Validating Measurement System Using Measurement Systems Analysis (MSA)

Author:
Hemant Urdhwareshe
Director, Institute of Quality and Reliability

On the Six Sigma roadmap, we have so far seen how to define projects and use graphical tools for preliminary analysis. In this article, we will discuss how to validate measurement system using Measurement Systems Analysis (MSA).
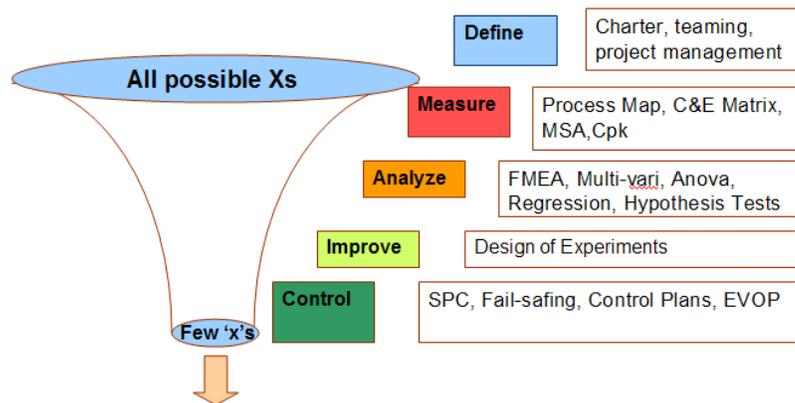
We frequently do not realize how our decisions depend on the results of a measurement system. We also do not adequately understand the consequences of incorrect results of a measurement system. Imagine that a doctor prescribes medicines to reduce blood pressure (BP) when it is actually normal! Doctors use BP measuring apparatus with mercury column or with a dial type gauge. If the mercury column type BP apparatus is used, it is vulnerable to "cosine error" when it is kept tilted instead of vertical. Result can artificially "increase" the BP of the person and can result in to incorrect prescription! This looks like a simple example but the incorrect decision can create major problem for the patient. In statistical terms, the doctor is making a "type I" error. Many of us may have more than one clock in our houses. We know that each clock shows a different time! This may not be a serious problem in routine decisions. However, we will not like such differences when critical decisions are made.

The least count of our wrist watches is one second. But this is not the accuracy with which we can estimate time elapsed between events. If we want to find time taken for an athlete to run a distance of 100 meters, error using a wrist watch may be about a second. Thus we will use a stop watch instead. If we want to decide winner of an Olympic race, our stop watch also may be unacceptable measurement system. We will then need a more precise system and a fast video camera to decide winner in case of very narrow difference in time. Thus we need to use a measurement system that is appropriate for the objective and is able to adequately resolve "process variation" to take correct decisions.

Process and Measurement variation

Our objective in most situations is to measure process variation (PV). But what we *actually* measure is a combined total of process as well as measurement variation. In

measurement system analysis (MSA), we use statistical methods to estimate how much of the total study variation (SV) is due measurement system. An ideal measurement system should not have any variation. However, this is impossible and we have to be satisfied with a measurement system that has variation less than 10% of the process variation. As the portion of variation due to measurement system increases, the value of measurement system goes on reducing. If this proportion is more than 30%, the measurement system is unacceptable. If we are buying vegetables, we may not be worried about 10 gm less and shopkeeper may not be worried about 10gm excess. However, we will need a balance that can distinguish between 1 gm if we are controlling composition of a chemical that can affect the process outcome or yield. Following table summarizes criteria as per MSA Manual published by Automotive Industry Action Group (AIAG).

| Measurement variation as % of Total Study Variation	Decision Guideline
<10 %	Acceptable measurement system.
Between 10 and 30%	Marginally acceptable. However, this needs to be agreed with the customer
>30%	Unacceptable Measurement System

Repeatability and Reproducibility:

We further estimate components of measurement variation. Total measurement variation (MV) is sum of variation due to repeatability and reproducibility. Following figure explains this. (σ is the standard deviation)

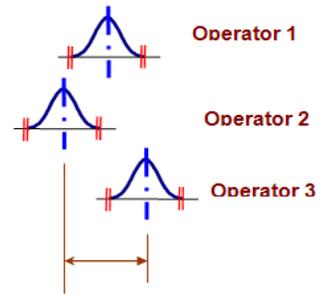**Repeatability**   **Reproducibility**   **Gauge R & R (GRR)**
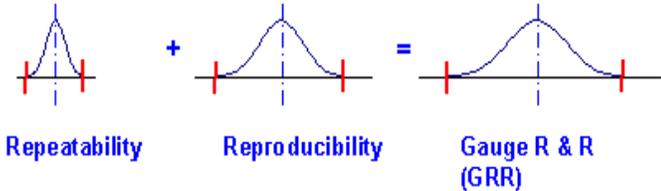
$$\sigma^2_{\text{GRR}} = \sigma^2_{reproducibility} + \sigma^2_{repeatability}$$

