Designation: E2782 – 17

An American National Standard

Standard Guide for Measurement Systems Analysis (MSA)[1]

This standard is issued under the fixed designation E2782; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This guide presents terminology, concepts, and selected methods and formulas useful for measurement systems analysis (MSA). Measurement systems analysis may be broadly described as a body of theory and methodology that applies to the non-destructive measurement of the physical properties of manufactured objects.

1.2 *Units*—The system of units for this guide is not specified. Dimensional quantities in the guide are presented only as illustrations of calculation methods and are not binding on products or test methods treated.

1.3 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

E177 Practice for Use of the Terms Precision and Bias in ASTM Test Methods

E456 Terminology Relating to Quality and Statistics

E2586 Practice for Calculating and Using Basic Statistics

E2587 Practice for Use of Control Charts in Statistical Process Control

3. Terminology

3.1 *Definitions:*

3.1.1 Unless otherwise noted, terms relating to quality and statistics are defined in Terminology E456.

3.1.2 *accepted reference value, n*—a value that serves as an agreed-upon reference for comparison, and which is derived as: *(1)* a theoretical or established value, based on scientific principles, *(2)* an assigned or certified value, based on experimental work of some national or international organization, or *(3)* a consensus or certified value, based on collaborative experimental work under the auspices of a scientific or engineering group. **E177**

3.1.3 *calibration, n*—process of establishing a relationship between a measurement device and a known standard value(s).

3.1.4 *gage, n*—device used as part of the measurement process to obtain a measurement result.

3.1.5 *measurement process, n*—process used to assign a number to a property of an object or other physical entity.

3.1.5.1 *Discussion*—The term "measurement system" is sometimes used in place of measurement process. (See 3.1.7.)

3.1.6 *measurement result, n*—number assigned to a property of an object or other physical entity being measured.

3.1.6.1 *Discussion*—The word "measurement" is used in the same sense as measurement result.

3.1.7 *measurement system, n*—the collection of hardware, software, procedures and methods, human effort, environmental conditions, associated devices, and the objects that are measured for the purpose of producing a measurement.

3.1.8 *measurement systems analysis (MSA), n*—any of a number of specialized methods useful for studying a measurement system and its properties.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *appraiser, n*—the person who uses a gage or measurement system.

3.2.2 *discrimination ratio, n*—statistical ratio calculated from the statistics from a gage R&R study that measures the number of 97 % confidence intervals, constructed from gage R&R variation, that fit within six standard deviations of true object variation.

3.2.3 *distinct product categories, n*—alternate meaning of the discrimination ratio.

3.2.4 *gage consistency, n*—constancy of repeatability variance over a period of time.

3.2.4.1 *Discussion*—Consistency means that the variation within measurements of the same object (or group of objects) under the same conditions by the same appraiser behaves in a state of statistical control as judged, for example, using a control chart. See Practice E2587.

[1] This guide is under the jurisdiction of ASTM Committee E11 on Quality and Statistics and is the direct responsibility of Subcommittee E11.20 on Test Method Evaluation and Quality Control.

Current edition approved Jan. 1, 2017. Published February 2017. Originally approved in 2011. Last previous edition approved in 2011 as E2782 – 11$^{\varepsilon 1}$. DOI: 10.1520/E2782-17.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

3.2.5 *gage performance curve, n*—curve that shows the probability of gage acceptance of an object given its real value or the probability that an object's real measure meets a requirement given the measurement of the object.

3.2.6 *gage R&R, n*—combined effect of gage repeatability and reproducibility.

3.2.7 *gage resolution, n*—degree to which a gage can discriminate between differing objects.

3.2.7.1 *Discussion*—The smallest difference between two objects that a gage is capable of detecting is referred to as its finite resolution property. For example, a linear scale graduated in tenths of an inch is not capable of discriminating between objects that differ by less than 0.1 in. (0.25 cm).

3.2.8 *gage stability, n*—absence of a change, drift, or erratic behavior in bias over a period of time.

3.2.8.1 *Discussion*—Stability means that repeated measurements of the same object (or average of a set of objects) under the same conditions by the same appraiser behave in a state of statistical control as judged for example by using a control chart technique. See Practice E2587.

3.2.9 *linearity, n*—difference or change in bias throughout the expected operating range of a gage or measurement system.

3.2.10 *measurement error, n*—error incurred in the process of measurement.

3.2.10.1 *Discussion*—As used in this guide, measurement error includes one or both of R&R types of error.

3.2.11 *repeatability conditions, n*—in a gage R&R study, conditions in which independent measurements are obtained on identical objects, or a group of objects, by the same operator using the same measurement system within short intervals of time.

3.2.11.1 *Discussion*—As used in this guide, repeatability is often referred to as equipment variation or EV.

3.2.12 *reproducibility conditions, n— in a gage R&R study*, conditions in which independent test results are obtained with the same method, on identical test items by different operators.

3.2.12.1 *Discussion*—As used in this guide, reproducibility is often referred to as appraiser variation or AV. This term is also used in a broader sense in Practice E177.

4. Significance and Use

4.1 Many types of measurements are made routinely in research organizations, business and industry, and government and academic agencies. Typically, data are generated from experimental effort or as observational studies. From such data, management decisions are made that may have wide-reaching social, economic, and political impact. Data and decision making go hand in hand and that is why the quality of any measurement is important—for data originate from a measurement process. This guide presents selected concepts and methods useful for describing and understanding the measurement process. This guide is not intended to be a comprehensive survey of this topic.

4.2 Any measurement result will be said to originate from a measurement process or system. The measurement process will consist of a number of input variables and general conditions that affect the final value of the measurement. The process variables, hardware and software and their properties, and the human effort required to obtain a measurement constitute the measurement process. A measurement process will have several properties that characterize the effect of the several variables and general conditions on the measurement results. It is the properties of the measurement process that are of primary interest in any such study. The term "measurement systems analysis" or MSA study is used to describe the several methods used to characterize the measurement process.

NOTE 1—Sample statistics discussed in this guide are as described in Practice E2586; control chart methodologies are as described in Practice E2587.

5. Characteristics of a Measurement System (Process)

5.1 Measurement has been defined as "the assignment of numbers to material objects to represent the relations existing among them with respect to particular properties. The number assigned to some particular property serves to represent the relative amount of this property associated with the object concerned." (**1**)[3]

5.2 A measurement system may be described as a collection of hardware, software, procedures and methods, human effort, environmental conditions, associated devices, and the objects that are measured for the purpose of producing a measurement. In the practical working of the measurement system, these factors combine to cause variation among measurements of the same object that would not be present if the system were perfect. A measurement system can have varying degrees of each of these factors, and in some cases, one or more factors may be the dominant contributor to this variation.

5.2.1 A measurement system is like a manufacturing process for which the product is a supply of numbers called measurement results. The measurement system uses input factors and a sequence of steps to produce a result. The inputs are just varying degrees of the several factors described in 5.2 including the objects being measured. The sequence of process steps are that which would be described in a method or procedure for producing the measurement. Taken as a whole, the various factors and the process steps work collectively to form the measurement system/process.

5.3 An important consideration in analyzing any measurement process is its interaction with time. This gives rise to the properties of stability and consistency. A system that is stable and consistent is one that is predictable, within limits, over a period of time. Such a system has properties that do not deteriorate with time (at least within some set time period) and is said to be in a state of statistical control. Statistical control, stability and consistency, and predictability have the same meaning in this sense. Measurement system instability and inconsistency will cause further added overall variation over a period of time.

5.3.1 In general, instability is a common problem in measurement systems. Mechanical and electrical components may wear or degrade with time, human effort may exhibit increasing fatigue with time, software and procedures may change

[3] The boldface numbers in parentheses refer to the list of references at the end of this standard.

with time, environmental variables will vary with time, and so forth. Thus, measurement system stability is of primary concern in any ongoing measurement effort.

5.4 There are several basic properties of measurement systems that are widely recognized among practitioners. These are repeatability, reproducibility, linearity, bias, stability, consistency, and resolution. In studying one or more of these properties, the final result of any such study is some assessment of the capability of the measurement system with respect to the property under investigation. Capability may be cast in several ways, and this may also be application dependent. One of the primary objectives in any MSA effort is to assess variation attributable to the various factors of the system. All of the basic properties assess variation in some form.

5.4.1 Repeatability is the variation that results when a single object is repeatedly measured in the same way, by the same appraiser, under the same conditions (see Fig. 1). The term "precision" also denotes the same concept, but "repeatability" is found more often in measurement applications. The term "conditions" is sometimes combined with repeatability to denote "repeatability conditions" (see Terminology E456).

5.4.1.1 The phrase "intermediate precision" is also used (for example, see Practice E177). The user of a measurement system shall decide what constitutes "repeatability conditions" or "intermediate precision conditions" for the given application. Typically, repeatability conditions for MSA will be as described in 5.4.1.

5.4.2 Reproducibility is defined as the variation among average values as determined by several appraisers when measuring the same group of objects using identical measurement systems under the same conditions (see Fig. 2). In a broader sense, this may be taken as variation in average values of samples, either identical or selected at random from one homogeneous population, among several laboratories or as measured using several systems.

5.4.2.1 Reproducibility may include different equipment and measurement conditions. This broader interpretation has attached "reproducibility conditions" and shall be defined and interpreted by the user of a measurement system. (In Practice E177, reproducibility includes interlaboratory variation.)

5.4.3 Bias is the difference between a standard or accepted reference value for an object, often called a "master," and the average value of a sample of measurements of the object(s) under a fixed set of conditions (see Fig. 1).

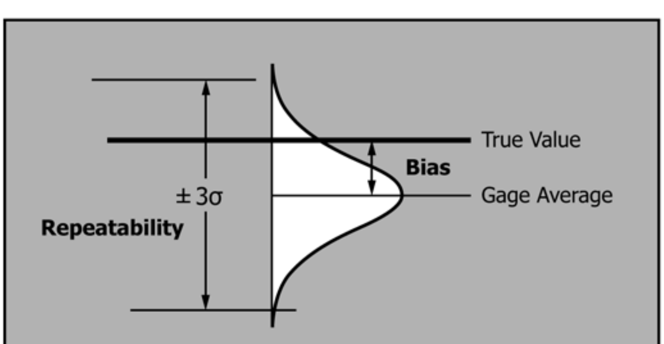

FIG. 1 Repeatability and Bias Concepts

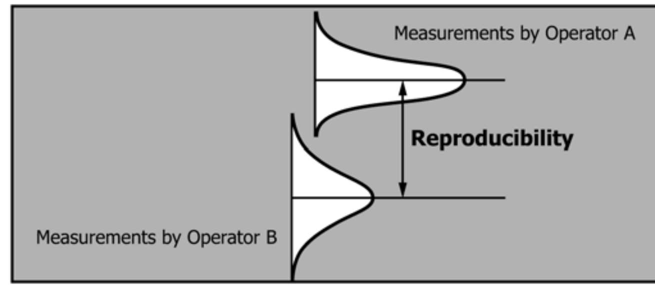

FIG. 2 Reproducibility Concept

5.4.4 Linearity is the change in bias over the operational range of the measurement system. If the bias is changing as a function of the object being measured, we would say that the system is not linear. Linearity can also be interpreted to mean that an instrument response is linearly related to the characteristic being measured.

5.4.5 Stability is variation in bias with time, usually a drift or trend, or erratic behavior.

5.4.6 Consistency is the change in repeatability with time. A system is consistent with time when the standard deviation of the repeatability error remains constant. When a measurement system is stable and consistent, we say that it is a state of statistical control.

5.4.7 The resolution of a measurement system has to do with its ability to discriminate between different objects. A system with high resolution is one that is sensitive to small changes from object to object. Inadequate resolution may result in identical measurements when the same object is measured several times under identical conditions. In this scenario, the measurement device is not capable of picking up variation as a result of repeatability (under the conditions defined). Poor resolution may also result in identical measurements when differing objects are measured. In this scenario, the objects themselves are too close in true magnitude for the system to distinguish among.

5.4.7.1 Resolution plays an important role in measurement in general. We can imagine the output of a process that is in statistical control and follows a normal distribution with mean, μ, and standard deviation, σ. Based on the normal distribution, the natural spread of the process is 6σ. Suppose we measure objects from this process with a perfect gage except for its finite resolution property. Suppose further that the gage we are using is "graduated" as some fraction, $1/k$, of the 6σ natural process spread (integer k). For example, if $k = 4$, then the natural process tolerance would span four graduations on the gage; if $k = 6$, then the natural process spread would span six graduations on the gage. It is clear that, as k increases, we would have an increasingly better resolution and would be more likely to distinguish between distinct objects, however close their magnitudes; at the opposite extreme, for small k, fewer and fewer distinct objects from the process would be distinguishable. In the limit, for large k, every object from this process would be distinguishable.

5.4.7.2 In using this perfect gage, the finite resolution property plays a role in repeatability. For very large k, the resulting standard deviation of many objects from the process would be nearly the magnitude of the true object standard deviation, σ. As k diminishes, the standard deviation of the measurements would increase as a result of the finite resolution property. Fig. 3 illustrates this concept for a process centered at 0 and having σ = 1 for k = 4.

5.4.7.3 The illustration from Fig. 3 is a system capable of discriminating objects into groups no smaller than 1.5σ in width so that a frequency distribution of measured objects from this system will generally have four bins. This means four distinct product values can be detected. Using Fig. 3 and the theoretical probabilities from the normal distribution, it is possible to calculate the variance of the measured values for various values of k. In this case, the variance of the measured values is approximately 1.119 or 11.9 % larger than the true variance. The standard deviation is, therefore, 1.058 or 5.8 % larger.

5.4.7.4 This illustrates the important role that resolution plays in measurement in general and an MSA study in particular. There is a subtle interaction between the degree of resolution and more general repeatability and other measurement effects. In extreme cases of poor resolution, an MSA study may not be able to pick up a repeatability effect (all objects measured yield the same value). For an ideal system, for varying degrees of finite resolution as described in 5.4.7, there will be a component of variance as a result of resolution alone. For positive integer value, k, when the smallest measurement unit for a device is 1/kth of the 6σ true natural process range, the standard deviation as a result of the resolution effect may be determined theoretically (assuming a normal distribution). Table 1 shows the effect for selected values of k.

