LA-11190-MS

Received by OSTI

JAN 12 1988

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36.

DO NOT MICROFILM
COVER

***On Precision and Accuracy (Bias)
Statements for Measurement Procedures***

Los Alamos Los Alamos National Laboratory
Los Alamos, New Mexico 87545

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

DO NOT MICROFILM
COVER

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

LA-11190-MS

UC-32

Issued: January 1988

LA--11190-MS

DE88 004434

On Precision and Accuracy (Bias) Statements for Measurement Procedures

Lawrence A. Bruckner

Merril W. Hume*

Wayne L. Delvin**

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

*Rocky Flats Plant, P.O. Box 464, Golden, CO 80401.

**Hanford Engineering Development Laboratory, P.O. Box 1970, Richland, WA 99353.

MASTER

Los Alamos

Los Alamos National Laboratory
Los Alamos, New Mexico 87545

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

jsu

ON PRECISION AND ACCURACY (BIAS) STATEMENTS FOR
MEASUREMENT PROCEDURES

by

Lawrence A. Bruckner, Merrill W. Hume, and Wayne L. Delvin

ABSTRACT

Measurement procedures are often required to contain precision and accuracy or precision and bias statements. This paper 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 paper 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.

To characterize the uncertainties of a measurement procedure, precision and accuracy statements are often included in the description of the procedure. The formulation of these statements has been addressed from time to time, and at least two standards practices [1,2] have been issued by the American Society for

Testing and Materials (ASTM). The 1982 Compilation of ASTM Standard Definitions [3] devotes several pages to usage of these terms as well as the term bias, which is now recommended by ASTM in lieu of accuracy. (See Glossary entry ACCURACY for the rationale behind this change. This recommendation will be followed in this paper.) With regard to precision and accuracy, Kendall and Buckland [4] 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. In general, the precision of an estimator varies with the square root of the number of observations upon which it is based.

(Note that this statement does not provide a definition of the terms.) A forthcoming document from the American National Standards Institute (ANSI) [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 paper 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. Thus, the following statement is useless.

The precision is 1.54 g.

Even a precision statement such as the following leaves much to be desired.

The standard deviation is 1.54 g.

For example, was the standard deviation based on 5 or 50 measurements? What sources of variation (analyst, instrument, calibration, environment) are included? The following precision and bias statement is much better.

The standard deviation based on 12 measurements by the same analyst using the same instrument was 1.54 g. The measurements were made on the same day one after the other on a standard whose accepted value was 83.721 g. The sample mean was 87.747 g, which does not indicate a statistically significant bias.

Now, compare the above statement with the following statement.

The standard deviation based on three measurements at each of four different laboratories was 1.54 g. The measurements were made on a standard whose accepted value was 83.721 g. The sample mean was 87.747 g, which does not indicate a statistically significant bias.

These statements say something different. The second statement addresses variation from laboratory, analyst, and instrument effects, as well as analytical error. The first statement would appear to be an indicator of analytical error only. Both statements, however, do describe how the measurements were obtained.

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

It is essential to realize that measurements are subject to error and that the ways in which the errors affect the measurements are important. Thus a key section in what follows is the section on ERROR MODELS. 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.

The following glossary is intended as an aid to understanding the statistical concepts found in typical precision and bias statements; there is no

intention that this be a self-contained introduction to statistics. Appendix I provides additional insight into the statistical foundations, and Reference 1 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 II. Reference to this example will be found throughout the glossary.

One final comment is necessary. 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 the glossary 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. Thus, although 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.

INDEX

ACCURACY.....	6
ANALYSIS OF VARIANCE (ANOVA).....	6
BIAS.....	6
COEFFICIENT OF VARIATION.....	7
CONFIDENCE INTERVAL.....	7
CONFIDENCE LEVEL.....	8
ERROR MODEL.....	8
PRECISION.....	9
RANDOM ERROR.....	11
RANDOM SAMPLE (SIMPLE).....	11
RANGE.....	11
RELATIVE STANDARD DEVIATION (RSD).....	11
REPEATABILITY.....	12
REPRESENTATIVE SAMPLE.....	12
REPRODUCIBILITY.....	13
STANDARD DEVIATION (SAMPLE).....	13
STANDARD DEVIATION OF THE MEAN.....	13
SYSTEMATIC ERROR.....	13
UNCERTAINTY.....	14
VARIANCE (SAMPLE).....	14
REFERENCES.....	16
APPENDICES	
I CONCEPTS OF STATISTICS.....	17
II EXAMPLE.....	18
III CONFIDENCE INTERVAL.....	23
IV ERROR MODELS.....	25

GLOSSARY

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.

