Designation: C1215 – 92 (Reapproved 2012)ε1

Standard Guide for Preparing and Interpreting Precision and Bias Statements in Test Method Standards Used in the Nuclear Industry[1]

This standard is issued under the fixed designation C1215; 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.

ε[1] NOTE—Changes were made editorially in June 2012.

INTRODUCTION

Test method standards are required to contain precision and bias statements. This guide contains a glossary that explains various terms that often appear in these statements as well as an example illustrating such statements for a specific set of data. Precision and bias statements are shown to vary according to the conditions under which the data were collected. This guide emphasizes that the error model (an algebraic expression that describes how the various sources of variation affect the measurement) is an important consideration in the formation of precision and bias statements.

1. Scope

1.1 This guide covers terminology useful for the preparation and interpretation of precision and bias statements. This guide does not recommend a specific error model or statistical method. It provides awareness of terminology and approaches and options to use for precision and bias statements.

1.2 In formulating precision and bias statements, it is important to understand the statistical concepts involved and to identify the major sources of variation that affect results. Appendix X1 provides a brief summary of these concepts.

1.3 To illustrate the statistical concepts and to demonstrate some sources of variation, a hypothetical data set has been analyzed in Appendix X2. Reference to this example is made throughout this guide.

1.4 It is difficult and at times impossible to ship nuclear materials for interlaboratory testing. Thus, precision statements for test methods relating to nuclear materials will ordinarily reflect only within-laboratory variation.

1.5 No units are used in this statistical analysis.

1.6 This guide does not involve the use of materials, operations, or equipment and does not address any risk associated.

2. Referenced Documents

2.1 *ASTM Standards:*[2]

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

E691 Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method

2.2 *ANSI Standard:*

ANSI N15.5 Statistical Terminology and Notation for Nuclear Materials Management[3]

3. Terminology for Precision and Bias Statements

3.1 *Definitions:*

3.1.1 *accuracy (see bias)* —(*1*) bias. (*2*) the closeness of a measured value to the true value. (*3*) the closeness of a measured value to an accepted reference or standard value.

3.1.1.1 *Discussion*—For many investigators, accuracy is attained only if a procedure is both precise and unbiased (see *bias*). Because this blending of precision into accuracy can result occasionally in incorrect analyses and unclear statements of results, ASTM requires statement on bias instead of accuracy.[4]

3.1.2 *analysis of variance (ANOVA)*—the body of statistical theory, methods, and practices in which the variation in a set of data is partitioned into identifiable sources of variation.

[1] This guide is under the jurisdiction of ASTM Committee C26 on Nuclear Fuel Cycle and is the direct responsibility of Subcommittee C26.08 on Quality Assurance, Statistical Applications, and Reference Materials.

Current edition approved June 1, 2012. Published June 2012. Originally approved in 1992. Last previous edition approved in 2006 as C1215–92(2006). DOI: 10.1520/C1215-92R12E01.

[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] Available from American National Standards Institute (ANSI), 25 W. 43rd St., 4th Floor, New York, NY 10036, http://www.ansi.org.

[4] Refer to *Form and Style for ASTM Standards*, 8th Ed., 1989, ASTM.

Sources of variation may include analysts, instruments, samples, and laboratories. To use the analysis of variance, the data collection method must be carefully designed based on a model that includes all the sources of variation of interest. (See Example, X2.1.1)

3.1.3 *bias (see accuracy)*—a constant positive or negative deviation of the method average from the correct value or accepted reference value.

3.1.3.1 *Discussion*—Bias represents a constant error as opposed to a *random error*.

(a) A method bias can be estimated by the difference (or relative difference) between a measured average and an accepted standard or reference value. The data from which the estimate is obtained should be statistically analyzed to establish bias in the presence of random error. A thorough bias investigation of a measurement procedure requires a statistically designed experiment to repeatedly measure, under essentially the same conditions, a set of standards or reference materials of known value that cover the range of application. Bias often varies with the range of application and should be reported accordingly.

(b) In statistical terminology, an estimator is said to be unbiased if its expected value is equal to the true value of the parameter being estimated. (See Appendix X1.)

(c) The bias of a test method is also commonly indicated by analytical chemists as *percent recovery*. A number of repetitions of the test method on a reference material are performed, and an average percent recovery is calculated. This average provides an estimate of the test method bias, which is multiplicative in nature, not additive. (See Appendix X2.)

(d) Use of a single test result to estimate bias is strongly discouraged because, even if there were no bias, random error alone would produce a nonzero bias estimate.

3.1.4 *coefficient of variation*—see *relative standard deviation*.

3.1.5 *confidence interval*—an interval used to bound the value of a population parameter with a specified degree of confidence (this is an interval that has different values for different random samples).

