An Honest Gauge R&R Study

Donald J. Wheeler

January 2009 Revision

In a 1994 paper Stanley Deming reported that when he did his literature search on measurement error he had expected to find a single, original paper followed by many extensions and additions in a logical stream of development. Instead, he found not one stream but many different rivulets, each developed without reference to each other, and most of which were done without the aid of any mathematical rationale or support. In short, the problem of measurement error is so widely recognized that many different solutions, in different fields of endeavor, have been proposed. These various solutions cover the spectrum from naive to theoretical, from simple to complex, and from wrong to right.

One such solution that has been widely promoted in the automotive world and beyond is known as a Gauge R&R Study. While this technique can be traced back to at least 1962, over the years it has been subject to many generations of changes. After being subjected to serious revisions over the past 20 years some of the more egregious mistakes have been eliminated. However, fundamental problems remain. This paper will illustrate these problems and propose an alternative that is both simple and correct.

1. The AIAG Gauge R&R Study

An example will be helpful in our discussion of a Gauge R&R Study. Generally a Gauge R&R Study will have two or more operators, one gauge, and up to ten parts. Each operator will then measure each part two or three times, resulting in a fully-crossed data structure with subgroups of size two or three. For our purposes let us say that we have three operators who will use a micrometer to measure the thickness of each of five gaskets twice. Thus, this study has $n = 2$, $o = 3$, and $p = 5$. The 30 data (in mils) are shown in Table 1.

Table 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

Since 1990 the Gauge R&R Study has been tended by the Auto Industry Action Group of the American Society for Quality. References to the current version refer to the Gauge R&R Study as it is presented and explained in the *Third Edition* of the AIAG's *Measurement System Analysis.* The following is a step-by-step discussion of the current version.

Step 1. The AIAG Gauge R&R Study begins by calculating the Average Range for the $k = 15$ subgroups of size $n = 2$ of Table 1 and finding the corresponding Upper Range Limit. Then the observed ranges are compared to this Upper Range Limit to see if any exceed this limit. For Table 1 the Average Range is 4.267 mils and the $URL = 13.9$ mils. None of the ranges in Table 1 exceed this limit.

Step 2. Next the Average Range from Step 1 is divided by the appropriate Bias Correction Factor to obtain an estimate of the standard deviation for measurement error. This estimate is called the **Repeatability** or the Equipment Variation (EV).

$$EV = \hat{\sigma}_{pe} = \frac{\bar{R}}{d_2} = \frac{4.267}{1.128} = 3.783 \text{ mils}$$

Step 3. Next the **Reproducibility** or Appraiser Variation (AV) is estimated. The formula used here has evolved over the years. The current version uses the range of the Operator Averages. For our example the $o = 3$ Operator Averages are 181.0, 172.5, and 173.9. The range of these three averages is $R_o = 8.5$ mils. The Bias Correction Factor for ranges used here is the Bias Correction Factor for estimating variances which is commonly known as d_2^*. For the range of three values this value is $d_2^* = 1.906$, and each of the Operator Averages was based upon $[(n\,o\,p)\,/\,o\,] = 10$ original data. The formula for the Reproducibility is:

$$AV = \hat{\sigma}_o = \sqrt{\left[\frac{R_o}{d_2^*}\right]^2 - \frac{o}{n\,o\,p}\,\hat{\sigma}_{pe}^2} = \sqrt{\left[\frac{8.5}{1.906}\right]^2 - \frac{3}{30}\,3.783^2} = 4.296 \text{ mils}$$

Step 4. Next the **Combined Repeatability and Reproducibility** (Gage R&R) is found by squaring the results of Steps 2 and 3, adding them together, and taking the square root:

$$GRR = \hat{\sigma}_e = \sqrt{EV^2 + AV^2} = \sqrt{3.783^2 + 4.296^2} = 5.724 \text{ mils}$$

Step 5. Next the **Product Variation** (PV) is estimated using the range of the $p = 5$ Part Averages. The Part Averages are 158.0, 206.167, 182.0, 184.833, and 148.0. Their range is $R_p = 58.167$. Once again the Bias Correction Factor for the range of five values is the one for estimating variances, $d_2^* = 2.477$. Thus the Product Variation is estimated to be:

$$PV = \hat{\sigma}_p = \frac{R_p}{d_2^*} = \frac{58.167}{2.477} = 23.483 \text{ mils}$$

Step 6. Finally the **Total Variation** (TV) is estimated by combining the Product Variation with the Repeatability and the Reproducibility to get:

$$TV = \hat{\sigma}_x = \sqrt{EV^2 + AV^2 + PV^2} = \sqrt{3.783^2 + 4.296^2 + 23.483^2} = 24.171 \text{ mils}$$

Up to this point everything is okay. While the estimators defined here are not the only estimators that could have been used, and while they are not always unbiased estimators, they do provide reasonable estimates for the quantities described. In fact, the estimator in Step 2 is Estimator 1 from Table 4. The estimator in Step 3 is the square root of Estimator 11 from Table 4, and the estimator in Step 5 is the square root of Estimator 6 from Table 4. (While the AIAG Gauge R&R computational worksheet simplifies the formulas by defining several "*K*-factors" the AIAG formulas are algebraically equivalent to those given on the preceding page.)

The train wreck begins when the AIAG Gauge R&R Study tries to use the estimates from the preceding steps to characterize relative utility. In the current version the quantities from Steps 2, 3, 4, and 5 are all expressed as a percentage of the last value, the Total Variation.

