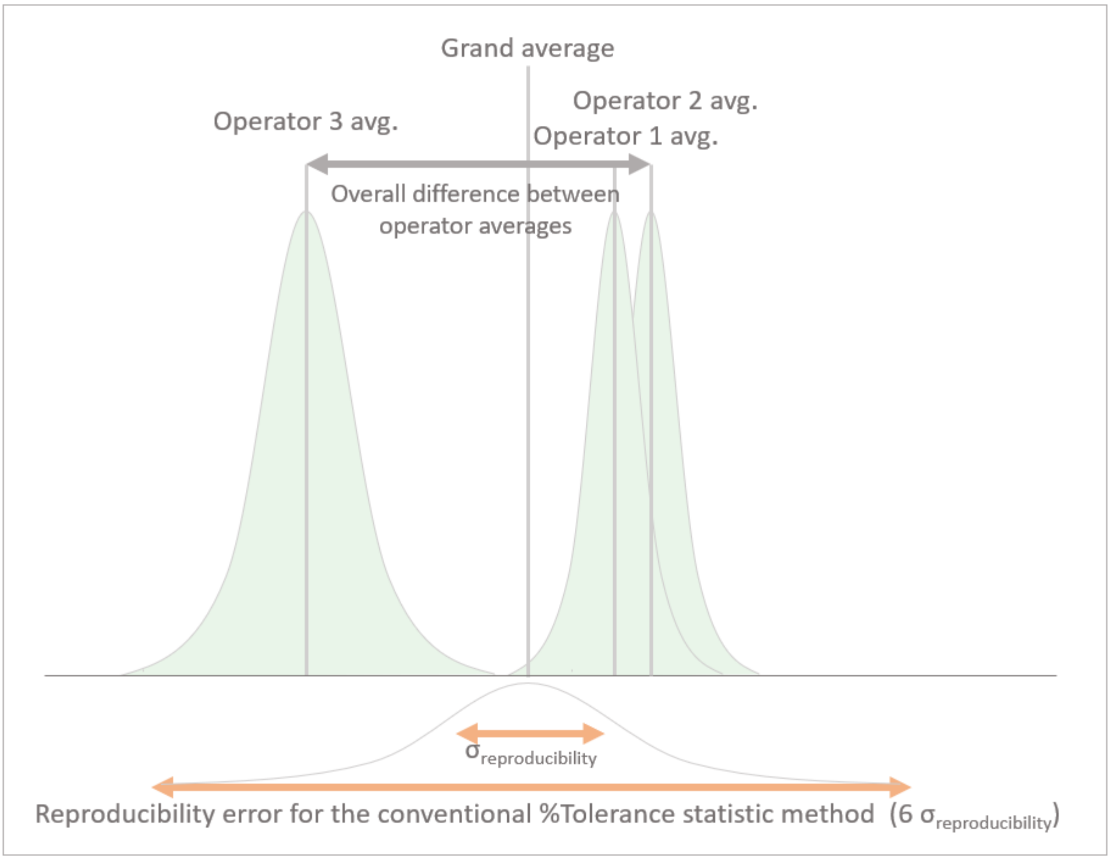
- The estimation error of these summary statistics can be significant. Let me say it one more time so it sinks in: *significant*. What that means is that if, say, the %Study Variation summary statistic for your Gage R&R study puts you in the "May be acceptable for some applications" bucket (between 10% and 30%), you could, in reality, very well be in any of the three buckets, that is either in the *"Generally considered an acceptable measurement system"* or in the *"May be acceptable for some applications"* or in the *"Considered to be unacceptable"* bucket. This is because the sample sizes recommended by the AIAG MSA Reference Manual – 5-10 parts measured 2-3 times by 2-3 operators – render these summary statistics, especially the %Study Variation statistic, highly questionable.

- If the product tolerance does not reflect actual product function, the %Tolerance summary statistic is useless.

- The guideline acceptance criteria for the %Tolerance and %Study Variation summary statistics are the same: 10-20-30%, which doesn't make any practical sense.

- Reproducibility error in the conventional Gage R&R method is calculated as a standard deviation of the operator averages. Since actual, noticeable differences between operators are more often than not *systemic* differences (if one operator is doing something different than the rest of the operators, that noticeable difference will always be there), what we are really talking about is *bias* rather than random error, which the calculation method suggests. Needless to say, calculating a standard deviation from bias doesn't make too much sense.

  Let me illustrate this to you through an example: The below figure shows averages for three operators. Apparently, operator 3 is averaging considerably lower than the rest of the operators (not to mention the spread of his measurements being larger than that of the other operators). There is a clearly noticeable operator bias here. Since this difference would show up as statistically significant in the conventional calculation method (ANOVA Gage R&R), a standard



Difference between operator averages vs. reproducibility error as a standard devitation deviation would be calculated from the three operator averages, and the standard deviation would be reported as $\sigma_{reproducibility}$.

At this point I need to add that I am not the first person to take note of these things; in fact, several quality gurus, statisticians (for example Don Wheeler) and other practitioners, whom I highly esteem, have written about this same issue, but I felt the need to elaborate on it in this eBook to set the tone for the next phases of the discussion.

All in all, meeting these recommended guideline metrics *should not be used* as the determining factor when making a decision, as is the case with these conventional Gage R&R guidelines a lot of the time.

These summary statistics – and any summary statistics in general – tend to hide information; in the case of MSA summary statistics, information about the measurement system being evaluated. When MSA activities are boiled down to one summary statistic and compared against guidelines or acceptance criteria, one is essentially making a decision based on a single number that they may not even fully understand.

Metric-based thinking, or I should say, solely relying on metrics hinders our ability to see what is really going on in the activity of interest.

Another example of a metric widely, and wildly(!), used by practitioners is p-values.

What are p-values? They are a commonly accepted metric for statistical significance. Note the word statistical… nowhere does it say practical. Yet, important, practical business decisions are often based on a single p-value.

In 2016 the American Statistical Association (ASA) released a statement on Statistical Significance and p-values, in which the ASA determines that "The p-value was never intended to be a substitute for scientific reasoning. Well-reasoned statistical arguments contain much more than the value of a single number and whether that number exceeds an arbitrary threshold."

Below are the six guiding principles established in the paper:

1. P-values can indicate how incompatible the data are with a specified statistical model.
2. P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
3. Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.
4. Proper inference requires full reporting and transparency.
5. A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.
6. By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis

Those of you interested in reading the full statement, click here.

It is also worth mentioning that p-values play an important role in the widely-used ANOVA Gage R&R calculation method, which, again, puts statistical significance above practical importance.

Deming coined two terms: *enumerative studies* and *analytic studies*.[2] They are terms for studies that are focused on judgement of results (*enumerative*), and studies that are focused on assessing the behavior of a system/process that created the results being evaluated, with the ultimate goal of improving it (*analytic*).

Based on Deming's definition, measurement system analysis activities are combination of enumerative and analytic studies; we interrogate the measurement process so that we get actionable information that we can use to improve the system itself. Of course, sometimes what you will find after you do the analytic part of the study is that the measurement system is good to use as is, so you end up making a judgement on it, which is an attribute of enumerative studies, but you did go through the analytic part!

My point is that in order to make a judgement on measurement error and the measurement system itself, it is necessary to carry out an analytic study.

[2] W. Edwards Deming: Some Theory of Sampling, Chapter 7, 1950

A PRACTICAL APPROACH

Now that we have discussed metric-based thinking, its major drawbacks and the need for measurement systems analysis to be a truly analytic study, let's talk about an actual approach to how to do just that.

What is the practitioner to do?

The missing piece here is critical thinking coupled with practical study design, planning and analysis of data, so let's begin to discuss what the foundations of a practical approach look like.

In order for one to evaluate, let alone judge, a measurement system not relying solely on summary statistics, one needs to start the planning and analysis at the most essential level and move up from there.

Here are the three levels of analysis:

CALCULATE METRICS

VISUALIZE YOUR DATA

APPLY COMMON SENSE

These three levels follow each other sequentially, which means that only when you are done at the lowest level of analysis should you proceed to the next level. With that said, the sequence to follow is first apply common sense to planning the assessment, then visualize your data to gain insight and learn about process behavior, and finally calculate metrics to complement findings.

Let's see what this looks like in practice!

APPLY COMMON SENSE

com·mon sense
/ˌkämən ˈsens/
noun
1. good sense and sound judgment in practical matters.

This is the first and most fundamental step. Applying common sense, in this context, means asking intelligent, practical questions that are relevant to the measurement system being evaluated.

The following are typical *practical* questions one should be asking planning or going into the study:

- What is the intended use of the measurement? Wheeler[3] puts these into four categories: describe, characterize, represent, predict.

  DESCRIBE Identification of what how much, how many. Simply measuring something.

  CHARACTERIZE Assessing something against specific criteria or a specification.

  REPRESENT Characterizing things *in the past* against specific criteria or a specification and making a determination on them.
  Example: batch acceptance sampling (receiving inspection).

  PREDICT Characterizing things *in the future*.
  Example: process monitoring using SPC.

- How big of a difference do you need to be able to detect? Think of it as “Signal to Noise”. In this context, the signal is the difference you need to be able to detect, and the noise is the measurement error. Knowing this can be useful in experimental studies where one is trying to assess whether the observed difference between treatments is real.