3.1.5.1 *Discussion*—When providing a confidence interval, analysts should give the number of observations on which the interval is based. The specified degree of confidence is usually 90, 95, or 99 %. The form of a confidence interval depends on underlying assumptions and intentions. Usually, confidence intervals are taken to be symmetric, but that is not necessarily so, as in the case of confidence intervals for variances. Construction of a symmetric confidence interval for a population mean is discussed in Appendix X3.

It is important to realize that a given confidence-interval estimate either does or does not contain the population parameter. The degree of confidence is actually in the procedure. For example, if the interval (9, 13) is a 90 % confidence interval for the mean, we are confident that the procedure (take a sample, construct an interval) by which the interval (9, 13) was constructed will 90 % of the time produce an interval that does indeed contain the mean. Likewise, we are confident that 10 % of the time the interval estimate obtained will not contain the mean. Note that the absence of sample size information detracts from the usefulness of the confidence interval. If the interval were based on five observations, a second set of five might produce a very different interval. This would not be the case if 50 observations were taken.

3.1.6 *confidence level*—the probability, usually expressed as a percent, that a confidence interval will contain the parameter of interest. (See discussion of *confidence interval* in Appendix X3.)

3.1.7 *error model*—an algebraic expression that describes how a measurement is affected by error and other sources of variation. The model may or may not include a sampling error term.

3.1.7.1 *Discussion*—A measurement error is an error attributable to the measurement process. The error may affect the measurement in many ways and it is important to correctly model the effect of the error on the measurement.

(*a*) Two common models are the additive and the multiplicative error models. In the additive model, the errors are independent of the value of the item being measured. Thus, for example, for repeated measurements under identical conditions, the additive error model might be

$$X_i = \mu + b + \varepsilon_i \tag{1}$$

where:

X_i = the result of the i^{th} measurement,
μ = the true value of the item,
b = a bias, and
ε_i = a random error usually assumed to have a normal distribution with mean zero and variance σ^2.

In the multiplicative model, the error is proportional to the true value. A multiplicative error model for percent recovery (see *bias*) might be:

$$X_i = \mu b \varepsilon_i \tag{2}$$

and a multiplicative model for a neutron counter measurement might be:

$$X_i = \mu + \mu b + \mu \cdot \varepsilon_i \tag{3}$$

$$= \mu(1 + b + \varepsilon_i)$$

(*b*) Clearly, there are many ways in which errors may affect a final measurement. The additive model is frequently assumed and is the basis for many common statistical procedures. The form of the model influences how the error components will be estimated and is very important, for example, in the determination of measurement uncertainties. Further discussion of models is given in the Example of Appendix X2 and in Appendix X4.

3.1.8 *precision*—a generic concept used to describe the dispersion of a set of measured values.

3.1.8.1 *Discussion*—It is important that some quantitative measure be used to specify precision. A statement such as, "The precision is 1.54 g" is useless. Measures frequently used to express precision are *standard deviation, relative standard deviation, variance, repeatability, reproducibility, confidence interval*, and *range*. In addition to specifying the measure and the precision, it is important that the number of repeated

measurements upon which the precision estimated is based also be given. (See Example, Appendix X2.)

(*a*) It is strongly recommended that a statement on precision of a measurement procedure include the following:

(1) A description of the procedure used to obtain the data,

(2) The number of repetitions, *n*, of the measurement procedure,

(3) The sample mean and standard deviation of the measurements,

(4) The measure of precision being reported,

(5) The computed value of that measure, and

(6) The applicable range or concentration.

The importance of items (*3*) and (*4*) lies in the fact that with these a reader may calculate a confidence interval or relative standard deviation as desired.

(*b*) Precision is sometimes measured by repeatability and reproducibility (see Practice E177, and Mandel and Laskof (**1**)). The ANSI and ASTM documents differ slightly in their usages of these terms. The following is quoted from Kendall and Buckland (**2**):

"In some situations, especially interlaboratory comparisons, *precision* is defined by employing two additional concepts: *repeatability* and *reproducibility*. The general situation giving rise to these distinctions comes from the interest in assessing the variability *within* several groups of measurements and *between* those groups of measurements. *Repeatability*, then, refers to the within-group dispersion of the measurements, while *reproducibility* refers to the between-group dispersion. In interlaboratory comparison studies, for example, the investigation seeks to determine how well each laboratory can repeat its measurements (repeatability) and how well the laboratories agree with each other (reproducibility). Similar discussions can apply to the comparison of laboratory technicians' skills, the study of competing types of equipment, and the use of particular procedures within a laboratory. An essential feature usually required, however, is that repeatability and reproducibility be measured as variances (or standard deviations in certain instances), so that both within- and between-group dispersions are modeled as a random variable. The statistical tool useful for the analysis of such comparisons is the analysis of variance."