5.4.7.5 A common rule of thumb is for a measurement device to have a resolution no greater than 0.6σ, where σ is the true natural process standard deviation. This would give us k = 10 graduation divisions within the true 6σ natural process limits. In that particular case, the resulting variance of all measurements would have increased by approximately 1.9 % (Table 1, k = 10).

5.5 MSA is a broad class of activities that studies the several properties of measurement systems, either individually, or some relevant subset of properties taken collectively. Much of this activity uses well known methods of classical statistics, most notably experimental design techniques. In classical statistics, the term variance is used to denote variation in a set of numbers. It is the square of the standard deviation. One of the primary goals in conducting an MSA study is to assess the several variance components that may be at play. Each factor will have its own variance component that contributes to the overall variation. Components of variance for independent variables are additive. For example, suppose y is the result of a measurement in which three independent factors are at play. Suppose that the three independent factors are x_1, x_2, and x_3. A simple model for the linear sum of the three components is $y = x_1 + x_2 + x_3$. The variance of the overall sum, y, given the variances of the components is:

$$\sigma_y^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 \quad (1)$$

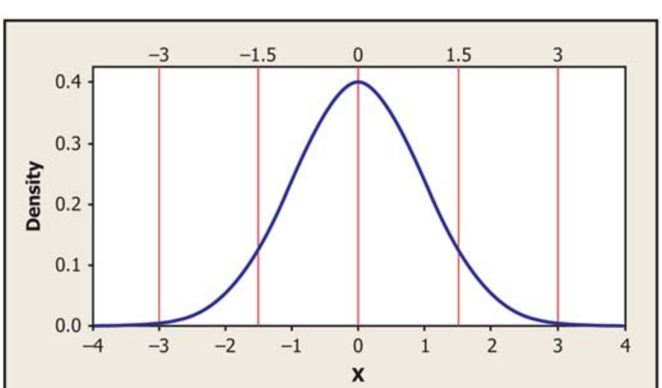

FIG. 3 Finite Resolution Property of a Measurement System where Four "Graduations Fit within the Natural 6σ Process Spread"

TABLE 1 Behavior of the Measurement Variance and Standard Deviation for Selected Finite Resolution Property, *k*, True Process Variance is 1

k	total variance	resolution component	stdev due to resolution
2	1.36400	0.36400	0.60332
3	1.18500	0.18500	0.43012
4	1.11897	0.11897	0.34492
5	1.08000	0.08000	0.28284
6	1.05761	0.05761	0.24002
8	1.04406	0.04406	0.20990
9	1.03549	0.03549	0.18839
10	1.01877	0.01877	0.13700
12	1.00539	0.00539	0.07342
15	1.00447	0.00447	0.06686

5.5.1 We say that each variance on the right is a component of the overall variance on the left. This model is theoretical; in practice, we do not know the true variances and have to estimate their values from data.

5.5.2 Statistical methods allow one to estimate the several variance components in MSA. Sometimes the analyst may only be interested in one of the components, for example, repeatability. In other cases, it may be two or more components that may be of interest. Depending on how the MSA study is designed, the variance components may be estimable free and clear of each other or combined as a composite measure. Several widely used basic models and associated statistical techniques are discussed in Section 6.

6. Basic MSA Methods

6.1 *Simple Repeatability*—Simple repeatability may be evaluated using at least two measurements of each of several objects by a single appraiser under identical conditions. The simplest such experiment is to use a number of distinct objects, say n, and two measurements of each object. Let y_{i1} and y_{i2} be the two measurements of object i. Each measurement is contaminated with a repeatability error component, e. This model may be written as:

$$y_{ij} = x_i + e_{ij} \quad (2)$$

6.1.1 In this model, the y_{ij} values are the observed measurements of object, i, measurement, j; the x_i values are considered the "true" or reference values of the objects being measured; and e_{ij} is the repeatability error associated with object, i, and measurement, j. The difference, d_i, between two measurements of the same object may be written as:

$$d_i = y_{i1} - y_{i2} = x_i + e_{i1} - x_i - e_{i2} = e_{i1} - e_{i2} \quad (3)$$

6.1.2 If the error terms can be considered normally distributed, then the paired differences, the *d*'s, will possess a normal distribution. Generally, the repeatability error term, *e*, is assumed to have a mean of 0 and some unknown variance σ^2. This is the repeatability variance. Under the model assumptions and further assuming that the errors are uncorrelated; the variable, d_i, will be normally distributed with mean 0 and variance $2\sigma^2$. The variance σ^2 may be analyzed using standard statistical theory as follows. The estimate of σ^2 is formed as:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} d_i^2}{2n} \quad (4)$$

The square root of quantity Eq 4 is the estimate of the repeatability standard deviation.

6.1.3 With the previous assumptions, the sum of the squared deviations, the d^2 terms, divided by $2\sigma^2$ will have a chi-square distribution with *n* degrees of freedom. From this fact, a confidence interval for σ^2 may be constructed.

$$\chi^2 = \frac{\sum_{i=1}^{n} d_i^2}{2\sigma^2} \quad \text{Chi-square with } n\ df \quad (5)$$

6.1.4 It may be important to check that the mean of the variable *d* is zero. For this purpose, we can use a classical confidence interval construction for the true mean of the differences. The form of the confidence interval is:

$$\bar{d} \pm \frac{tS_d}{\sqrt{n}} \quad (6)$$

6.1.5 Here, *t* is selected from the Student's *t* distribution, using a two-sided 100(1 – α)% confidence level and degrees of freedom *n* – 1, and S_d is the ordinary sample standard deviation of the differences. If the interval includes 0, then the assumption of a mean equal to 0 cannot be refuted at significance level, α. The normal distribution assumption may be checked using the several values of d_i, and a normal probability plotting technique (see Practice E2586).

6.1.6 The paired comparison (two measurements per each object) scenario is convenient and very common in practice; however, it is also possible to modify the methodology using more than two measurements per object measured. When this approach is used, the formulas for the resulting estimates and confidence interval formulation will be different. An analysis of variance (ANOVA) may be used for the more than two measurements per object case. The ANOVA technique also allows for differences from one object to another in the number of times each object is measured (see 6.4 for details).

6.2 *Use of the Range Control Chart in Evaluating Repeatability*—The range control chart may be used to evaluate consistency of the measurement system and resolution issues. In addition, the control chart gives an alternative measure of repeatability that, under perfect stability and consistency conditions, should be very close to Eq 4. Suppose there are *n* objects to be measured. For each pair of repeated measurements, calculate the range as:

$$R_i = |y_{i1} - y_{i2}| \quad (7)$$

6.2.1 The absolute value bars simply indicate that we are looking at the absolute difference between measurements or the range in each pair. The average range is:

$$\bar{R} = \frac{\sum_{i=1}^{n} R_i}{n} \quad (8)$$

6.2.2 The range estimate of the standard deviation of repeatability is:

$$\hat{\sigma} = \frac{\bar{R}}{d_2} \quad (9)$$

6.2.3 A short table of the constant d_2 appears in Appendix X6 or see Practice E2587. The constant d_2 converts[4] the average range into an unbiased estimate of σ. It is a function of the subgroup size, *k*. Here, *k* = 2. The range control chart is constructed as a series of the *n* sample range values with center line equal to the average range and control limits (upper and lower control limits) calculated from the formulas:

$$UCL_R = D_4\bar{R} \quad (10)$$

$$LCL_R = D_3\bar{R} \quad (11)$$

6.2.4 The values D_3 and D_4 may be found in Appendix X6, Practice E2587, or any text on statistical process control. When the subgroup size is less than seven, the constant D_3 will be 0. A sequence of such values, exhibiting good statistical control, will give every indication of a random sequence of observations with all points falling within the control limits. This kind of chart is always done first when performing MSA studies on repeatability. The principle signs of inconsistency are points outside of the control limits or other nonrandom patterns such as runs above (below) the center line or trends of increasing (decreasing) direction. Such signs indicate inconsistency and out-of-control conditions.

6.2.5 When zeros appear on a range control chart, this is a sign of either a resolution problem or that the repeatability error is small enough to be considered negligible. In any event, it is still resolution that is at issue. Poor resolution in the presence of modest repeatability error may yield identical values in repeating a measurement. Too many zeros appearing in the range chart will reduce the estimate of the repeatability standard deviation and perhaps underestimate its real effect. One way to counteract this problem is to replace zeros with *q* as:

$$q = ud_2/\left(2\sqrt{3}\right) \quad (12)$$

6.2.6 The quantity, *u*, is defined as the smallest unit of resolution the measurement device is capable of discriminating and d_2 is as previously defined. For example, if one uses a ruler

[4] Formally, the constant d_2 is equal to the mathematical expectation of the sample range divided by σ, when sampling from a normal distribution. The value, d_2, is a function only of the subgroup size, *k*. Some writers prefer to use the constant d_2^*. Dividing the average range (Eq 8) by this constant and squaring makes the resulting number an unbiased estimate of σ^2. The value, d_2^*, is a function of the subgroup size, *k*, as well as the number of subgroups, *n*. See Appendix X6 for tables.

graduated in eighths of an inch, then $u = 0.125$. The reason for this is that the standard deviation of a uniform distribution between 0 and u is $u/(2\sqrt{3})$. Post multiplication by d_2 gives an estimate of the expected range in a sample of size, k (the subgroup size). An alternative method is to estimate the expected range based on the subgroup sample size, k. For this method, we would replace a zero range with $u(k - 1)/(k + 1)$, which is precisely the expected range in a sample of size, k, from a uniform distribution between 0 and u. Still, another method is to replace zero ranges with simulated ranges from a uniform distribution on the interval $[0,u]$. In each method, these pseudo ranges replace zeros on the range chart.

6.3 *Use of the Average Control Chart in Evaluating Repeatability*—The averages are formed from each pair of repeated measurements (each pair is a subgroup). These can be plotted in time order using a control chart for averages. The center line for such a chart will be the overall average, $\bar{\bar{x}}$, of the several subgroup averages. Note that the subgroup size in this case is two; but this method is general and any subgroup size may be used. The range chart, having already been constructed, is used in constructing the control limits for the average chart. The control limits are calculated from the following classical formulas based on the subgroup average range:

$$UCL_{\bar{x}} = \bar{\bar{x}} + A_2\bar{R} \quad (13)$$

$$LCL_{\bar{x}} = \bar{\bar{x}} + A_2\bar{R} \quad (14)$$

6.3.1 The constant, A_2, is defined as:

$$A_2 = \frac{3}{d_2\sqrt{n}} \quad (15)$$

6.3.2 For $n = 2$, the constant A_2 is 1.88. The overall average $\bar{\bar{x}}$ is defined as:

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{n} \bar{x}_i}{n} \quad (16)$$

6.3.3 For subgroup sizes other than two, tables of the constant, A_2, can be found in Appendix X6 or Practice E2587. The stability of the system may be assessed from the control chart for averages. Individual plotted points should indicate a random pattern about the center line. The control limits for the average chart are constructed from the range chart and the range chart is measuring repeatability variation only; therefore, if the object to object variation is much greater than the repeatability variation, most points on the average chart will fall outside of the control limits. Points falling within the control limits are said to be indistinguishable from one another. The region between the control limits is a kind of noise band (noise being a repeatability error) and object averages are like the "signals" of real object-to-object variation. A fair benchmark for this kind of chart is to have at least 50 % of the average points fall outside of the control limits. Anything less indicates that repeatability variation is dominant over object variation. This method is more powerful when the subgroup sample size is greater than two. Also, if the objects were handpicked and not random samples from a process, interpretation of this type of chart may be incorrect.

6.4 *Repeatability Using More than Two Observations per Object Measured*—When more than two measurements can be made for each of several sample objects, the analysis of variance (ANOVA) with random effects may be used. It is best to use a fixed number of repeat measurements per object, although this method still works when the subgroup size varies. Whenever possible, measurements should be made in randomized order. If there are n objects to be measured and m measurements per object, randomize the numbers 1, 2, … mn. The randomized numbers should then form the basis of the sequence for how the measurements would be obtained. Upon completion, there will be n sets of m repeated measurements, m for each of the n objects measured. Let y_{ij} be the jth repeated measurement of object, i, where i varies from 1 to n and j from 1 to m. The quantity $\bar{y}_{i.}$ represents the average of the measurements from the i^{th} object (the "dot" notation in the subscript signifies that we are averaging over the second index, j). The following statistic is an unbiased estimate of the repeatability variance σ^2.

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij} - \bar{y}_{i.})^2}{n(m-1)} = \frac{SSE}{n(m-1)} \quad (17)$$

6.4.1 Table X1.1 in Appendix X1 contains a complete ANOVA table for this type of model. The quantity SSE/σ^2 possesses a chi-square distribution with $n(m - 1)$ degrees of freedom, and from this fact, a confidence interval for the repeatability standard deviation, σ, can be obtained. In the case where there are varying numbers of repeat measurements for each of the several objects measured, Eq 17 would be modified. Suppose for n objects, there are m_i measurements for the ith object. The estimate of repeatability variance becomes:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i}(y_{ij} - \bar{y}_{i.})^2}{\sum_{i=1}^{n}(m_i - 1)} = \frac{SSE}{\sum_{i=1}^{n}(m_i - 1)} \quad (18)$$

6.4.2 Again, the quantity SSE/σ^2 possesses a chi-square distribution with degrees of freedom as shown in the denominator, and from this fact, a confidence interval for the repeatability standard deviation, σ, can be obtained. Eq 17 and 18 come from an ANOVA approach to repeatability analysis. Table X1.1 in Appendix X1 contains the details for this model in which n objects are measured m times each by a single appraiser.

6.5 *Repeatability Using Known or Standard Reference Values*—An MSA study may be conducted using several known or standard objects. Let y be the measurement of an object whose standard value is x. The model is:

$$y_i = x_i + \varepsilon_i \quad (19)$$

6.5.1 The ε term is assumed to have mean 0 and some unknown variance σ^2 representing repeatability. The goal is to estimate σ. If we have n objects to measure, then form the paired differences $d_i = y_i - x_i$. The d_i values are equivalent to the ε_i values. In this model, we do not have to assume a distribution for the variable, x. We only need one consistent distribution for the paired differences. In theory, this type of study could be carried out using a single object measured m times (see 6.5.4).

6.5.2 The following quantity is used as the point estimate of the repeatability variance:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} d_i^2}{n} \tag{20}$$

6.5.3 If the d_i terms can be considered normally distributed, the sum of squared differences divided by σ^2 will possess a chi-square distribution with n degrees of freedom, and from this fact, a confidence interval for σ^2 can be constructed.