3 Donald J. Wheeler: EMP III Evaluating The Measurement Process & Using Imperfect Data, SPC Press, 2006

- What are your specification limits? Do they reflect actual product function?
- Do you have information on the spread and the position of the product variation (where the process is running)?
- Being knowledgeable of the measurement process and the physics of the measurement, do you expect stability, repeatability error, differences between operators or any other sources of variability to be an issue? Having a good idea of what to expect can point you to the right study design.
- What is the ideal study design for the situation at hand? This depends on the purpose of the study. In other words, what questions is the study trying to answer?

Below are typical questions one might consider asking after performing the study:

- Is there anything about the measurement data that stands out? Clues such as significant differences between operators or other factors being assessed, or significant repeatability error could be apparent just by looking at the measurement data itself.
- Is there anything that you notice right off the bat just by looking at the data?
  For example, if you can tell there is a noticeable difference between operators just by looking at the measurement data, that could be a telltale sign of a large amount of operator bias or other operator effects at play. A lot of the time, those differences are not significant enough to be noticeable. That is when data visualization comes into play.

VISUALIZE YOUR DATA

vis·u·al·ize
/ˈviZH(o͞o)əˌlīz/
verb
1. make (something) visible to the eye.

In this part of the analysis, you plot your data in different charts – I discuss some of these charts in later chapters – with the goal of making observations on your data. Let me tell you, this part of the analysis is *money*; most observations on the behavior of the measurement system are typically made through visualizing your data, that is looking at it graphically, and not through metrics.

Considerations:

- Are there any clues (shifts, drifts, patterns) that stand out?
- How much random measurement error is observed?
- Do the data appear stable, or there are special causes present?
- Is the amount of random measurement error constant, or it changes over the range of results?
- Does the measurement resolution appear adequate?
- In general, is there anything about the data that graphically stands out?
- Is there operator bias present?
- Is the repeatability error consistent between operators?

There are a variety of graphical tools for this part of the analysis. What is important is that you use the appropriate tool. Among other graphical tools, the Multi-Vari Chart and the Youden Plot are great tools for graphical analysis. I discuss the Multi-Vari Chart in detail later in this book and also provide a free version of the tool on my website practicalmsa.com.

CALCULATE METRICS

met·rics
/ˈmetriks/
noun
1. a method of measuring something, or the results obtained from this.

This should be the last step of your analysis. Only when you have gone through the first two stages of analysis should you consider looking at metrics, summary statistics. In this stage, you calculate metrics or summary statistics to quantify the amount of measurement error and, in some cases, to determine if the observed measurement error is significant as to whether or not to take action.

Consider what summary statistic(s) you choose.

In addition to the conventional Gage R&R summary statistics of %Tolerance, %Study Variation and NDC, which often hide information, there are other summary statistics that often offer more insight; I discuss some of them in detail later in this eBook and in future publications.

Now, let's talk about what kind of tools are available for you to use for practical measurement analysis!

TOOLS FOR PRACTICAL MSA

As you already know, evaluating your measurement system involves designed experiments to assess the behavior of the measurement system, so coming up with the appropriate experimental design that pertains to the problem at hand is very important.

Like I said earlier, applying common sense (in other words, good engineering) and asking those practical questions will get you closer to what questions you want your study to answer. Once you know what questions you want answered, you need to come up with the right experimental design.

TREE DIAGRAMS

Tree Diagrams are a great tool to help you visualize the study design you have in mind; it gives you the ability to plan out the study in a way that is easy to follow, even by others (yes, at times you will need to be able to communicate to others what the study consisted of, and Tree Diagrams, among other tools, are a tool to do just that).

Typically, a study involves what is called multiple *study factors (or study levels)*.

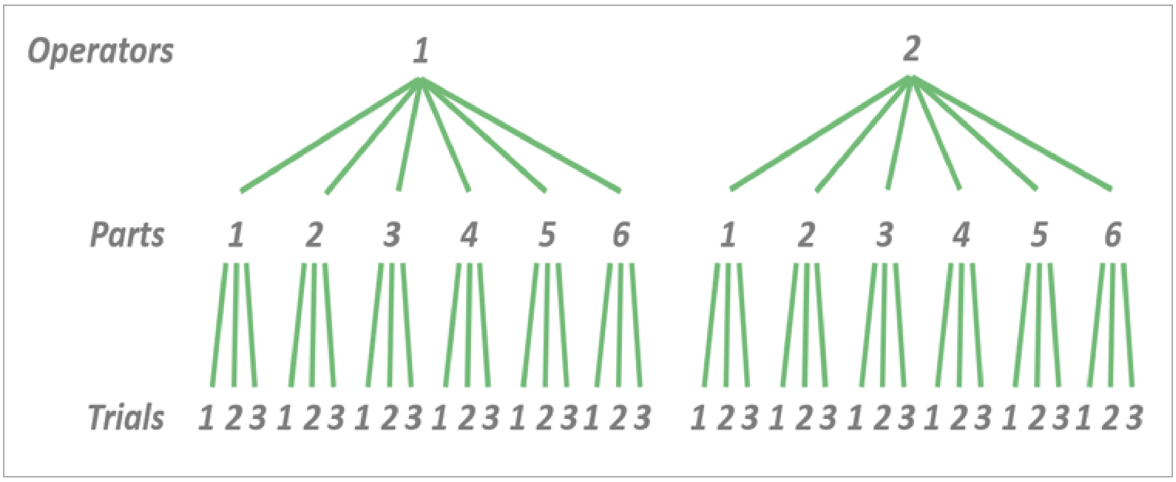
Let me walk you through you an example:

You decide to assess a measurement system that measures, say, the length of a plastic component.

The gage itself is an optical measurement system operated by inspection personnel. You want to assess repeatability error, and also want to see if there are any differences between operators. This study sure sounds simple, right? Well, this is in fact a simple study design, but I want to use this example to show you how to make a Tree Diagram.

1. You plan to include two operators in the study to assess whether differences exist between them. This will be a factor in your study.
2. You also plan to include six parts in your study, each measured by both operators. This is the second factor in your study.
3. Last but not least, you will assess repeatability error through repeated measurements of the same parts. You plan to have each part measured three times by each operator. This is the third factor in your study.

Below is what the Tree Diagram for this study design looks like.



What is great about this tool is that you have all three factors (or study levels) – that is Operators, Parts and Trials – and their relationship shown here. You can also think of these factors as different levels of the study. At the lowest level you have Trials, one level up you have Parts, and at the top you have Operators.

Note that within Parts you have six experimental units (parts), and they are denoted 1-6 for both operators, which indicates that the same parts 1-6 are measured by both operators.

Note that at the Trials level, each set of trials is also denoted 1-3. This is because that factor of the design is *crossed*. Crossed just means that experimental units at of one factor co-occur or have something *in common* with units of another factor (study level). In the case of the Parts factor, this commonality means that parts 1-6 are *the exact same parts* measured by both operators. In the case of the Trials factor, the sequence 1-2-3 across all parts and operators (the higher study levels) suggests that there is a *logical relationship*, and that relationship is *time-related (chronological or temporal)*– the measurement that first occurs is always denoted 1, the second measurement 2, and the third 3. This specific design allows for gaining time-related information from your study, which I discuss in more detail in later sections of this eBook.

In a *nested* study design, experimental units of one factor are nested within another factor at a higher level, meaning they are *not identical*. The emphasis here is really on them being *not identical*. In the case of MSA, nested study designs are typical when the measurement process either destroys the parts (destuctive tests) or permanently alters them. What that means is that those commonalities at the Parts and Trials levels described in the previous section are not always possible to achieve or even assume, so nested studies require a different approach.

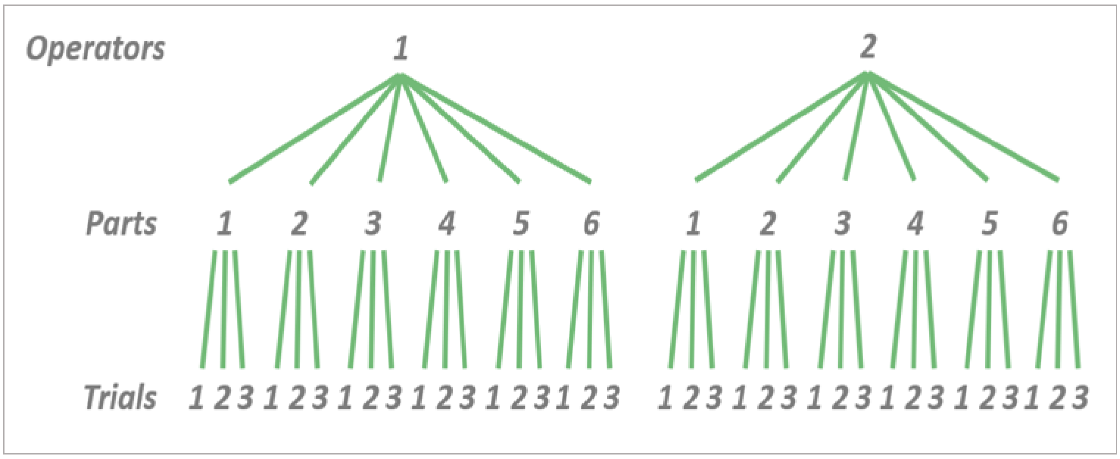
Now that you are familiar with how tree diagrams work and how to construct them, you should be able to start making use of them. You can just jot them down on a piece of paper to facilitate planning or communicating your activity, or if you want to go all out, you may also create them electronically.

THE MULTI-VARI CHART

Developed by Leonard Seder[4] back in the 1950's, Multi-Vari Charts are overall just a tool of great utility. They are excellent for graphical analysis of experimental data and generating clues from them.