NOTE: 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 now requires a statement on bias instead of accuracy.

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 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, Appendix II.)

BIAS (See ACCURACY)

1. The difference between the expected value of an estimator (See Appendix I) and the true value being estimated (theory).
2. A persistent or systematic error that remains constant over a series of replicated measurements.

NOTE: Bias represents a constant error as opposed to a RANDOM ERROR. An estimator is said to be unbiased if its expected value is equal to the true value of the parameter being estimated. (See Appendix I.)

A bias can be estimated by the difference between a measured average and an accepted standard or reference value. Such data 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 that cover the range of application. Bias often varies with the range of application and should be reported accordingly.

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 II.)

Use of a single test run to estimate bias is strongly discouraged because, even if there were no bias, random error would most likely produce a nonzero bias estimate.

COEFFICIENT OF VARIATION

See RELATIVE STANDARD DEVIATION.

CONFIDENCE INTERVAL

An interval used to bound the value of a population parameter with a specified degree of confidence. (This is an interval that assumes different values for different random samples.)

NOTE: 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 III.

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 $n = 5$ observations, a second set of 5 might produce a very different interval. This would not be the case if $n = 50$.

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 III.)

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.

NOTE: 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 process by which the error enters.

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

$$X_i = \mu + b + \epsilon_i ,$$

where X_i is the result of the i^{th} measurement,

μ is the true value of the item,

b is a bias, and

ϵ_i is a random error usually assumed to have a normal distribution with mean zero and variance σ_ϵ^2 .

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

$$X_i = \mu b \epsilon_i ,$$

and a multiplicative model for a neutron counter measurement might be

$$X_i = \mu + \mu b + \epsilon_i / \sqrt{\mu}$$

$$= \mu (1 + b) + \epsilon_i / \sqrt{\mu} .$$

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 II and in Appendix IV.

PRECISION

A generic concept employed to describe the dispersion of repeated measurements with respect to a measure of location of central tendency, usually the mean.

NOTE: 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 II.)

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

- (a) a description of the procedure used to obtain the data;
- (b) the number of repetitions, n , of the measurement procedure;
- (c) the sample mean and standard deviation of the measurements;
- (d) the measure of precision being reported;
- (e) the computed value of that measure; and
- (f) the applicable range or concentration.

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

Precision is sometimes measured by repeatability and reproducibility. The ANSI and ASTM documents differ slightly in their usages of these terms. The following is found in Ref. 5.

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.

ASTM [2] recommends reserving repeatability for the intrinsic variation due solely to the measurement procedure, excluding all variation from factors such as analyst, time and laboratory and 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.

Two additional terms are recommended by ASTM. 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 $1.96\sqrt{2} s_R$." [2] 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.

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.

RANDOM SAMPLE (SIMPLE)

A sample of specified size (e.g., n elements, x grams, y liters) selected from a population in such a way that every possible sample of the same size has the same probability of being selected.

RANGE

The largest minus the smallest of a set of numbers.

RELATIVE STANDARD DEVIATION (RSD)

The sample STANDARD DEVIATION expressed as a percentage of the sample mean. The RSD is calculated using the following equation:

$$RSD = 100 \frac{s}{\bar{x}}$$

where s = sample standard deviation and

$\bar{x}$ = sample mean.

NOTE: The use of the RSD to describe precision implies that the uncertainty is a function of the measurement values. An appropriate error model might then be $x_i = \mu (1 + b + \epsilon_i)$. (See EXAMPLE, Appendix II.)

The RSD is also known as the coefficient of variation, although sometimes the coefficient of variation is expressed as a ratio ($s/|\bar{x}|$) instead of as a percentage [$100(s/|\bar{x}|)$]. The following practices are recommended:

1. Do not use the $\pm$ symbol with the RSD.
2. If a standard with true value μ is measured n times and results in sample mean $\bar{x}$ and sample standard deviation s , then
$$\text{RSD} = 100(s/|\bar{x}|), \text{ not } 100(s/|\mu|)$$
3. Because RSD has been used for $s/|\bar{x}|$, some writers used %RSD or RSD(%) when expressing the relative standard deviation as a percentage.