6.5.4 If several objects are each measured a variable number of times, say k_i times for object i, the formulation of Eq 20 is the same. Let N = the total number of paired observations including all repeated measurements of all n objects. The estimate of repeatability variance is Eq 20 with n replaced by N. The sum of all N squared differences possesses a chi-square distribution with N degrees of freedom.

6.6 *Reproducibility—Appraiser Component of Variance*—In using a measurement system, it is always possible for different people to get different results when identical objects are measured in the same manner. Two sources of variation are responsible for the difference among appraisers: *(1)* simple repeatability error will cause differences among appraisers and *(2)* overall differences among appraisers may be due to individual biases on the average. The second component is the subject of reproducibility. This is shown in Fig. 2. The difference between means of the two appraisers in Fig. 2 is the effect of reproducibility. In practice, it is the difference in sample means of the two groups as measured by differing appraisers that estimates reproducibility.

6.6.1 Reproducibility can be considered as a random bias component assigned to every appraiser. A bias simply means that the appraiser tends to measure every object either higher or lower on average than the "true" measure of the object. We can think of appraisers coming from a population of all such appraisers, each with a unique and fixed bias. The distribution of these biases is assumed to be normal with mean 0 and some unknown variance θ^2. The parameter, θ, is the reproducibility standard deviation. We may think of the random variable u as denoting the random bias (reproducibility) component. When several specific appraisers are used in a measurement systems study, we are effectively picking several random values of u.

6.6.2 For several appraisers, the model for the measurement of the ith object by appraiser j at the kth repeat is:

$$y_{ijk} = x_i + u_j + e_{ijk} \tag{21}$$

6.6.2.1 The quantity e_{ijk} continues to play the role of the repeatability error term which is assumed to have mean 0 and variance σ^2. Quantity x_i represents the standard or "true" value of the object being measured and quantity u_j is a random reproducibility term associated with appraiser j. This last quantity is assumed to come from a distribution having mean 0 and some variance θ^2. If the objects being measured can be considered a random sample from a population of objects, then the x_i are random variables with some mean, the true population mean, and variance v^2.

6.6.2.2 Eq 21 assumes no part-operator interaction term, which might sometimes be a reasonable assumption in practice. An interaction between part and operator means that for increasing (decreasing) values of some objects, some appraisers follow an opposite trend—that is, they measure smaller (larger) values. If interaction is to be considered, an additional term, w_{ij}, would have to be included in Eq 21. The model including interaction is:

$$y_{ijk} = x_i + u_j + w_{ij} + e_{ijk} \tag{22}$$

6.6.3 *Variance Components*—For Eq 21, assuming independence of the three terms, the variance of the measurements is simply:

$$\mathrm{var}(y_{ijk}) = \nu^2 + \theta^2 + \sigma^2 \tag{23}$$

6.6.3.1 For Eq 22, including the interaction term, the variance of the measurements becomes:

$$\mathrm{var}(y_{ijk}) = \nu^2 + \theta^2 + \alpha^2 + \sigma^2 \tag{24}$$

6.6.3.2 Each of ν^2, θ^2, α^2, and σ^2 are the components of the overall variance with α^2 playing the role of interaction. One principle objective of a measurement systems analysis is to obtain estimates of these variance components. The combined variance components, θ^2 and σ^2, are often referred to as the gage R&R variance. Many software packages will perform this type of analysis.

6.6.3.3 When several appraisers each measure a group of objects once only (no repeats), it is still possible to estimate R&R but not interaction. Appendix X2 and Appendix X3 give complete ANOVA tables for Eq 21 and 22, respectively.

6.7 *Bias*—Reproducibility variance may be viewed as coming from a distribution of the appraiser's personal measurement bias. In addition, there may be a global bias present in the measurement system that is shared equally by all appraisers. Bias is the difference between the mean of the overall distribution of all measurements by all appraisers and a "true" or reference average of all objects. Whereas reproducibility refers to a distribution of appraiser averages, bias refers to a difference between the average of a set of measurements and a known or reference value. The measurement distribution may itself be composed of measurements from differing appraisers or it may be a single appraiser that is being evaluated. Thus, it is important to know what the objective is in evaluating bias.

6.7.1 Bias may also vary as a function of the reference value. For example, bias may be larger for "larger objects" and smaller for "smaller objects." This concept is referred to as linearity. See 6.8 for further details on this concept.

6.7.2 For single appraiser, and a single object, bias is evaluated as the difference between the average of several measurements and the known reference value. This is Eq 25, where x is the known reference value and b is the observed empirical bias.

$$b = \bar{y} - x \tag{25}$$

6.7.2.1 Eq 25 represents a point estimate for the bias. It might or might not be significant, because quantity b is also contaminated with repeatability error. We can determine if the observed bias is significant by constructing a confidence interval for the real bias B. If the confidence interval includes

0 as a plausible value, then we may conclude that a significant non-zero bias has not been detected. To understand how the confidence interval is constructed, we shall consider the model for this scenario and its assumptions.

6.7.3 The model for simple bias is:

$$y_i = B + x + \varepsilon_i \tag{26}$$

6.7.3.1 The value x is the fixed known reference value, quantity B is the unknown bias, the ε terms are random repeatability errors assumed to be normally distributed with mean 0 and unknown variance σ^2, and the y terms are the actual measurements. A series of n measurements will have an average given by:

$$\bar{y}. = B + x + \bar{\varepsilon}. \tag{27}$$

6.7.3.2 The empirical bias b is therefore equal to:

$$b = \bar{y}. - x = B \pm \bar{\varepsilon}. \tag{28}$$

6.7.3.3 Quantity b therefore possesses a normal distribution with mean B and variance σ^2/ n. If the repeatability variance were known, then we could create a confidence interval for B in the usual way using critical values from the standard normal distribution. Typically, σ^2 is not known and must be estimated from the sample data. The estimate of the σ^2 under the assumptions of this model is:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n - 1} \tag{29}$$

6.7.3.4 A confidence interval for the bias B may then be constructed using Student's t distribution with $n – 1$ degrees of freedom. The confidence interval is:

$$\bar{y}. - x \pm \frac{t_{\alpha/2} S_y}{\sqrt{n}} \tag{30}$$

6.7.3.5 Here, $t_{\alpha/2}$ is a positive constant chosen using confidence 1 – α from Student's t distribution with $n – 1$ degrees of freedom such that $P(-t_{\alpha/2} \leq t \leq t_{\alpha/2}) = 1 - \alpha$. If the confidence interval includes 0 as a plausible value, then we cannot conclude that the bias, B, is non-zero. Note that this does not mean that the bias is 0; it simply indicates that we have not detected a significant non-zero bias. This may mean that our sample size was not adequate to detect the bias.

6.8 *Linearity*—A closely related concept to bias is linearity. Bias may vary as a function of the reference value. For example, bias may be larger for "larger objects" and smaller for "smaller objects." A measurement process has a significant linearity effect if the bias changes in linear manner over the operational range of a set of reference standards. Linearity may be measured using a linear regression analysis of measurements, y, on reference values, x. The measure of linearity is the slope of the least squares best fit line or some function thereof. A simple model for linearity is:

$$y_{ij} = mx_i + B + \varepsilon_{ij} \tag{31}$$

6.8.1 The concept of linearity is often applied in calibration problems. In such cases, a measurement y is related to a set of standard values (x) according to Eq 31. One objective is to determine the range for y that makes the probability of conforming x values very high, say 90 %.

6.8.2 The y_{ij} term is the jth measurement of object x_i, and the ε_{ij} term is the random repeatability error term normally distributed with mean 0 and variance σ^2 associated with the jth measurement of object, i. The parameter, B, represents a global bias; the parameter m represents linearity. When $m = 1$ and $B = 0$, the measurement model reverts to Eq 2. The system is then unbiased and perfectly linear. When $m \neq 1$, then the system possesses a linearity effect.

6.8.3 The linear regression proceeds with a selection of several reference objects (x) to be used for measurement several times each. It is important that the reference objects represent the range of possible objects that the system may see in practice. A linear regression analysis of y on x is then carried out on the data. Typically, when a simple regression analysis is implemented using a software package, the results will include point estimates for the model parameters (m and B). Confidence intervals may be constructed to determine if $B \neq 0$ or if $m \neq 1$. The estimate of the repeatability error standard deviation σ^2 is also output from any good statistics software package when a simple linear regression analysis is performed.

6.8.3.1 For a set of n (x,y) data pairs, the regression analysis results in a pair of estimates calculated using Eq 32 and Eq 33. Let S_x and S_y be the ordinary standard deviations of the x and y values, respectively; let r be the Pearson correlation coefficient between x and y; and let $\bar{x}$ and $\bar{y}$ be the sample averages for the x and y values. The point estimates of m and B are:

$$\hat{m} = \frac{rS_y}{S_x} \tag{32}$$

$$\hat{B} = \bar{y} - \hat{m}\bar{x} \tag{33}$$

6.8.3.2 Confidence intervals for the model parameters may be constructed from well-known formulas. See, for example, Ref (**2**). The predicted value of y given x is calculated as $\hat{y}_i = \hat{m}x_i + \hat{B}$. The estimate of the standard error (repeatability standard deviation) is:

$$\hat{\sigma} = = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n - 2}} \tag{34}$$

6.8.4 In some cases, the bias term may be known to be 0 at the outset, and linearity is the main concern. The model then becomes:

$$y_{ij} = mx_i + \varepsilon_{ij} \tag{35}$$

6.8.4.1 The least squares estimate of the parameter, m, is:

$$\hat{m} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \tag{36}$$

6.8.4.2 The predicted value of y given x is then $\hat{y}_i = \hat{m}x_i$. The error of any particular measurement is specified as the residual: $e_i = y_i - \hat{y}_i$. If the error term can be considered normally distributed, then the estimate of σ is:

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n - 1}} \tag{37}$$

6.8.4.3 In Eq 37, $n - 1$ is used in the denominator because there is but a single parameter of interest.

6.9 *Stability and Consistency*—Stability refers to the absence of special causes of variation affecting the mean of a process. Consistency refers to the absence of special causes affecting the variation in a process. When a process is stable and consistent, we refer to it as in a state of statistical control and predictable, within limits.

6.9.1 Where a measurement process is concerned, a stable and consistent process means that the distribution of measurements from the process does not depend on time order. Such a process typically follows some theoretical model such as the normal distribution.

6.9.2 The properties of stability and consistency are best studied using a control chart technique. A set of control samples are measured over time and plotted on a control chart. Aside from the inherent variation in the several control objects, variation in the individual samples and the mean and range are due entirely to the measurement system. This includes both R&R effects as well as other effects such as linearity. Variation patterns visible within the control chart and other analytical techniques are then used to judge the degree of stability and consistency.

6.10 *Gage Performance Curves*—Gage performance curves may be defined in several ways. One way to do this is to define performance as the probability that a given object actually meets a specification requirement given the value of its measurement. In this use, we require that the objects being measured are selected from a process in statistical control and normally distributed. Further, the mean, μ, and standard deviation, ν, of the true object measure are assumed known as well as the gage R&R standard deviation, σ. These assumptions usually are backed by some kind of previous data for the measurement system being used. The alternate type of performance curve looks at the probability of gage acceptance (or rejection) given the true object measurement, *x*.

7. Planning the Measurement System Study

7.1 All measurement system studies are application dependent; however, several good practices apply to any type of study. Start by selecting the several objects to be measured. There are *n* of these. Preferably, the objects should be randomly selected from the population of all such objects and the process that produced them should have been in a state of statistical control. Several appraisers are selected from a population of all potentials appraisers. There are *p* appraisers. Often, there may only be two or three appraisers available. These are then chosen for the study and assumed to represent the population of all such appraisers that may be thought to exist. In some quarters, differing measurement devices or laboratories may play the role of the appraiser. A number, *m*, of repeat measurements shall be decided. It is possible to do an experiment with only one measurement, but in such a case, it is not possible to estimate an interaction effect. Even when it is believed that interaction between object and appraiser is not a concern, it is advisable to plan for at least two repeated measurements, when possible. When these numbers are decided, the size of the experiment is simply $(n)(m)(p)$. For example, if $p = 3$, $n = 10$, and $m = 3$, the experiment will consist of 90 observations (3 appraisers, 10 objects, and 3 repeated measurements). This configuration is very common in manufacturing organizations.

7.2 *Conducting the Experiment*—In conducting an experiment, care should be taken in obtaining the observations. Since any experiment will be conducted in time order, and since time may introduce additional nuisance variation in the results, any experiment should be conducted in random order. A random order may be obtained using commercially available software or using some mechanical randomization process such as random numbers selected from a table.

7.2.1 During the course of running the experiment, every effort should be made to preserve the constancy of the measurement process. Do not introduce changes to the system such as recalibrations, changes to hardware/software, changes to procedures, or other stability upsetting changes. This practice will give assurance that the measurement system/process remains stable during the experiment. Measurements made by one appraiser should not be revealed to other appraisers during the experiment. Care should also be taken by individuals who may be observing the experiment for this may introduce variation as a result of the so-called "Hawthorne" effect.[5]

8. Analysis of Test Results

8.1 Several statistical methods are useful for assessing R&R. Among these the Analysis of Variance (ANOVA) is the principle tool. There are numerous types of ANOVA. Statistics based on sample ranges are also in wide use. Several models that have found wide use in industrial quarters are portrayed in the appendices. More traditional gage R&R, as for example found in manufacturing applications, are summarized in Ref (**3**) and forms for these calculations are given in Appendix X5.

8.2 *Case 1 ANOVA*—Simple repeatability using one appraiser, *n* objects, and *m* repeats per object. Refer to the sections on repeatability, 6.3 and 6.4, and Appendix X1.

8.3 *Case 2 ANOVA*—Repeatability using *p* appraisers, *n* objects, and one repeat only. The model is Eq 21 with one repeat. See Appendix X2.

8.4 *Case 3 ANOVA*—Repeatability using *p* appraisers, *n* objects, and *m* repeats per each object. The model is Eq 22. See Appendix X3.

8.5 *Repeatability with Given Standards*—The model is Eq 22. See Appendix X4.

8.6 Standard methodology based on sample ranges for gage R&R. See Appendix X5.

8.7 For control chart factors, see Appendix X6.

[5] The Hawthorne effect refers to the possibility that subjects in an experiment improve or modify their behavior in response to the fact that they are being studied, not in response to any particular experimental manipulation.