2. The "Percentages" of the Total Variation

Step 7. The Repeatability (EV) is divided by the Total Variation (TV) and multiplied by 100 to be expressed as a percentage. This ratio is labeled $\%EV$ and is interpreted as the percentage of the Total Variation that is consumed by Repeatability or Equipment Variation. For the data of Table 1 this computation will yield a value of:

$$\%EV = 100\,[\,3.783/24.171\,] = 15.65\%$$

Step 8. The Reproducibility (AV) is divided by the Total Variation (TV) and multiplied by 100 to get a value denoted by $\%AV$ that is interpreted as that percentage of the Total Variation that is consumed by Reproducibility. For the data of Table 1 this computation will yield a value of:

$$\%AV = 100\,[\,4.296/24.171\,] = 17.77\%$$

Step 9. The Combined Repeatability and Reproducibility (GRR) is divided by the Total Variation and multiplied by 100 to get a value denoted by $\%GRR$ that is interpreted as that percentage of the Total Variation that is consumed by Combined R&R. For the data of Table 1 this computation will yield a value of:

$$\%GRR = 100\,[\,5.724/24.171\,] = 23.68\%$$

Step 10. The Product Variation is divided by the Total Variation and multiplied by 100 to get a value denoted by $\%PV$ that is interpreted as that percentage of the Total Variation that is consumed by the Product Variation. For the data of Table 1 this computation will yield a value of:

$$\%PV = 100\,[\,23.483/24.171\,] = 97.15\%$$

Following these formulas in the current version of the AIAG manual there is 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 $\%EV$ and $\%AV$ do not add up to the $\%GRR$ because they are not proportions. Likewise, the $\%GRR$ and the $\%PV$ do not add up to 100% because they are not proportions. They are instead trigonometric functions. Figure 1 shows how the five estimates found in Step 2 through Step 6 are related.

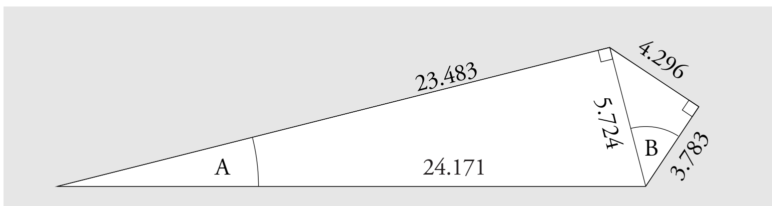


Figure 1: The Quantities Used for the Ratios in Steps 7, 8, 9, & 10

Before multiplication by 100, the ratio computed in Step 7 can be expressed in terms of Angles A and B as:

$$\frac{\%EV}{100} = \text{(Sine A) (Cosine B)} = \frac{5.724}{24.171}\,\frac{3.783}{5.724} = 0.1565$$

Before multiplication by 100, the ratio computed in Step 8 can be expressed in terms of Angles A and B as:

$$\frac{\%AV}{100} = \text{(Sine A) (Sine B)} = \frac{5.724}{24.171}\,\frac{4.296}{5.724} = 0.1777$$

Before multiplication by 100, the ratio computed in Step 9 can be expressed in terms of Angle A as:

$$\frac{\%GRR}{100} = \text{(Sine A)} = \frac{5.724}{24.171} = 0.2368$$

Before multiplication by 100, the ratio computed in Step 10 can be expressed in terms of Angle A as:

$$\frac{\%PV}{100} = \text{(Cosine A)} = \frac{23.483}{24.171} = 0.9715$$

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

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

A set of ratios are 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 the formula in Step 6. The Total *Variance* is the sum of the Repeatability *Variance*, the Reproducibility *Variance*, and the Product *Variance*.

$$\sigma_x^2 = \sigma_{pe}^2 + \sigma_o^2 + \sigma_p^2$$

Since these variances are additive, we know from the Pythagorean Theorem that the standard deviations *cannot be additive*. However, when the ratios of Steps 7, 8, 9, and 10 are expressed as percentages and interpreted as proportions they 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 Steps 7, 8, 9, and 10. It is why the "percentages" 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. They sound like nonsense because they are trigonometric functions being interpreted as proportions when they are *not* proportions.

SO WHAT ARE THE ACTUAL PROPORTIONS?

What proportion of the Total Variation is attributable to Repeatability or Equipment Variation? The equations above suggest that a reasonable estimate would be:

$$\frac{\hat{\sigma}_{pe}^2}{\hat{\sigma}_x^2} = \frac{EV^2}{TV^2} = \frac{3.783^2}{24.171^2} = 0.0245$$

or 2.45% rather than the 15.65% erroneously found in Step 7.

What proportion of the Total Variation is attributable to Reproducibility or Appraiser Variation? The equations above suggest that a reasonable estimate would be:

$$\frac{\hat{\sigma}_o^2}{\hat{\sigma}_x^2} = \frac{AV^2}{TV^2} = \frac{4.296^2}{24.171^2} = 0.0316$$

or 3.16% rather than the 17.77% erroneously found in Step 8.

What proportion of the Total Variation is attributable to Combined R&R? The equations above suggests that a reasonable estimate would be:

$$\frac{\hat{\sigma}_e^2}{\hat{\sigma}_x^2} = \frac{GRR^2}{TV^2} = \frac{5.724^2}{24.171^2} = 0.0561$$

or 5.61% rather than the 23.68% erroneously found in Step 9. Note that this value is the sum of the two preceding values.

