Designation: D 2906 – 97 (Reapproved 2002)

Standard Practice for Statements on Precision and Bias for Textiles[1]

This standard is issued under the fixed designation D 2906; 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.

INTRODUCTION

Work was begun in August 1966 on recommendations for statements on precision and accuracy. The first recommendations were issued as ASTMD-13 White Paper, Statements on Precision and Accuracy, MARK I, December 1968, prepared by Subcommittee C-6 on Editorial Policy and Review. After a decision that the recommendations should be a recommended practice under the responsibility of Subcommittee D13.93, Sampling, Presentation and Interpretation of Data, the recommendations were revised and published as Practice D 2906 – 70 T.

Information was added in Practice D 2906 – 73 on methods (*1*) for which precision has not been established, (*2*) for which test results are not variables, and (*3*) for which statements are based on another method. Practice D 2906 – 74 was expanded to include test methods in which test results are based on the number of successes or failures in a specified number of observations or on the number of defects or instances counted in a specified interval of time or in a specified amount of material. The present text provides for a nontechnical summary at the beginning of recommended texts based on normal distributions or on transformed data and for a more positive statement on accuracy when the true value of a property can be defined only in terms of a test method.

In 1984, changes were introduced to replace the term "accuracy" with "bias" as directed in the May 1983 edition of *Form and Style for ASTM Standards.*

1. Scope

1.1 This practice serves as a guide for using the information obtained as directed in Practice D 2904 or obtained by other statistical techniques from other distributions, to prepare statements on precision and bias in ASTM methods prepared by Committee D-13. The manual on form and style for standards specifies that statements on precision and bias be included in test methods.[2] Committee D-13 recommends at least a statement about single-operator precision in any new test method, or any test method not containing a precision statement that is put forward for 5-year approval, in both instances with a complete statement at the next reapproval. If a provisional test method is proposed, at least a statement on single-operator precision is expected.

1.2 The preparation of statements on precision and bias requires a general knowledge of statistical principles including the use of components of variance estimated from an analysis of variance. Instructions covering such calculations are available in Practice D 2904 or in any standard text **(1,2,3,4,** and **5).**[3]

1.3 The instructions in this practice are specifically applicable to test methods in which test results are based (*1*) on the measurement of variables, (*2*) on the number of successes or failures in the specified number of observations, (*3*) on the number of defects or incidents counted in a specified interval or in a specified amount of material, and (*4*) on the presence or absence of an attribute in a test result (a go, no-go test). Instructions are also included for methods of test for which precision has not yet been estimated or for which precision and accuracy have been reported in another method of test. For observations based on the measurement of variables, the instructions of this practice specifically apply to test results that are the arithmetic average of individual observations. With qualified assistance, the same general principles can be applied to test results that are based on other functions of the data such as standard deviations.

[1] This practice is under the jurisdiction of ASTM Committee D13 on Textiles and is the direct responsibility of Subcommittee D13.93 on Statistics.

Current edition approved Nov. 10, 1997. Published August 1998 . Originally published as D 2906 – 70 T. Last previous edition D 2906 – 91.

[2] *Form and Style for ASTM Standards*, May 1983, available from American Society for Testing and Materials, 100 Barr Harbor Drive, West Conshohocken, PA 19428.

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

1.4 This standard includes the following sections:

2. Referenced Documents

2.1 *ASTM Standards:*

D 123 Terminology Relating to Textiles[4]

D 2904 Practice for Interlaboratory Testing of a Textile Test Method that Produces Normally Distributed Data[4]

D 2905 Practice for Statements on Number of Specimens for Textiles[4,5]

E 456 Terminology Relating to Quality and Statistics[6]

E 691 Practice for Conducting An Interlaboratory Study to Determine the Precision of a Test Method[6]

2.2 *ASTM Adjuncts:*

TEX-PAC[7]

NOTE 1—Tex-Pac is a group of PC programs on floppy disks, available through ASTM Headquarters, 100 Barr Harbor Drive, West Conshohocken, PA 19428, USA. The calculations of critical differences and confidence limits described in the various sections of this practice can be performed using some of the programs in this adjunct.

3. Terminology

3.1 *Definitions:*

3.1.1 *accuracy*, *n—of a test method*, the degree of agreement between the true value of the property being tested (or accepted standard value) and the average of many observations made according to the test method, preferably by many observers. See also *bias* and *precision*.

3.1.1.1 *Discussion*—Increased accuracy for a test method is associated with decreased bias relative to an accepted reference value. Although the total bias of a test method is equivalent to the accuracy of the test method, the present edition of *Form and Style for ASTM Standards* recommends using the term "bias" since the accuracy of individual observed values is sometimes defined as involving both the precision and the bias of the method.

