Appraiser Variation in Gage R&R Measurement

by **Donald S. Ermer**

In part one of this column,[1] I said if our data analysis is inaccurate, it does not represent the true quality characteristics of the part or product being measured, even if we're using quality improvement tools correctly.

Therefore, it is important to have a valid quality measurement study beforehand to ensure the part or product is accurate and the power of statistical process control and design of experiments is fully used. Accuracy—in other words, the absence of bias—is the function of calibration, which is performed before the precisions of the gage and its operators are measured.

In part one, I reviewed the gage R&R study in the Automotive Industry Action Group (AIAG) manual[2] for its weakness in determining the true capability of the different parts of a measurement system and used a geometrical approach to describe the components of total measurement variance. This showed why the standard deviations or measurement errors of the equipment, appraiser and part in the AIAG method are not additive and cannot be compared directly in a ratio.

In part two, I'm providing a worksheet for correctly executing a measurement process capability study (Tables 1 and 2, p. 76).

Introduction of Appraiser Variation

Although the proper variances in part one indicated the capability of the measurement process, they are good only when one operator is measuring the product, which makes appraiser variance insignificant. In actual situations, it is difficult to isolate or eliminate appraiser error in the measurement process. Therefore,

Part two of a two-part series on a geometrical approach.

it is necessary to include appraiser variance. This is shown as BC in Figure 1 (p. 77) and BE in Figure 2 (p. 77). It is calculated as

$$\hat{\sigma}_a = \frac{R_0}{d^*_{2,a}} = \text{appraiser variation}$$

in which R_0 = range of the appraiser averages and $d^*_{2,a}$ = 1.41 and 1.91 (see first row of Table 3 of part one) for two and three appraisers respectively.

With the addition of appraiser variation, the relationship among all the variances is changed to:

$$\sigma^2_{m'} = \sigma^2_p + \sigma^2_e + \sigma^2_a$$

in which $\sigma^2_{m'}$ = new total product measurement variation (assuming no interaction between parts and appraiser).

The relationships among all the variations can be illustrated in Figure 1 or 2. The total product measurement error ($\sigma^2_{m'}$) will respond with one unit change when there is a unit change in gage, appraiser or true product variance—in other words,

TABLE 1 Raw Data for Comparison of Different R&R Studies

I. Data set A										
Operator	A	A	WR(A)	B	B	WR(B)	C	C	WR(C)	Part average
Trial	1	2		1	2		1	2		
Part one	67	62	5	55	57	2	52	55	3	58.0
Part two	110	113	3	106	99	7	106	103	3	106.2
Part three	87	83	4	82	79	3	80	81	1	82.0
Part four	89	96	7	84	78	6	80	82	2	84.8
Part five	56	47	9	43	42	1	46	54	8	48.0
Within range average ($\overline{WR}$)			5.6			3.8			3.4	
Appraiser average		81.0			72.5			73.9		
Overall within range average ($\overline{\overline{WR}}$)			4.267							
Range of part average (R_p)			58.167							
Range of appraiser average, $\bar{X}_{diff}$ (R_0)			8.5							

WR = within range R&R = repeatability and reproducibility

Note: See "Improved Gage R&R Measurement Studies," Table 2, *Quality Progress*, March 2006, p. 78, for results of studies using these data.

TABLE 2 Gage R&R Output for New Method

General Information

Part number and name: Data set A	Gage name:	Date:
Characteristics:	Gage number:	Performed by:
Specification:	Gage type:	Plant:
Tolerance:	Gage calibration expiration date:	Gage resolution:

Part Information from Data Sheet in Table 1

Number of trials (r) = 2	Number of appraisers (k) = 3	Number of parts (n) = 5	
$\overline{\overline{WR}} = 4.267$	$R_o = 8.5$	$R_p = 58.167$	$\overline{\overline{X}} = 75.8$

Measurement Unit Analysis

Repeatability—equipment variation (EV) $EV = \frac{\overline{\overline{WR}}}{d^*_{2,e}}$ $= 4.267/1.15$ $= 3.71 = \sigma_e$	$\frac{(EV)^2}{(TV)^2} \times 100 = \frac{(3.710)^2}{(24.11)^2} \times 100 = 2.45\%$
Reproducibility—appraiser variation (AV) $AV = \sqrt{\left(\frac{R_o}{d^*_{2,o}}\right)^2 - \left(\frac{E.V.^2}{n \times r}\right)}$ $= \sqrt{\left(\frac{8.5}{1.91}\right)^2 - \left(\frac{13.76}{10}\right)}$ $= 4.293 = \sigma_a$	$\frac{(AV)^2}{(TV)^2} \times 100 = \frac{(4.293)^2}{(24.11)^2} \times 100 = 3.09\%$
Repeatablilty and reproducibility (R&R) $R\&R = \sqrt{EV^2 + AV^2}$ $= \sqrt{(3.710)^2 + (4.293)^2}$ $= 5.674$	$\frac{(R\&R)^2}{(TV)^2} \times 100 = \frac{(5.674)^2}{(24.11)^2} \times 100 = 5.54\%$
Reproducibility—part or product variation (PV) $PV = \sqrt{\left(\frac{R_p}{d^*_{2,m}}\right)^2 - \left(\frac{EV^2}{k \times r}\right)}$ $= \sqrt{\left(\frac{58.167}{2.48}\right)^2 - \left(\frac{13.76}{6}\right)}$ $= 23.43 = \sigma_a$	$\frac{(PV)^2}{(TV)^2} \times 100 = \frac{(23.43)^2}{(24.11)^2} \times 100 = 94.46\%$
Total variation (TV) $TV = \sqrt{(R\&R)^2 + (PV)^2}$ $= \sqrt{(5.674)^2 + (23.43)^2}$ $= 24.11$	Check: $94.46\% + 3.09\% + 2.45\% = 100\%$

