

Problems with the AIAG Gauge R&R Study

And how to fix them so as to avoid the nonsense

Donald J. Wheeler

Measurement error is ubiquitous. As a result, over the past 250 years, different areas of science and engineering have come up with many different ways to deal with the problem of measurement error. One approach to the problem of measurement error was developed during the 1960s within General Motors. Over the years it was modified and revised until, in 1989, it was turned over to the Automotive Industry Action Group (AIAG). Since that time the AIAG Gauge R&R Study has been promoted throughout many different industries. Unfortunately, the original procedure contained some fundamental problems that have not been corrected over the years. This column will address these historic problems and suggest solutions.

THE AIAG GAUGE R&R STUDY

The Gauge R&R Study starts out with a sound strategy for collecting data. A simple fully crossed experiment is performed where two or more operators measure each of three to ten parts two or three times apiece. To have an example to use we shall use the data shown in Figure 1 where three operators measure each of five parts two times apiece.

Figure 1 The Gasket Thickness Data

Operator	A					B					C				
Part	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1st Value	167	210	187	189	156	155	206	182	184	143	152	206	180	180	146
2nd Value	162	213	183	196	147	157	199	179	178	142	155	203	181	182	154
Averages	164.5	211.5	185.0	192.5	151.5	156.0	202.5	180.5	181.0	142.5	153.5	204.5	180.5	181.0	150.0
Ranges	5	3	4	7	9	2	7	3	6	1	3	3	1	2	8

The average and range are shown for each of the 15 pairs of measurements in Figure 1. Since each of these pairs represents the same operator measuring the same part, the ranges characterize the basic component of measurement error. This component has many names. Among these are test-retest error, repeatability, and equipment variation. This component may be estimated by dividing the average range by the appropriate bias correction factor d_2 :

$$Repeatability = \frac{\bar{R}}{d_2} = \frac{4.267}{1.128} = 3.783 \text{ mils} \quad [1]$$

Next, under the assumption that the operators are different, an estimate of the reproducibility is obtained by computing an average value for each operator and finding the range of these averages. Here the operator averages are 181.0, 172.5, and 173.9 and the range of these three averages is $R_o = 8.5$ mils. This range value is divided by a bias correction factor for the range of three values, $d_2^* = 1.906$, and then, with o = number of operators, p = number of parts, and n = number of repeated measurements, the reproducibility is estimated to be:

$$\text{Reproducibility} = \sqrt{\left[\frac{R_o}{d_2^*}\right]^2 - \frac{o}{n o p} [\text{Repeatability}]^2} = 4.296 \text{ mils} \quad [2]$$

Once we have an estimate of both the repeatability and the reproducibility we can combine them to estimate the combined R&R component as:

$$\text{Combined R\&R} = \sqrt{\text{Repeatability}^2 + \text{Reproducibility}^2} = 5.724 \text{ mils} \quad [3]$$

Next the product variation is estimated by computing averages for each of the p parts and using the range, R_p . Here the five part averages are 158.0, 206.167, 182.0, 184.833, and 148.0. The range for these part averages is $R_p = 58.167$. The bias correction factor used is $d_2^* = 2.477$, which is for one range of five values. The product variation is estimated to be:

$$\text{Product Variation} = \frac{R_p}{d_2^*} = \frac{58.167}{2.477} = 23.483 \text{ mils} \quad [4]$$

Finally, the total variation for the set of product measurements is estimated by combining the repeatability, the reproducibility, and the product variation to get:

$$\begin{aligned} \text{Total Variation} &= \sqrt{\text{Repeatability}^2 + \text{Reproducibility}^2 + \text{Product Variation}^2} \\ &= 24.171 \text{ mils} \end{aligned} \quad [5]$$

Up to this point everything is okay. While the estimators defined here are not the only formulas that could have been used, and while they are not always unbiased estimators, they do provide reasonable estimates for the quantities described. (While the AIAG Gauge R&R computational worksheet simplifies the formulas by defining several “K-factors” the AIAG formulas are algebraically equivalent to formulas [1] through [5].)