(*c*) In Practice E177 it is recommended that the term *repeatability* be reserved for the intrinsic variation due solely to the measurement procedure, excluding all variation from factors such as analyst, time and laboratory and reserving *reproducibility* for the variation due to all factors including laboratory. Repeatability can be measured by the standard deviation, σ_r, of *n* consecutive measurements by the same operator on the same instrument. Reproducibility can be measured by the standard deviation, σ_R, of *m* measurements, one obtained from each of *m* independent laboratories. When interlaboratory testing is not practical, the reproducibility conditions should be described.

(*d*) Two additional terms are recommended in Practice E177. These are *repeatability limit* and *reproducibility limit*. These are intended to give estimates of how different two measurements can be. The repeatability limit is defined as $1.96\sqrt{2}s_r$, and the reproducibility limit is defined as $1.96\sqrt{2}s_R$, where s_r is the estimated standard deviation associated with repeatability, and s_R is the estimated standard deviation associated with reproducibility. Thus, if normality can be assumed, these limits represent 95 % limits for the difference between two measurements taken under the respective conditions. In the reproducibility case, this means that "approximately 95 % of all pairs of test results from laboratories similar to those in the study can be expected to differ in absolute value by less than $1.96\sqrt{2}s_R$." It is important to realize that if a particular s_R is a poor estimate of σ_R, the 95 % figure may be substantially in error. For this reason, estimates should be based on adequate sample sizes.

3.1.9 *propagation of variance*—a procedure by which the mean and variance of a function of one or more random variables can be expressed in terms of the mean, variance, and covariances of the individual random variables themselves (Syn. *variance propagation, propagation of error*).

3.1.9.1 *Discussion*—There are a number of simple exact formulas and Taylor series approximations which are useful here (**3, 4**).

3.1.10 *random error*—(*1*) the chance variation encountered in all measurement work, characterized by the random occurrence of deviations from the mean value. (*2*) an error that affects each member of a set of data (measurements) in a different manner.

3.1.11 *random sample (measurements)*—a set of measurements taken on a single item or on similar items in such a way that the measurements are independent and have the same probability distribution.

3.1.11.1 *Discussion*—Some authors refer to this as a simple random sample. One must then be careful to distinguish between a simple random sample from a finite population of *N* items and a simple random sample from an infinite population. In the former case, a simple random sample is a sample chosen in such a way that all samples of the same size have the same chance of being selected. An example of the latter case occurs when taking measurements. Any value in an interval is considered possible and thus the population is conceptually infinite. The definition given in 3.1.11 is then the appropriate definition. (See *representative sample* and Appendix X5.)

3.1.12 *range*—the largest minus the smallest of a set of numbers.

3.1.13 *relative standard deviation (percent)*—the sample *standard deviation* expressed as a percent of the sample mean. The %RSD is calculated using the following equation:

$$\%RSD = 100\,\frac{s}{\left|\bar{x}\right|} \tag{4}$$

where:

s = sample standard deviation and
$\bar{x}$ = sample mean.

3.1.13.1 *Discussion*—The use of the %RSD (or RSD(%)) to describe precision implies that the uncertainty is a function of the measurement values. An appropriate error model might then be $X_i = \mu(1 + b + \varepsilon_i)$. (See Example, Appendix X2.)

Some authors use RSD for the ratio, $s/|x|$, while others call this the *coefficient of variation*. At times authors use RSD to mean %RSD. Thus, it is important to determine which meaning is intended when RSD without the percent sign is used. The recommended practice is $\%RSD = 100\ (s/|\bar{x}|)$ and $RSD = s/|\bar{x}|$.

3.1.14 *repeatability*—see Discussion in 3.1.8.

3.1.15 *representative sample*—a generic term indicating that the sample is typical of the population with respect to some specified characteristic(s).

3.1.15.1 *Discussion*—Taken literally, a representative sample is a sample that represents the population from which it is selected. Thus, "representative sample" has gained considerable colloquial acceptance in discussions involving the concepts of sampling. However, its use is avoided by most sampling methodologists because the concept of representative does not lend itself readily to definition or theoretical treatment. In particular, the concept is almost meaningless in describing a sample or its method of selection (see ANSI N15.5). Kendall and Buckland **(2)** suggest: "On the whole, it seems best to confine the word 'representative' to samples which turn out to be so, however chosen, rather than apply it to those chosen with the objective of being representative." "Representative sample" is not synonymous with "random sample." A random sample from a well-mixed material is probably representative; a random sample from an inhomogeneous material probably is not. It is likely many scientists mean random sample when using the term representative sample. If so, then the term *random sample* should be used to avoid possible confusion. In Appendix X5, an example relating to random and representative samples is given.

3.1.16 *reproducibility*—see Discussion in 3.1.8.

3.1.17 *standard deviation*—the positive square root of the *variance*.

3.1.17.1 *Discussion*—The use of the standard deviation to describe precision implies that the uncertainty is independent of the measurement value.