9. Measurement System Performance

9.1 When an MSA study is complete, estimates for the various properties will result depending on the type of study that was done. The purpose of this section is to elaborate some of the uses to which these statistics may be put. It is assumed throughout that the derived results stand for the associated parameter values. In this sense, the results discussed are theoretical; however, in practice, these measures will have associated standard errors. In the discussion that follows, the standard deviation of measurement error will be denoted as σ and the standard deviation of the error-free object variation is denoted as ν. Note that the process (measurements) variance is $\nu^2 + \sigma^2$; also, that the standard deviation of measurement error, σ, may have been derived from repeatability, reproducibility, or both. This section also assumes that interaction effects are negligible.

9.2 *Discrimination Ratio*—We have seen that the finite resolution property (*u*) of a measurement system places a restriction on the discriminating ability of the measurement system (see 5.4.7). This property is a function of the hardware and software system components; we shall refer to it as "mechanical" resolution. The several factors of measurement variation discussed in this guide, particularly R&R, contribute to further restrictions on object discrimination. This aspect of resolution will be referred to as the statistical resolution.

9.2.1 The effects of mechanical and statistical resolution can be combined as a single measure of discriminating ability. When the true object variance is ν^2, and the measurement error variance is σ^2, the following quantity describes the discriminating ability of the measurement system.[6]

$$D = \sqrt{\frac{2\nu^2}{\sigma^2}+1} \approx \frac{1.414\nu}{\sigma} \tag{38}$$

9.2.2 The right-hand side of Eq 38 is the approximation formula found in many texts and software packages (see, for example, Refs **(4)** or **(5)**). The interpretation of the approximation formula is as follows. Multiply the top and bottom of the right-hand member of Eq 38 by 6 and rearrange and simplify. This gives:

$$D \approx \frac{6(1.414)\nu}{6\sigma} = \frac{6\nu}{4.24\sigma} \tag{39}$$

9.2.3 The denominator, 4.24σ, in Eq 39 is the span of an approximate 97 % interval for a normal distribution centered on its mean. The numerator is a similar 99.7 % (6-sigma) span for a normal distribution. The numerator represents the real object variation and the denominator, variation caused by a measurement error (including mechanical resolution). Then *D*, referred to as the discrimination ratio, represents the number of non-overlapping 97 % confidence intervals that fit within the true object variation. This is referred to as the number of distinct product categories or effective resolution within the true object variation for the process.

9.2.4 The theoretical basis for the left-hand side of Eq 38 is as follows. Suppose *x* and *y* are measurements of the same object. If each is normally distributed, then *x* and *y* have a bivariate normal distribution. If the measurement error has variance, σ^2, and the true object has variance, ν^2, then it may be shown that the bivariate correlation coefficient for this case is $\rho = \nu^2/(\nu^2 + \sigma^2)$. The expression for *D* in Eq 38 is the square root of the ratio $(1 + \rho)/(1 - \rho)$. This ratio is related to the bivariate normal density surface, a function $z = f(x,y)$. Such a surface is shown in Fig. 4.

9.2.5 When a plane cuts this surface parallel to the x,y plane, an ellipse is formed. Each ellipse has a major and minor axis. The ratio of the major to the minor axis for the ellipse is the expression for *D*, Eq 38. The mathematical details of this theory have been sketched by Shewhart **(6)**. Now consider a set of bivariate *x* and *y* measurements from this distribution. Plot the *x*,*y* pairs on coordinate paper. First plot the data as pairs (*x*,*y*). In addition, plot the pairs (*y*,*x*) on the same graph. The reason for the duplicate plotting is that there is no reason to use either the *x* or the *y* data on either axis—thus, we use both. This plot will be symmetrically located about the line, $y = x$. If *r* is the sample correlation coefficient, an ellipse may be constructed and centered on the data. Construction of the ellipse and its relation to *D* is also described in Refs **(6)** and **(7)**. Fig. 5 shows such a plot with the ellipse superimposed and the number of distinct product categories shown as squares of side equal to *D* in Eq 38.

9.2.6 What we see is an elliptical contour at the base of the bivariate normal surface where the ratio of the major to the minor axis is approximately three. This may be interpreted from a practical point of view in the following way. From Fig. 5, the length of the major axis is due principally to the real object variance, while the length of the minor axis is due to the repeatability variance alone. To put an approximate length measurement on the major axis, we realize that the major axis is the hypotenuse of an isosceles triangle whose sides we may measure as 6ν (real object standard deviation) each. It follows from simple geometry that the length of the major axis is approximately 1.414(6ν). The length of the minor axis we can characterize simply as 6σ (error variation). The approximate ratio of the major to the minor axis is, therefore, approximated by discarding the "1" under the radical sign in Eq 38 and 39.

9.2.7 Care should be taken in calculating and using the ratio *D* in practice. First, the values of ν and σ are not typically known with certainty and are estimated from the results of a measurement system study; second, the estimate of ν is based

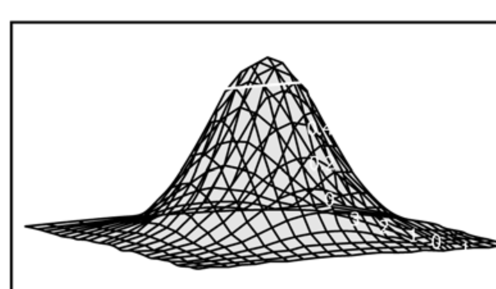

FIG. 4 Bivariate Normal Density Function

[6] An alternative definition for the discrimination, algebraically equivalent to Eq 38 is as follows. Note that σ_m is the standard deviation of the measurement:

$$D = \sqrt{\frac{2\sigma_m^2}{\sigma^2} - 1}$$

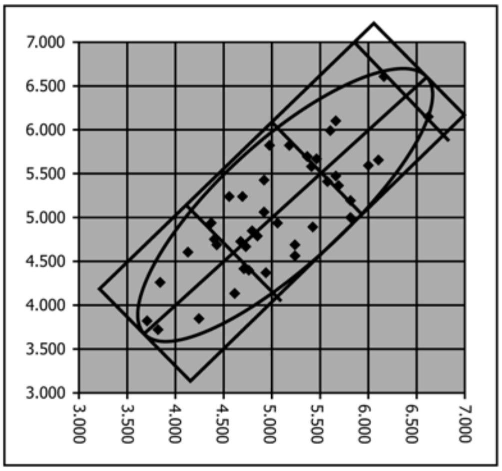

FIG. 5 Bivariate Normal Surface Cross Section with Superimposed Data

on the objects selected for the study. If the several objects used for the study were specially selected and not a random selection, then the estimate of ν will not represent the standard deviation of the distribution of real object variation biasing the calculation of D (it may be an over or an under estimate).

9.3 Wherever a measurement is used, the question may always be asked: "What is the error in the measurement?" If y is the measurement, the answer is $y \pm e$, where e is an estimate of the "error" in the measurement process. The interval $(y - e, y + e)$ is assumed to enclose or capture the "real" measure of the object represented by the measurement, y. Usually, this statement is made with some level of confidence (probability). Suppose the measurement error standard deviation is σ. The value of σ is referred to as one standard error of the measurement, y. Many quarters use one standard error as the error in a single measurement. If the object measured comes from a normal distribution, then the interval, $y \pm \sigma$, is an approximate 68 % confidence interval for the "real" object measure. Some quarters use the so-called probable error and this carries 50 % confidence. The associated interval is $y \pm 0.67\sigma$. Still, some quarters demand higher confidence such as 90 or 95 %. In these cases, the intervals are $y \pm 1.64\sigma$ and $y \pm 1.96\sigma$, respectively.

9.3.1 A frequent question is to ask how far apart we might expect two measurements of the same object to be, determined under the same conditions, when the measurement error standard deviation is σ. For two independent measurements, the standard deviation of their difference may be shown to be approximately 1.414σ. Using this theory and the confidence interval idea, approximate 95 % confidence bounds for the difference between two measurements is $\pm 2(1.414)\sigma$ or approximately $\pm 2.8\sigma$. For any two measurements, y_1 and y_2, this essentially means that the absolute value of their difference is not more than 2.8σ with 95 % confidence. This is embodied in:

$$P\{|y_1 - y_2| \le 2.8\sigma\} = 0.95 \quad (40)$$

9.3.2 Other confidence levels may be used, for example, for 90 % confidence, the interval is $\pm 2.33\sigma$.

9.4 Suppose a practitioner finds that the resulting interval is too wide for his use. In this case, additional measurements can reduce the resulting interval. If n measurements are made, the interval takes the following form:

$$\bar{y} \pm \frac{k\sigma}{\sqrt{n}} \quad (41)$$

Where k = chosen from the standard normal distribution at the chosen confidence level (k = 1.96 for 95 % and k = 1.645 for 90 % confidence).

9.4.1 This interval is particularly important when a measurement is very near (either above or below) to a specification limit. In such cases, the interval becomes tighter as n increases shrinking error variation around the real object measure (as Eq 41 indicates). It may sometimes happen that the reduced interval finally falls entirely within the specification limit requirement rendering the object acceptable with some stated confidence.

9.5 *Gage Performance Curve*—When the real object and measurement error standard deviations are approximately known, a performance curve may be constructed that describes the probability of the real object measure, X, given an actual measurement, y. Several additional themes are also possible using this technique.

9.5.1 Let x be the true object measure, and y the actual measurement. Let these be related through the linear relation $y = mx + B + e$. Quantity B is a possible bias component and m is a possible linearity component. Parameters B and m may take any value and are as previously defined. The random variable, e, represents the measurement error with mean 0. Variables, x and e, are each normally distributed with variances, ν^2 and σ^2, respectively. The covariance between x and y may be shown to be ν^2. The bivariate correlation coefficient between x and y may be shown to be:

$$\rho = \frac{\nu}{\sqrt{\nu^2 + \sigma^2}} \quad (42)$$

9.5.1.1 Let μ be the mean of the real object distribution, x. The mean of the measured objects, y, is $m\mu + B$. Assume that upper and lower specification limits for an object are a and b, respectively. We want to calculate the probability $P(x > a|y)$ or $P(x < b|y)$. These are statements of the probability that the true object measure meets the limiting requirement given the actual measurement, y. From the foregoing facts, the key result may be developed and shown to be:

9.5.1.2 For the lower limit:

$$P(X > a|y) = P\left(Z > \frac{a - \mu - (\rho^2/m)(y - m\mu - B)}{\nu\sqrt{1 - \rho^2}}\right) \quad (43)$$

9.5.1.3 For the upper limit:

$$P(X < b|y) = P\left(Z < \frac{b - \mu - (\rho^2/m)(y - m\mu - B)}{\nu\sqrt{1 - \rho^2}}\right) \quad (44)$$

9.5.1.4 The variable Z has the standard normal distribution. In Eq 43 and 44, y is the actual measurement of an object from a normal distribution. A plot of either $P(x < b|y)$ or $P(x > a|y)$ versus y is one version of a simple gage performance curve. An example of such a curve is shown in Section 10.

9.5.2 Another type of performance curve looks at the probability that an object's measurement, *y*, meets a requirement given the true reference value, *x*, of the object. This measures the extent to which the gage can render a correct disposition of any true object dimension. The two-sided theoretical formulation for this case is:

$$P(a<y<b|x) = P\left\{\frac{a-mx-B}{\sigma} \le Z \le \frac{b-mx-B}{\sigma}\right\} \quad (45)$$

9.5.2.1 Often, in Eq 43, Eq 44, or Eq 45, we might have $B = 0$ and $m = 1$. In that case, the system is unbiased and perfectly linear. The model becomes: $y = x + e$. When a linearity parameter, $m \neq 1$, exists, the performance curve is sometimes used as a calibration mechanism (see examples).

10. Examples

10.1 *Example 1: Simple Repeatability*—Consider Table 2. These data are surface measurements made under identical conditions by one analyst on fifteen bearing races. Two measurements per unit are recorded. In replicating each race measurement, care was taken to remove the race from the gage and then reset the race for its second measurement. This captures the setup-to-setup part of gage repeatability. The data are shown along with the differences, *d*, and squared differences for each race.

10.1.1 Eq 4 was used to calculate the estimate of the repeatability. The estimate of the repeatability standard deviation is the square root of this quantity.

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} d_i^2}{2n} = \frac{0.2652}{2(15)} = 0.00884 \quad (46)$$

$$\hat{\sigma} = \sqrt{0.00884} = 0.094 \quad (47)$$

10.1.2 This calculation assumes that the variable, d_i, has a mean of 0. To check this assumption, use Eq 6. The average difference is –0.036; the standard deviation of the differences is: 0.1325. The sample size is $n = 15$ making 14 degrees of freedom. The associated *t* value for a two-sided 95 % confidence level is 2.145. The confidence interval using Eq 6 is:

$$-0.036 \pm \frac{2.145(0.1325)}{\sqrt{15}} \rightarrow -0.036 \pm 0.0734 \quad (48)$$

10.1.3 This confidence interval includes 0; therefore, we cannot reject the hypothesis of a mean of 0. We can also check for the normal distribution assumption on the variable *d*. If a probability plot and associated normal distribution test is used (for example the Anderson-Darling test), the results would show that the normal distribution assumption is valid.

10.2 *Example 2: Repeatability from Comparison to a Reference Measurement*—Suppose that, instead of two separate measurements, we had a set of known reference standards. Typically, reference standard values are obtained on the best equipment available for the purpose. For this example we suppose that the *x* values in Table 2 represent the reference value and *y* is the measurement on the equipment being evaluated. These data are reproduced in Table 3. Eq 20 is used to calculate the repeatability variance. The calculation is:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} d_i^2}{n} = \frac{0.2652}{15} = 0.0177 \quad (49)$$

$$\hat{\sigma} = \sqrt{0.0177} = 0.1330 \quad (50)$$

10.2.1 A confidence interval for the repeatability standard deviation may be constructed from either Eq 46 or Eq 49. Here, Eq 46 is used. Eq 5 is used to construct the confidence interval. For a 95 % confidence interval, find values, *a* and *b*, from the chi-square distribution with 15 degrees of freedom such that $P(a \le \chi^2 \le b) = 0.95$. There are many ways this can be done. Here, we choose to place 2.5 % probability in each tail of the distribution. The values are $a = 6.262$ and $b = 27.488$. Then, noting that the expression in Eq 5 has this distribution, substitute this expression for χ^2 and solve the resulting inequality for σ. The resulting confidence interval has the form:

$$\sqrt{\frac{\sum_{i=1}^{n} d_i^2}{2b}} \le \sigma \le \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{2a}} \quad (51)$$

10.2.2 For this example, the sum of the squared differences is 0.413 45 and using the values for *a* and *b* described in 10.2.1, the 95 % confidence interval for σ is: $0.0867 \le \sigma \le 0.1817$.