not with standard deviations.

In the AIAG study, the constants $d_{2,e}$, $d_{2,m}$ and $d_{2,o}$ are all assumed equal to d_2 for the different sample sizes in the subgroup. However, these three values may be equal to either d_2 or d_2^*, depending on the number of subgroups and sample size. If the number of subgroups is greater than or equal to 25, then d_2 should be used in the calculation. Otherwise, d_2^* is used.

The number of subgroups and subgroup size depend on the number of parts, operators and trials used in the R&R study. For $d_{2,e}$, the total number of within ranges used to calculate the average is the number of subgroups (nxk), while the number of trials (r) of each part will be the sample or subgroup size. For $d_{2,m}$ and $d_{2,o}$, the number of subgroups is always equal to 1, and the sample size is the number of parts (n) tested or the number of operators (k) in the measurement study.

For example, if a measurement study used five parts, as in Table 1 (p. 75), with each part measured twice by each of the three operators, then $d_{2,e}$ would be based on only k = 15 subgroups for the sample size equal to 2 (and $d_{2,e}^* = 1.15$ in Table 3 of part one). The $d_{2,m}$ value would be based on only one subgroup (k = 1) and a sample size of 5 (and $d_{2,m}^* = 2.48$), while $d_{2,o}$ would also be based on only one subgroup (k = 1) and a sample size of 3 (and $d_{2,o}^* = 1.91$). Therefore, for this example, $d_{2,e}$, $d_{2,m}$ and $d_{2,o}$ should all use d_2^* instead of d_2. The values of d_2^* (and d_2) are given in Table 3 of part one.

In addition, a more accurate estimate of appraiser variance should be obtained. Use a correction factor to eliminate the contamination caused by the measurement equipment variance in the data. The modified equation is:

$$\hat{\sigma}_a^2 = \left(\frac{R_o}{d^*_{2,a}}\right)^2 - \frac{\hat{\sigma}_e^2}{(n)(r)}$$

in which $\frac{\hat{\sigma}_e^2}{(n)(r)}$ is the correction factor (CF 1).

The estimation of the true product or part variation can be improved by also including a correction factor in its calculation, although it will not be large. The correction factor is similar to CF 1 but with a different denominator in the last term. The improved estimation of the part or product variation is:

$$\hat{\sigma}_p^2 = \left(\frac{R_p}{d^*_{2,m}}\right)^2 - \frac{\hat{\sigma}_e^2}{(k)(r)}$$

in which $\frac{\hat{\sigma}_e^2}{(k)(r)}$ is the relatively small correction factor (CF 2).

Given these changes, the new measurement study will be more accurate and correct. Therefore, the new method should be used for a proper measurement process study, as in Table 2.

Solving Identified Problem Areas in the Measurement

For a measurement process with a problem in the equipment/gage variation area, there are several steps to check to find the root cause of the problem. The first is to check whether the measurement system has an adequate number of decimal places, meaning it has a resolution good enough for measuring the product variation.[3] If a problem of resolution occurs, consider using a measurement unit smaller than the gage standard division. For example, if the measurement unit of a data set is 0.01, then the gage standard division must be smaller than 0.01, such as 0.001, to have a resolution good enough for the R&R study.

Another way to improve gage accuracy is to calibrate the gage regularly. Although most measurement gage manufacturers provide calibration services to their customers, it is the gage user's responsibility to make sure the

FIGURE 1 Relationships of the Measurement Standard Deviation in Three-Dimensional Space