The train wreck begins when the AIAG Gauge R&R Study tries to use the estimates from formulas [1] through [5] to characterize relative utility. In the current version the first four quantities are all expressed as a percentage of the total variation from formula [5].

THE “PERCENTAGES” OF THE TOTAL VARIATION

The repeatability [1] is divided by the total variation [5] and multiplied by 100 to be expressed as a percentage. This ratio is said to represent the percentage of the total variation that is consumed by repeatability. Here we find:

$$\frac{100 \text{ Repeatability}}{\text{Total Variation}} = \frac{100 (3.783)}{24.171} = 15.65\% \quad [6]$$

Next the reproducibility [2] is divided by the total variation [5] and multiplied by 100. This ratio is said to represent the percentage of the total variation that is consumed by reproducibility. Here we find:

$$\frac{100 \text{ Reproducibility}}{\text{Total Variation}} = \frac{100 (4.296)}{24.171} = 17.77\% \quad [7]$$

The combined repeatability and reproducibility [3] is divided by the total variation [5] and multiplied by 100. This ratio is said to represent the percentage of the total variation that is consumed by the combined repeatability and reproducibility. Here we find:

$$\frac{100 \text{ Combined R\&R}}{\text{Total Variation}} = \frac{100 (5.724)}{24.171} = 23.68\% \quad [8]$$

Finally, the product variation [4] is divided by the total variation [5] and multiplied by 100. This ratio is said to represent the percentage of the total variation that is consumed by the product variation. Here we find:

$$\frac{100 \text{ Product Variation}}{\text{Total Variation}} = \frac{100 (23.483)}{24.171} = 97.15\% \quad [9]$$

Following these ratios, the current version of the AIAG manual has a simple statement to the effect the "The sum of the percent consumed by each factor will not equal 100%." This statement has no explanation attached. There is no guidance offered on how to proceed now that common sense and every rule in arithmetic have been violated. Just a simple statement that these numbers do not mean what they were just interpreted to mean, and the user is left to his or her own devices. Unfortunately, unlike the Red Queen in Wonderland, when it comes to arithmetic we do not get to say that things mean whatever we want them to mean.

WHY THESE "PERCENTAGES" DO NOT ADD UP

The repeatability "percentage" of formula [6] and the reproducibility "percentage" of formula [7] do not add up to the combined R&R "percentage" of formula [8] because they are not based on proportions. Likewise, the combined R&R "percentage" of formula [8] and the product variation "percentage" of formula [9] do not add up to 100 percent because they are not based on proportions. Instead of being proportions, it turns out that the ratios computed above are all trigonometric functions. Figure 2 shows how the five estimates of variation are related.

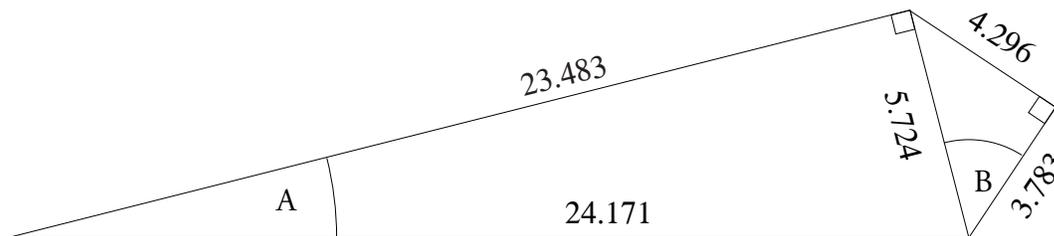


Figure 2: The Five Estimates of Variation

Thus, based on Figure 2, and recalling some of our high school trigonometry, we see that the ratio of the repeatability to the total variation is:

$$\frac{\text{Repeatability}}{\text{Total Variation}} = 0.1565 = \frac{3.783}{5.724} \frac{5.724}{24.171} = (\text{Cosine B}) (\text{Sine A})$$

The ratio of the reproducibility to the total variation is:

$$\frac{\text{Reproducibility}}{\text{Total Variation}} = 0.1777 = \frac{4.296}{5.724} \frac{5.724}{24.171} = (\text{Sine B}) (\text{Sine A})$$