REPEATABILITY

See discussion under PRECISION.

REPRESENTATIVE SAMPLE

A generic term indicating that the sample is typical of the population with respect to some specified characteristic(s).

NOTE: 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. Consider, for instance, the problem of selecting a sample of 5 playing cards from an ordinary 52-card poker deck in such a manner that the sample represents the deck. Kendall and Buckland...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." [5]

The random sample from a well-mixed material is probably representative, a random sample from a solid material (metal) may or may not be. Probably many scientists really mean random sample when using the term representative sample. If so, then the term RANDOM SAMPLE should be used to avoid possible confusion.

REPRODUCIBILITY

See discussion under PRECISION.

STANDARD DEVIATION (SAMPLE)

The positive square root of the VARIANCE.

NOTE: The use of the standard deviation to describe precision implies that the uncertainty is independent of the measurement value. An appropriate error model might be $x_i = \mu + b + \epsilon_i$. (See EXAMPLE, Appendix II.)

The practice of associating the $\pm$ symbol with standard deviation (or RSD) is not recommended. The $\pm$ symbol denotes an interval. The standard deviation is not an interval and it should not be treated as such. If the $\pm$ 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?

STANDARD DEVIATION OF THE MEAN

The STANDARD DEVIATION divided by the square root of the number of measurements.

The equation is $s_{\bar{x}} = \frac{s}{\sqrt{n}}$,

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.

NOTE: Synonym: Standard Error of the Mean. (See EXAMPLE, Appendix II.)

SYSTEMATIC ERROR

The term systematic error should not be used unless defined carefully.

NOTE: 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, Appendix II.)

UNCERTAINTY

A generic term indicating the inability of a measurement process to measure the correct value.

NOTE: Uncertainty is a concept which has been used to encompass both precision and accuracy. 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." [5]

See PRECISION and BIAS for preferred terms and Ku [6] for additional discussion.

VARIANCE (SAMPLE)

A measure of the dispersion of a set of measurements. It is the sum of the squares of the individual deviations from the sample mean divided by one less than the number of measurements involved.

NOTE: The equation that expresses this definition is as follows:

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

where s^2 = sample variance of the measurements,

n = number of measurements obtained,

x_i = i^{th} individual measurement, and

$\bar{x}$ = sample mean of the measurements $\left(\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \right)$

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

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

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 measurements that are independent and identically distributed. (See EXAMPLE, Appendix II.)

ACKNOWLEDGEMENTS

The authors would like to thank the many people who by their comments and criticisms, both solicited and unsolicited, have refined this document. In particular we thank Gary L. Tietjen for the material in Appendix I.

REFERENCES

- [1] ASTM, "Standard Practice for Conducting an Interlaboratory Test Program to Determine the Precision of Test Methods," E691-79 (ASTM, Philadelphia, PA, 1979).
- [2] ASTM, "Standard Practice for Use of the Terms Precision and Bias in ASTM Test Methods," E177-86 (ASTM, Philadelphia, PA, 1986).
- [3] ASTM, Compilation of ASTM Standard Definitions, 5th Ed. (ASTM, Philadelphia, 1982).
- [4] Kendall, M. G., and Buckland, W. R., A Dictionary of Statistical Terms, 3rd Ed. (Hafner Publishing Co., Inc., New York, 1971).
- [5] ANSI (DRAFT), "Statistical Terminology and Notation for Nuclear Materials Management," ANSI N15.5-1982 (New York, 1982).
- [6] Ku, H. H., "Statistical Concepts in Metrology," in Precision Measurements and Calibration, Special Publication 300, Vol. 1 (National Bureau of Standards, Washington, DC, 1969), pp. 296-330.

ADDITIONAL BIBLIOGRAPHY

Bainbridge, T. R., "The Committee on Standards: Precision and Bias," ASTM Standardization News, January 1985, pp. 44-46.

Enright, C. F., "E691 is Precisely the Method," ASTM Standardization News, January 1985, p. 60.

ASTM, "ASTM Standards on Precision and Bias for Various applications," PCN:03-511085-34, (ASTM, Philadelphia, PA, 1985).