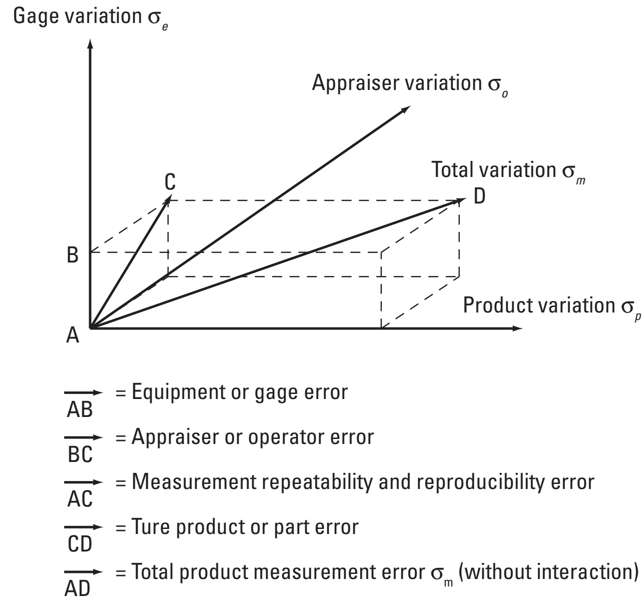
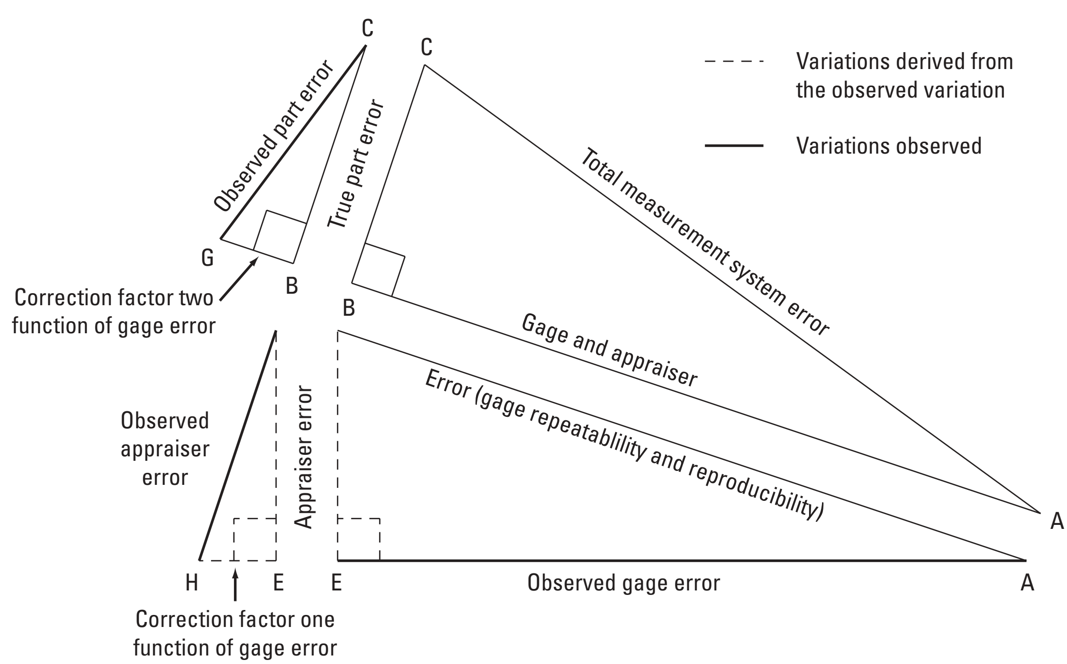


FIGURE 2 Alternate Illustration of Total Measurement System Errors



gage is calibrated before a gage R&R study. The user also should make sure the gage is performing at the standard claimed by the manufacturer.

When appraiser bias effect is detected, the problem can be temporarily solved by offsetting the amount of bias to all the measurements made by that appraiser. However, the long-term solution is to understand why that appraiser has a bias on all the measurements. When appraiser inconsistency is detected, the appraiser is usually having problems using the equipment properly.

For example, he or she may not align the product correctly before taking a measurement or may have a problem reading the fine marks on the gage. Also, the appraiser may not have clear instructions on which part of the product should be measured. Many of these problems are the result of ineffective training. Either the appraiser needs to undergo a training program or a new training program needs to be developed.

Change Current Method

A graphical analysis helps in understanding the components of the measurement system and their relative importance. Current AIAG R&R methods may be misleading and should be modified according to the methods given in this column. Also, appropriate software could be used to calculate the correct d2* values and the correction factor for part variation as a basis for a more precise variable measurement study.

This two-part column has shown the importance of reliable measurement data and their analysis. I hope it will help all quality conscious organizations further improve their products and the productivity of their processes.

REFERENCES

1. Donald S. Ermer, "Improved Gage R&R Measurement Studies," part one, *Quality Progress*, March 2006.
2. L.A. Brown, B.R. Daugherty and V.W. Lowe, *Measurement Systems Analysis*, third edition, Auto Industry Action Group, 2003.
3. Graeme G. Payne, "Calibration: Who Does It," *Quality Progress*, July 2005.

DONALD S. ERMER *is the Procter & Gamble professor emeritus of total quality at the University of Wisconsin-Madison, where he also earned his doctorate in mechanical engineering. Ermer is an ASQ Fellow, is chair of Madison Section 1217 and received ASQ's 1997 Edward Oakley Award and 2000 Grant Medal.*

Please comment

If you would like to comment on this article, please post your remarks on the *Quality Progress* Discussion Board at www.asq.org, or e-mail them to editor@asq.org.