What proportion of the Total Variation is attributable to Product Variation? The equations above suggest that a reasonable estimate would be:

$$\frac{\hat{\sigma}_p^2}{\hat{\sigma}_x^2} = \frac{PV^2}{TV^2} = \frac{23.483^2}{24.171^2} = 0.9438$$

or 94.38% rather than the 97.15% erroneously found in Step 10. When this value is added to the value attributable to Combined R&R above we effectively get 100%. This ratio of the Product Variance to the Total Variance is the traditional measure of relative utility introduced by Sir Ronald Fisher in 1921. It is commonly known as the Intraclass Correlation Coefficient.

Therefore, the ratios of Steps 7, 8, 9, and 10 of the AIAG Gauge R&R Study are completely and totally wrong. They are a triumph of computation over content. The numbers obtained do not represent what they are said to represent.

But what does the ratio of σ_e to σ_x actually represent? To understand this ratio we need to think about our model for a product measurement: Let X denote the observed value, let P denote the value of the thing being measured, and let E denote the measurement error associated with this particular observation: $X = P + E$. If something happened to shift the measurement system by an amount equal to 3 σ_e, then we would expect this shift to show up in future product measurements, changing the model to:

$$X = P + E + 3\,\sigma_e$$

Now if the ratio of σ_e to σ_x is, say, 0.10, then we could rewrite the expression above as:

$$X = P + E + 0.3\,\sigma_x$$

Thus, when the ratio of σ_e to σ_x is 0.10, a three sigma shift in the measurement system will look like a three-tenths of a sigma shift in the product measurements. Therefore the ratio that the AIAG Gauge R&R Study interprets as the Combined R&R actually characterizes the strength of a signal coming from the ***measurement system*** (rather than one from the production process.)

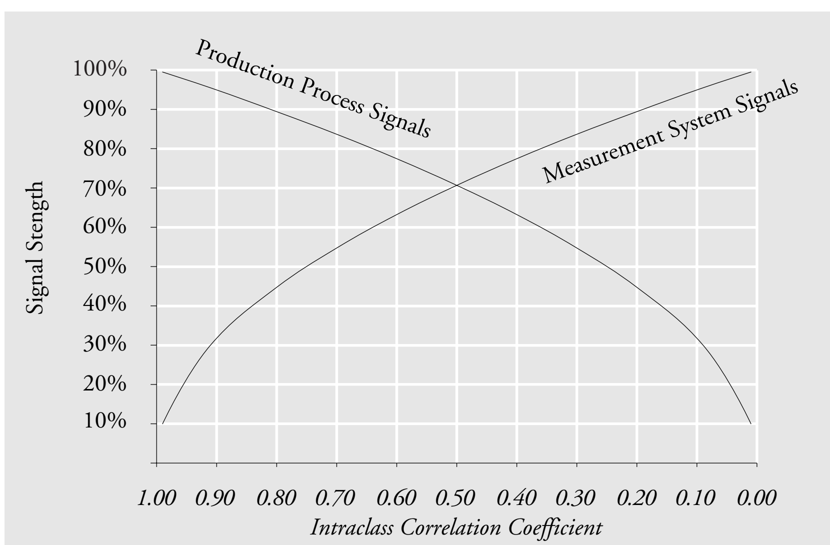


Figure : The Attenuation of Signals from Production and Measurement Systems

When plotted against the Intraclass Correlation Coefficient we have the ascending curve in Figure 2 as a representation of how *measurement system* signals are attenuated before they show up in the product measurements.

At the same time, and by the same logic, the ratio of σ_p to σ_x in Step 10 represents the signal strength for signals originating in the production process. In fact, the two signal strength curves are mirror images of each other, as shown in Figure 2. So when the $\%GRR$ value is 10%, any signal coming from the *production process* will show up at 99.5 percent of full strength, while signals coming from the measurement system will show up at 10 percent of full strength. Likewise, when the $\%GRR$ value is 30%, any signal coming from the *production process* will show up at 95.4 percent of full strength, while signals from the measurement system will only show up at 30 percent of full strength.

3. The "Percentages" of the Specified Tolerance

Prior to 1984 the Gauge R&R Study compared the estimates in Steps 2, 3, and 4 with the Specified Tolerance. Each of the estimates were multiplied by 5.15, divided by the Specified Tolerance, and multiplied by 100%. The current form of the Gauge R&R Study has a variation on these original ratios: instead of multiplying by 515%, the estimates are now multiplied by 600% and divided by the Specified Tolerance. To illustrate these ratios we will need to know that the specifications for the data of Table 1 are 175 mils –30/+50 mils.

Step 11. The Repeatability is multiplied by 6, divided by the Specified Tolerance, and multiplied by 100 to be expressed as a percentage. This ratio is still labeled $\%EV$ and is interpreted to be that percentage of the Specified Tolerance which is consumed by Repeatability or Equipment Variation. For the data of Table 1 this computation will yield a value of:

$$\%EV \;=\; 600\,\frac{3.783}{80} \;=\; 28.4\%.$$

This value is interpreted to mean that Repeatability consumes 28.4% of the Specified Tolerance.

Step 12. The Reproducibility is multiplied by 6, divided by the Specified Tolerance, and multiplied by 100 to be expressed as a percentage. This value is denoted by $\%AV$ and is interpreted to be that percentage of the Specified Tolerance that is consumed by Reproducibility. For the data of Table 1 this computation will yield a value of:

$$\%AV \;=\; 600\,\frac{4.296}{80} \;=\; 32.2\%.$$

This value is interpreted to mean that Reproducibility consumes 32.2% of the Specified Tolerance.