(a) An appropriate error model might be $X_i = \mu + b + \varepsilon_i$. (See Example, Appendix X2.)

(b) The practice of associating the ± symbol with standard deviation (or RSD) is not recommended. The ± symbol denotes an interval. The standard deviation is not an interval and it should not be treated as such. If the ± notation is used as in, "The fraction of uranium was estimated as 0.88 ± 0.01," a footnote should be added to clearly explain what is meant. Is 0.01 one standard deviation, two standard deviations, the standard deviation of the mean, or something else? Is the interval a confidence interval?

3.1.18 *standard deviation of the mean (sample)*— the sample *standard deviation* divided by the square root of the number of measurements used in the calculation of the mean (Syn. *standard error of the mean*).

3.1.18.1 *Discussion*—The equation for standard deviation of the mean is

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \tag{5}$$

where:

$s_{\bar{x}}$ = standard deviation of the mean of a set of measurements,
s = standard deviation of the set, and
n = number of measurements in the set.

3.1.19 *systematic error*—the term systematic error should not be used unless defined carefully.

3.1.19.1 *Discussion*—Some consider systematic error as a synonym for bias and treat it as a constant, whereas others make a distinction between the two terms. Some publications have used systematic error to refer to both a fixed and a random error. If the term is used, it should be clearly defined, preferably by specifying the error model. (See *bias* and Example, X2.1.1.)

3.1.20 *uncertainty*—a generic term indicating the inability of a measurement process to measure the correct value.

3.1.20.1 *Discussion*—Uncertainty is a concept which has been used to encompass both precision and bias. Thus, one measurement process (or a set of measurements based on the process) is sometimes referred to as "more uncertain" than another process. But, just as with precision, it is important that a quantitative *measure* be used to specify uncertainty. Thus, a phrase like, "The uncertainty is 5.2 units," should be avoided. Unfortunately, no single quantitative measure to specify uncertainty is universally accepted. Thus, "the quantification of uncertainty is itself an uncertain undertaking" (ANSI N15.5).

See *precision* and *bias* for preferred terms and Ku **(5)** for additional discussion.

3.1.21 *variance (sample)*—a measure of the dispersion of a set of results. Variance is the sum of the squares of the individual deviations from the sample mean divided by one less than the number of results involved.

3.1.21.1 *Discussion*—The equation that expresses this definition is as follows:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{6}$$

where:

s^2 = sample variance,
n = number of results obtained,
x_i = ith individual result, and
$\bar{x}$ = sample mean

$$\left(\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\right)$$

.

The following is an equation that is sometimes used to calculate sample variance:

$$s^2 = \frac{1}{n-1}\left[\sum x_i^2 - n\bar{x}^2\right] \tag{7}$$

Although this equation is mathematically exact, in practice it can lead to appreciable errors because of computer round-off problems. This can occur especially if the %RSD is small. The definition formula is, in general, to be preferred. To be useful, the variance must be based on results that are independent and identically distributed. (See Example, X2.1.1.)

4. Significance and Use

4.1 To describe the uncertainties of a standard test method, precision and bias statements are required.[4] The formulation of these statements has been addressed from time to time, and at least two standards practices (Practices E177 and E691) have been issued. The 1986 *Compilation of ASTM Standard Definitions*(**6**)[5] devotes several pages to these terms. This guide should not be used in cases where small numbers of test results do not support statistical normality.

4.2 ANSI N15.5 attempts to provide "a standard on statistical terminology and notation [that] can benefit communication" among nuclear materials managers. Precision, accuracy, and bias are all discussed. Although these various documents are quite valuable, a simpler document written for analysts appears needed. The intent of this guide is to help analysts prepare and interpret precision and bias statements. It is essential that, when the terms are used, their meaning should be clear and easily understood.

4.3 Appendix X1 provides the theoretical foundation for precision and bias concepts and Practice E691 addresses the problem of sources of variation. To illustrate the interplay between sources of variation and formulation of precision and bias statements, a hypothetical data set is analyzed in Appendix X2. This example shows that depending on how the data was collected, different precision and bias statements are possible. Reference to this example will be found throughout this guide.

4.4 There has been much debate inside and outside the statistical community on the exact meaning of some statistical terms. Thus, following a number of the terms in Section 3 is a list of several ways in which that term has been used. This listing is not meant to indicate that these meanings are equivalent or equally acceptable. The purpose here is more to encourage clear definition of terms used than to take sides. For example, use of the term *systematic error* is discouraged by some. If it is to be used, the reader should be told exactly what is meant in the particular circumstance.

4.5 This guide is intended as an aid to understanding the statistical concepts used in precision and bias statements. There is no intention that this be a self-contained introduction to statistics. Since many analysts have no formal statistical training, it is advised that a trained statistician be consulted for further clarification if necessary.