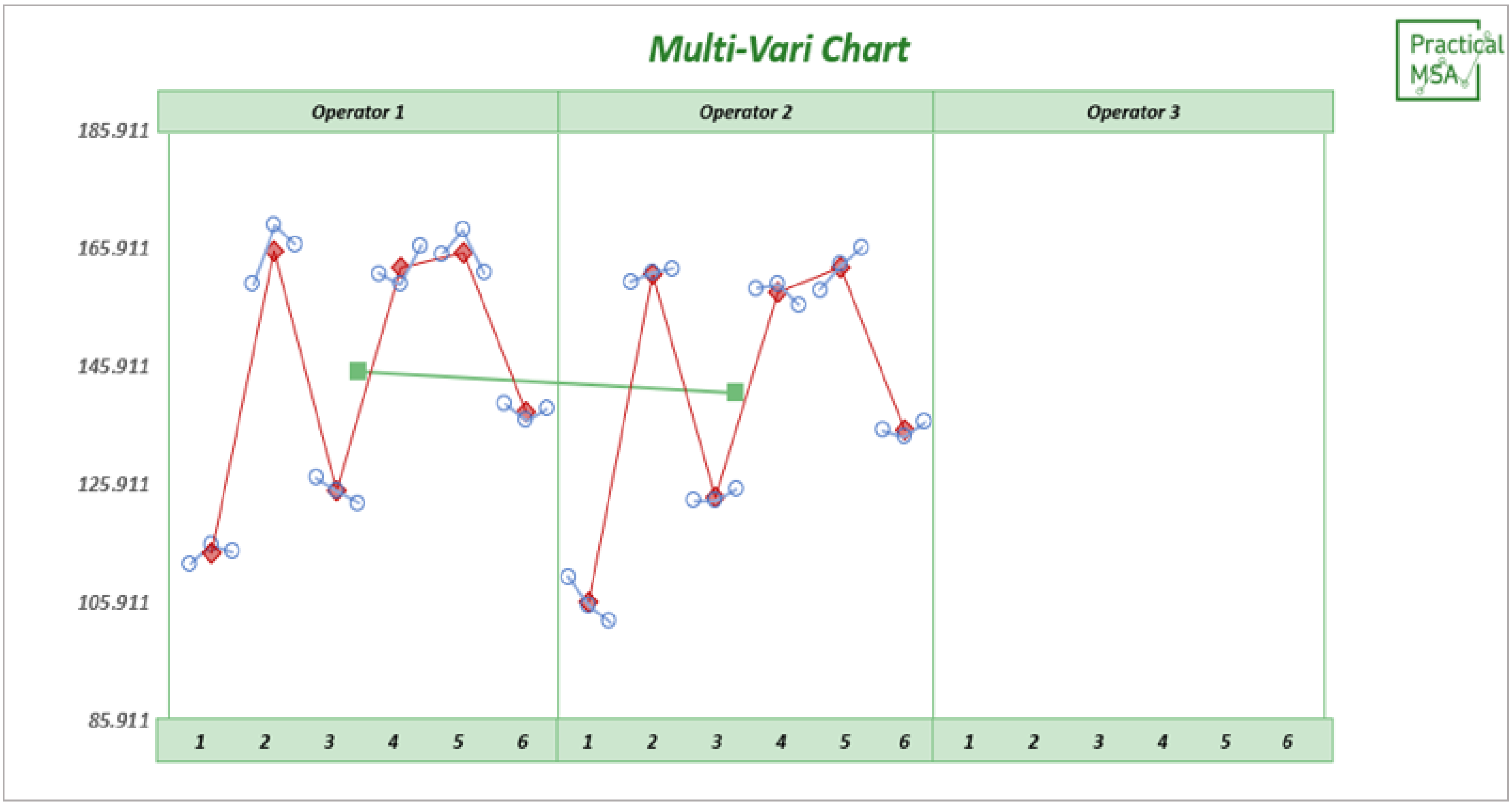
The good news is that I have developed a Multi-Vari Chart for you to use; click here to download it and start using it right away, or bear with me while I give you the details on how to interpret the chart.

In a Multi-Vari Chart, the factors from your experimental study are plotted in a way that facilitates graphical analysis and gaining valuable insight from your data. Every data point from your study along with averages are plotted in the chart. But before I go into detail on what that looks like, let's go back to the Tree Diagram we created for our scenario. We had 6 parts measured 3 times by 2 operators for a total of 6 x 3 x 2 = 36 measurements.



Now, let's see what this looks like when plotted on a Multi-Vari Chart.

[4] Leonard Seder: "Diagnosis with Diagrams – Part I", Industrial Quality Control (1950)



It looks like there are quite a few dots and lines on the chart, so I have some explaining to do here.

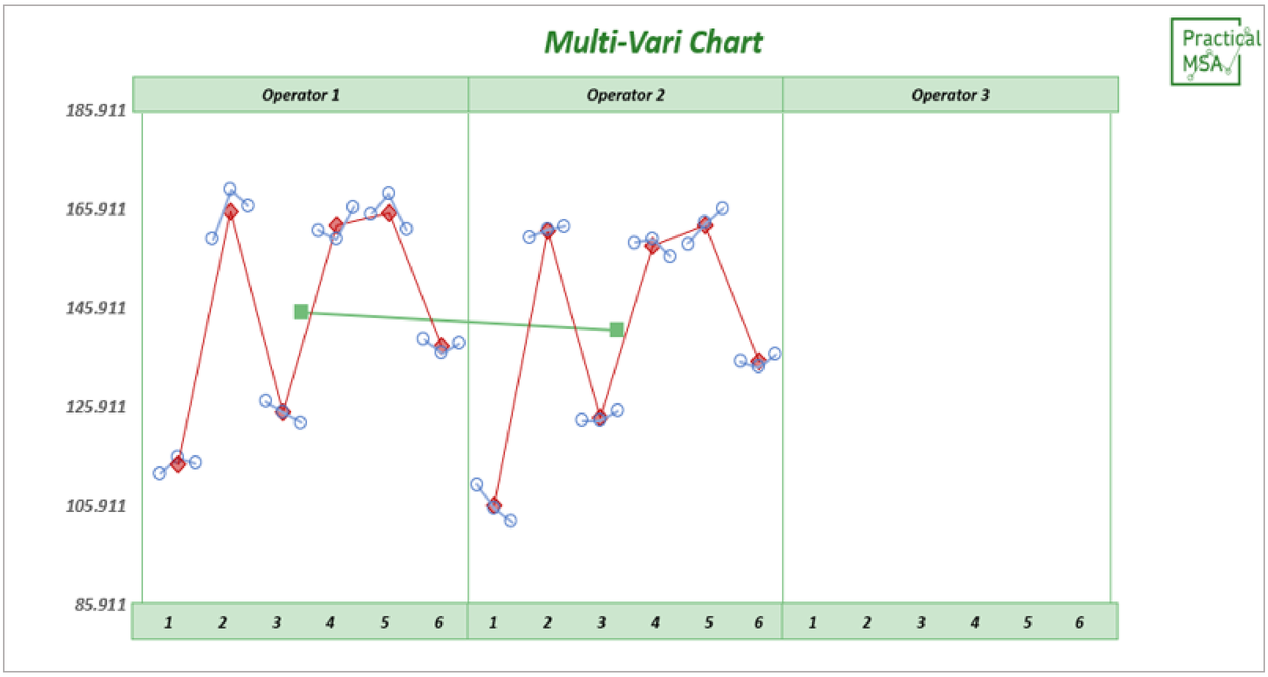
All 36 measurements are plotted on the charts as individual data points.

The chart is split into three *panels*, one for each operator (Operator 1, 2 and 3). Since only two operators were used in this study, the panel for Operator 3 contains no data points.

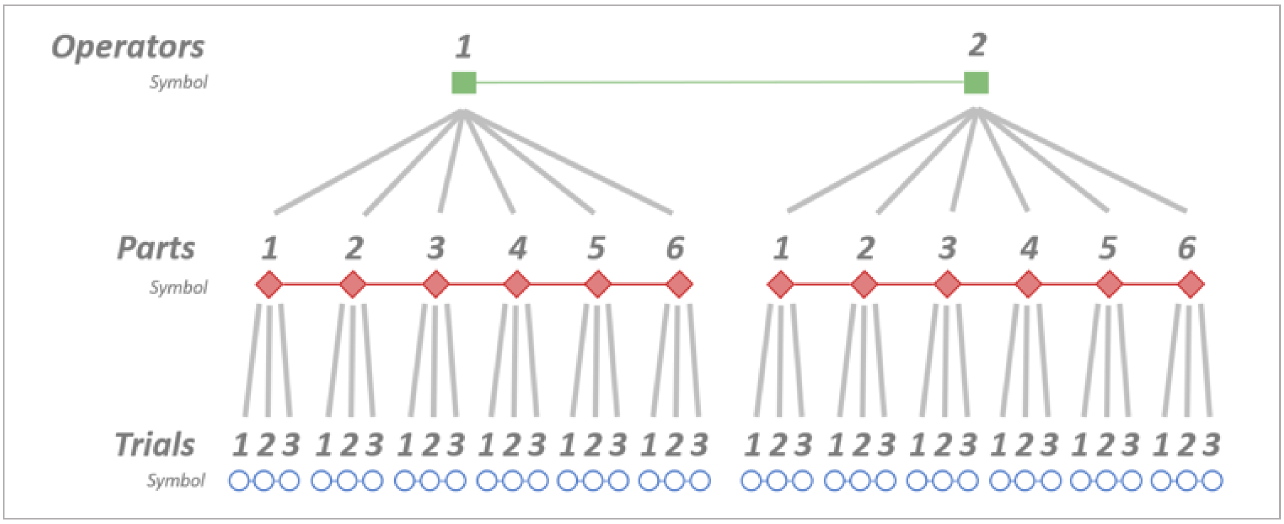
All of the data points are plotted in a way that enables you to look at and compare data points from the three levels of the study (Trial, Part, Operator). Think of it as: for any given part (experimental unit), the average of the measurements from the three trials results in the average for that specific part, and the average of all six parts results in the average for the operator.