10.3 *Example 3: Range Estimate of Repeatability*—The data recorded in Table 2 are plotted on a range control chart in Fig. 6.

TABLE 2 Bearing Race Data—Two Independent Measurements, without Reference Standards

i	x_i	y_i	d_i	$(y_i–x_i)^2$
1	3.220	3.110	0.110	0.012
2	9.830	10.020	-0.190	0.036
3	2.340	2.290	0.050	0.002
4	12.120	12.040	0.080	0.006
5	7.730	7.800	-0.070	0.005
6	7.910	8.100	-0.190	0.036
7	5.290	5.420	-0.130	0.017
8	3.940	3.960	-0.020	0.000
9	11.850	11.620	0.230	0.053
10	3.770	4.060	-0.290	0.084
11	9.110	9.050	0.060	0.004
12	5.880	5.910	-0.030	0.001
13	6.370	6.440	-0.070	0.005
14	6.890	6.940	-0.050	0.003
15	11.330	11.360	-0.030	0.001

TABLE 3 Independent Measurement (*y*) with Reference Standards (*x*)

x	y	$(y–x)^2$
3.220	3.110	0.012
9.830	10.020	0.036
2.340	2.290	0.002
12.120	12.040	0.006
0.730	0.800	0.005
0.910	1.100	0.036
5.290	5.420	0.017
3.940	3.960	0.000
1.850	1.620	0.053
3.770	4.060	0.084
9.110	9.050	0.004
5.880	5.910	0.001
6.370	6.440	0.005
11.890	11.940	0.002
11.330	11.360	0.001

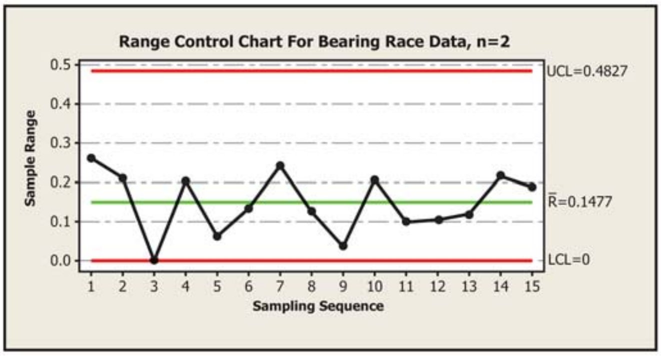

FIG. 6 Range Control Chart for Table 2 Data

10.3.1 This chart shows the behavior of the sample range in groups of $n = 2$ of the bearing race data. The center line is the average of the 15 sample ranges. The upper control limit was constructed using Eq 10. For $n = 2$, the lower control limit defaults to 0. This chart gives every reason to believe that the time order of the data was in a state of statistical control as the experiment was performed, and that, therefore, the measurement system exhibits consistency for this time period.

10.3.2 Using the average range and Eq 9, another estimate of the repeatability standard deviation may be obtained. The estimate based on the range is:

$$\hat{\sigma} = \frac{\bar{R}}{d_2} = \frac{0.1477}{1.128} = 0.1309 \quad (52)$$

10.3.3 Note that this estimate compares favorably with the previous two estimates (Eq 46, Eq 47, Eq 49, and Eq 50). Using the average range, we can construct a control chart for the subgroup averages, again using the 15 pairs of race measurements. This is shown in Fig. 7.

10.3.4 Control limits for Fig. 7 were constructed using Eq 13 and 14, and in these equations, the average range (from the range chart) as the estimate of variability is used.

10.3.4.1 Thus, the control limits are based entirely on the repeatability variation, not the actual object-to-object variability. What one wants to see is a chart in which the majority of the points fall outside the control limits—such as is illustrated in Fig. 6. The area between the control limits represents the span of the repeatability error. This acts like a noise region; the plotted averages represent real object-to-object variation and act as a "signal." Points falling within the control limits are said to be indistinguishable from one another since variation between such points falls within the limits of the repeatability error.

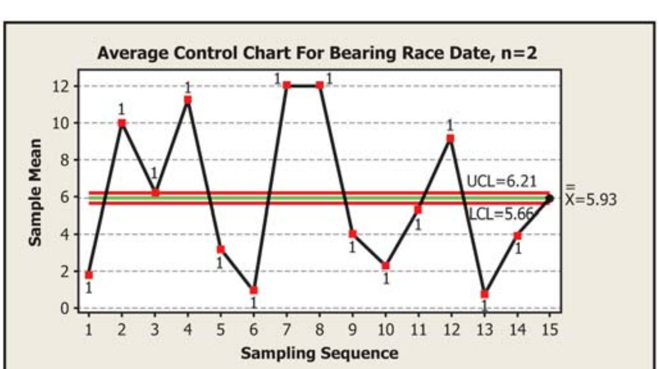

FIG. 7 Average Control Chart for Table 2 Data

10.4 *Example 4: Repeatability Using Varying Numbers of Observations per Object Measured*—In Table 4, measurements of a cylindrical shaft diameter are shown. Twelve parts were used as representatives from a process in statistical control. Parts were measured by an automatic in-line device and have a variable number of repeated measurements as a result of some data having been lost. There are 41 measurements in total.

10.4.1 Eq 18 is used to calculate the repeatability of the in-line measurement device. To do this, we calculate the sample variance for each shaft and then multiply each variance by the associated $n - 1$ for subgroup sample size, n. The sum of these results is the numerator of Eq 18. This sum is then divided by $41 - 12 = 29$. This is the sum of the degrees of freedom for each group. The shaft variance sums of squares calculation results are shown in Table 5.

10.4.2 The sum of the SS column in Table 5 is the numerator in Eq 18. This is 0.003 021, the sums of squares as a result of repeatability. The sum of the "$n - 1$" column (degrees of freedom) is 29. The repeatability variance estimate is 0.003 021⁄29 = 0.000 104 2. The square root of the variance is the estimate of repeatability. This result is 0.010 21. For the real object variance, use the equation in Appendix X1, Note X1.4. This gives us 0.004 25 as the variance component estimate of the real part. The standard deviation is the square root of this or 0.0652.

10.4.3 We can use the repeatability estimate and the real part estimate in Eq 39 to determine the distinct product categories for this example, Eq 53.

$$D \approx \frac{6(0.0652)}{4.24(0.0102)} \cong 9 \quad (53)$$

10.4.4 The gage is capable of discriminating nine distinct categories of real part dimension. Note that some of the

TABLE 4 Shaft Diameters

shaft	y	shaft	y
1	1.3237	8	1.3660
1	1.3171	8	1.3521
1	1.3145	8	1.3449
1	1.3213	8	1.3634
2	1.3697	8	1.3739
2	1.3943	8	1.3461
3	1.3686	8	1.3653
3	1.3453	8	1.3503
3	1.3370	9	1.3677
3	1.3385	9	1.3767
3	1.3275	9	1.3823
4	1.3183	10	1.3195
4	1.3275	10	1.2960
5	1.1846	11	1.4108
5	1.1950	11	1.4089
5	1.2063	12	1.4585
6	1.4249	12	1.4493
6	1.4151	12	1.4517
6	1.4276	12	1.4469
7	1.3307	12	1.4419
7	1.3341		

TABLE 5 Variance and Sums of Squares for Shaft Diameter Data

shaft Variance	n	n–1	SS
0.000017	4	3	0.000052
0.000302	2	1	0.000302
0.000239	5	4	0.000957
0.000042	2	1	0.000042
0.000118	3	2	0.000236
0.000043	3	2	0.000087
0.000006	2	1	0.000006
0.000115	8	7	0.000805
0.000054	3	2	0.000107
0.000276	2	1	0.000276
0.000002	2	1	0.000002
0.000037	5	4	0.000149

repeatability variations may have come from the variation within the object itself, as placement of the shaft on the measurements device was random. If these shafts are not perfectly round, then there is an additional component to be accounted for of within-part variation.

10.5 *Example 5: Appraiser and Object Components of Variance*—The following example is a standard gage R&R experiment. In such an experiment, we have *p* appraisers, *n* parts, and *m* repeats. In this example, $p = 3$, $n = 10$, and $m = 3$ making a total of 90 observations. The data are weight measurements in grams of residue taken from an industrial air-filtering system. Data are presented in Table 6.

10.5.1 The randomized order of the measurements is shown in the column labeled order, the appraiser is indicated in the column labeled Ap, and the object is shown in the column labeled object. A single scale was used for the weighing. The data are first summarized in preparation for the range method of analysis. This is shown in Tables 7 and 8. Subgroup ranges are plotted in Fig. 8. This figure gives a reasonable picture of a process in statistical control, except for the one point by Appraiser 2 above the upper control limit. Such points should

TABLE 7 Gage R&R Summarized Data

Ap.	object	avg	Range
1	1	791.11	9.60
1	2	798.92	12.20
1	3	815.45	5.61
1	4	755.35	1.22
1	5	789.64	20.56
1	6	825.69	6.32
1	7	790.59	4.27
1	8	824.06	2.14
1	9	798.25	1.58
1	10	830.68	9.58
2	1	791.92	6.91
2	2	803.27	14.53
2	3	817.54	5.74
2	4	764.23	25.50
2	5	792.19	8.81
2	6	827.04	5.05
2	7	796.32	14.14
2	8	834.91	12.27
2	9	799.60	6.35
2	10	834.06	15.51
3	1	785.91	5.92
3	2	793.49	9.43
3	3	811.62	0.88
3	4	754.72	1.31
3	5	788.62	4.01
3	6	822.84	10.51
3	7	795.29	6.27
3	8	821.41	8.08
3	9	792.74	9.82
3	10	825.59	7.82
		Rbar	8.40

TABLE 6 Data for Standard Gage R&R Experiment

order	Ap.	object	y	order	Ap	object	y	order	Ap	object	y
32	1	1	794.81	51	2	1	794.90	9	3	1	789.80
50	1	1	785.20	53	2	1	787.99	57	3	1	783.88
73	1	1	793.32	76	2	1	792.86	63	3	1	784.04
12	1	2	806.65	37	2	2	811.79	3	3	2	789.54
16	1	2	794.45	47	2	2	800.74	61	3	2	791.96
88	1	2	195.66	54	2	2	797.27	70	3	2	798.97
8	1	3	811.93	7	2	3	816.80	41	3	3	811.35
69	1	3	816.88	72	2	3	815.04	48	3	3	812.20
84	1	3	817.54	89	2	3	820.79	81	3	3	811.31
6	1	4	755.86	31	2	4	775.29	11	3	4	754.40
24	1	4	754.63	36	2	4	749.79	60	3	4	755.54
34	1	4	155.56	66	2	4	767.62	71	3	4	754.22
17	1	5	798.35	5	2	5	788.43	22	3	5	790.52
75	1	5	192.79	23	2	5	790.88	74	3	5	788.81
87	1	5	111.79	38	2	5	797.25	82	3	5	786.51
4	1	6	823.38	30	2	6	824.82	42	3	6	822.83
44	1	6	824.01	55	2	6	829.87	49	3	6	817.59
62	1	6	829.69	86	2	6	826.43	59	3	6	828.10
20	1	7	789.03	19	2	7	797.72	10	3	7	798.78
28	1	7	793.30	45	2	7	788.56	52	3	7	794.57
33	1	7	789.45	64	2	7	802.70	68	3	7	792.51
13	1	8	823.24	46	2	8	827.49	2	3	8	825.60
27	1	8	823.56	65	2	8	837.47	40	3	8	817.52
58	1	8	825.38	19	2	8	839.76	78	3	8	821.10
14	1	9	797.61	1	2	9	803.46	18	3	9	799.13
56	1	9	797.96	26	2	9	797.11	80	3	9	789.76
61	1	9	799.19	77	2	9	798.22	83	3	9	789.32
21	1	10	835.97	29	2	10	836.31	15	3	10	820.68
25	1	10	829.67	43	2	10	825.18	39	3	10	828.50
35	1	10	826.40	85	2	10	840.69	90	3	10	827.58

TABLE 8 Summarized Data by Appraiser

Appraiser	Average	Rbar
1	801.98	7.31
2	806.11	11.43
3	799.22	6.41
R_A	6.89	

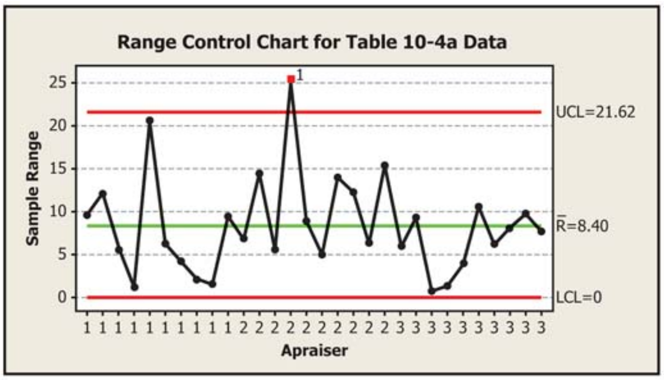

FIG. 8 Range Control Chart for Table 6 Data

be carefully examined to understand possible root causes. The overall average range in Fig. 8 is 8.4.

10.5.2 Using the equations outlined in Appendix X5, the gage R&R components are calculated. These are shown in Eq 54-57. The estimate of repeatability uses the overall average range from the control chart, Fig. 8, divided by the appropriate constant for a subgroup size of 3.

$$\hat{\sigma} = \frac{\bar{R}}{d_2} = \frac{8.4}{1.693} = 4.96 \tag{54}$$

10.5.3 The estimate of reproducibility is Eq 55 using the formula from Appendix X5.

$$\hat{\theta} = \sqrt{\left(\frac{R_A}{d_2^*}\right)^2 - \frac{(\hat{\sigma})^2}{nm}} \tag{55}$$

$$= \sqrt{\left(\frac{6.89}{1.912}\right)^2 - \frac{(4.96)^2}{(10)\ (3)}} = \sqrt{12.16} = 3.49 \tag{56}$$

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

$$S_G = \sqrt{\hat{\sigma}^2 + \hat{\theta}^2} = \sqrt{4.96^2 + 3.49^2} = 6.06 \tag{57}$$

10.5.5 To compute the estimate of real object variation, we need the individual object averages. An individual object average is the average of all measurements of an object by all appraisers. These are shown in Table 9.