The ratio of the combined repeatability and reproducibility to the total variation is:

$$\frac{\text{Combined R\&R}}{\text{Total Variation}} = 0.2368 = \frac{5.724}{24.171} = \text{Sine A}$$

And the ratio of the product variation to the total variation is:

$$\frac{\text{Product Variation}}{\text{Total Variation}} = 0.9715 = \frac{23.483}{24.171} = \text{Cosine A}$$

In this form we can begin to see why these quantities do not add up. *While they were dressed up to look like percentages, and while they were interpreted as percentages, they are, and always have been, nothing more than trigonometric functions.* And trigonometric functions do not satisfy the conditions required for a set of ratios to be interpreted as proportions or percentages.

$$\frac{a}{a+b+c} + \frac{b}{a+b+c} + \frac{c}{a+b+c} = 1$$

A set of ratios will be proportions only when the denominator is the sum of the numerators. This additivity of the numerators is the essence of proportions. So what is additive in a Gauge R&R Study? Look at the structure of formula [5]. It is the variances that are additive:

$$\text{Total Variation}^2 = \text{Repeatability}^2 + \text{Reproducibility}^2 + \text{Product Variation}^2$$

Since these variances are additive, we know from the Pythagorean Theorem that the standard deviations *cannot be additive*. However, the ratios computed by the AIAG Gauge R&R Study implicitly assume that the standard deviations are additive. This implicit assumption of additivity is a violation of the Pythagorean Theorem, and is what makes it impossible to make sense of the ratios in the AIAG Gauge R&R Study. This is why the “percentages” of ratios [6] through [9] do not add up, and this is why engineers have told me they never could figure out exactly what the final numbers in a Gauge R&R Study represented. The ratios sound like nonsense simply because it is nonsensical to interpret trigonometric functions as proportions.

SO WHAT ARE THE ACTUAL PROPORTIONS?

Formula [5] above suggests the solution to the problem of how to characterize the contribution of the various components to the total variation in the product measurements. For example, the repeatability contribution to the total variation could be estimated by:

$$\frac{\text{Repeatability}^2}{\text{Total Variation}^2} = \frac{3.783^2}{24.171^2} = 0.0245 \quad [\text{Honest Ratio 6}]$$

which is 2.45 percent rather than the 15.65 percent erroneously found earlier. The reproducibility contribution to the total variation could be estimated by:

$$\frac{\text{Reproducibility}^2}{\text{Total Variation}^2} = \frac{4.296^2}{24.171^2} = 0.0316 \quad [\text{Honest Ratio 7}]$$

which is 3.16 percent rather than the 17.77 percent erroneously found earlier. The combined repeatability and reproducibility contribution to the total variation could be estimated by:

$$\frac{\text{Combined R\&R}^2}{\text{Total Variation}^2} = \frac{5.724^2}{24.171^2} = 0.0561 \quad [\text{Honest Ratio 8}]$$

which is 5.61 percent rather than the 23.68 percent erroneously found earlier. We also should note that this 5.61 percent is the sum of the 2.45 percent and the 3.16 percent, which is what we expect when computing proportions. The product variation contribution to the total variation could be estimated by:

$$\frac{\text{Product Variation}^2}{\text{Total Variation}^2} = \frac{23.483^2}{24.171^2} = 0.9438 \quad [\text{Honest Ratio 9}]$$

which is 94.38 percent rather than the 97.15 percent erroneously found earlier. When this value is added to the combined R&R percentage we effectively get 100 percent. Now we have correctly accounted for the various components of the variation in the product measurements.

Honest Ratio [9] is the traditional measure of relative utility introduced by Sir Ronald Fisher in 1921. In this context it is commonly known as the intraclass correlation coefficient (*ICC*). Here the value of 0.9438 tells us that over 94 percent of the variation in the product measurements is attributable to the variation in the product stream, and less than 6 percent of the variation in the product measurements is attributable to variation in the measurement system (combined R&R). Thus, this one number summarizes the essential information contained in all four of the honest ratios above and provides a characterization of the relative utility of the measurement system for measuring a particular product.