Now, if this still doesn't quite add up, here's a step-by-step rundown of the elements of the chart for you.

- The measurements from the three trials for each part are represented by **blue** dots, and they are connected by a **blue** line.
- The average of the measurements from the three trials is represented by a **red** tilted square, which then denotes the average for that specific part.
- All six parts for each operator are connected by a **red** line, and their average is represented by a **green** square.
- The **green** squares denoting operator averages are connected by a **green** line.



See below for a Tree Diagram of the same study design with the corresponding elements elements of the Multi-Vari Chart shown:



So how is all this helpful?

The Multi-Vari Chart uncovers a lot of information on the behavior of the measurement system and gives you the ability to compare different sources of measurement error.

Let's start at the lowest level, the Trial level. By graphically assessing those three measurements for each part, you get an idea of the consistency of the measurements, that is the repeatability error present. Remember, what you are looking for is *random* error. What does random error look like? Well, it looks random! Thinking directionally, that means that you have a 50% chance of getting a measurement value above the average, and a 50% chance of getting one below the average. By this logic and knowing that the three data points on the chart follow a chronological order, you would expect the three time-ordered measurement results for each part to follow an *up-down-up* or *down-up-down* pattern, *at least the majority of the time.*

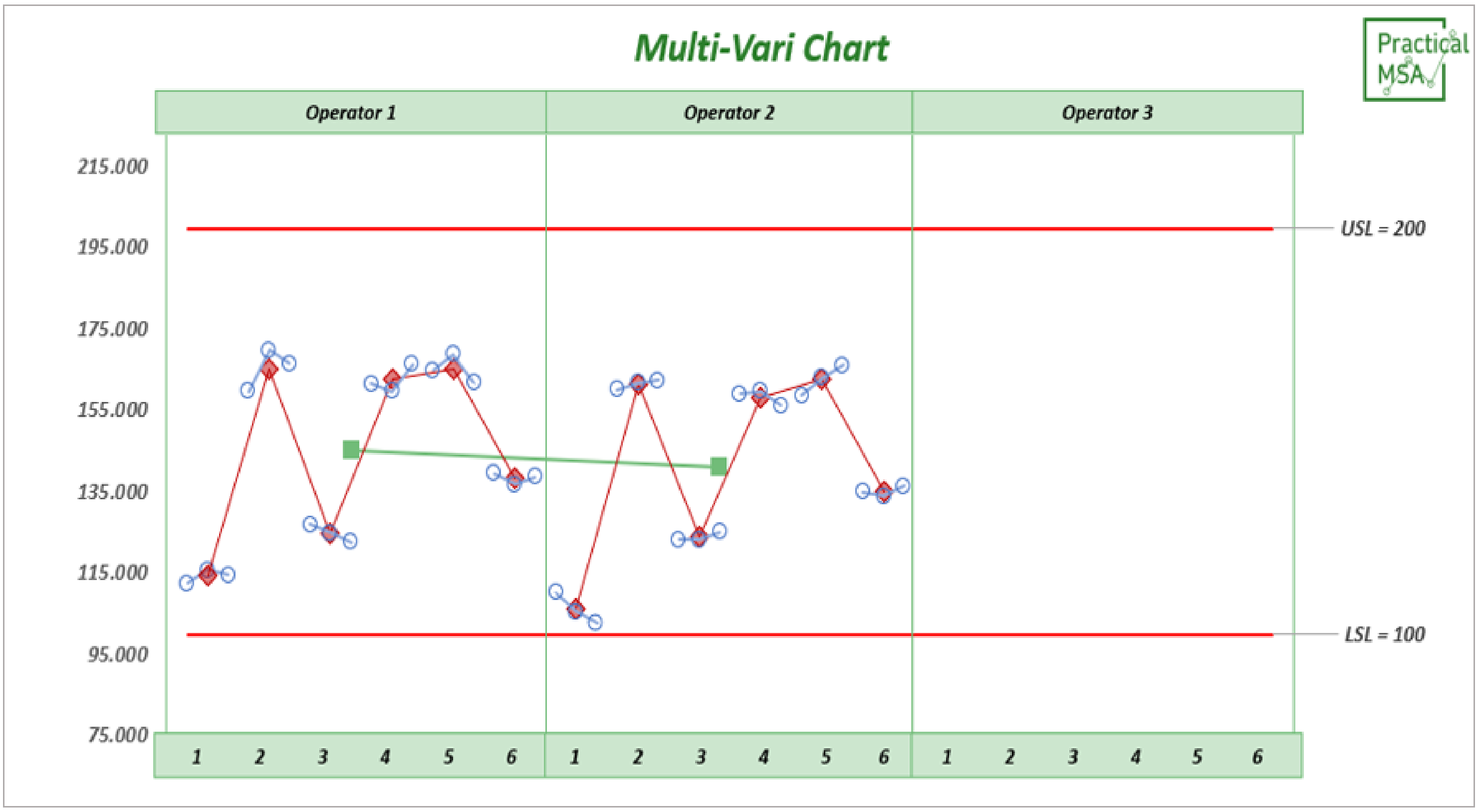
Take a look at the chart again; does it look like that the measurement error you see there is random? It looks like *the majority of the time* it is indeed random. Granted, there are a few exceptions such as part 3 for Operator 1 and part 1 for Operator 2, but random error is, again, random, so there is a chance that you will see measurements from three consecutive trials behave like that, even if the probability of that happening is significantly lower than an up-down-up or a down-up-down pattern. As long as there is not a consistent pattern where the three consecutive measurements are exhibiting a downward or upward trend, the assumption of random measurement error should hold true.

Now, what if your data is truly exhibiting a downward or upward trend? That could be indicative of the repeated measurements physically affecting the parts rendering the measurement non-repeatable or even destructive. Investigating and finding the causes could take some time and resources – a lot of times it really comes down to *the physics of the measurement* though –, but these are the kinds of practical discoveries one can make on the measurement process using the Multi-Vari Chart and other graphical tools and not by looking at summary statistics alone.

All in all, the Multi-Vari chart is an extremely useful tool for graphical analysis. Here is a list of potential uses for the chart:

ASSESS AND COMPARE REPEATABIITY ERROR

Graphically assess the amount of repeatability error and compare it to the range of the part averages (red tilted squares). If you enter the specification limits for your product (if such limits exist), you can compare the repeatability error to those limits as well. Below is an example of what that looks like on the Multi-Vari Chart (in this particular case, the specifications limits entered are 100 and 200).



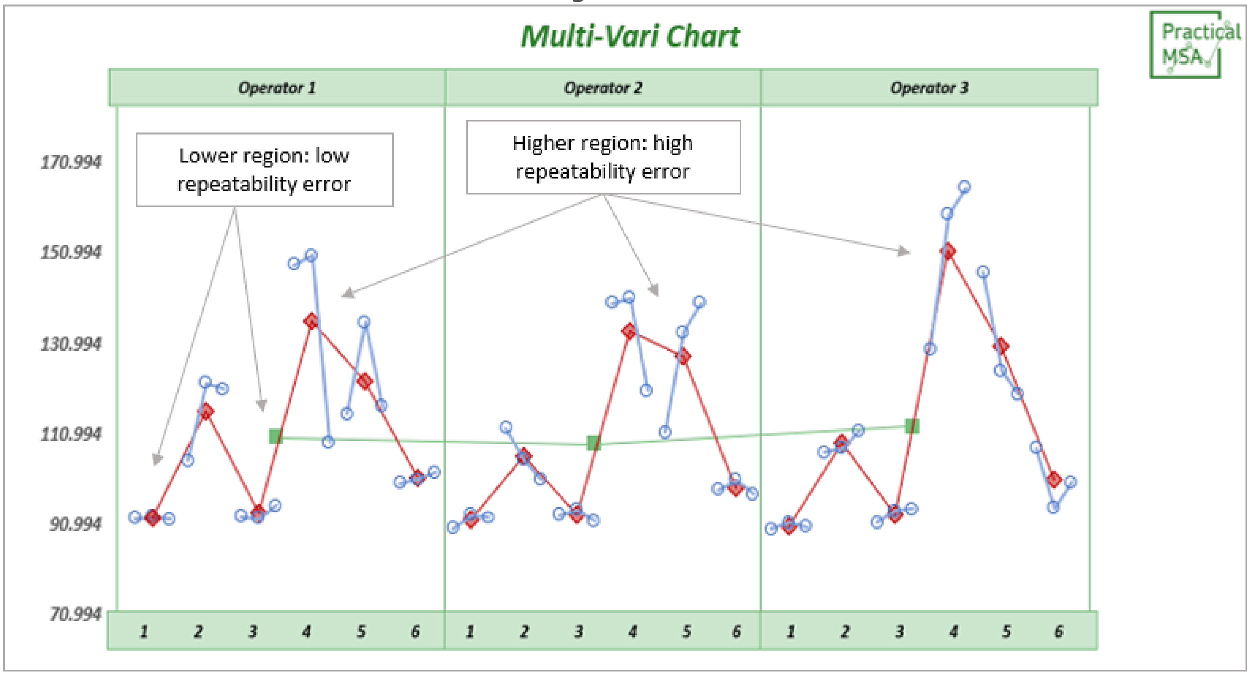
ASSESS OPERATOR BIAS

Graphically assess operator bias, that is the difference between operator averages (green squares). You want the green line connecting the operator averages to be horizontal; if there is a departure from that, that could mean operator bias (a systemic difference) is present. The free downloadable Multi-Vari tool automatically calculates the operator averages for you so you can compare them not only graphically but also through comparing the averages and considering practical significance.