10.5.6 The estimate of the real object standard deviation is:

$$\hat{\nu} = \frac{\bar{R}_n}{d_2^*} = \frac{72.01}{3.18} = 22.64 \tag{58}$$

10.5.7 Control charts should also be generated for these statistics. The data are arranged by appraiser and object with subgroups formed from the repeated measurements. In this manner, control limits for averages will be formed from the repeatability variation alone, and this allows for graphical estimation of the discriminating power of the gage as in the example in Fig. 7. The average and range control chart for this example is shown in Fig. 9.

TABLE 9 Summarized Object Averages

object	average
1	789.65
2	798.56
3	814.87
4	758.10
5	790.15
6	825.19
7	794.07
8	826.79
9	796.86
10	830.11
R_n	72.01

10.5.8 The method of the ANOVA uses the formulation outlined in Appendix X3. To carry out this analysis, there are a great many software packages that could be used. Any time-tested package will do. The formulas given in Appendix X3 could also be implemented in a spreadsheet-type program. In Table 10, a typical ANOVA output for data such as we are considering in this example is shown.

10.5.9 When the *p*-value for the interaction term is greater than the specified significance level (typically 0.05 or 0.1), then the SS term for interaction shall be added to the SS term for error. This can be done directly or the ANOVA model can be run without an interaction term. This has been done for Table 10. Results are shown in Table 11.

10.5.10 Notice that the sum of squares (SS) term as a result of error has changed, while the others have not. From the mean sum of squares (MSS) column and using the formulas given in Appendix X3, the variance components may be calculated. These are shown in Table 12.

10.5.11 Note that these values are similar to that obtained using the range method outlined previously. From the estimates in Table 12, the total gage R&R variance and standard deviation is computed as:

$$S_G = \sqrt{\hat{\sigma}^2 + \hat{\theta}^2} = \sqrt{5.02^2 + 3.34^2} = 6.03 \tag{59}$$

10.5.11.1 This is in good agreement with the range method.

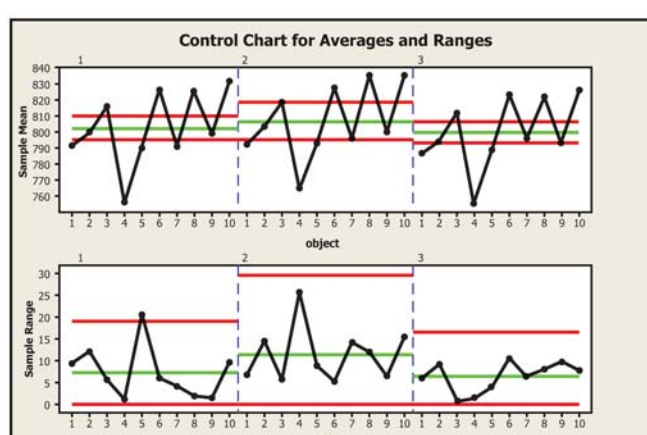

FIG. 9 Control Chart for Table 6 Data

TABLE 10 ANOVA Output for Gage R&R Experiment

Source	df	SS	MSS	F	p
object	9	39850.1	4427.80	257.18	0.00
appraiser	2	720.7	360.30	20.93	0.00
obj*Ap	18	309.9	17.20	0.62	0.87
Error	60	1656..9	27.60		
Total	89	42537.5			

TABLE 11 ANOVA Output for Gage R&R Experiment—No Interaction Term

Source	DF	SS	MSS	F	P
part	9	39850.10	4427.80	175.60	0.00
op	2	720.70	360.30	14.29	0.00
Error	78	1966.80	25.20		
Total	89	42537.50			

TABLE 12 Variance Components and Standard Deviations for Gage R&R Experiment

Source	Variance	Stdev
part	489.17	22.12
op	11.17	3.34
Error	25.22	5.02

10.5.12 It is common practice to measure the effect that gage R&R variation will have in the application for the measurement system under study. There are several methods in wide use. One method compares the range of gage R&R variation to either the specification range for the object being measured or to six times the standard deviation of the process output. A multiple of the gage R&R standard deviation is used for this comparison, and one typically finds 5.15 or 6 as the multiplier for the R&Rstandard deviation. The factor 5.15 is just the number of standard deviations that would consume a 99 % probability if centered on a normal distribution (that is, $\pm 2.575\sigma$ from the mean). The comparison using the factor 6 represents a 99.73 % probability in the same sense. The selection of either factor is a matter of industry preference. If T represents a specification range, the measure that compares the R&R variation to T is kS_G/T, where k is either 5.15 or 6. The final result is expressed as a percent.

10.5.13 When the alternative comparison to the process output is required, one has to be careful to use a reasonable estimate of the total process variation. For this purpose, an independent estimate of the total process standard deviation is recommended. If the objects used for the gage R&R study are used, it has to be assumed that the objects represent a random sample of the process under study. If w is an independent estimate of the process standard deviation, when this method is used and, assuming that we are using the factor of $k = 6$ as a multiplier of the standard deviation, the formula for the final result becomes S_G/w. In the present example, suppose that an independent estimate of the process standard deviation is $w = 40$. The gage R&R standard deviation is 6.03. The measure of total process consumption by gage R&R is therefore:

$$\frac{S_G}{w} = \frac{6.03}{40} = 0.151 \qquad (60)$$

10.5.14 This further means that the gage R&R standard deviation is approximately 15.1 % the size of the total process standard deviation. This method of comparison could also be carried out using the standard deviations of repeatability and reproducibility alone as well as the true object component. When this method is used, however, the several proportions of contribution to the total variation do not add up to 1. This has been a major criticism of this method.

10.5.15 In the classical method of comparison, the theory that the overall process variance is the sum of the variances of the independent components is used. For the model we are considering, the variance components are related as shown in Eq 1. The ratio of any variance component to the total variance is the measure of that component's contribution to the total variance. In this method, the several component measures add to one. For the variance components reported previously, Table 13 shows this analysis.

10.5.16 Using Eq 38 and quantities from Table 13, the discrimination ratio, D, is calculated as:

$$D \approx \frac{6\hat{v}}{4.24S_G} = \frac{6(22.12)}{4.24(6.03)} \approx 5 \qquad (61)$$

10.5.16.1 This gives us approximately five distinct object categories as the capability of this measurement system.

10.6 *Example 6: Gage Bias*—Table 14 contains hardness measurements and summary statistics of a known standard Rockwell C hardness sample at a force of 331 lb (150 kg). The known value of the standard is 54.5. Three measurements were made at twelve time periods over the course of several days.

10.6.1 From this data, we wish to determine the degree of bias present in the gage and the device repeatability. Generally, it is best to start with a control chart subgrouped by test in time order. The control chart would show if the twelve sets of measurement behave in a state of statistical control with respect to both the mean and the sample variation. That is, the control chart shows if the measurements process was stable and consistent during the experiment. The control chart is shown in Fig. 10.

10.6.2 The control chart gives every indication of a process in statistical control with constant mean and standard deviation. We can use the range method for a quick estimate of the repeatability standard deviation. This estimate is:

$$\hat{\sigma} = \frac{\bar{R}}{d_2} = \frac{6.55}{1.69} = 3.876 \qquad (62)$$

10.6.2.1 Here we have used the constant $d_2 = 1.69$ associated with a subgroup size of $m = 3$. For comparison, we can also compute the pooled variance and standard deviation using Eq 17. This result is $\hat{\sigma} = 3.564$. This value is what would have resulted for the root mean sums of square for error if we had performed an analysis of variance using "test" as a factor. The

TABLE 13 Variance Component % Ratios and Standard Deviations for Gage R&R Experiment

Source	Stdev	VarComp	% of Total
Repeatability	5.02	25.22	4.80
Reproducibility	3.34	11.17	2.13
object to object	22.12	489.18	93.08
Total Variation	22.93	525.56	100.00
Total Gage R&R	603	36.39	6.92

TABLE 14 Hardness Data

test	y1	y2	y3	average	S
1	49.18	50.75	47.85	49.26	1.45
2	53.90	54.96	52.50	53.79	1.23
3	58.38	51.63	56.48	55.50	3.48
4	46.20	50.59	56.27	51.02	5.05
5	56.14	52.22	49.10	52.49	3.53
6	49.08	54.72	57.53	53.77	4.31
7	49.61	55.26	50.78	51.88	2.98
8	48.80	58.45	55.21	54.15	4.91
9	56.57	52.20	52.38	53.72	2.47
10	54.06	48.10	51.32	51.16	2.99
11	51.29	48.44	53.78	51.17	2.67
12	47.07	53.02	57.03	52.37	5.01

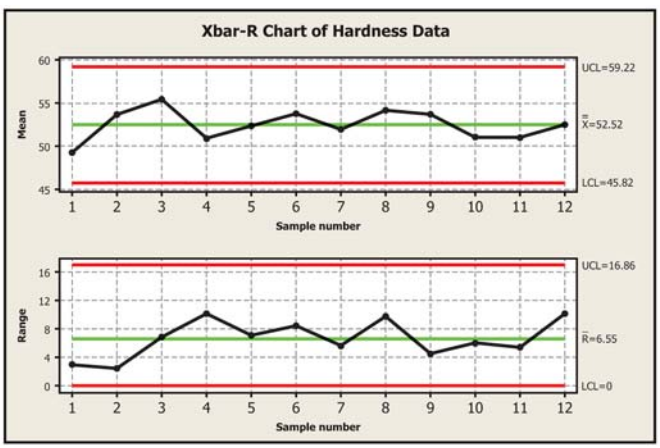

FIG. 10 Control Chart for Hardness Data

two estimates are in reasonable agreement. Since the ANOVA would show that the means are equal, we opt to use an estimate of variance based on all of this data. The overall standard deviation is 3.40. Here, again, this value does not depart radically from the two previous estimates.

10.6.3 To test for bias against the standard value of 54.5, use Eq 30.

10.6.4 Using 35 degrees of freedom and a confidence level of 95 %, the value of t is 2.03. The overall average is 52.523. Using x = 54.5, the confidence interval estimate for the bias is:

$$52.523.-54.5\pm\frac{2.03(3.40)}{\sqrt{36}} \qquad (63)$$

10.6.4.1 This bias interval is –0.827 to –3.127. This measurement device, therefore, possesses a negative bias that may be as large as –3.127.

10.7 *Example 7: Probability of Acceptance Given Measurement Value*—In the present example, we will use one side of a specification, but the application is readily adaptable to a two-sided specification. Let x be the true measure of an object and suppose that this variable has a mean of μ = 800 and a standard deviation of ν = 12. The gage R&R standard deviation is approximately σ = 4 and the gage is known to give unbiased measurements (B = 0). Also, the gage does not possess a significant linearity effect (m = 1). With these values, use Eq 42 and find the value of ρ.

$$\rho=\frac{12}{\sqrt{12^2+4^2}}\approx 0.949 \qquad (64)$$

10.7.1 The quantity, ρ, is the bivariate normal correlation between x and y. Also, note that the denominator in Eq 64 is the process (measurements) standard deviation. Suppose that the lower specification requirement for the characteristic being measured is a = 760. If a measure y = 762 is recorded, does it appear that this object is really greater than 760? The performance curve will show this probability. Using the lower limit version (Eq 43), we can calculate this probability.

$$P(x>a|y=762)=P(Z>-1.5284)=0.9368 \qquad (65)$$

10.7.2 For y = 762, the probability is 0.937. Thus, there is reasonably high assurance that this object meets the lower limit requirement. We can calculate this probability for other possible measurements, y, then plot the performance curve, $P(x > 760|y)$, against y. Fig. 11 shows the resulting curve.

10.7.3 Suppose that the gage possesses a positive bias, B = 2. This means that the gages tends to measure objects two units larger then they really are. The calculation is the same except that B = 2. For the same measurement, y = 762, the probability that this object really meets the lower limit requirement of a = 760 is 0.854. Fig. 12 shows the resulting performance curve.

10.7.4 Note that in this last example we are assuming that there is no linearity effect in the measurement process. The model is, therefore, $y = x + B$.

10.7.5 When we take the limit as ρ^2 approaches one (that is, gage R&R disappears), we find that $P(Z > -\infty)$ or $P(Z < \infty)$, so long as $y > a$ or $y < b$, respectively. This is certain acceptance of a certain in-spec object dimension. If y is just under a or just over b, we get $P(Z < -\infty)$ = 0 or certain detection of an out-of-spec part. If $y = a$ or $y = b$ exactly, we get the probability that the part meeting the limiting requirement is 0.5, since, when $y = a$, there is no more reason to believe $y < a$ than $y > a$, likewise for b.

10.8 *Example 8: Probability of Gage Acceptance Given True Measurement*—Take the previous example and suppose the true object dimension is x = 762. Suppose further that B = 0 and m = 1. Using the lower limit version of Eq 45, the calculation is:

$$P(x>760|y=762)=P(Z>-0.5)=0.691 \qquad (66)$$

10.8.1 Thus, there is an approximate 69 % chance that this object meets the lower limit requirement.

10.8.2 When a linearity parameter ($m \neq 1$) is used, as is often the case when we want to calibrate the measurement system to a set of standards (x), the performance curve may be used to good advantage. Continue to use a lower limit requirement of a = 760 and suppose that the gage R&R standard deviation is σ = 2. The true object continues to be normally distributed with mean, μ = 800, and standard deviation, ν = 12. Suppose the relation between x and y is $y = 1.1\ x + 3$. Here, we have introduced a linearity parameter, m = 1.1. Note also the bias, B = 3. With these assumptions, we want to determine a value for y such that there is at least 90 % probability that x meets the required minimum, 760.

10.8.3 This can be done by constructing a gage performance curve of the type previously discussed. Use Eq 43 with the given inputs. The curve is shown in Fig. 13.