Repeatability of the instrument: This is the inherent ability to show the same value of measurement when used by the same person. Remember when we check our weight 3 or 4 times, we often get different results. If the *same operator* measures *same part* number of times, the closeness is a measure of repeatability. Improving repeatability frequently requires better instrument.
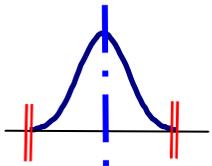
Reproducibility is variation of measurement due to operators, usually called appraisers. Improving reproducibility frequently requires training for operators and better work instructions. Sometimes reproducibility depends on type of instrument. In such cases, we say that there is interaction between operators and measuring instrument.

Repeatability and Reproducibility (R&R)

In a Gauge R & R study, we try to quantify measurement variation MV as a % of process Variation PV. We also estimate repeatability and reproducibility components of measurement variation. The corresponding standard deviation σ is used to evaluate *"% of study variation"*. We compare standard deviation of measurement system σ_m with standard deviation σ_s for the study variation to find Gauge Repeatability and Reproducibility (GRR). We can also compare the MV with Tolerance. In this case, the ratio is sometimes called *"Precision to Tolerance" or P/T ratio*. To compare the tolerance with measurement variation, we multiply standard deviation by 6. This corresponds to a ratio of "Zone of 99.73% measurements" with Tolerance. The multiplier 6 is used to for calculating "Zone of 99.73%" considering normal distribution.

$$\%R\&R = \frac{\sigma_m}{\sigma_s} \times 100$$

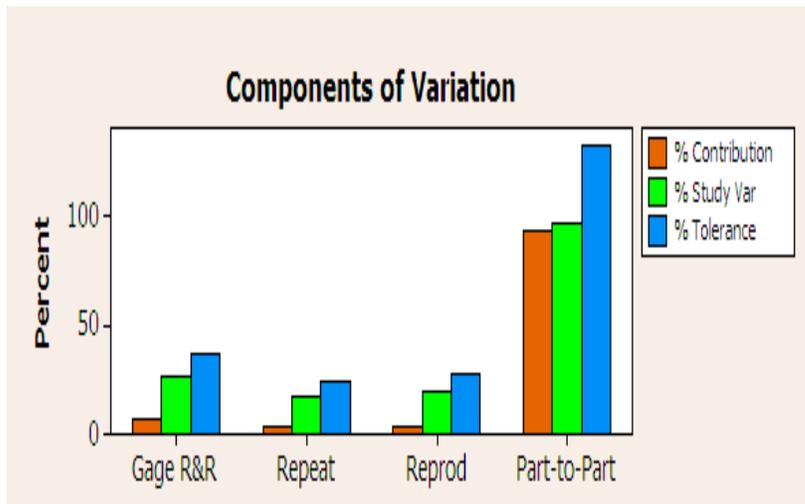$$\%Tolerance = \frac{6\sigma_m}{Tolerance} \times 100$$

Now we may try to relate GRR to our decision making process. In the Olympic race, we want to decide the rankings of the participants accurately. We cannot afford to declare a Gold Medal to a participant who actually reached second because our measurement system has higher variation! Similarly, in our Six Sigma Projects and business decisions, we need to ensure that measurement system is appropriate considering process variation and tolerance.

Most of the statistical softwares such as Minitab or SigmaXL are capable of analyzing MSA data. To conduct an MSA, we can select 10 parts. These 10 parts should have spread comparable to the process variation. We then ask 3 different operators (called appraisers) to measure each part thrice. While doing this, we ensure that the parts are given to the appraisers in a random order so that they do not remember their earlier result and each measurement is "independent" of the earlier measurement. They also should not see other appraisers measuring the parts. Once the measurements are completed, the data is analyzed using any of the softwares. A typical output from Minitab is shown here for reference. The software also provides useful

| Source	StdDev (SD)	Study Var (6 * SD)	%Study Var (%SV)	%Tolerance (SV/Toler)
Total Gage R&R	0.30589	1.83536	26.70	36.71
Repeatability	0.20181	1.21087	17.61	24.22
Reproducibility	0.22988	1.37925	20.06	27.59
Part-To-Part	1.10412	6.62474	96.37	
	132.49			
Total Variation	1.14571	6.87428	100.00	137.49

Number of Distinct Categories = 5

graphical output that helps us in deciding actions to be taken if Gauge R&R is not acceptable. In the bar chart, the height of Gauge R&R should be small compared to Study Variation SV and tolerance. Most of the softwares also calculate *Number of Distinct Categories* (NDC). NDC indicate the number of groups within your process data that your measurement system can discern. These should be more than 5. NDC of 2 means the gauge can group the data in to "good" and "bad" only and is only like a "Go" and "Not Go" gauge! NDC value of < 1 makes the gauge completely useless for any decision making.