ASSESS LINEARITY OF REPEATABILITY ERROR (CV EFFECT)

Graphically assess if the repeatability error changes over the range. In extreme, also fairly rare, cases, the repeatability error could grow exponentially towards either the lower or upper end of the range resulting in significantly more repeatability error in that region. Now, I can't say that non-linear repeatability error happens all the time, but I can say that the Multi-Vari Chart is a tool that can detect it.

Take a look at the below example where parts 4 and 5, which are averaging higher than the rest of the parts, see significantly more repeatability across all three operators than the rest of the parts in the lower region of the overall range.



ASSESS TRENDS AND PATTERNS

Like I mentioned earlier, the Multi-Vari Chart is just great at picking up patterns and trends thanks to how the data are structured.

OK, so how are the data structured?

BETWEEN-WITHIN COMPARISON

The Multi-Vari Chart is structured in a way that it compares variation coming from within and between the experimental units. Since the experimental units in the MSA activity described earlier in the book are the parts themselves, the following statements hold true:

1. The variation *within* an experimental unit is the variation contained in the three consecutive measurements, that is the repeatability error.

2. The variation *between* experimental units is then the variation seen between the parts, that is the product variation.

   A word of caution: If you were to compare the within variation with the between variation, since the typical samples sizes (i.e. the number of parts) involved in a conventional gage study render product variation estimates very inaccurate (pretty much useless), instead of calculating the product variation (as a standard deviation), you are better off just either looking at it practically (that's exactly what the Multi-Vari Chart is for!) or using a more accurate estimate, such as using historical product variation or utilizing a larger sample size to estimate product variation from your study; I will elaborate on some of these specific techniques in another book. In the meantime, remember: instead of looking at only summary statistics (%Study Variation or %Tolerance), which could lead you down the wrong path, assess behavior graphically using the Multi-Vari Chart.

THE RANGE CHART

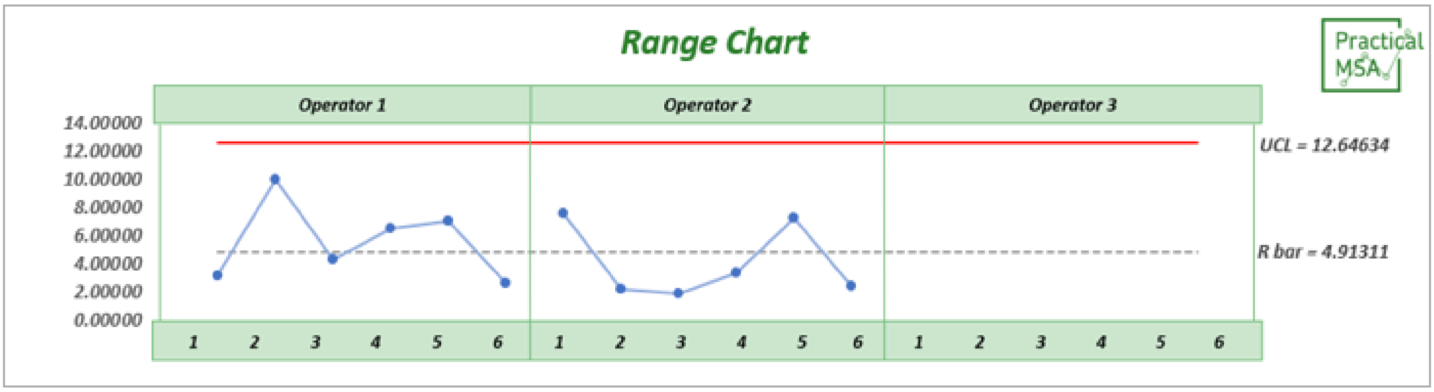
There is a tool that augments the Multi-Vari Chart, and that is the Range Chart, which is the R part of the good old Xbar-R process behavior chart invented by Walter Shewhart.[5]

In fact, the Multi-Vari and Range Charts complement each other so well that they are best used together, and that is why I have developed a Multi-Vari/Range Chart tool for you to use. If you would like to download the tool at this point, you can do so by clicking here.

Range Charts are used to estimate within-subgroup variation. In a Range Chart, the range of individual values within a subgroup is plotted for each subgroup. In the case of a gage study, subgroups are the equivalent of parts.

Along with subgroups ranges, the average range across all subgroups and the upper control limit are also plotted on the chart.

Let's take a look at what the Range Chart looks like for the gage study example we discussed earlier. As a refresher, the study involved 6 parts, 3 trials, 2 operators.



OK, let's see what's going on here.

5 W. A. Shewhart: Economic Control of Quality of Manufactured Product (1931)

Similarly to the Multi-Vari Chart, the Range Chart is staged by operator.

The blue dots represent the range for each subgroup (the range of measurements from three trials for each part). This is the consistency of measurements, i.e. repeatability error.

The dotted line represents the average range or *R bar*, that is the expected range across all subgroups. You can think of R bar as a measure of average repeatability error.

The red line represents the Upper Control Limit (UCL). The UCL is automatically calculated from the multiply of the R bar and the bias correction factor D4, which depends on the subgroup size, in the following fashion:

$$UCL = D_4 \times R\ bar$$

Subgroup size	D_4
2	3.268
3	2.574

This provides a pretty robust range-based three-sigma upper control limit that is remains unbiased for even for smaller sample sizes, such as the one in this example study (6 parts, 3 trials, 2 operators).

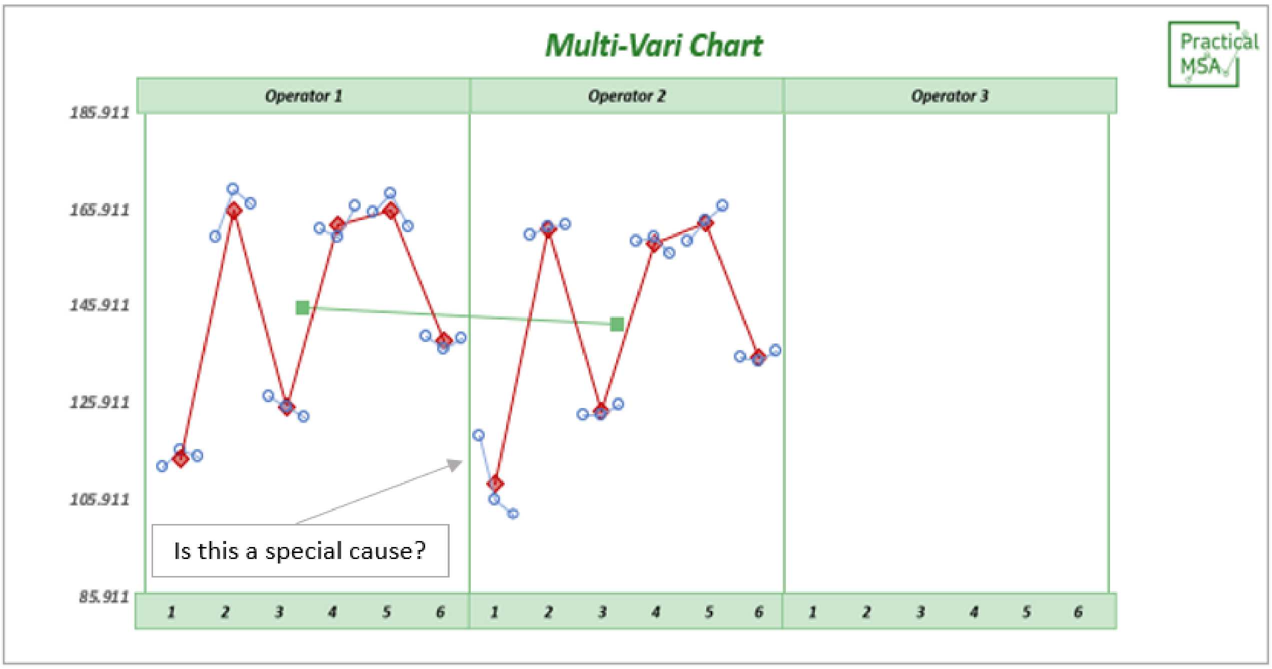
What do the three-sigma limit mean? Of course, they are tied to probabilities, but they are really practically-defined action limits[6]; if any individual subgroup ranges happen to land outside the three-sigma limit, that is a good indication of a special cause present.

