

What Are Repeatability and Reproducibility?

Part 3: Their Meaning in Gage r&R Methodology

BY STEPHEN LUKO

Q. How are repeatability and reproducibility applied in gage r&R methodology?

A: Gage r&R methodology was developed in the 1960s to address the estimation of measurement system variation as applied to manufacturing. The automotive industry led in developing and applying this technique; today, gage r&R is a standard practice in many quarters.

Here, we compare and contrast the use of repeatability (lowercase r) and reproducibility (uppercase R) as used in traditional manufacturing with its use in ASTM International standards. The most important difference is that the latter use probably contains far more applications to raw material testing while the former probably contains more applications in metal fabrication, molding and machining, assembly of subsystems and other manufacturing- and fabrication-type gaging. This distinction is important because materials-type tests are often subject to more sources of variation than industrial processes.

For ASTM materials testing, the key terms for r&R are *repeatability conditions* and *reproducibility conditions*. ASTM E177, Practice for Use of the Terms Precision and Bias in ASTM Test Methods, defines repeatability conditions as “conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time.” To adapt this to gage r&R methodology, we interpret *test results* simply as *gage measurements* and *laboratory* as *facility*. The reason is that the term test result is more general than measurement, and facility implies that the measurements are made in one location – not necessarily a lab. With these distinctions, there is no difference between *r* as used in ASTM or in manufacturing.

With reproducibility there is a noticeable interpretation difference in using this term in

ASTM materials testing and its use in manufacturing. E177 defines reproducibility conditions as “conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment.” Also, “a different laboratory of necessity means a different operator, different equipment, a different location and under different supervisory control.” In manufacturing, r&R methodology generally does not use different labs, equipment or supervisory control. Rather, these aspects of gage r&R methodology are more likely to be carefully controlled than allowed to vary.

The role of reproducibility in gage r&R methodology is to define the variation among operator averages. Here, the term operator, or appraiser, refers to the person who uses the gage. Operator averages are thought to vary from person to person, and the R effect in gage r&R is attempting to estimate the variance due to operator differences in practice.

The statistical model for r and R in gage r&R methodology (without interaction) is Equation 1:

$$y_{ijk} = \chi_i + \alpha_j + \epsilon_{ijk} \quad (1)$$

In Equation 1, y_{ijk} is the k^{th} repeat measurement of the i^{th} part by the j^{th} operator. The χ_i component is the true value of the i^{th} part dimension, the α_j component is the reproducibility effect associated with operator j and the ϵ_{ijk} is the random repeatability error that occurs with each measurement. Each measurement, y , is composed of these three components. The reproducibility term (α) may be thought of as a kind of personal bias associated with an operator, i.e., each operator measures the various parts somewhat differently than the true value x , and this is the individual's α effect. When we use several operators in a gage r&R study, we effectively pick a random sample (of operators) from a potentially infinite universe of all such

possible operators. The α terms are assumed to have a mean of 0 and an unknown variance θ^2 .

The total variance of all measurements, y , has a variance equal to the sum of the individual variance components as in Equation 2.

$$\text{var}(y) = v^2 + \theta^2 + \sigma^2 \quad (2)$$

A gage r&R study is a designed experiment used to estimate the individual variance components. Typically, θ^2 and σ^2 are the main components of interest. The method, based on sample ranges, has enjoyed continued popularity, particularly for small to modest sample sizes, for many years. Today, many computer packages will perform gage r&R using the analysis of variance (or ANOVA) technique as well as the range technique. The following simple illustrations exemplify the method where sample ranges are used.

ILLUSTRATIONS

Five repeat measurements of a cylindrical shaft diameter were made by a single appraiser using the same measurement system under the same conditions. The resulting data were: 3.158, 3.157, 3.161, 3.165 and 3.151. The range of the five measurements is $R = 0.010$. This is converted into the repeatability standard deviation by division by a constant d_2 , in this case, 2.326. The resulting estimate of the repeatability standard deviation is $\hat{\sigma} = 0.010/2.326 = 0.0043$. For several data sets, use the average range in this calculation.

More generally, gage r&R experiments will have p appraisers, n parts and m repeats. One standard plan is to use $p = 3$, $n = 10$ and $m = 3$, making a total of 90 observations. Suppose we have performed such an experiment. The 90 measurements comprise 30 sets of repeated

measurements. Each set of three will have a range. Denote the average of these ranges by \bar{R} and suppose this is equal to 8.4. For $m = 3$ and $np = 30$ measurements, the d_2^* constant is approximately 1.693. Accordingly, the estimate of the repeatability standard deviation is:

$$\hat{\sigma} = \frac{\bar{R}}{d_2^*} = \frac{8.4}{1.693} = 4.96 \quad (3)$$

For reproducibility, we need the appraiser averages range. For three appraisers, the range is calculated as the maximum average minus the minimum average. Denote this as R_A and suppose this is equal to 6.89. Using standard formulas,¹ the reproducibility standard deviation may be calculated as shown. The conversion constant $d_2^* = 1.912$, appropriate for $p = 3$ appraisers, is used.

$$\hat{\sigma} = \sqrt{\left(\frac{R_A}{d_2^*}\right)^2 - \frac{(\hat{\sigma})^2}{nm}} = \sqrt{\left(\frac{6.89}{1.912}\right)^2 - \frac{(4.96)^2}{(10)(3)}} = \sqrt{12.17} = 3.49 \quad (4)$$

The total gage r&R standard deviation is computed as:

$$S_G = \sqrt{\hat{\sigma}^2 + \hat{\sigma}^2} = \sqrt{4.96^2 + 3.49^2} = 6.06 \quad (5)$$

REFERENCE

1. *Measurement Systems Analysis Reference Manual*, 3rd edition, Automotive Industry Action Group, Southfield, Mich., 2005.

STEPHEN LUKO, Pratt & Whitney Aircraft, is chair of Committee E11 on Quality and Statistics, and a fellow of ASTM.

DEAN NEUBAUER is the DataPoints column coordinator and E11.90.03 publications chair.

Statistics play an important role in the ASTM International standards you write. A panel of experts is ready to answer your questions about how to use statistical principles in ASTM standards. Please send your questions to SN Editor in Chief Maryann Gorman at mgorman@astm.org or ASTM International, 100 Barr Harbor Drive, P.O. Box C700, West Conshohocken, PA 19428-2959.