5. Precision and Bias Considerations

5.1 With regard to precision and accuracy, Kendall and Buckland (**2**) include this generic statement in their dictionary:

"In exact usage precision is distinguished from accuracy. The latter refers to closeness of an observation to the quantity intended to be observed. Precision is a quality associated with a class of measurements and refers to the way in which repeated observations conform to themselves; and in a somewhat narrower sense refers to the dispersion of the observations, or some measure of it, whether or not the mean value around which the dispersion is measured approximates to the 'true' value."

5.2 A fundamental question is, "What sources of measurement variation are being estimated?" The measurement should be taken in such a way as to include all the desired sources of variation. The results should be stated so that it is clear which sources of variation have been included and which measure of precision is used. It is best to report precision and bias in the most complete manner possible so that the reader can properly interpret the results. Statements such as "The precision is 1.54 g" are useless. (See 3.1.8, *precision*, for a discussion of what is desired.)

5.3 It is essential to realize that measurements are subject to error and that the ways in which the errors affect the measurements are important. This is discussed in the sections on error models (3.1.7 and Appendix X4). It is only in the presence of a specified error model that such concepts as precision, bias, random error, and systematic error become completely meaningful. The error model describes how the different sources of variation enter into the measurement process. Once the model is specified, these generic concepts should be defined relative to the model and their value estimated. Enough information should be given to allow proper statistical evaluation of the resultant estimates.

6. Keywords

6.1 bias; error models; precision; statistics

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

APPENDIXES

(Nonmandatory Information)

X1. CONCEPTS OF STATISTICS

X1.1 Parameters are constants used to index a family of distributions. The family of normal distributions, for example, is indexed by the mean, μ, and the standard deviation, σ. Specifying values for these two constants yields a particular member of the family. Of particular interest is the estimation of the parameters by means of a random sample, $X_1, \ldots, X_n$, of size n. We use capital letters to denote random variables and corresponding lower-case letters for their realizations, so that X_i is the symbol for the i^{th} sample value (before the sample is taken) and x_i is the actual observed value of X_i. A (simple) random sample means that the X_i are statistically independent and identically distributed.

X1.2 To estimate a parameter θ, a function $T = f(X_1, \ldots, X_n)$ of the sample values is used. T is said to be a statistic and is a random variable. More specifically, T is an estimator of θ. Use the observed values of the sample to get an estimate, $t = f(x_1, \ldots, x_n)$, of θ that is a number rather than a random variable. If $E(T)$ denotes the population average or expected value of T, $E(T) - \theta$ is the bias in T, and T is an unbiased estimator of θ only if $E(T) = \theta$. Accuracy is a general term referring to the closeness of a measured value to the "true" value. One measure of accuracy is bias. Another measure is the absolute value of the bias. In practice, one does not know the true value of θ, so the bias is estimated by using a reference value of θ or an accepted or standard or target value in place of θ. The bias is then described as relative to this reference value. Precision is a general term used to describe the dispersion (scatter, variability) in an estimator. There are many measures of precision of which the variance, $E\,(T - E\,(T))^2$, and its positive square root, the standard deviation, are just two. A measure that combines precision and bias is the mean square error, $E(T - \theta)^2$, which is equal to the variance plus the square of the bias.

NOTE X1.1—These and many other statistical concepts are more fully explained in Ref (7).

X2. EXAMPLE OF STATISTICAL CONCEPTS AND SOURCES OF VARIATION

X2.1 The following example illustrates that data from a measurement procedure should never be merely collected. Factors of interest—time, laboratory, analyst, instrument, calibration—that may affect the results should first be identified and an experiment designed to allow estimation of the effects of these factors over the appropriate range of values.

X2.1.1 *Example*—Write a precision and bias statement based on the following 24 hypothetical test measurements on a material whose reference value is μ = 64.23 g.

	Column			
Row	1	2	3	4
1	53	67	64	44
2	61	80	82	55
3	57	86	67	38
4	66	71	60	53
5	45	74	52	59
6	84	66	65	57

X2.1.2 How these data are analyzed and the nature of the precision and bias statement associated with the measurement procedure depend on how the data were collected and what assumptions on error models and probability distributions are made. For simplicity, all errors will be assumed to have a normal probability distribution. Of course, in practice this should be verified.

X2.1.3 Consider the following data collection possibilities:

X2.1.3.1 *Case 1*—All 24 measurements come from the same analyst using the same instrument on the same day. The results are assumed to be statistically independent. Thus, the 24 results represent a simple random sample (see discussion under *random sample* (measurements)) from a single population.

X2.1.3.2 *Case 2*—The measurements come from the same analyst using the same instrument on four successive Mondays, denoted by the four columns. The results within each column are assumed to be statistically independent. Thus, the measurements represent four simple random samples of size six from four populations. For later discussions, it is assumed that whatever effect is experienced on Mondays influences all measurements within the week. (The four columns could also represent four different laboratories.)