SO WHAT SHOULD YOU DO?

Since any attempt to use ratios [6] through [9] from the AIAG Gauge R&R Study is a violation of the Pythagorean Theorem, you will need to convert the ratios into the correct quantities shown as honest ratios [Honest Ratio 6] through [Honest Ratio 9] before attempting to interpret or use them in any manner. No exceptions. No loopholes. No excuses.

THE “PERCENTAGES” OF THE SPECIFIED TOLERANCE

Ratios [6] through [9] were added to the GM Gage [sic] R&R Study in its 1984 revision. Prior to 1984 there were only three ratios. These compared the repeatability [1], the reproducibility [2], and the combined R&R [3] to the specified tolerance.

Assume that the specifications for the gaskets of Figure 1 are 145 mils to 225 mils. How much of this specification range is “lost” due to measurement error? In an attempt to answer this question the GM Gage R&R Study multiplied the repeatability by 5.15 (presumably to account for 99% of the measurement error), divided by the specified tolerance, and then multiplied by 100 to express the result as a percentage. In an attempt to account for more than 99% of the measurement error, the AIAG Gauge R&R Study eventually changed the initial multiplier from 5.15 to 6.00. Thus, the following ratio is said to represent that amount of the specified tolerance that is consumed by repeatability:

$$\frac{600 \text{ Repeatability}}{\text{Specified Tolerance}} = \frac{600 (3.783)}{80} = 28.4\% \quad [10]$$

In a similar manner, reproducibility is said to consume the following amount of the specified tolerance:

$$\frac{600 \text{ Reproducibility}}{\text{Specified Tolerance}} = \frac{600 (4.926)}{80} = 32.2\% \quad [11]$$

However, the combined repeatability and reproducibility is said to consume the following amount of the specified tolerance:

$$\frac{600 \text{ Combined R\&R}}{\text{Specified Tolerance}} = \frac{600 (5.724)}{80} = 42.9\% \quad [12]$$

Once again, these three ratios do not add up. Moreover, since reproducibility cannot occur without repeatability, formula [11] is merely an academic exercise without any practical use.

Formulas [10] and [12] are versions of what are popularly known as precision to tolerance (P/T) ratios. The values of 28.43% and 42.9% found here would erroneously be classed by the AIAG study as being respectively “marginal” and “unacceptable.” However, as we learned last month in “More About the Precision to Tolerance Ratio” (*Quality Digest* June 1, 2021), we cannot use these P/T ratios to condemn any measurement process. While P/T ratios that are less than 10% are good, the converse is not true. Here, with an interclass correlation of 0.94, a P/T ratio of 28% corresponds to a potential process yield of 90% good product produced and shipped. With the larger P/T ratio of 43% we would estimate a potential process yield of 88% good product produced and shipped.

THE “NUMBER OF DISTINCT CATEGORIES”

Finally, the AIAG Gauge R&R Study computes a quantity it calls the “number of distinct categories” (ndc) using the formula:

$$ndc = 1.41 \frac{\text{Product Variation}}{\text{Combined R\&R}} = 1.41 \frac{23.483}{5.724} = 5.8 \quad [13]$$

The AIAG manual suggests rounding this value off to an integer, and that values of 5 or greater are “good.” (Simple algebra will show that this value is inconsistent with the guidelines given below, but inconsistencies of this sort are found throughout the AIAG manual.)

The citation given in the AIAG Manual for formula [13] is the First Edition of *Evaluating the Measurement Process* by this author and Richard Lyday. In that text we defined a quantity which we called the classification ratio, and the formula above does provide an estimate of this classification ratio. However, nowhere in that text did we ever suggest that this ratio would define the number of distinct categories.

Unfortunately, in practice there is no simple interpretation for the classification ratio. While the classification ratio does approximate the relative sizes of the major and minor axes in the intraclass correlation plot, this plot does not lend itself to any practical interpretation. So while the ratio given in [13] does describe one aspect of a particular plot of the data, neither that plot, nor ratio [13], provide any practical characterization of the number of distinct categories for the product.