Step 13. The Combined Repeatability and Reproducibility is multiplied by 6, divided by the Specified Tolerance, and multiplied by 100 to be expressed as a percentage. This value is denoted by %*GRR* and is interpreted to be that percentage of the Specified Tolerance that is consumed by Combined R&R. For the data of Table 1 this computation will yield a value of:

$$\%GRR \;=\; 100\;\frac{6\,\hat{\sigma}_e}{USL - LSL} \;=\; 600\;\frac{5.724}{80} \;=\; 42.9\%.$$

This value is interpreted to mean that Combined Repeatability and Reproducibility consumes 42.9% of the Specified Tolerance.

Inspection of the ratios above will reveal that the numerators are the same as the numerators of Steps 7, 8, and 9. Since these numerators still do not add up, these "percentages" will not add up, and the interpretations above are flawed in the same way that those of Steps 7 through 10 were flawed.

THE PRECISION TO TOLERANCE RATIO

The ratio in Step 13, expressed as a proportion, is commonly known as the Precision to Tolerance, or *P/T*, ratio. If you inverted this ratio you would have that Capability Ratio which corresponds to an Intraclass Correlation of zero, which could be denoted as C_{p00}. This value is an unattainable upper bound for the Capability Ratio.

The question addressed by Steps 11, 12, and 13 is "How does measurement error affect the use of the specifications?" This question is carefully and completely answered in Wheeler [2004 and 2006]. There I rigorously demonstrate that 99% Manufacturing Specifications would consume an amount equal to 6 Probable Errors. Since a Probable Error is defined as 0.675 σ_e, six Probable Errors is equivalent to 4.04 σ_e of the Watershed Tolerance rather than the 6 σ_e of Step 13. This results in the following proportions.

To get 99% Manufacturing Specifications we tighten the Watershed Specifications by 3 PE on each end. The Watershed Specifications are 144.5 to 225.5. If there were no Operator Bias present, and measurement error consisted of Repeatability alone, then this adjustment for the uncertainty introduced by measurement error would consume:

$$\frac{4.04\;\hat{\sigma}_{pe}}{\text{Watershed Tolerance}} \;=\; \frac{4.04\,(3.783)}{81} \;=\; 0.189$$

or 18.9% of the Watershed Tolerance rather than the 28.4% found in Step 11.

If we do not eliminate the Operator Bias, then measurement error will consist of both Repeatability and Reproducibility. In this case 99% Manufacturing Specifications would consume:

$$\frac{4.04\;\hat{\sigma}_{e}}{\text{Watershed Tolerance}} \;=\; \frac{4.04\,(5.724)}{81} \;=\; 0.285$$

or 28.5% of the Watershed Tolerance rather than the 42.9% found in Step 13.

Comparing the two values above, we see that the incremental degradation of the Watershed Tolerance due to the presence of the Operator Bias is:

$$0.285 \;-\; 0.189 \;=\; 0.096$$

which is 9.6% of the Watershed Tolerance, rather than the 32.2% found in Step 12.

Thus, the values computed in Steps 11, 12, and 13 of the AIAG Gauge R&R Study overstate the impact of measurement error upon the specifications.

4. The "Number of Distinct Categories"

Step 14. The current version of the Gauge R&R Study includes a quantity known as the "number of distinct categories:"

$$\textit{number of distinct categories} \;=\; ndc \;=\; 1.41\,\frac{PV}{GRR}$$

For the data of Table 1 this value is *ndc* = 1.41 (23.483/5.724) = 5.8. The AIAG manual suggests rounding this value off to an integer, and that values of 5 or greater are "good."

This *ndc* value is an estimate of the Classification Ratio:

$$\textit{Classification Ratio} \;=\; C_R \;=\; \sqrt{2}\,\frac{\sigma_p}{\sigma_e}$$

The Classification Ratio was originally used in the First Edition of *Evaluating the Measurement Process* as a numerical summary for the relative utility displayed by the Average Chart in an EMP Study. It also provided a simple approximation for the Discrimination Ratio, since:

$$C_R \;=\; \sqrt{D_R{}^2 - 1.0}$$

Unfortunately, as this author has discovered after much effort, neither the Discrimination Ratio nor the Classification Ratio has a simple interpretation in practice. While the Discrimination Ratio does characterize the relative sizes of the major and minor axes in the Intraclass Correlation Plot, this plot does not lend itself to any practical interpretation. Moreover, neither the Discrimination Ratio nor the Classification Ratio defines the number of distinct categories. As a result, like everything else after Step 6 of the AIAG Gauge R&R Study, the "number of distinct categories" value is not what it claims to be.

5. The Guidelines

In spite of the many revisions made in the Gauge R&R Study over the years the guidelines for interpreting the proportions that are not really proportions has remained unchanged. Whether they were applied to the ratios from Steps 7, 8, and 9, or applied to the ratios from Steps 11, 12, and 13, 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.

These guidelines give the results noted in Table 2. There we find that the measurement system in Table 1 to be simultaneously good, marginal, and unacceptable! Any questions?

Table 2 The "Percentages" Obtained from the Gauge R&R Study for Gasket Thickness.

	As % of Specified Tolerance	As % of Total Variation
Repeatability	28.4% = **marginal**	15.7% = **marginal**
Reproducibility	32.2% = **unacceptable**	17.8% = **marginal**
Combined R & R	42.9% = **unacceptable**	23.7% = **marginal**
Part-to-Part Variation		97.2%
no. distinct categories		5.8 = **good**

Prior to 1984 only the first column of Table 2 was computed by the Gauge R&R Study. After 1984, when the values in the second column began to be included, the guidelines that had always been used for the first column were simply applied to the second column as well. This was done in spite of the fact that the values obtained in the two columns are not even remotely alike. This is presumably one of the benefits of using a guideline that has 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.