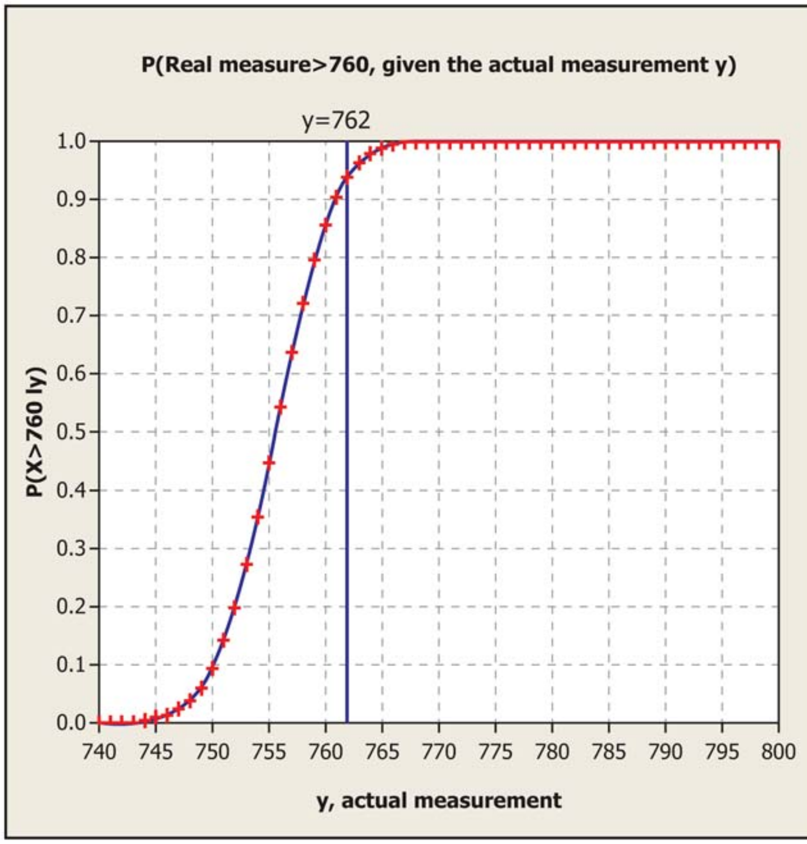

FIG. 11 Gage Performance Curve, Lower Limit, and Bias = 0

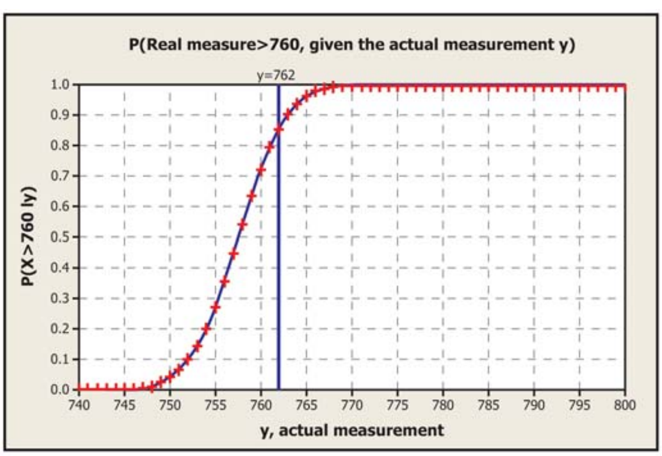

FIG. 12 Gage Performance Curve, Lower Limit, and Bias = 2

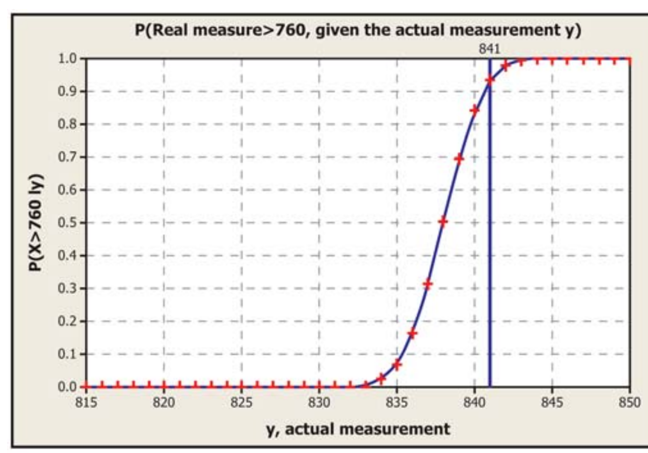

FIG. 13 Calibration of Measurement *y* to True Object *x* Lower Limit (760)—Functional Relation is $y = 1.1x + 3$

10.8.4 From the data used to construct this curve, we may deduce that as long as y is at least 841, then x will be at least 760. The actual probability at $y = 841$ is 0.932.

11. Keywords

11.1 analysis of variance; ANOVA; bias; discrimination ratio; gage consistency; gage performance; gage R&R; gage stability; linearity; repeatability; reproducibility; resolution; variance components

APPENDIXES

(Nonmandatory Information)

X1. MODEL FOR SIMPLE REPEATABILITY, ONE "APPRAISER," *n* OBJECTS, *m* REPEATS PER OBJECT

$$y_{ij} = x_i + \varepsilon_{ij},\ i = 1...n,\ j = 1...m,\ x_i \Rightarrow N(\mu, \nu^2),\ \varepsilon_{ij} \Rightarrow N(0, \sigma^2),$$
$$\mathrm{cov}(\nu, \varepsilon) = 0,\ \mathrm{cov}(\varepsilon_{i,h}, \varepsilon_{i,r}) = 0$$

NOTE X1.1—This study design includes n objects to be measured and m repeated measurements per object. There is one appraiser. The x_i, represents true object values and the ε_i, represent repeatability errors. MSSE indicates "mean sums of squares for error" and MSSO indicates "mean sums of squares for objects."

NOTE X1.2—Confidence interval for σ^2: Quantity SSE/σ^2 has a chi-square distribution with $n(m-1)$ degrees of freedom. Choose a confidence coefficient, $C = 1 - \alpha$. Select points a and b from this distribution such that $P(a <$ chi-square $< b) = 1 - \alpha$. The confidence interval for σ is:

$$\sqrt{\frac{SSE}{b}} \leq \sigma \leq \sqrt{\frac{SSE}{a}} \quad \text{at confidence } C = 1 - \alpha$$

NOTE X1.3—Confidence interval for ν/σ: Let R = MSSE/MSSO. The following confidence interval for the S/N ratio (ν/σ) results. Use the $F[(n(m-1), n-1]$ distribution, find values a and b for q such that $P(a < q < b) = 1 - \alpha$, where $1 - \alpha$ signifies the appropriate confidence level. The confidence interval for the ratio ν/σ is:

$$\sqrt{\frac{1-bR}{mbR}} \leq \frac{\nu}{\sigma} \leq \sqrt{\frac{1-aR}{maR}} \quad \text{at confidence } C = 1 - \alpha$$

NOTE X1.4—When this design is unbalanced, there are variable m_i measurements per object measured where $\Sigma m_i = N$, the total number of all measurements. Use the following modified formulas for the variance components estimates.

For the estimate of σ^2, use Eq 18.

For the estimate of ν^2, use Eq 18 for MSSE and the following as the estimate of ν^2. See Ref (7) for details.

$$\frac{(n-1)(MSSO - MSSE)}{N - \left(\sum_{i=1}^{k} m_i^2\right)/N}$$

where:

$$MSSO = \frac{\sum_{i=1}^{n} m_i(\bar{y}_{i.} - \bar{y}_{..})^2}{n-1}$$

TABLE X1.1 Model for Simple Repeatability

Source of Variance	d.f	Sum of Squares Formula (SS)	Mean SS (MSS)	Variance Component Estimate	Expected Value of Variance Component Estimate
Objects	$n-1$	$m\sum_{i=1}^{n}(\bar{y}_{i.}-\bar{y}_{..})^2$	$\frac{SSO}{n-1}$	$\frac{MSSO - MSSE}{m}$	ν^2
Repeatability Error	$n(m-1)$	$\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-\bar{y}_{i.})^2$	$\frac{SSE}{n(m-1)}$	$\frac{SSE}{n(m-1)}$	σ^2
Total	$nm-1$	$\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-\bar{y}_{..})^2$	$\frac{SST}{nm-1}$	$\frac{SST}{nm-1}$	$\nu^2 + \sigma^2$

X2. MODEL FOR SIMPLE REPEATABILITY, *p* APPRAISERS, *n* OBJECTS, AND ONE REPEAT ONLY

$$y_{ij} = x_i + u_j + \varepsilon_{ij},\ i = 1...n,\ j = 1...\ p,$$
$$\varepsilon_{ij} \Rightarrow N(0, \sigma^2),\ u_i \Rightarrow N(0, \theta^2),\ x_j \Rightarrow N(0, \nu^2),\ x, u, \varepsilon \Rightarrow \text{independent}$$

NOTE X2.1—This study design includes *n* objects to be measured, *p* appraisers, and one measurement per object. All appraisers measure all objects once. There are a total of *np* observations. The x_i represent real part values and the ε_i represent repeatability errors. SSE indicates "sums of squares for error," SSO indicates "sums of squares for objects," SSA indicates "sum of squares for appraisers" and SST indicates "sum of squares total". There is no separate interaction term in this model. The Mean Sum of Squares for each term is calculated by dividing the Sum of Squares term by its associated degrees of freedom.

NOTE X2.2—Analytical formula for computing SSE:

$$SSE = \sum_{i=1}^{n} \sum_{j=1}^{p} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$$

NOTE X2.3—When $p = 2$ appraisers is used, this layout is similar to a "paired" sampling design. There are *n* pairs of data in two columns. Let column 1 be labeled *x* and column 2, *y*. Let S_x and S_y be the sample standard deviations of the *x* and *y* values, respectively; let S_p be the pooled standard deviation between the *x* and *y* values; let *r* be the sample correlation coefficient between the *x* and *y* columns; and let *d* be the difference in matched pairs, $d = x - y$. Let S_d be the standard deviation of the variable, *d*. The MSSE and MSSO terms may be shown to be:

$$MSSE = \frac{S_d^2}{2} = S_p^2 - rS_xS_y$$

$$MSSO = S_p^2 + rS_xS_y$$

NOTE X2.4—Confidence interval for σ^2: use the chi-square distribution with $(n - 1)(p - 1)$ degrees of freedom. The following form is used.

$$\frac{SSE}{b} \le \sigma^2 \le \frac{SSE}{a} \quad \text{at confidence } C = 1 - \alpha$$

Let *q* have a chi-square distribution. Constants *a* and *b* are selected from the chi-square distribution with $(n - 1)(p - 1)$ degrees of freedom such that:

$$P(a < q < b) = 1 - \alpha$$

NOTE X2.5—Confidence Interval for θ^2/σ^2. Use the *F* distribution to construct a confidence interval for the ratio θ^2/σ^2 or some function thereof. Let *q* have an $F[(p - 1),(n - 1)(p - 1)]$ distribution. For confidence $C = 1 - \alpha$, find *a* and *b* for which $P(a < F < b) = 1 - \alpha$. The resulting confidence interval is:

$$\frac{MSSA - bMSSE}{nbMSSE} \le \frac{\theta^2}{\sigma^2} \le \frac{MSSA - aMSSE}{naMSSE}$$

NOTE X2.6—Confidence Interval for ν^2/σ^2. Use the *F* distribution to construct a confidence interval for the ratio ν^2/σ^2 or some function thereof. Let *q* have an $F[(n - 1),(n - 1)(p - 1)]$ distribution. For confidence $C = 1 - \alpha$, find *a* and *b* for which $P(a < F < b) = 1 - \alpha$. The resulting confidence interval is:

$$\frac{MSSO - bMSSE}{pbMSSE} \le \frac{\nu^2}{\sigma^2} \le \frac{MSSO - aMSSE}{paMSSE}$$

TABLE X2.1

Source of Variance	d.f	Sum of Squares Formula (SS)	Mean SS	Variance Component Estimate	Expected Value of Variance Estimate
Object	$n - 1$	$p\sum_{i=1}^{n} (\bar{y}_{i..} - \bar{y}_{...})^2$	$\frac{SSO}{n-1}$	$\frac{MSSO - MSSE}{p}$	ν^2
Appraiser	$p - 1$	$n\sum_{j=1}^{p} (\bar{y}_{.j} - \bar{y}_{..})^2$	$\frac{SSA}{p-1}$	$\frac{MSSA - MSSE}{n}$	θ^2
Error	$(n - 1)(p - 1)$	By subtraction SSE = SST-SSO-SSA (see Note X2.2)	$\frac{SSE}{(n-1)(p-1)}$	$MSSE$	σ^2
Total	$np - 1$	$\sum_{i=1}^{n}\sum_{j=1}^{p} (y_{ij} - \bar{y}_{..})^2$	$\frac{SSP}{np-1}$	$MSST$	$\nu^2 + \theta^2 + \sigma^2$

X3. MODEL 3: *p* APPRAISERS, *n* OBJECTS, *k* REPEATS PER OBJECT

$$y_{ijk} = x_i + u_j + w_{ij} + \varepsilon_{ijk}\ ,\ i = 1...n,\ j = 1...p,\ k = 1...m,$$
$$\varepsilon_{ijk} \Rightarrow N(0,\ \sigma^2),\ u_i \Rightarrow N(0,\ \theta^2),\ x_j \Rightarrow N(\mu,\ \nu^2),\ w_{ij} \Rightarrow N(0,\ \alpha^2)$$

NOTE X3.1—This study design includes n objects to be measured, p appraisers, and m repeated measurements per object. All appraisers measure all objects m times. The model is, therefore, a balanced design. There are a total of nmp observations. The x_i represents part true values; the u_j represents appraiser variations; the w_{ij} represents part-appraiser interactions; and the ε_{ijk} represents repeatability errors. MSSE indicates "mean sums of squares for error," MSSO indicates "mean sums of squares for objects," MSSA indicates "mean sum of squares for appraisers," and MSSI indicates "mean sums of squares for interaction." For treatment of unbalanced cases, see Ref (**7**).