THE GUIDELINES

In spite of the many revisions made in the Gauge R&R Study over the years the guidelines for interpreting the “percentages that are not really percentages” has remained unchanged. Whether they were applied to ratios [6], [7], or [8], or were applied to ratios [10], [11], or [12], the guidelines have always been:

- Ratios that are less than 10% are said to be good;
- Ratios that between 10% and 30% are said to be marginal; and
- Ratios that exceed 30% are said to be unacceptable.

Using these guidelines we would interpret the ratios computed earlier as shown in Figure 3.

Figure 3 Interpreting the Results for the Gauge R&R Study for Gasket Thickness.

AIAG Ratio	Quantities Involved	Value	AIAG Interpretation
[6]	Repeatability / Total Variation	15.7%	marginal
[7]	Reproducibility / Total Variation	17.8%	marginal
[8]	Combined R&R / Total Variation	23.7%	marginal
[10]	Repeatability / Specified Tolerance	28.4%	marginal
[11]	Reproducibility / Specified Tolerance	32.2%	unacceptable
[12]	Combined R&R / Specified Tolerance	42.9%	unacceptable
[13]	Number of Distinct Categories	5.8	good

Here we find that the measurement system used to measure the gaskets in Figure 1 is simultaneously marginal, unacceptable, and good! Any questions?

While the ratios [6], [7], and [8] have the same numerators as ratios [10], [11], and [12] respectively, they have different denominators. In spite of these different denominators, the same guidelines are used for each ratio in the AIAG Gauge R&R Study. This is presumably one of the benefits of using guidelines that have absolutely no contact with reality. Since no justification for these guidelines exists, they can be adapted to fit any set of ratios that might be computed. After all, when attempting to interpret patent nonsense, it is always best to use arbitrary guidelines.

The inconsistency of the results in Figure 3 is an inherent feature of the AIAG Gauge R&R Study. Once you ignore the Pythagorean Theorem, mere guidelines are not going to remedy the situation. After finding reasonable estimates for repeatability, reproducibility, combined R&R, product variation, and the total variation with formulas [1] through [5], the eight ratios [6] through [13] computed in the AIAG Gauge R&R Study are erroneous, fallacious, naive, incorrect, and, to put it quite simply, wrong.



BUT HOW DO THESE GUIDELINES

MATCH UP WITH THE FOUR CLASSES OF PROCESS MONITORS?

In "The Intraclass Correlation Coefficient" (*Quality Digest* December 2, 2010) four classes of process monitors were defined. To see how the Gauge R&R Guidelines compare with the four classes of process monitors we need to begin by noting that the ratio in formula [8] is 100 times the following:

$$\frac{\text{Combined R\&R}}{\text{Total Variation}} = \sqrt{1 - \text{Intraclass Correlation}}$$

Thus, if we plotted values for ratio [8] on one scale (in ascending order), and plotted values for the intraclass correlation on another scale (in descending order) we could connect equivalent points on the two scales to get Figure 4.