X2.1.3.3 *Case 3*—The measurements come from six different analysts (the six rows) each working on a different instrument and each making one run on each of four successive Mondays. Then the results might represent 24 random samples of size 1 from 24 populations.

X2.1.4 Clearly there are many other collection possibilities involving such considerations as calibration, time of day, season of year, different analysts on the same instrument, or the same analyst on different instruments. In each of these cases different sources of variation may affect the data. In Case 1, the only source of variation would appear to be measurement random error; in Case 2 there may be an additional source of variation because of a weekly effect. The possible sources of variation in Case 3 include time and analyst/instrument. The reader might refer to Practice E691 for a fuller discussion of this topic. (Of course, some of the above-mentioned sources of variation may contribute little or nothing to the total variation. One of the functions of a statistically designed experiment is to identify and quantify the major sources of variation.)

X2.1.5 Consider Case 1 in which only random error affects the results. The following statistics are easily calculated:

Sample size (n)	24
Sample mean ($\bar{x}$)	62.8 g
Bias estimate ($\bar{x} - \mu$)	−1.5 g (see Note)
Range (high – low)	48.0 g
Standard deviation (s)	12.6 g
%RSD	20 %
Standard deviation of the mean	2.6 g

NOTE X2.1—A simple statistical test shows that this value is not significantly different from zero at any reasonable significance level. Hence, the data do not support a hypothesis of nonzero bias.

X2.1.5.1 If the following additive error model is assumed,

$$X_i = \mu + b + \varepsilon_i \quad i = 1, 2, \ldots 24, \varepsilon_i \sim (0, \sigma^2), \qquad (X2.1)$$

$$= 64.23 + b + \varepsilon_i$$

the data support the hypothesis $b = 0$ with an estimated random error variance, s^2, of $(12.6\ g)^2$. (The symbol $\sim(\mu, \sigma^2)$, indicates that ε_i is a random variable with mean μ and variance σ^2.) Had a multiplicative error model been appropriate,

$$X_i = \mu(1 + b + \varepsilon_i) \quad i = 1, 2, \ldots 24, \varepsilon_i \sim (0, \sigma^2), \qquad (X2.2)$$

$$= 64.23\,(1 + b + \varepsilon_i)$$

then the random error standard deviation, σ, would be estimated by the RSD expressed as a fraction, 0.201. Again, the hypothesis $b = 0$ would be supported.

X2.1.6 A test method statement on precision and bias in the latter case might then be as follows:

X2.1.6.1 The test method was independently run 24 times in a row by the same analyst on the same instrument under virtually the same conditions on a material whose reference value was 64.23 g. The sample mean of the 24 measurements was 62.8, which is not indicative of bias. The precision (repeatability) of the test method, as measured by the %RSD, was estimated to be 20 %. (Had the data come from 24 independent laboratories, the 20 % would have been a measure of reproducibility.)

X2.1.6.2 The reader will probably feel more comfortable if several materials that covered a range of interest were measured and if some evidence of verification of assumptions (for example, normal errors, multiplicative error model) were presented in the write-up.

X2.1.6.3 In Case 2 an appropriate error model might be:

$$X_{ij} = \mu + W_i + \varepsilon_{ij} \quad i = 1, \ldots, 4, j = 1, \ldots, 6, \qquad (X2.3)$$

$$= 64.23 + (W_i + \varepsilon_{ij})$$

where:

X_{ij} = test result of the j^{th} run in the i^{th} week,
W_i = effect due to the i^{th} week (assume W_i is a normal random variable with mean zero and common variance σ^2_w), and
ε_{ij} = random error effects (assume the ε_{ij} are also normal random variables with mean zero and common variance σ^2_ε).

It is assumed that the W_i and the ε_{ij} are mutually independent.

X2.1.7 A precision and bias statement should include information on how the results were affected by the weekly effect. A one-way ANOVA yields the following estimates of σ^2_w and σ^2_ε, respectively:

$$s^2_w = 74.07 = (8.61)^2 \text{ and} \qquad (X2.4)$$

$$s^2_\varepsilon = 100.90 = (10.04)^2 \qquad (X2.5)$$

Thus, the variance of an individual result is:

$$Var(X) = \sigma^2_w + \sigma^2_\varepsilon \qquad (X2.6)$$

$$= 174.97$$

$$= (13.23)^2$$

X2.1.7.1 This result is greater than the $(12.6\ g)^2$ obtained in Case 1. The ANOVA shows that there is a statistically significant weekly effect, that is, not all weeks have the same mean. (In a real situation one might want to discover the cause of this effect and remove it.) This weekly effect represents a bias or systematic error that varies from week to week. It is being assumed that the effect remains constant within a week. This would need to be verified. Perhaps it could be due to a weekly calibration. (As mentioned earlier, the columns might represent data from different laboratories. Then σ_w measures interlaboratory variation.)