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

APPENDIX I

CONCEPTS OF STATISTICS

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, $X_1, \dots, X_n$, 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 sample $x_1, \dots, x_n$ is chosen in such a way that every possible sample of the same size has the same probability of being chosen. This is equivalent to saying that the X_i are statistically independent and identically distributed.

To estimate a parameter θ , a function $T = f(X_1, \dots, X_n)$ 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, $t = f(x_1, \dots, x_n)$, of θ that is a number rather than a random variable. If $E(T)$ denotes the 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 accuracy 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.

APPENDIX II

EXAMPLE

The following example illustrates that data from a measurement procedure should never be just 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.

EXAMPLE: Write a precision and bias statement based on the following 24 hypothetical test measurements on a material whose reference value is $\mu = 64.23$ g.

Row	Column			
	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

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.

Consider the following data collection possibilities:

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 from a single population.

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.)

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.

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 ASTM Practice E691-79 [1] 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.)

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*
Range (high - low)	48.0 g
Standard deviation (s)	12.6 g
%RSD	20%
Standard deviation of the mean	2.6 g

$$(s_{\bar{x}} = s/\sqrt{n})$$

* A simple statistical test shows that this value is not statistically significantly different from zero at any reasonably significance level. Hence, the data do not support a hypothesis of nonzero bias.

If the following additive error model is assumed,

$$\begin{aligned} X_i &= \mu + b + \epsilon_i & i = 1, 2, \dots, 24, \quad \epsilon_i \sim (0, \sigma_\epsilon^2), \\ &= 64.23 + b + \epsilon_i \end{aligned}$$

the data support $b = 0$ and an estimated random error variance, $\hat{\sigma}_\epsilon^2$, of $(12.6 \text{ g})^2$. (The symbol $\sim (a, b)$, indicates that ϵ_i is a random variable with mean a and variance b .) Had a multiplicative error model been appropriate,

$$\begin{aligned} X_i &= \mu (1 + b + \epsilon_i) & i = 1, 2, \dots, 24, \quad \epsilon_i \sim (0, \sigma_\epsilon^2), \\ &= 64.23 (1 + b + \epsilon_i) \end{aligned}$$

then the random error standard deviation, σ_ϵ , would be estimated by the RSD expressed as a fraction, 0.201. Again b would be zero.

The following might then be a statement on precision and bias in the latter case.

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. Note that 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 (e.g., normal errors, multiplicative error model) were presented in the write-up.

In Case 2 an appropriate error model might be

$$\begin{aligned} X_{ij} &= \mu + w_i + \epsilon_{ij} & i = 1, \dots, 4, \quad j = 1, \dots, 6, \\ &= 64.23 + w_i + \epsilon_{ij} \end{aligned}$$

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_w^2 .

ϵ_{ij} = random error effects. Assume the ϵ_{ij} are also normal random variables with mean zero and common variance σ_ϵ^2 .

It is assumed that the w_i and the ϵ_{ij} are independent among themselves and from each other.

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_w^2 and σ_ϵ^2 , respectively:

$$s_w^2 = 74.07 - (8.61)^2 \text{ and}$$

$$s_\epsilon^2 = 100.90 - (10.04)^2.$$

Thus, the variance of an individual result is

$$\text{Var}(X) = \sigma_w^2 + \sigma_\epsilon^2$$

$$= 174.97$$

$$= (13.23)^2.$$

Note that this result is greater than the $(12.6 \text{ g})^2$ 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.)

The following might be a statement of precision and bias for this case.

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.

$$s_w^2 = 77.07 = (8.61)^2 \text{ and}$$

$$s_e^2 = 100.90 = (10.04)^2$$

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.

Further Comment

If normality is assumed, a 95% confidence interval (see Appendix III) for the mean of the population in Case 1 is

$$62.75 \pm 2.07 (12.6/\sqrt{24}) \text{ or } (57.43, 68.07).$$

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

$$51.00 \pm 2.57 (8.22/\sqrt{6}) \text{ or } (42.40, 59.60).$$

This interval does not contain the reference value and supports the analysis presented.

APPENDIX III

CONFIDENCE INTERVAL

Construct a $100(1 - \alpha)\%$ symmetric confidence interval for a population mean, μ .