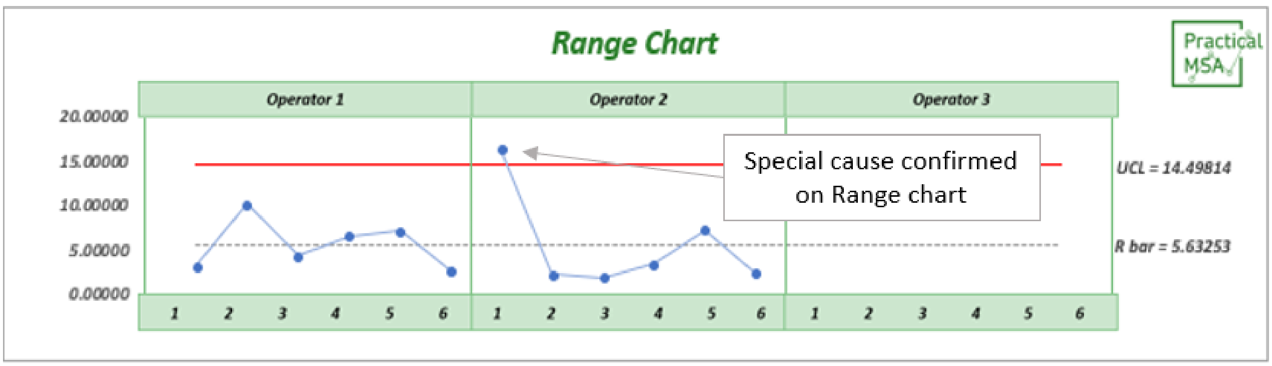
Again, the guiding principle here is learning from your data and taking appropriate action. The reason why the Range Chart works so well in conjunction with the Multi-Vari Chart is because the Multi-Vari Chart allows for picking up clues (patterns, trends), while the Range Chart provides information as to the significance of those clues.

6 Donald J. Wheeler: Normality And The Process Behavior Chart (2000)

Think of a single data point on the Multi-Vari Chart indicating an unusually high within-subgroup range (the range of three trials). With the clue identified, one can go to the Range chart and confirm whether or not what appears to be a high within-subgroup range on the Multi-Vari is actually showing up as significant on the Range Chart.

Here is an actual example for you: look at the below Multi-Vari-Range Chart combo. If you look at the individual data points on the Multi-Vari, there is one subgroup that appears to see more *within* variation, (part 1 for operator 2). Looking at that subgroup range on the Range chart confirms that this is indeed a special cause that is above the UCL. Should you investigate this issue? Yes, you should probably at least take a closer look.





To reiterate what I said earlier, a measurement system is considered reliable when it is consistent, *predictable* over time. Special causes like the one identified in the above Range chart are not an attribute of a predictable measurement process, so such special causes should be identified and ultimately eliminated.

A lot of the time, such special causes have to do with *the physics of the measurement process* itself. Examples include part orientation and clamping in the test fixture, measurement location etc.

Staying with the part clamping example, if the clamping of the part does not prevent movement of the part in the fixture, that could cause inconsistencies in the measurements results themselves, and that's what ends up showing up on the Multi-Vari and Range Charts.

THE MULTI-VARI/RANGE CHART TOOL

Now that we have gone over both the Multi-Vari Chart and the Range Chart individually, let's summarize the features of the tool:

ASSESS AND QUANTIFY REPEATABIITY ERROR

While the Mult-Vari Chart compares repeatability error to the specification limits or the range of units involved in the study, the Range Chart estimates average repeatability error through the average subgroup range (*R bar).*

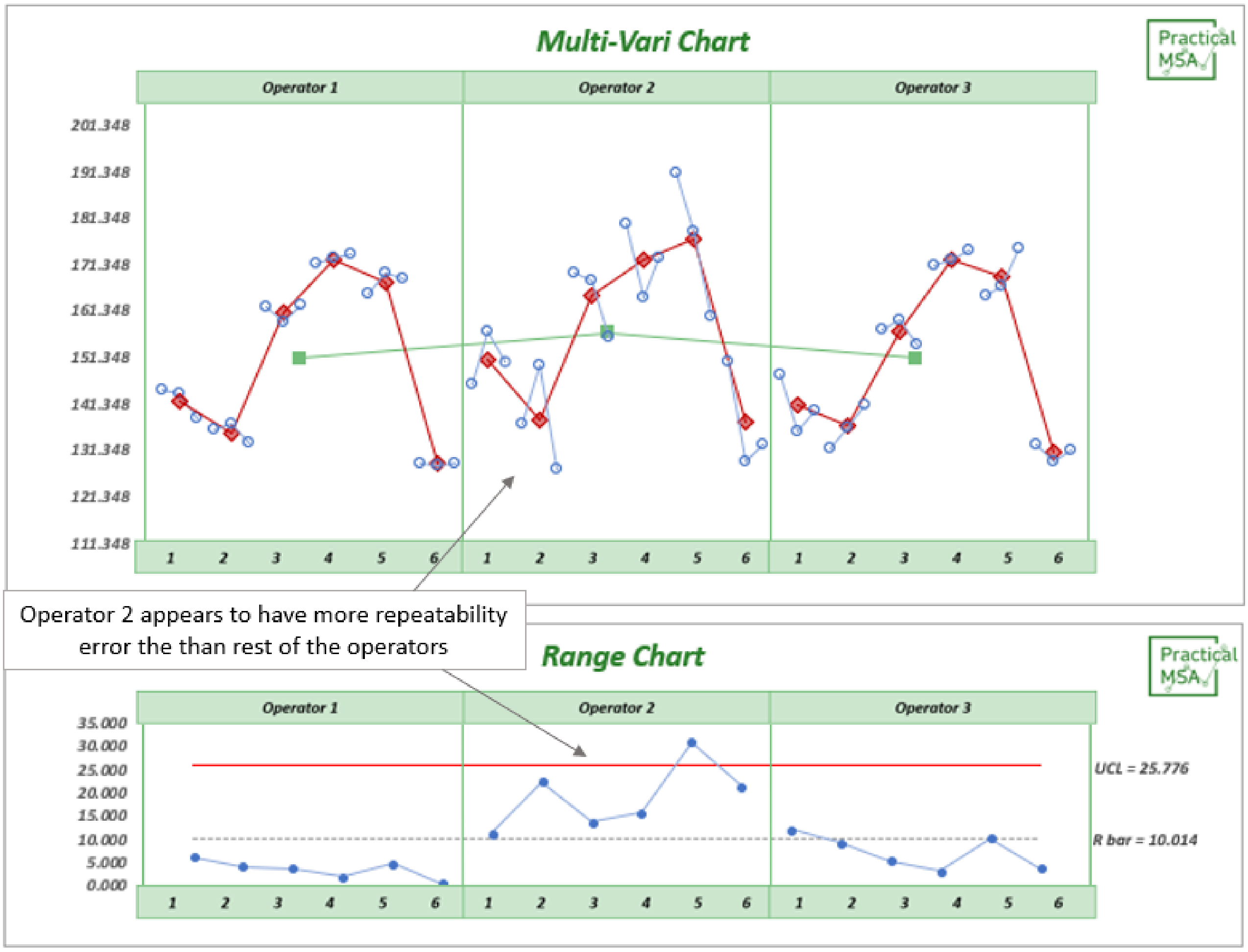
Also, an unbiased estimate for the standard deviation of the repeatability error is calculated by dividing R bar by d2, which is a bias correction factor and depends on the subgroup size.

$$\hat{\sigma}_{repeatability} = \frac{R\ bar}{d_2}$$

Subgroup size	d_2
2	1.128
3	1.693

ASSESS CONSISTENCY OF REPEATABILITY ERROR BETWEEN OPERATORS

The Mult-Vari/Range Chart combo gives you the ability to compare repeatability error between operators. In the below example, operator 2 sees significantly more repeatability error than the rest of the operators. As a general rule, *the repeatability error should be consistent among operators* before other sources of measurement error are assessed and eliminated.