X2.1.7.2 A statement of precision and bias for this case might be the test method was run by the same analyst on the same instrument six times on each of four successive Mondays on a material whose reference value was 64.23 g. A statistically significant bias that varied from week to week was found. An ANOVA yielded the following estimates of variances of the weekly and random error effects, respectively:

$$s^2_w = 74.07 = (8.61)^2 \text{ and } s^2_\varepsilon = 100.90 = (10.04)^2 \qquad (X2.7)$$

X2.1.7.3 The analysis of Case 3 requires a two-way ANOVA and will not be discussed here. Suffice it to say that the data allow estimation of the effects from different analysts/instruments and time, as well as the random effects.

X2.1.8 *Additional Information:*

X2.1.8.1 If normality is assumed, a 95 % confidence interval (see Appendix X3) for the mean of the population in Case 1 is:

$$62.8 \pm 2.07\,(12.6/\sqrt{24}) \text{ or } (57.4, 68.1) \qquad (X2.8)$$

X2.1.8.2 This interval contains the reference value. However, if just the fourth week's data were available, a 95 % confidence interval for the mean of that population would be:

$$51.0 \pm 2.57\,(8.22/\sqrt{6}) \text{ or } (42.4, 59.6) \qquad (X2.9)$$

This interval does not contain the reference value, thus supporting the conclusion that there is a weekly effect.

X3. CONFIDENCE INTERVAL

X3.1 Construct a 100(1 – α)% symmetric confidence interval for a population mean, μ.

X3.1.1 *Assumption*—The population of values under consideration has a normal (Gaussian) distribution with mean μ and standard deviation σ.

X3.2 Consider a random sample of *n*measurements. Let$\bar{X}$and *S* be the sample mean and standard deviation, respectively. These are random variables; they are estimators of μ and σ, respectively. Let $t_{k,\alpha/2}$ be the upper 100(1 – α/2)th percentile of the Student's *t*-distribution for $k = n-1$ degrees of freedom. Then,

$$\bar{X} \pm t_{k,\alpha/2} S/\sqrt{n} \tag{X3.1}$$

is a 100(1 – α)% confidence interval estimator for the population mean, μ. Of all possible such intervals (based on random samples of size *n*) that could be obtained, 100(1 – α)% of them will indeed contain μ; 100α % will not.

X3.2.1 Now suppose that the *n*measurements have been obtained. Let$\bar{x}$and*s*be the observed sample mean and standard deviation. These are estimates. Then,

$$\left(\bar{x} - t_{k,\alpha/2} s/\sqrt{n},\ \bar{x} + t_{k,\alpha/2} s/\sqrt{n}\right) \tag{X3.2}$$

is a 100(1 – α)% confidence interval estimate of μ. This interval is fixed. It either contains μ or it does not.

X3.2.2 If $n = 1$, this procedure does not work because *s* is not defined. In this case an independent estimate of the population standard deviation, σ, must be obtained. Call this estimate $\hat{\sigma}$. Let *k* be the degrees of freedom on which this estimate is based. Then if $t_{k,\alpha/2}$ is the appropriate *t*-value for α and *k* degrees of freedom,

$$\left(x - t_{k,\alpha/2}\hat{\sigma},\ x + t_{k,\alpha/2}\hat{\sigma}\right) \tag{X3.3}$$

is the desired confidence interval.

X3.2.3 If σ is known, the normal probability values may be used in place of the *t*-distribution values in X3.2.1 and X3.2.2.Then, for example, a 100(1 – α)% confidence interval for μ based on a single determination is $x \pm z_{\alpha/2}\sigma$, where $z_{\alpha/2}$ comes from the normal probability table.

X4. ERROR MODELS

X4.1 The importance of the model can be demonstrated by calculating the expected value and variance of the measured value for four different error models.

Suppose $X =$

$\mu + b + \varepsilon$	additive
$\mu b\varepsilon$	multiplicative (type I), $E(\varepsilon) = 1$
$\mu(1 + b + \varepsilon)$	multiplicative (type II)
$\mu(1+\varepsilon)+b+\varepsilon'/\sqrt{\mu}$	mixed

Then, it can be shown that

(X4.1)

$E(X) =$	$\mu + b$	additive
	μb	multiplicative (I)
	$\mu + \mu b$	multiplicative (II)
	$\mu + b$	mixed

and

(X4.2)

$\mathrm{Var}(X) =$	$\mathrm{Var}(\varepsilon)$	additive
	$\mu^2 b^2 \mathrm{Var}(\varepsilon)$	multiplicative (I)
	$\mu^2 \mathrm{Var}(\varepsilon)$	multiplicative (II)
	$\mu^2 \mathrm{Var}(\varepsilon) + \mathrm{Var}(\varepsilon')/\mu$	mixed

X4.1.1 For the mixed model it is assumed that both ε and ε' have a mean of zero and are independent. In the other cases, except as noted, ε has a mean of zero.