FOUR CLASSES OF PROCESS MONITORS

The two curves in Figure 2 suggest that we will need four classes to fully describe the relative utility of a measurement system for a given application. First Class Monitors will have only slight attenuation for signals coming from the production process (less than 10%) while signals from the measurement system are greatly attenuated (more than 55%). First Class Monitors will have an Intraclass Correlation between 1.00 and 0.80.

Second Class Monitors will have Intraclass Correlations between 0.80 and 0.50. In this region signals coming from the production process will be attenuated between 10% and 30%, while signals coming from the measurement system will be attenuated from 55% to 30%.

Third Class Monitors will have an Intraclass Correlation between 0.50 and 0.20. Signals from the production process will be attenuated between 30% and 55%, while signals from the measurement system will only be attenuated from 30% to 10%.

Fourth Class Monitors will have an Intraclass Correlation below 0.20. Here any signals coming from the production process will be very highly attenuated (more than 55%), while signals coming from the measurement system will come through at nearly full strength.

HOW DO THE GUIDELINES
MATCH UP WITH THE FOUR CLASSES OF MONITORS ?

For purposes of comparison let us consider the third ratio in the last column of Table 2. The Combined R&R divided by the Total Variation (expressed as a proportion) is an estimate of:

$$\frac{\sigma_e}{\sigma_x} = \sqrt{1-\rho}$$

If we plotted the values for the ratio of σ_e to σ_x on one scale, and plotted the Intraclass Correlation values, ρ, in descending order on another scale we could show the relationship between these two scales graphically as in Figure 3. Corresponding values on the two scales are connected by the diagonal lines. The AIAG Guidelines establish three categories: Good, Marginal, and Unacceptable.

Using the correspondence between the two scales we see that the AIAG Guidelines approve of those applications having an Intraclass Correlation of 0.99 or better. Applications with an Intraclass Correlation between 0.91 and 0.99 are marginal, and everything else is unacceptable.

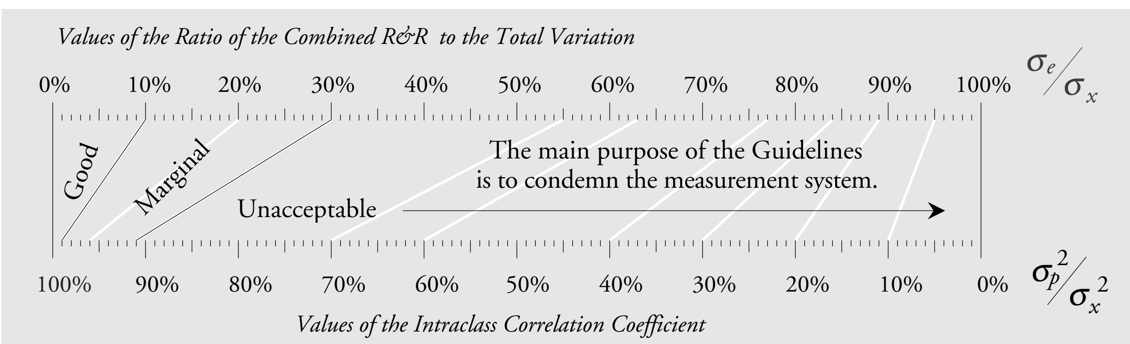


Figure 3: The Conservative Nature of the AIAG Guidelines

Using the same scales as Figure 3, Figure 4 shows the Four Classes of Monitors. There we see that whenever the σ_e to σ_x ratio is less than 45% we will have a First Class Monitor. Whenever the σ_e to σ_x ratio is between 45% and 71% we will have a Second Class Monitor. And whenever the σ_e to σ_x ratio is between 71% and 89% we will have a Third Class Monitor. Hence, the Guidelines from the AIAG manual are excessively conservative.

For those who are accustomed to using the AIAG Guidelines, it should be noted that First Class Monitors are those for which Production Process Signals will be attenuated by about 10% or less. Likewise, Second Class Monitors are those for which Production Process Signals will be attenuated by approximately 10% to 30% (see Figure 2). Thus, there is an actual congruence between the *concept* behind the AIAG Guidelines and the Four Classes of Monitors. Unfortunately, back in the 1960s, the original authors of the Gauge R&R Study simply did not know how to compute the right values.

Thus, the AIAG Guidelines are overly conservative and do not reflect any underlying reality. The Four Classes of Monitors can be used to describe what really happens in practice. Moreover, the Four Classes of Monitors allow you to use your less than perfect data to actually improve your production processes. Since money spent on measurement systems is an overhead expense, this ability to work with less than perfect data is a major reason to cease using the AIAG Guidelines.