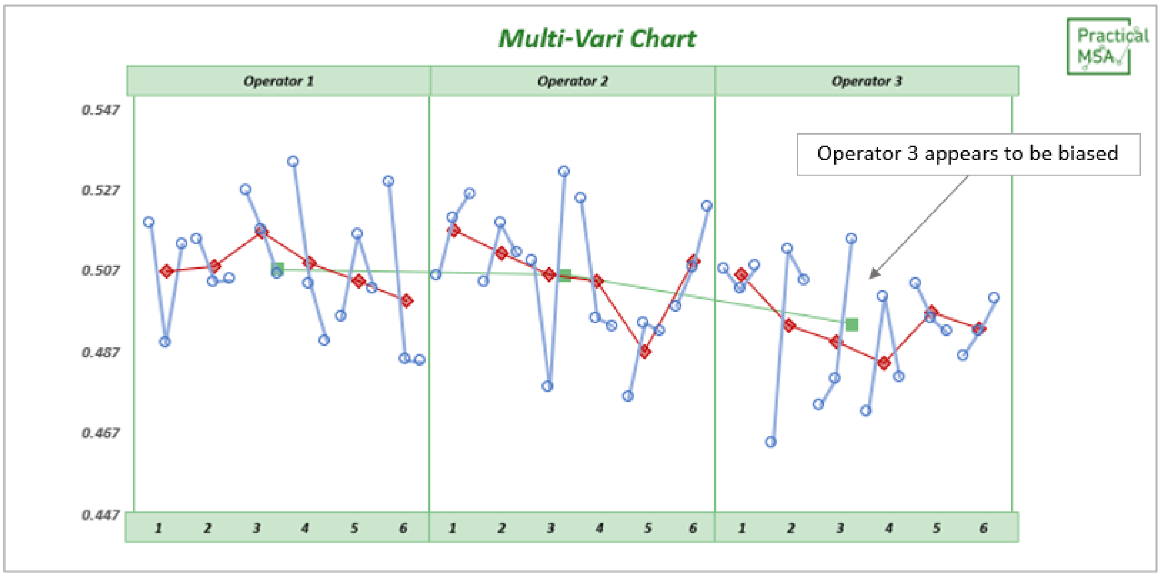


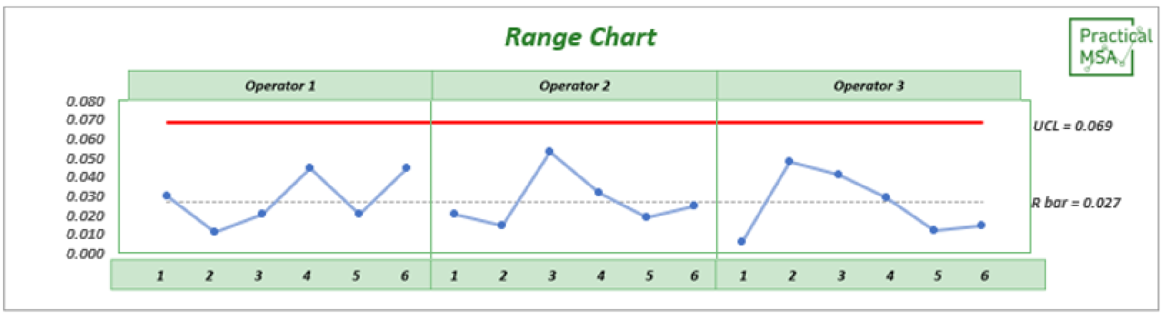
ASSESS OPERATOR BIAS

The Multi-Vari Chart can graphically assess operator bias. Operator averages are also calculated by the tool so that they can be compared from a practical perspective.

Again, true operator bias occurs when a systemic difference exists between operators, i.e. the way in which one operator (or multiple operators) is measuring the parts is inherently different than that of the rest of the operators.

A word of caution: since the individual trials within a subgroup are driving the part averages, which in turn are driving the operator averages, excess repeatability error (in relation to the product variation) can cause a statistically significant amount of reproducibility error, and even make it look like there is operator bias present. In reality, such phenomena *could be* caused merely by *sampling error*, more specifically the standard error of the mean (Central Limit Theorem). In the below example, there is clearly a significant amount of repeatability error present, which *could* cause one or more of the operator averages to appear biased.





All in all, it really comes down to this: If you have excess repeatability error, first focus on addressing that, and you will, at the same time, have fixed any reproducibility issues. If after addressing excess repeatability error one or more opeators still appear to be biased, chances are there is true, systemic operator bias present.

I want to add here that in addition to the Multi-Vari Chart as a graphical tool to assess operator bais, there are additional techniques available to assess whether or not there is true systemic operator bias present, which I will introduce in a future publication.

ASSESS TRENDS AND PATTERNS

The Multi-Vari Chart can detect clues, patterns, special causes.

SPECIAL CAUSES

The Multi-Vari Chart can detect suspect special causes. The Range chart can confirm is the identified data points are indeed special causes.

TRENDS, PATTERNS

The Multi-Vari Chart can detect trends such as an upward or downward pattern within a subgroup, which could be indicative of a number of things, for example the measurement process physically altering the experimental units or inadequate execution of the study etc.

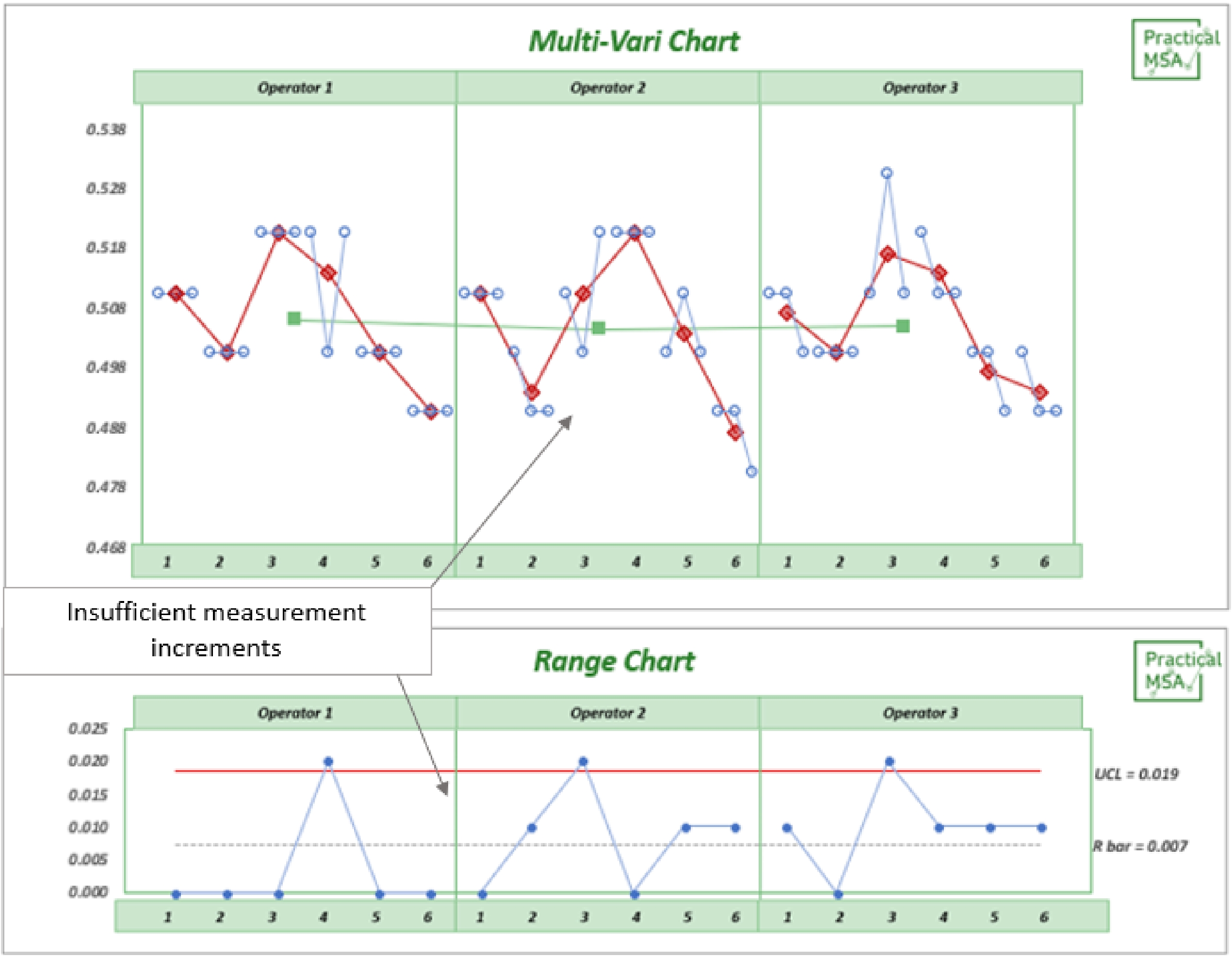
ASSESS LINEARITY OF REPEATABIITY ERROR (CV EFFECT)

The Multi-Vari/Range chart combo can graphically assess and detect non-linear repeatability error. See more on this on page 32.

DETECT MEASUREMENT RESOLUTION ISSUES

Both the Multi-Vari Chart and the Range chart can detect resolution issues.

In the below example, the smallest measurement increment is *0.01* units, which is a clear indication of insufficient measurement resolution in relation to the product variation (the range of units covered in the study).



If you look at the above Multi-Vari Chart, you will observe that the measurements from the three trials for each part either all come in at the same value, or they are "bucketed" into two or three values 0.01 or 0.02 units from each other. You will also notice that the entire product range is no more than 5-6 "buckets" wide.

The Range Chart also has telltale signs of the same phenomenon; all of the ranges fall into "buckets" of either 0 units or 0.010 units or 0.020 units.

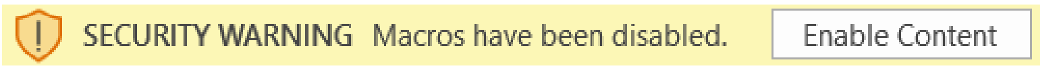
It is also important to note that, due to this "bucketing" of the ranges, insufficient measurement resolution can significantly bias the estimate for R bar and therefore the repeatability error.

USING THE MULTI-VARI/RANGE CHART TOOL

DOWNLOADING AND OPENING THE TOOL

The free Multi-Vari/Range Chart tool (download it here) can handle up to 6 parts, 3 trials and 3 operators.

After you download the tool from the website in the form of an Excel file (*Practial_MSA-Multi-Vari_Range Chart.xlsm*), open it. The following warning message will appear:



Make sure to click *Enable Content* as the tool uses macros for certain features.

The Excel file contains two tabs: the *Data Input* and *Analysis* tabs.

THE DATA INPUT TAB

The Data Input tab is used to capture the measurement data from your study.

You can either capture the measurement data directly in the Data Input tab or copy and paste it from another spreadsheet.

Use the *Activity* and *Characteristic* fields to capture information pertaining the study.