Apart from repeatability and reproducibility, we also need to consider and evaluate some other aspects of the measurement system.

Bias: It can have a different value than the master value. This is called bias. For example, if a watch is 5 minutes faster than standard time (standard time being the master value), 5 minutes is the bias.
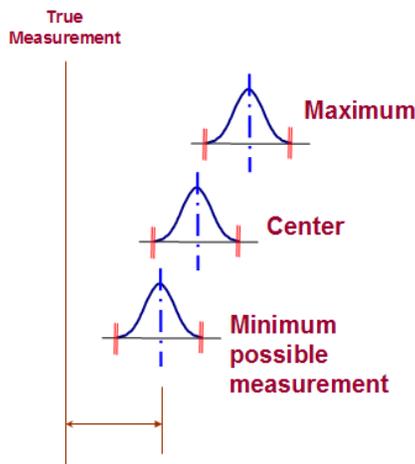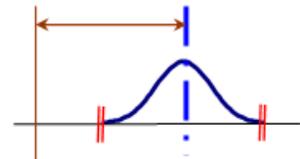
Linearity: Consistency over the measurement range. We can consider that Fuel indicator gauge on a car dashboard may not be linear. Thus when it shows 1/3 full, it may actually be ¼ full.
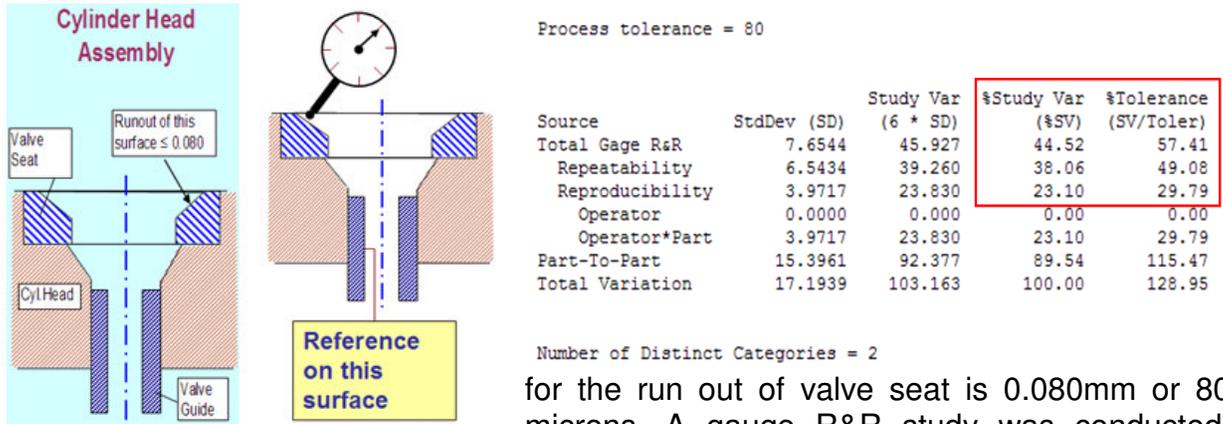
Stability is measurement variation over time. We should calibrate a gauge or instrument to ensure its stability. This is sometimes called drift
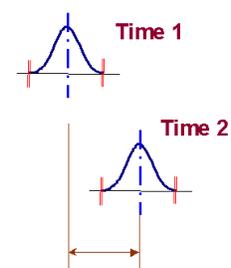
The stability and linearity errors are to be addressed during calibration of the instrument.