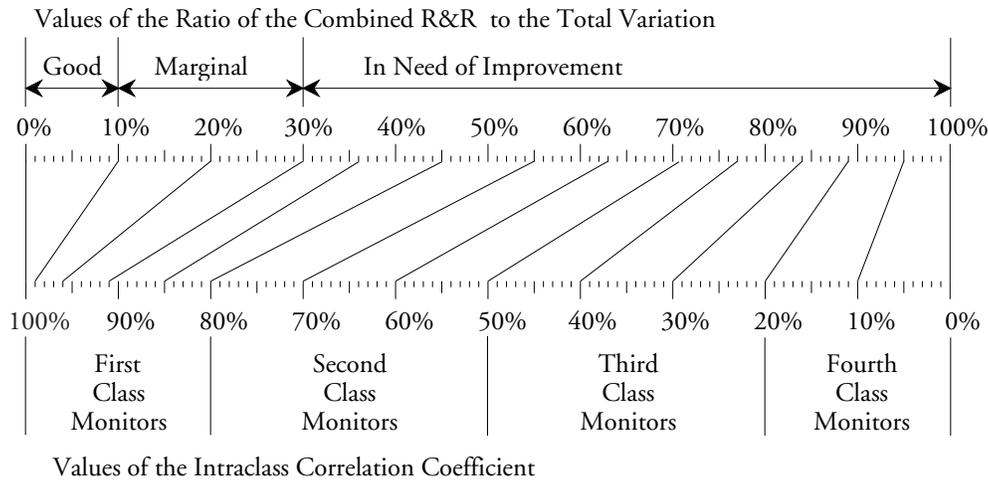


Figure 4: The AIAG Guidelines and the Four Classes of Process Monitors

Intraclass Correlation	Attenuation of Process Signals	Chance of Detecting a 3 Std. Error Shift	Ability to Track Process Improvements
1.00			
First Class Monitors	Less than 10 Percent	More than 99% with Rule One	Up to Cp80
0.80			
Second Class Monitors	From 10 % to 30 %	More than 88% with Rule One	Up to Cp50
0.50			
Third Class Monitors	From 30% to 55%	More than 91% w/ Rules 1, 2, 3, 4	Up to Cp20
0.20			
Fourth Class Monitors	More than 55 Percent	Rapidly Vanishes	Unable to Track
0.00			

Figure 5: The Four Classes of Process Monitors Defined by the Intraclass Correlation

The AIAG Gauge R&R Guidelines will pronounce a measurement system to be “good” only when it has an intraclass correlation in excess of 0.99; “marginal” measurement systems will have an intraclass correlation between 0.91 and 0.99; and everything else is “in need of improvement.” Contrast this to the results from the December 2010 column which are summarized in Figure 5. There we see that even third class monitors can be used to track process improvements.

When we compare the results in Figures 4 and 5 it becomes clear that the main purpose of the AIAG Gauge R&R Study is to condemn the measurement process. To achieve this end the summary ratios [6] through [12] are inflated so as to overstate the damage of measurement error while the guidelines used to interpret these inflated ratios are excessively conservative.

So what can you learn from the AIAG Gauge R&R Study? Virtually nothing that is true, correct, or useful! You have taken the time and gone to the trouble to collect good data, and then

you have wasted the information contained in those data by performing a hopelessly flawed analysis.

SUMMARY

The Pythagorean Theorem would need to be repealed before the summary ratios [6] through [12] of the AIAG Gauge R&R Study could even begin to make sense. Since that is unlikely to happen anytime soon, you need to stop using the erroneous ratios of the AIAG Gauge R&R Study.

The number of distinct categories value from ratio [13] does not represent anything that can be expressed in practical terms. So even though I may be the author of this ratio, it is useless in practice. I personally quit using it back in the 1980s. I suggest that you do the same, starting immediately.

The erroneous ratios [10], [11], and [12] overstate the impact of measurement error upon specifications. As shown in last month's column the actual relationship between measurement error and specifications is much more complex than these ratios imply. Large values for ratios [10], or [12] cannot be used to judge the utility of the measurement process.

Honest ratios [Honest Ratio 6] through [Honest Ratio 9] replace the erroneous ratios [6] through [9]. Honest ratio [Honest Ratio 9] is an estimate of the traditional, theoretically sound, intraclass correlation coefficient which correctly defines the relative utility of a measurement process for a given application..

Since many software packages currently give the nonsense AIAG ratios as part of their "measurement system analysis" you will need to identify these fallacious numbers on the output in order to avoid being misled by them.

This paper and my earlier papers provide you with practical, theoretically sound, and easy to use alternatives to the nonsense ratios of the AIAG Gauge R&R Study. They allow you to learn and use the correct formulas for evaluating the measurement process. Further documentation and explanation can be found in my book *EMP III: Evaluating the Measurement Process and Using Imperfect Data*.

However, if you still want to obfuscate and confuse the issues, if you want to needlessly condemn measurement systems, and if you want to continue to beat vendors over the head, then by all means continue to use the AIAG Gauge R&R Study where all the results are untrue, inappropriate, and wrong.

