INTERNATIONAL

Designation: ~~C1215 – 92 (Reapproved 2012)$^{\varepsilon 1}$~~ C1215 – 18

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.

1.7 This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.

2. Referenced Documents

2.1 *ASTM Standards:*[2]

C859 Terminology Relating to Nuclear Materials

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]~~

[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~~July 1, 2018. Published ~~June 2012~~July 2018. Originally approved in 1992. Last previous edition approved in ~~2006~~2012 as ~~C1215–92(2006).~~C1215 – 92 (2012)$^{\varepsilon 1}$. DOI: ~~10.1520/C1215-92R12E01.~~10.1520/C1215-18.

[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. Terminology ~~for Precision and Bias Statements~~

3.1 For definitions of terms used in this guide but not defined herein, see Terminology C859.

3.2 ~~*Definitions:*~~*Terminology for Precision and Bias Statements*

3.2.1 *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.

3.2.1.1 *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.[3]

3.2.2 *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. 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)

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

3.2.3.1 *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.

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

3.2.5 *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).

3.2.5.1 *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~~ 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

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

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.

3.2.6 *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.)

3.2.7 *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.

3.2.7.1 *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:

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

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

(*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.

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

3.2.8.1 *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:

~~(*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.~~

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.

3.2.9 *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*).

3.2.9.1 *Discussion—*

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

3.2.10 *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.

3.2.11 *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.

3.2.11.1 *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~~3.2.11 is then the appropriate definition. (See *representative sample* and Appendix X5.)

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

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

$$\%RSD = 100 \frac{s}{\left|\bar{x}\right|} \quad (4)$$

where:

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

3.2.13.1 *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 = $s/|\bar{x}|$.

3.2.14 *repeatability*—see Discussion in ~~3.1.8~~3.2.8.

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

3.2.15.1 *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).~~ selection. 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.

3.2.16 *reproducibility*—see Discussion in ~~3.1.8~~3.2.8.

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

3.2.17.1 *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?

3.2.18 *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*).

3.2.18.1 *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.

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

3.2.19.1 *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.)

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

3.2.20.1 *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”~~undertaking.”

See ~~(*ANSI*~~*precision* ~~N15.5).~~and *bias* for preferred terms and Ku (**5**) for additional discussion.

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

3.2.21 *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.

3.2.21.1 *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.)