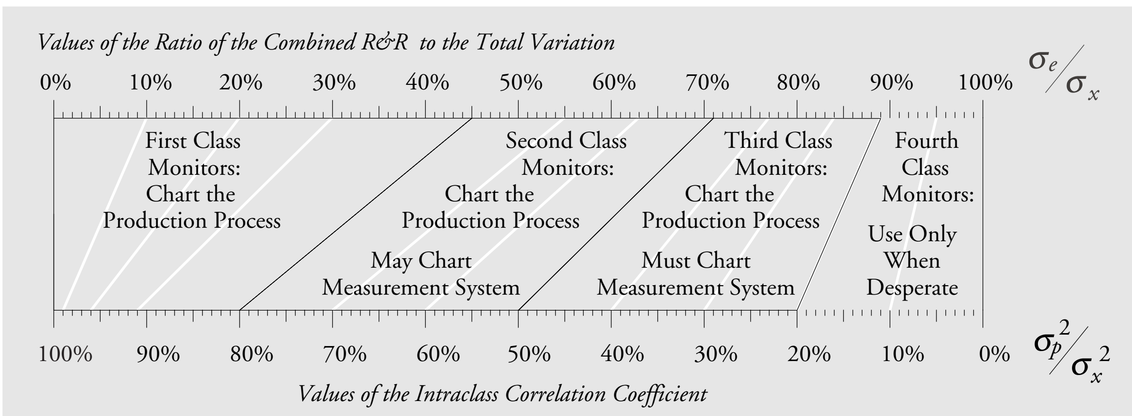


Figure 4: The Four Classes of Process Monitors

6. What Can You Learn From a Gauge R&R Study?

The inconsistency of the results in Table 2 is an inherent feature Gauge R&R Studies. Once you ignore the Pythagorean Theorem, mere guidelines are not going to remedy the situation. After Step 6, every aspect of the Gauge R&R Study is erroneous, fallacious, naive, incorrect, and, to put it quite simply, wrong.

Table 3 What Can You Learn from a Gauge R&R Study?

For the Gasket Thickness Data of Table 1:	Gauge R&R	The Truth
Step 7: % Total Variation due to Repeatability	15.7 %	2.4 %
Step 8: % Total Variation due to Reproducibility	17.8 %	3.2 %
Step 9: % Total Variation due to Combined R&R	23.7 %	5.6 %
Step 10: % Total Variation due to Product Variation	97.2 %	94.4 %
Step 11: % Specified Tolerance used by Repeatability	28.4 %	18.9 %
Step 12: % Specified Tolerance used by Reproducibility	32.2 %	9.6%
Step 13; % Specified Tolerance used by Combined R&R	42.9 %	28.5 %

Clearly, the main purpose of a Gauge R&R Study is to condemn the measurement process, and to achieve this end the summary measures are inflated so as to overstate the damage of measurement error while the guidelines used 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.

So what should you do? The least you can do is to stop using the AIAG Gauge R&R Study and start using an Honest Gauge R&R Study, as described in the next section.

7. An Honest Gauge R&R Study

I. DATA COLLECTION

Let o operators measure each of p parts n = 2, 3, or 4 times each.
Arrange these [$n\ o\ p$] data into $k = o\ p$ subgroups of size n,
and compute the range for each of these k subgroups.

Honest Step 1. Use the Average Range for the k subgroups of size n to find the Upper Range Limit. If any of the subgroup ranges exceed this upper limit you need to find out why.

$$\textit{Upper Range Limit} = D_4 \bar{R}$$

II. ESTIMATES OF VARIANCE COMPONENTS

Honest Step 2. Use the Average Range from Honest Step 1 to estimate the Repeatability Variance Component:

$$\hat{\sigma}_{pe}^2 = \left[\frac{\bar{R}}{d_2} \right]^2$$

Honest Step 3. Use the range of the *o* Operator Averages to estimate the Reproducibility Variance Component:

$$\hat{\sigma}_o^2 = \left\{ \left[\frac{R_o}{d_2^*} \right]^2 - \frac{o}{n\,o\,p}\ \hat{\sigma}_{pe}^2 \right\}$$

Honest Step 4. Add the estimates above to get the estimated Combined R&R Variance Component:

$$\hat{\sigma}_e^2 = \hat{\sigma}_{pe}^2 + \hat{\sigma}_o^2$$

Honest Step 5. Use the range of the *p* Part Averages to estimate the Product Variance Component:

$$\hat{\sigma}_p^2 = \left[\frac{R_p}{d_2^*} \right]^2$$

Honest Step 6. Add the estimates in Honest Steps 4 and 5 to get the estimated Total Variance :

$$\hat{\sigma}_x^2 = \hat{\sigma}_p^2 + \hat{\sigma}_e^2$$

III. CHARACTERIZING RELATIVE UTILITY

Honest Step 7. That proportion of the Total Variance that is consumed by Repeatability is:

$$\text{Repeatability Proportion} = \frac{\hat{\sigma}_{pe}^2}{\hat{\sigma}_x^2}$$

Honest Step 8. That proportion of the Total Variance that is consumed by Reproducibility is:

$$\text{Reproducibility Proportion} = \frac{\hat{\sigma}_o^2}{\hat{\sigma}_x^2}$$

Honest Step 9. That proportion of the Total Variance that is consumed by the Combined Repeatability and Reproducibility is:

$$\text{Combined R\&R Proportion} = \frac{\hat{\sigma}_e^2}{\hat{\sigma}_x^2}$$

Honest Step 10. That proportion of the Total Variance that is consumed by Product Variation is an estimate of the Intraclass Correlation Coefficient:

$$\text{Proportion due to Product Stream} = \text{Intraclass Correlation} = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_x^2}$$

IV. INTERPRETING THE RESULTS

Honest Step 11. Characterize this application of the measurement system as a First, Second, Third, or Fourth Class Monitor.
Describe the attenuation of process signals,
the chances of detecting a three standard error shift,
and the ability to track process improvements by computing
the appropriate Crossover Capabilities.