TABLE X3.1

Source	d.f	Sum of Squares Formula, SS	Mean SS Formula	Estimate of Variance Component	Expected Value of Variance Estimate
Objects	$n-1$	$mp\sum_{i=1}^{n}(\bar{y}_{i.}-\bar{y}_{..})^2$	$\frac{SSO}{n-1}$	$\frac{MSSO-MSSI}{mp}$	ν^2
Appraisers	$p-1$	$nm\sum_{i=1}^{p}(\bar{y}_{.i.}-\bar{y}_{..})^2$	$\frac{SSA}{p-1}$	$\frac{MSSA-MSSI}{mn}$	θ^2
Interaction	$(n-1)(p-1)$	By subtraction SSI = SST-SSO-SSA-SSE	$\frac{SSI}{(n-1)(p-1)}$	$\frac{MSSI-MSSE}{m}$	α^2
Error	$np(m-1)$	$\sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{k=1}^{m}(y_{ijk}-\bar{y}_{ijk.})^2$	$\frac{SSE}{np(m-1)}$	$MSSE$	σ^2

X4. MODEL FOR SIMPLE REPEATABILITY, ONE "APPRAISER," *n* OBJECTS, *m* REPEATS PER OBJECT, STANDARDS USED

$$y_{ij} = x_i + \varepsilon_{ij},\ i = 1...n,\ j = 1...m,\ \varepsilon_{ij} \Rightarrow N(0, \sigma^2),$$
$$\mathrm{cov}(x, \varepsilon) = 0,\ \mathrm{cov}(\varepsilon_{i,h}, \varepsilon_{i,r}) = 0$$

NOTE X4.1—*Repeatability Estimate*—The value x represents the standard value and y represents the measurement. See Note X4.5.

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(x_i - y_{ij})^2}{nm}}$$

NOTE X4.2—A confidence interval estimate for σ may be constructed by noting that $\nu = nm\hat{\sigma}^2/\sigma^2$ has a chi-square distribution with nm degrees of freedom. Find a and b for which $P(a < \nu < b) = 1 - \alpha$. The following equation is the $100(1 - \alpha)\%$ confidence interval for unknown σ^2.

$$\frac{nm\hat{\sigma}^2}{b} \le \sigma^2 \le \frac{nm\hat{\sigma}^2}{a}$$

NOTE X4.3—This study design includes n objects to be measured and m repeated measurements per object. There is one appraiser. The x_i represents known standard values and the ε_i represents repeatability errors.

NOTE X4.4—The value $\mathrm{cov}(x,\varepsilon)$ indicates covariance for random variables x and ε. This is assumed to be 0. This further means that these random variables are independent.

NOTE X4.5—In this analysis, the system is assumed to be without bias as well as not possessing a significant linearity effect. A bias and linearity effect would imply a model of the form: $y = Mx + B + \varepsilon$ where $M \neq 1$ and $B \neq 0$. The squared difference, $y - x$, in this model would not represent the pure repeatability error. When using this model, a study should first be undertaken to demonstrate that $B = 0$ and $M = 1$. This may be approached by first performing a simple linear regression of y on x and showing, using hypothesis tests or confidence intervals, that $B = 0$ and $M = 1$ cannot be rejected.

TABLE X5.1 Factors d_2^*

Number of Groups	2	3	4	5	6	7	8	9	10
d_2^*	1.414	1.912	2.239	2.481	2.672	2.829	2.963	3.078	3.180

X5. LONG METHOD GAGE R&R BASED ON SAMPLE RANGES *n* OBJECTS, *m* REPEATS PER OBJECT, AND *p* APPRAISERS

X5.1 The range method requires a balanced study design. There are n objects and p appraisers where each appraiser measures each object m times. The total number of measurements in this design is $N = nmp$.

X5.1.1 *Calculation of the Repeatability Standard Deviation*—For each appraiser, there are n subgroups of size m. This makes a total of np subgroups. Each subgroup is comprised of repeated measurements of the same object. The subgroup range statistic is, therefore, measuring repeatability variation; and the behavior of the subgroup range is an indicator of the consistency of the measurement process. The estimate of the repeatability standard deviation is: $\hat{\sigma} = \bar{R}/d_2$ where $\bar{R}$ is the average of the np subgroup ranges. A control chart for the subgroup range can be constructed to study variance constancy or consistency of repeatability. See Practice E2587 or Ref (**8**) for additional tables of d_2 and control chart factors.

X5.1.2 *Calculation of the Reproducibility Standard Deviation:*

X5.1.2.1 For each appraiser, calculate the overall average of all mn measurements. There are p such averages. Let R_A be the range of the p appraiser averages. The estimate of the reproducibility standard deviation is:

$$\hat{\theta} = \sqrt{\left(\frac{R_A}{d_2^*}\right)^2 - \frac{(\hat{\sigma})^2}{nm}}$$

X5.1.2.2 The constant d_2^* is a function of the number of appraisers, p (number of groups). See Table X5.1.

X5.1.3 *Calculation of the Real Object Standard Deviation*—The average value of an object is the average of all measurements by all appraisers for each object. There are n such averages each comprising mp measurements. Let R_n be the range of these n averages. The estimate of the real object standard deviation is $R_n/\ d_2^*$ where the constant in the denominator is taken from Table X5.1 (the number of groups is the number of objects measured).

X5.1.4 *Calculation of Total Gage R&R*—Sum the squared estimates from X5.1.1 and X5.1.2. This is the gage R&R variance. The gage R&R standard deviation is the square root of the variance.

X5.1.5 *Attachments*—Gage R&R long form, range method; data collection, and analysis forms.

TABLE X5.2 Long Form Gage R&R, Range Method, Data Collection Form

Operator	Part										Averages
Trial #	1	2	3	4	5	6	7	8	9	10	
A_1											
A_2											
A_3											
Averages											$Xbar_A$ =
Ranges											$Rbar_A$ =
B_1											
B_2											
B_3											
Averages											$Xbar_B$ =
Ranges											$Rbar_B$ =
C_1											
C_2											
C_3											
Averages											$Xbar_C$ =
Ranges											$Rbar_C$ =
Part Avg[A,B] ($Xbar_p$)											R_p =
Rbar[C]											
$Xbar_{diff}$	$Xbar_{diff}$ = (Max Xbar – Min Xbar) =										
UCL_R[D]	$UCL_R = D_4$(Rbar) =			D_4 =							

[A] A "part average" is the average of all measurements of a part, by all operators.
[B] The range of part averages is R_p.
[C] Rbar is the average of the three operator ranges.
[D] D_4 = 3.27 for 2 replications and 2.58 for 3 replications. UCL_R represents the limit of individual R's. Identify the cause and correct. Repeat these readings using the same measurement system and operator as originally used or discard values and adjust R&R analysis for the reduced sample.

<table>
<tr><th>Measurement Unit Analysis</th><th>% Total Variation</th></tr>
<tr><td>Repeatability: Equipment Variation (EV)
$EV = R(k_1)$
=
<table><tr><th>Trials</th><th>k_1</th></tr><tr><td>2</td><td>4.56</td></tr><tr><td>3</td><td>3.05</td></tr></table></td><td>% EV = 100(EV/TV)</td></tr>
<tr><td>Reproducibility: Appraiser Variation (AV)
$AV = \sqrt{k_2^2(X_{diff})^2 - ((EV)^2)/nr)}$
=
n = number of parts
r = number of trials
<table><tr><th>Operator</th><th>2</th><th>3</th></tr><tr><td>k_2</td><td>3.65</td><td>2.7</td></tr></table></td><td>% AV = 100(AV/TV)

% R&R = 100(R&R/TV)</td></tr>
<tr><td>Gage R&R: Repeatability & Reproducibility
$RR = \sqrt{(EV)^2 + (AV)^2}$
=

Part Variation (PV)
$PV = \sqrt{(k_3)(R_p)}$
=

Total Variation (TV)
$TV = \sqrt{(RR)^2 + (PV)^2}$
=
<table><tr><th>Parts</th><th>k_3</th></tr><tr><td>2</td><td>3.65</td></tr><tr><td>3</td><td>2.70</td></tr><tr><td>4</td><td>2.30</td></tr><tr><td>5</td><td>2.08</td></tr><tr><td>6</td><td>1.93</td></tr><tr><td>7</td><td>1.82</td></tr><tr><td>8</td><td>1.74</td></tr><tr><td>9</td><td>1.67</td></tr><tr><td>10</td><td>1.62</td></tr></table></td><td>%PV = 100(PV/TV)

1. Part tolerance may be substituted for TV in % calculation
2. All calculations are based on 5.15*sigma
3. Normal Distribution of parts and R&R errors assumed

Data Analyst:
Gage #:
Part #:
Process:</td></tr>
</table>

1. All calculations are based upon predicting 5.15 sigma (99.0% of the area under the normal distribution curve).
   K_1 is 5.15/d2*, where d2* is dependent on the number of trials(m) and the number of parts times the number of appraisers.
2. Number of Operators (g) which is greater than 15, d_2 values are from tables.
3. AV - if a negative value is calculated under the square root sign, the appraiser variation (AV) defaults to zero (0).
4. K_2 is 5.15/d_2*, where d_2* is dependent on the number of operators (m) and (g) is 1, since there is only one range calculation.
5. K_3 is 5.15/d_2*, where d_2* is dependent on the number of parts (m) and (g) is 1, since there is only one range calculation.
6. d_2* is obtained from Table D_3, "Quality Control and Industrial Statistics," A.J. Duncan.

FIG. X5.1 Long Form Gage R&R, Range Method, Analysis Form

X6. STATISTICAL TABLES USEFUL FOR THE ANALYSIS OF MEASUREMENT DATA

NOTE X6.1—$\hat{\sigma}^2 = (\bar{R}/d_2^*)^2$, where $\bar{R}$ is the average range of k subgroups, each of size n.

TABLE X6.1 Control Chart Factors

subgroup size	factors for averages		factors for ranges		
	A	A2	d_2	D3	D4
1			1.1280		3.2670
2	2.1213	1.8806	1.1280		3.2686
3	1.7321	1.0231	1.6930		2.5735
4	1.5000	0.7285	2.0590		2.2822
5	1.3416	0.5768	2.3260		2.1144
6	1.2247	0.4833	2.5340		2.0039
7	1.1339	0.4193	2.7040	0.0758	1.9242
8	1.0607	0.3726	2.8470	0.1359	1.8641
9	1.0000	0.3367	2.9700	0.1838	1.8162
10	0.9487	0.3082	3.0780	0.2232	1.7768
11	0.9045	0.2851	3.1730	0.2559	1.7441
12	0.S660	0.2658	3.2580	0.2836	1.7164
13	0.8321	0.2494	3.3360	0.3076	1.6924
14	0.8018	0.2353	3.4070	0.3290	1.6710
15	0.7746	0.2231	3.4720	0.3476	1.6524
Id	0.7500	0.2123	3.5320	0.3638	1.6362
17	0.7276	0.2028	3.5880	0.3788	1.6212
18	0.7071	0.1943	3.6400	0.3918	1.6082
19	0.6882	0.1866	3.6890	0.4039	1.5961
20	0.6708	0.1796	3.7350	0.4145	1.5855
21	0.6547	0.1733	3.7780	0.4251	1.5749
22	0.6396	0.1675	3.8190	0.4344	1.5656
23	0.6255	0.1621	3.8580	0.4432	1.5568
24	0.6124	0.1572	3.8950	0.4516	1.5484
25	0.6000	0.1526	3.9310	0.4589	1.5411

TABLE X6.2 Factors for Converting an Average Range into an Unbiased Variance Estimate—Normal Distribution Assumed

	d_2^*, as related to n = subgroup sample size, and k = number of subgroups								
k	n=2	n=3	n=4	n=5	n=6	n=7	n=8	n=9	n=10
1	1.414	1.912	2.239	2.481	2.672	2.829	2.963	3.078	3.180
2	1.279	1.806	2.151	2.405	2.604	2.767	2.906	3.024	3.129
3	1.231	1.769	2.121	2.379	2.581	2.746	2.886	3.006	3.112
4	1.206	1.750	2.105	2.366	2.569	2.736	2.877	2.997	3.104
5	1.191	1.739	2.096	2.358	2.562	2.730	2.871	2.992	3.099
6	1.180	1.731	2.090	2.353	2.557	2.725	2.867	2.988	3.095
7	1.173	1.726	2.086	2.349	2.554	2.722	2.864	2.986	3.093
8	1.168	1.722	2.082	2.346	2.552	2.720	2.862	2.984	3.091
9	1.163	1.719	2.080	2.344	2.550	2.718	2.860	2.982	3.089
10	1.160	1.716	2.078	2.342	2.548	2.717	2.859	2.981	3.088
11	1.157	1.714	2.076	2.340	2.547	2.716	2.858	2.980	3.087
12	1.154	1.712	2.075	2.339	2.546	2.715	2.857	2.979	3.087
13	1.153	1.711	2.073	2.338	2.545	2.714	2.856	2.978	3.086
14	1.151	1.709	2.072	2.337	2.544	2.713	2.855	2.978	3.086
15	1.149	1.708	2.071	2.337	2.543	2.712	2.855	2.977	3.085
16	1.148	1.707	2.071	2.336	2.543	2.712	2.854	2.977	3.084
17	1.147	1.707	2.070	2.335	2.542	2.712	2.854	2.977	3.084
18	1.146	1.706	2.069	2.335	2.542	2.711	2.854	2.976	3.084
19	1.145	1.705	2.069	2.334	2.541	2.711	2.853	2.976	3.083
20	1.144	1.705	2.068	2.334	2.541	2.710	2.853	2.975	3.083
infinity	1.128	1.693	2.059	2.326	2.534	2.704	2.847	2.970	3.078

REFERENCES

(1) Eisenhart, Churchill, "Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems," *Journal of Research*, Vol 67C, No. 2, 1963.

(2) Montgomery, Douglas, C, Peck, Elizabeth, A., and Vining, Geoffrey, G., *Introduction to Linear Regression Analysis, Wiley Series in Probability and Statistics* , 4th edition, 2006.

(3) *Measurement Systems Analysis, MSA, Reference Manual*, Automotive Industry Action Group (AIAG), 3rd edition 2004.

(4) Wheeler, Donald, *Evaluating the Measurement Process*, 2nd edition, SPC Press, 1990.

(5) *Measurements Systems Analysis Manual*, MSA-3, Automotive Industry Action Group, Southfield, MI, 2002.

(6) Shewhart, W.A., *Economic Control of Quality of Manufactured Product*, D. Van Nostrand Company, Inc., 1931.

(7) *Handbook of Probability and Statistics*, CRC Press, Boca Raton, FL.

(8) Duncan, A. J., *Quality Control and Industrial Statistics*, 5th edition, Irwin, Homewood, IL, 1986.