MSA Example:

We consider a measurement system for concentricity of valve seat with valve guide. This is an important characteristic. See figure for a sketch of the system. The tolerance


```
Process tolerance = 80

                                      Study Var   %Study Var   %Tolerance
Source                StdDev (SD)   (6 * SD)       (%SV)      (SV/Toler)
Total Gage R&R            7.6544     45.927        44.52        57.41
  Repeatability           6.5434     39.260        38.06        49.08
  Reproducibility         3.9717     23.830        23.10        29.79
    Operator              0.0000      0.000         0.00         0.00
    Operator*Part         3.9717     23.830        23.10        29.79
Part-To-Part            15.3961     92.377        89.54       115.47
Total Variation         17.1939    103.163       100.00       128.95

Number of Distinct Categories = 2
```

for the run out of valve seat is 0.080mm or 80 microns. A gauge R&R study was conducted. Results As we can see, the Gauge R&R is 44.52% i.e. measurement variation is too large as compared to total variation. This shows an urgent need to improve the measurement system. The number of distinct categories is barely 2. Thus this is no better than a Go/Not Go gauge. Subsequently, the Six Sigma team made design changes in the gauge and R&R was improved to a much better value of 15%.

MSA for Attribute Data

There are many situations when we need to use attribute data. Attribute MSA is applicable when the measurement value is one of finite number of categories. Most common example is of "go/not go" gauge. Other examples are leak or no leak, crack or no crack, defect or no defect, pass or fail, complete or incomplete.

We will not discuss details of the methods for Attribute MSA. These methods are primarily used to evaluate their consistency with reference to a "master decision". Thus we take about 20 parts, some good and some bad. These parts are repeatedly measured by "master appraiser" and also other 2 or 3 appraisers. If the decisions by appraisers and masters are "reasonably" consistent, we can consider that the measurement system can be used. Else, we need to improve. This method is called "Attribute Agreement Analysis". The measure used for extent of attribute agreement is Kohen's Kappa value. This can range from −1 to +1. +1 Shows perfect agreement while as 0 shows that agreement is by chance. Kappa of <1 means agreement is less than chance. This is not usually seen. AIAG recommends Kappa>0.75 for good system and <0.4 as poor system.

Assessment Agreement

Date of study:
Reported by:
Name of product:
Misc:

Appraiser vs Standard
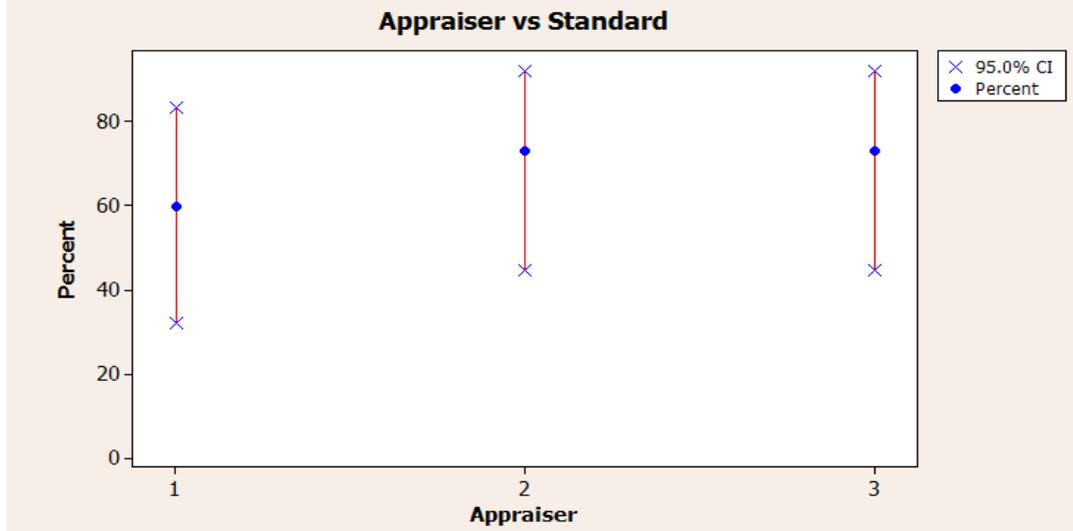
Measurement Systems Analysis needs to be performed before performing a capability study. If this is done after capability study and we find that the measurement system needs improvement, the earlier capability study results are not valid. We need to perform the capability study again after improving measurement system.

About Author:

Hemant Urdhwareshe is Director, Institute of Quality and Reliability, Pune. He is
  • *Qimpro Foundation Silver Standard 2003 Award Winner*
  • *First ASQ Certified Six Sigma Black Belt in India*
  • *ASQ Certified Quality Engineer, Quality Manager and Reliability Engineer.*
  • *The only person in India with 4 or more certifications from ASQ (as on Mar 2005)*
  • *Senior Member ASQ*

Email: hemant@world-class-quality.com   *Phone: 020-25367952*

Visit: www.world-class-quality.com