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
0.80				C_{p80}
	Second Class Monitors	From 10 % to 30 %	More than 88% with Rule One	Up to
0.50				C_{p50}
	Third Class Monitors	From 30% to 55%	More than 91% w/ Rules 1, 2, 3, 4	Up to
0.20				C_{p20}
	Fourth Class Monitors	More than 55 Percent	Rapidly Vanishes	Unable to Track
0.00				

$$C_{p80} = \frac{USL - LSL}{6\,\sigma_{pe}} \sqrt{1-.80} \qquad C_{p50} = \frac{USL - LSL}{6\,\sigma_{pe}} \sqrt{1-.50} \qquad C_{p20} = \frac{USL - LSL}{6\,\sigma_{pe}} \sqrt{1-.20}$$

Honest Step 12. Use the estimated Repeatability Variance Component from Honest Step 2 to find an estimate of the Probable Error:

$$Probable\ Error = 0.675 \sqrt{\hat{\sigma}_{pe}^{\,2}}$$

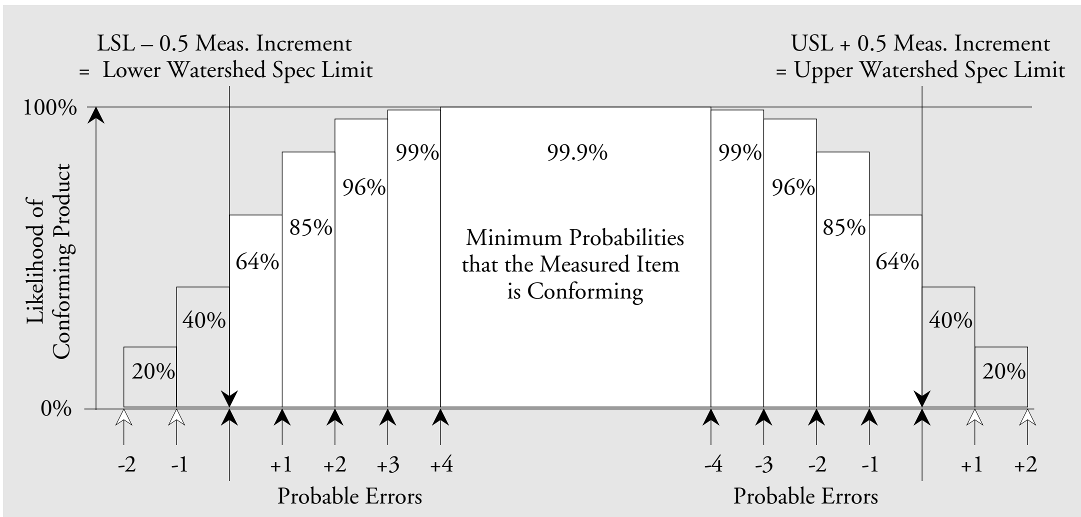
Honest Step 13. Compute the bounds for the Measurement Increment:

Smallest Effective Measurement Increment = 0.2 *Probable Errors* =

Largest Effective Measurement Increment = 2 *Probable Errors* =

Is the Measurement Increment within the interval above?
If not, add or drop digits as needed to change the
Measurement Increment used with future measurements.

Honest Step 14. 96% Manufacturing Specifications will consist of the Watershed Specifications tightened by 2 Probable Errors on each end.
If Manufacturing Specifications are needed,
you may choose your own trade-off between the degree of certainty and the width of the limits to compute appropriate Manufacturing Specifications.



For Table 1, o = 3, p = 5, and n = 2, and the steps of the Honest Gauge R&R Study yield:

H S 1. The Average Range is 4.267 mils, and no ranges fall above the *URL* of 13.9 mils.

H S 2. Repeatability: $\hat{\sigma}_{pe}^2 = (\ 3.7825\ )^2 = 14.307$

H S 3. Reproducibility: $\hat{\sigma}_o^2 = \left\{ \left[\frac{8.5}{1.906} \right]^2 - \frac{3}{30}\ (3.7825)^2 \right\} = 18.457$

H S 4. Combined R&R: $\hat{\sigma}_e^2 = \hat{\sigma}_{pe}^2 + \hat{\sigma}_o^2 = 32.765$

H S 5. Product Variance: $\hat{\sigma}_p^2 = \left[\frac{58.167}{2.477} \right]^2 = 551.444$

H S 6. Total Variation: $\hat{\sigma}_x^2 = \hat{\sigma}_p^2 + \hat{\sigma}_e^2 = 584.209$

H S 7. Repeatability Proportion of Total Variation: $\frac{14.309}{584.209} = 0.0245$

H S 8. Reproducibility Proportion of Total Variation: $\frac{18.457}{584.209} = 0.0316$

H S 9. Combined R&R Proportion of Total Variation: $\frac{32.765}{584.2093} = 0.0561$

H S 10. Intraclass Correlation Statistic: $r_o = \frac{551.444}{584.2093} = 0.9439$

H S 11. This gauge is a First Class Monitor for these gasket thicknesses.
Production process signals will be attenuated by $1 - \sqrt{.9439} = 0.028$ or 2.8%.
Better than 99% chance of detecting a three standard error shift.
Can track process improvement up to $C_{p80} = 1.58$ while a First Class Monitor.
Can track process improvement up to $C_{p50} = 2.49$ while a Second Class Monitor.
Can track process improvement up to $C_{p20} = 3.16$ while a Third Class Monitor.

H S 12. The Probable Error of a single measurement is $PE = 0.675\sqrt{14.31} = 2.55$ mils.

H S 13. The Smallest Effective Measurement Increment is 0.2 PE = 0.51 mils.
The Largest Effective Measurement Increment is 2 PE = 5.1 mils.
The data are recorded to the nearest 1.0 mil, which is appropriate.