Activity:

Characteristic:

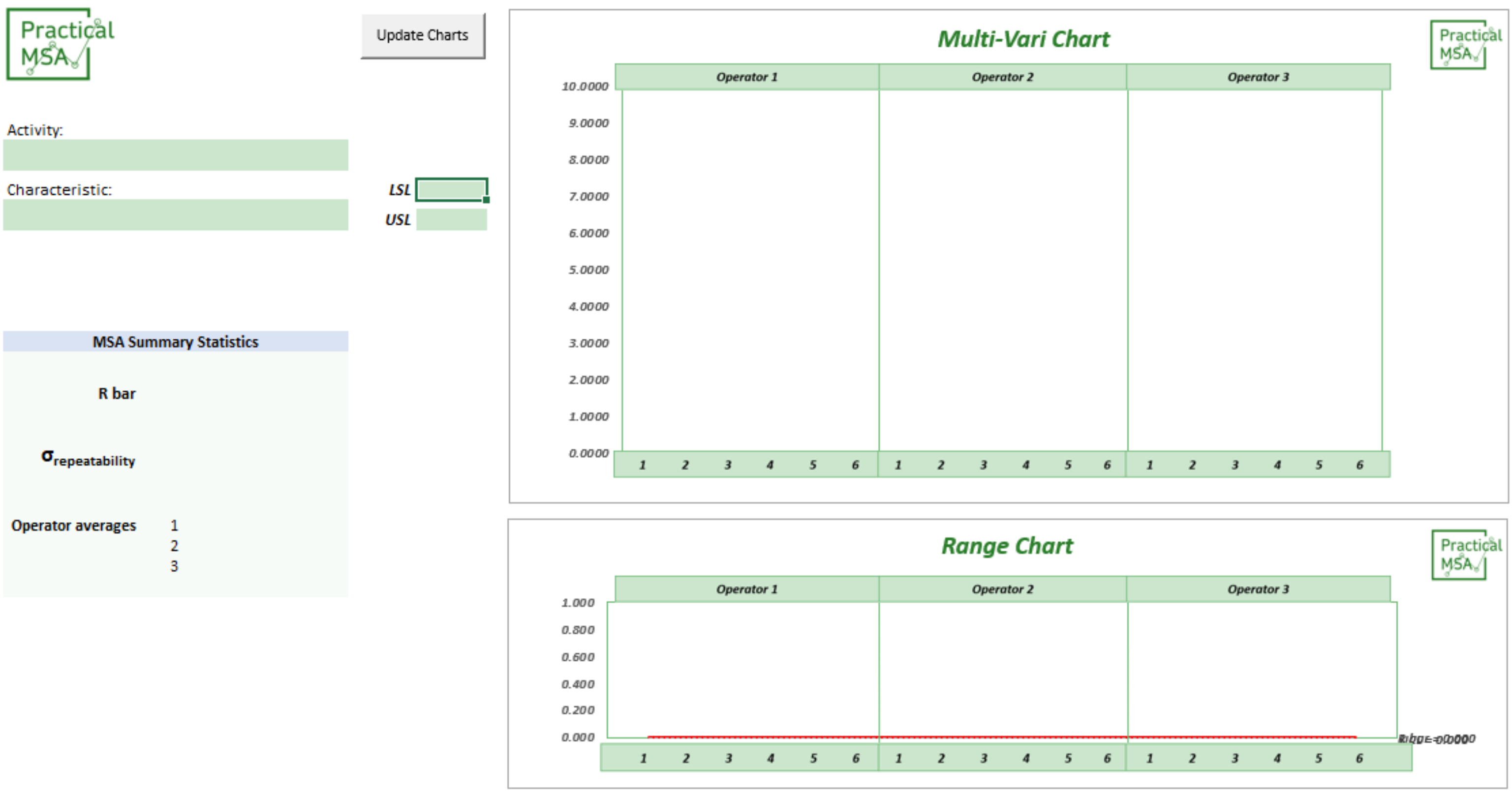
Operator	Part	Trial		
		1	2	3
1	1			
	2			
	3			
	4			
	5			
	6			
2	1			
	2			
	3			
	4			
	5			
	6			
3	1			
	2			
	3			
	4			
	5			
	6			

The Data Input tab

THE ANALYSIS TAB

The Analysis tab houses the Multi-Vari and Range Charts themselves as well as displays the following summary statistics:

- average range (*R bar*)
- the estimated standard deviation of repeatability error, $\hat{\sigma}_{repeatability}$
- operator averages



The Analysis tab

The *Activity* and *Characteristic* fields are carried over from the Data Input tab.

Use the *LSL* and *USL* fields to enter the lower and upper specification limits, if applicable.

Use the *Update Charts* button to refresh sheet and update chart axes

ANALYSIS STEPS

1 Iput measurement data into Data Input tab

2 Go to the Analysis tab

3 Click the Update Charts button. The data from the Data Input tab are now plotted on the Multi-Vari and Range Charts

4 Enter LSL and USL if needed, then click the Update Charts button.

5 Analyze charts and summary statistics based on the sections of this eBook

Activity:
MSA for P/N 12345
Characteristic:
overall length

Operator	Part	Trial 1	Trial 2	Trial 3
1	1	98.6128	97.1589	97.7582
	2	101.4390	100.6904	100.6328
	3	100.2754	100.3347	100.4689
	4	97.8587	98.1516	98.6516
	5	96.8026	97.2170	96.5445
	6	100.2582	100.5239	99.9435
2	1	99.3617	98.9239	98.7061
	2	102.1374	101.4895	101.2597
	3	100.1154	101.0679	99.8367
	4	99.2734	99.6029	100.0993
	5	97.6722	96.9144	97.3339
	6	101.0312	100.7847	101.6315
3	1	98.0197	98.0484	97.6322
	2	100.4385	100.2826	100.4609
	3	99.6820	99.6561	100.1370
	4	98.1357	98.2536	98.5629
	5	97.1855	96.8365	96.2157
	6	99.1050	100.5428	99.5701

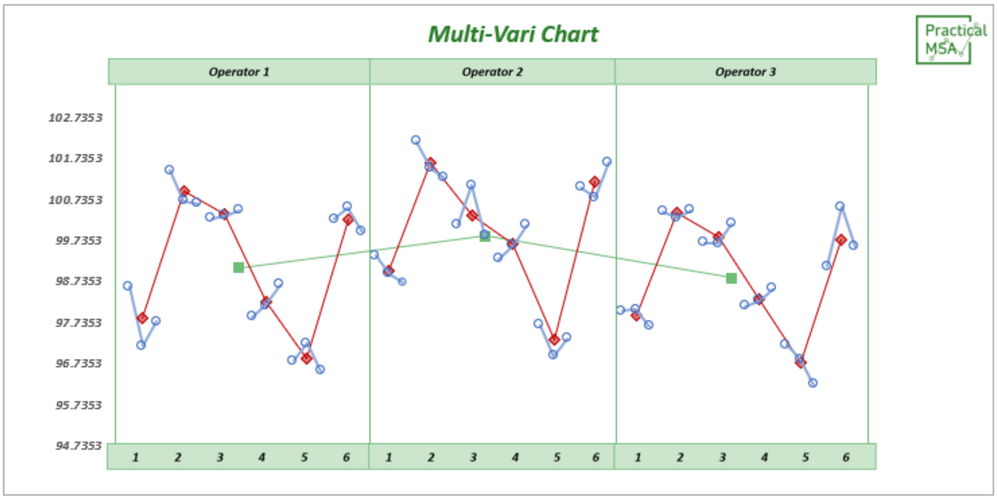
Activity:
MSA for P/N 12345
Characteristic:
overall length

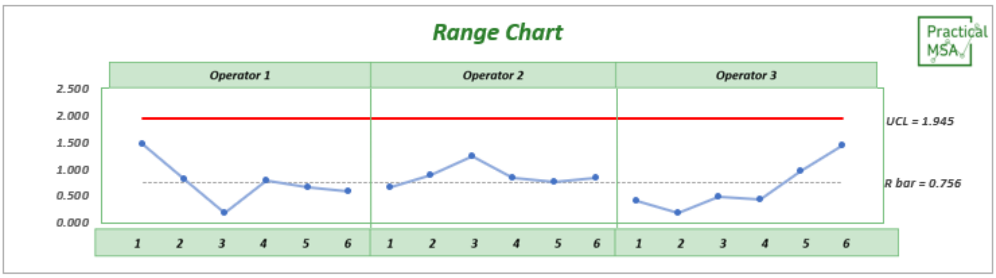
Update Charts

LSL
USL

MSA Summary Statistics

R bar		0.7558
$\sigma_{repeatability}$		0.4464
Operator averages	1	99.0735
	2	99.8468
	3	98.8203





REFERENCES

1 AIAG Measurement Systems Analysis Reference Manual, 4th Edition (2010)

2 W. Edwards Deming: Some Theory of Sampling, Chapter 7 (1950)

3 Donald J. Wheeler: EMP III Evaluating The Measurement Process & Using Imperfect Data (2006)

4 Leonard Seder: "Diagnosis with Diagrams—Part I", Industrial Quality Control (1950)

5 W. A. Shewhart: Economic Control of Quality of Manufactured Product (1931)

6 Donald J. Wheeler: Normality And The Process Behavior Chart (2000)

CLOSING REMARKS

I hope you enjoyed this eBook and the tools provided.

Although the book is intended to be a comprehensive overview of basic Measurement Systems Analysis concepts, there is so much more to be said on this very subject that I plan to release follow-up publications that will cover additional concepts or go deeper on some of the ideas discussed in this book and provide even more tools and techniques for you to apply.

If you are interested in reading future publications, please visit practicalmsa.com and sign up for our newsletter to receive information on release dates and other events.

Practical MSA was created with the goal of educating engineering and scientific professionals on the role of measurement systems analysis and emphasizing a practical approach to it.

For more information, visit practicalmsa.com.