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

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 $100(1 - \alpha/2)\%$ th percentile of the Student's t -distribution for $k = n - 1$ degrees of freedom. Then,

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

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

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,

$$(\bar{x} - t_{k, \alpha/2} s/\sqrt{n}, \bar{x} + t_{k, \alpha/2} s/\sqrt{n})$$

is a $100(1 - \alpha)\%$ confidence interval estimate of μ . This interval is fixed. It either contains μ or it does not.

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,

$$(x - t_{k, \alpha/2} \hat{\sigma}, x + t_{k, \alpha/2} \hat{\sigma})$$

is the desired confidence interval.

If k is large, say $k \geq 30$, or if σ is known, the normal probability tables may be used in place of the t -distribution tables. Then a $100(1 - \alpha)\%$ confidence interval for μ based on a single determination is $x \pm z_{\alpha/2} \hat{\sigma}$, where $z_{\alpha/2}$ comes from the normal probability table.

APPENDIX IV

ERROR MODELS

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

$$\text{Suppose } X = \begin{cases} \mu + b + \epsilon & \text{additive} \\ \mu b \epsilon & \text{multiplicative (type I), } E(\epsilon) = 1 \\ \mu(1 + b + \epsilon) & \text{multiplicative (type II)} \\ \mu(1 + \epsilon) + b + \epsilon' / \sqrt{\mu} & \text{mixed.} \end{cases}$$

Then, it can be shown that

$$E(X) = \begin{cases} \mu + b & \text{additive} \\ \mu b & \text{multiplicative (I)} \\ \mu + \mu b & \text{multiplicative (II)} \\ \mu + b & \text{mixed} \end{cases}$$

and

$$\text{Var } (X) = \begin{cases} \text{Var } (\epsilon) & \text{additive} \\ \mu^2 b^2 \text{Var } (\epsilon) & \text{multiplicative (I)} \\ \mu^2 \text{Var } (\epsilon) & \text{multiplicative (II)} \\ \mu^2 \text{Var } (\epsilon) + \text{Var } (\epsilon') / \mu & \text{mixed.} \end{cases}$$

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

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 II. 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

$$E(X) = \begin{cases} \mu & \text{additive} \\ \mu & \text{multiplicative (I), } E(b) = 1 \\ \mu & \text{multiplicative (II)} \\ \mu & \text{mixed} \end{cases}$$

and

$$\text{Var}(X) = \begin{cases} \text{Var}(b) + \text{Var}(\epsilon) & \text{additive} \\ \mu^2 [\text{Var}(b) + \text{Var}(\epsilon) + \text{Var}(b) \text{Var}(\epsilon)] & \text{multiplicative (I), } E(b) = 1, E(\epsilon) = 1 \\ \mu^2 [\text{Var}(b) + \text{Var}(\epsilon)] & \text{multiplicative (II)} \\ \mu^2 \text{Var}(\epsilon) + \text{Var}(b) & \text{mixed.} \\ + \text{Var}(\epsilon')/\mu & \end{cases}$$

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 $\text{Var}(X)$ are appropriate. One might write $E(X|b)$ and $\text{Var}(X|b)$ 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.

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

$$X_i = \mu_i + b + \epsilon_i,$$

where b is a constant and ϵ_i are 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

$$X_i = \mu_i + b_i + \epsilon_i.$$

In this case there would be no specific error term for calibration in the model. Note that in the first case $\text{Var}(X_i) = \sigma_\epsilon^2$ and in the second case $\text{Var}(X_i) = \sigma_b^2 + \sigma_\epsilon^2$, a larger quantity.

Printed in the United States of America
 Available from
 National Technical Information Service
 US Department of Commerce
 5285 Port Royal Road
 Springfield, VA 22161

Microfiche (A01)

Page Range	NTIS Price Code	Page Range	NTIS Price Code	Page Range	NTIS Price Code	Page Range	NTIS Price Code
001-025	A02	151-175	A08	301-325	A14	451-475	A20
026-050	A03	176-200	A09	326-350	A15	476-500	A21
051-075	A04	201-225	A10	351-375	A16	501-525	A22
076-100	A05	226-250	A11	376-400	A17	526-550	A23
101-125	A06	251-275	A12	401-425	A18	551-575	A24
126-150	A07	276-300	A13	426-450	A19	576-600	A25
						601-up*	A99

*Contact NTIS for a price quote.

Los Alamos

DO NOT MICROFILM
COVER