H S 14. The Specifications Limits are 145 and 225 mils.
The Watershed Specifications are 144.5 and 225.5 mils.
96% Manufacturing Specifications are thus 149.6 mils to 220.4 mils,
resulting in effective 96% Manufacturing Specifications of 150 mils to 220 mils.

Thus we have characterized this measurement system in six practical ways. It is instructive to compare the clarity of the statements in the last four steps of the Honest Gauge R&R Study with the superstitious nonsense obtained from the AIAG Gauge R&R Study.

The Honest Gauge R&R Study is every bit as easy to perform as the AIAG study, yet it gives the correct values from Table 3 rather than the incorrect values. In addition it also provides other useful information about the measurement system.

In the interest of simplicity the estimators used in Honest Steps 2, 3, 4, 5, and 6 paralleled those in the AIAG study. There are several alternate formulas that could be used for estimating the variance components in an R&R type of study. Several of these are listed in Table 4. The Honest R&R Study given here uses Formulas 2, 11, and 6 (under the condition that the number of determinations used in practice to obtain a single reported product measurement is $n_r = 1$).

Formula 2 could be replaced with Formula 3 or Formula 4 in practice and you would still have the right results.

Formula 6 could be replaced with either Formula 5, Formula 7, or Formula 8 without appreciable impact.

Finally Formula 11 could be replaced by Formula 12 without affecting the results of the study. (Formulas 9 and 10 are rarely used because they will commonly contain substantial bias.)

Table 4 Different Formulas for Estimating Variation in EMP Studies

	Parameter	Estimator	Comments	Value for Example
1.	σ_{pe}	$\frac{\bar{R}}{d_2}$	simple, unbiased, reasonably efficient	3.783
2.*	σ_{pe}^2	$\left[\frac{\bar{R}}{d_2}\right]^2$	simple, biased, commonly used	$(3.783)^2$
3.	σ_{pe}^2	$\left[\frac{\bar{R}}{d_2^*}\right]^2$	unbiased, rarely used	$(3.714)^2$
4.	σ_{pe}^2	Mean Square Within	unbiased, found in ANOVA	$(4.154)^2$
5.	σ_p^2	$\left[\frac{R_p}{d_2}\right]^2$	easy, biased, OK for 1st and 2nd Class Monitors	$(25.007)^2$
6.*	σ_p^2	$\left[\frac{R_p}{d_2^*}\right]^2$	simple, biased, OK for 1st and 2nd Class Monitors	$(23.483)^2$
7.	σ_p^2	$\left\{\left[\frac{R_p}{d_2^*}\right]^2 - \frac{p}{n\,o\,p}\hat{\sigma}_{pe}^2\right\}$	better, unbiased, use for 3rd, 4th Class Mon.	$(23.432)^2$
8.	σ_p^2	$\left\{s_p^2 - \frac{p}{n\,o\,p}\hat{\sigma}_{pe}^2\right\}$	best, unbiased, use for 3rd, 4th Class Mon.	$(23.034)^2$
9.	σ_o^2	$\left[\frac{R_o}{d_2}\right]^2$	biased and rarely used	$(5.020)^2$
10.	σ_o^2	$\left[\frac{R_o}{d_2^*}\right]^2$	biased and rarely used	$(4.460)^2$
11.*	σ_o^2	$\left\{\left[\frac{R_o}{d_2^*}\right]^2 - \frac{o}{n\,o\,p}\hat{\sigma}_{pe}^2\right\}$	better, unbiased, recommended	$(4.296)^2$
12.	σ_o^2	$\left\{s_o^2 - \frac{o}{n\,o\,p}\hat{\sigma}_{pe}^2\right\}$	best, unbiased, recommended	$(4.398)^2$

* Formulas 2, 11, and 6 were used in the Honest R&R Worksheet. The use of other formulas would only result in slight changes in the values found.

8. Summary

The Honest Gauge R&R Study uses the traditional, mathematically correct, measure of relative utility, the Intraclass Correlation to characterize the ability of a measurement system to capture meaningful information about a production stream. The estimated Intraclass Correlation can be used to establish Four Classes of Process Monitors.

The amount by which a signal coming from the production process is attenuated by the effects of measurement error may be estimated by:

$$\textit{Production Process Signal Attenuation} \; = \; 1 - \sqrt{\textit{Intraclass Correlation}}$$

The likelihood of detecting a shift in the production process can be approximated using the Four Classes of Process Monitors.

The ability of the measurement system to track process improvements can be quantified by using the Crossover Capabilities.

In addition, the Honest Gauge R&R Study also provides an absolute characterization of the quality of the values obtained from the measurement system. The Probable Error defines the essential uncertainty to attach to any measurement. With this number you can establish how many digits need to be recorded, quantify how aggressively to interpret any value, and establish reasonable and proper manufacturing specifications.

So if you 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.

However, if you want to learn how to use imperfect data to make better products, if you want to get the most out of your imperfect measurement systems, and if you want clear answers to straightforward questions, then you need to learn how to use an Honest Gauge R&R Study.

9. References

Wheeler, Donald J., ***Advanced Topics in Statistical Process Control**, **Second Edition***, SPC Press, Knoxville, Tenn., 2004.

Wheeler, Donald J., ***EMP III Using Imperfect Data***, SPC Press, Knoxville, Tenn., 2006.

———, ***Measurement Systems Analysis, Third Edition***, Automotive Industry Action Group, Automotive Division, American Society for Quality, Milwaukee, Wisc. 2002.

Published in ***SPC Ink March 2008***, revised Jan. 2009.
www.spcpress.com/pdf/DJW189.pdf