X4.1.2 It is also assumed that *b*is a bias and, hence, is a constant. Now suppose that the source of the bias is from calibration and that the calibration produces different biases at different times, as in Case 2 of Appendix X2. Then the *b* term might be considered as a random variable (assumed independent of ε and ε' and with mean zero, except as noted) so the above expressions become

(X4.3)

$E(X) =$	μ	additive
	μ	multiplicative (I), $E(b) = 1$
	μ	multiplicative (II)
	μ	mixed

and

(X4.4)

$\mathrm{Var}(X) =$	$\mathrm{Var}(b) + \mathrm{Var}(\varepsilon)$	additive
	$\mu^2[\mathrm{Var}(b) + \mathrm{Var}(\varepsilon) + \mathrm{Var}(b)\ \mathrm{Var}(\varepsilon)]$	multiplicative (I), $E(b) = 1$ $E(\varepsilon) = 1$
	$\mu^2[\mathrm{Var}(b) + \mathrm{Var}(\varepsilon)]$	multiplicative (II)
	$\mu^2\ \mathrm{Var}(\varepsilon) + \mathrm{Var}(b) + \mathrm{Var}(\varepsilon')/\mu$	mixed

X4.1.3 Note that the process now is, "Calibrate the instrument and make a measurement." Once the instrument is calibrated, *b*is fixed and the previously given expressions for$E(X)$ and $\mathrm{Var}(X)$ are appropriate. One might write $E(X \mid b)$ and $\mathrm{Var}(X \mid b)$ for these to emphasize that the value of *b* is fixed for a particular calibration. It should be clear now that knowledge of the bias and of the variance of ε alone does not suffice to determine the mean and variance of *X*; the error model must be known.

X4.1.4 As an example of the usage of models, suppose an electronic balance is calibrated and then used to determine the mass of *n* items individually. Suppose also that the measured weight of item *i*, X_i, can be written as:

$$X_i = \mu_i + b + \varepsilon_i \tag{X4.5}$$

where:

b = a constant and

ε_i = independent normally distributed random variables.

Then b is the bias (b might be due, for example, to imperfect calibration). However, if the n items were weighed on different days and if the balance was calibrated daily, the above model might become:

$$X_i = \mu_i + b_i + \varepsilon_i \quad \text{(X4.6)}$$

X4.1.5 In this case there would be no specific error term for calibration in the model. Note that in the first case $\mathrm{Var}(X_i) = \sigma^2_\varepsilon$ and in the second case $\mathrm{Var}(X_i) = \sigma^2_\varepsilon + \sigma^2_\varepsilon$, a larger quantity.

X5. AN EXAMPLE OF REPRESENTATIVE VERSUS RANDOM SAMPLING

X5.1 Suppose 100 g of PuO_2 and 100 g of UO_2 are mixed together in a container. A sample of 5 g is to be drawn and analyzed for Pu content.

X5.1.1 To draw a 5-g sample at random requires that all possible 5 g subsamples have the same chance of selection. If the material is first well-blended (homogeneous), it is likely that a 5-g random sample will be a representative sample. That is, the Pu content (%) of the sample will be approximately the same as the % Pu in the entire container. If the material is not well-blended (heteregeneous), it is likely that the sample will not be representative.

X5.1.2 Now consider the 5-g sample. Let this be well-blended and divided into five 1-g subsamples. If each subsample is analyzed for Pu content (%) by a specific technique, five assays will be observed. These five values will then be a simple random sample of measurements which are surely representative of the sample. They will be representative of the container contents if the 5-g sample is representative.

REFERENCES

(1) Mandel, J., and Laskof, T., "The Nature of Repeatability and Reproducibility," *Journal of Quality Technology*, Vol 19, Jan. 1987, pp. 29–36.

(2) Kendall, M. G., and Buckland, W. R., *A Dictionary of Statistical Terms*, 3rd Ed., Hafner Publishing Co., Inc., New York, NY, 1971 .

(3) Mood, A. M., Graybill, F. A., and Boes, D. C., *Introduction to the Theory of Statistics*, 3rd Ed., 1974, McGraw Hill, New York, NY, pp. 176–182.

(4) NRC, *Statistical Methods for Nuclear Material Management*, NUREG/CR-4604, Nuclear Regulatory Commission, Washington, DC, 1988, pp. 88–93.

(5) Ku, H. H., "Statistical Concepts in Metrology," *Precision Measurements and Calibration*, Special Publication 300, Vol 1, National Bureau of Standards, Washington, DC, 1969, pp. 296–330.

(6) ASTM, *Compilation of ASTM Standard Definitions*, 6th Ed., ASTM, Philadelphia, 1986.

(7) Tietjen, G. L., *A Topical Dictionary of Statistics*, Chapman and Hall, New York, NY, 1986.