3.1.2 *bias*, *n—in statistics*, a constant or systematic error in test results.

3.1.2.1 *Discussion*—Bias can exist between the true value and a test result obtained from one method, between test results obtained from two methods, or between two test results obtained from a single method, for example, between operators or between laboratories.

3.1.3 *characteristic*, *n*—a property of items in a sample or population which, when measured, counted, or otherwise observed, helps to distinguish between the items. (E 456)

3.1.4 *confidence level*, *n*—the stated proportion of times the confidence interval is expected to include the population parameter. (E 206)

3.1.4.1 *Discussion*—Statisticians generally accept that, in the absence of special considerations, 0.95 or 95 % is a realistic confidence level. If the consequences of not including the unknown parameter in the confidence interval would be grave, then a higher confidence level might be considered which would lengthen the reported confidence interval. If the consequences of not including the unknown parameter in the confidence interval are of less than usual concern, then a lower confidence level might be considered which would shorten the reported confidence interval.

3.1.5 *critical difference*, *n*—the observed difference between two test results which should be considered significant at the specified probability level.

3.1.5.1 *Discussion*—The critical difference is not equal to the expected variation in a large number of averages of observed values; it is limited to the expected difference between only two such averages and is based on the standard error for the difference between two averages and not on the standard error of single averages.

3.1.6 *laboratory sample*, *n*—a portion of material taken to represent the lot sample, or the original material, and used in the laboratory as a source of test specimens.

3.1.7 *lot sample*, *n*—one or more shipping units taken at random to represent an acceptance sampling lot and used as a source of laboratory samples.

3.1.8 *parameter*, *n—in statistics*, a variable that describes a characteristic of a population or mathematical model.

3.1.9 *percentage point*, *n*—a difference of 1 percent of a base quantity.

3.1.9.1 *Discussion*—A phrase such as "a difference of *X* %" is ambiguous when referring to a difference in percentages. For example, a change in the moisture regain of a material from 5 % to 7 % could be reported as an increase of 40 % of the initial moisture regain or as an increase of two percentage points. The latter wording is recommended.

3.1.10 *precision*, *n*—the degree of agreement within a set of observations or test results obtained as directed in a method.

3.1.10.1 *Discussion*—The term "precision", delimited in various ways, is used to describe different aspects of precision. This usage was chosen in preference to the use of "repeatability" and "reproducibility" which have been assigned conflicting meanings by various authors and standardizing bodies.

3.1.11 *precision*, *n—under conditions of single-operator precision*, the single-operator-laboratory-sample-apparatus-day precision of a method; the precision of a set of statistically independent observations all obtained as directed in the method and obtained over the shortest practical time interval in one laboratory by a single operator using one apparatus and randomly drawn specimens from one sample of the material being tested.

[4] *Annual Book of ASTM Standards*, Vol 07.01.
[5] *Annual Book of ASTM Standards*, Vol 07.02.
[6] *Annual Book of ASTM Standards*, Vol 14.02.
[7] PC programs on floppy disks are available through ASTM. Request ADJD2906.

3.1.11.1 *Discussion*—Results obtained under conditions of single-operator precision represent the optimum precision that can be expected when using a method. Results obtained under conditions of within-laboratory and between-laboratory precision represent the expected precision for successive test results when a method is used respectively in one laboratory and in more than one laboratory.

3.1.12 *precision, n—under conditions of within-laboratory precision with multiple operators*, the multi-operator, single-laboratory-sample, single-apparatus-day (within operator) precision of a method; the precision of a set of statistically independent test results all obtained in one laboratory using a single sample of material and with each test result obtained by a different operator, with each operator using one apparatus to obtain the same number of observations by testing randomly drawn specimens over the shortest practical time interval.

3.1.13 *precision, n—under conditions of between-laboratory precision*, the multi-laboratory, single-sample, single-operator-apparatus-day (within-laboratory) precision of a method; the precision of a set of statistically independent test results all of which are obtained by testing the same sample of material and each of which is obtained in a different laboratory by one operator using one apparatus to obtain the same number of observations by testing randomly drawn specimens over the shortest practical time interval.

3.1.14 *probability level*, *n*—a general term that reflects the stated proportion of times an event is likely to occur. (Compare to *confidence level* and *significance level*.)

3.1.15 *sample, n*—(*1*) a portion of material which is taken for testing or for record purposes. (See also *sample lot; sample, laboratory; and specimen.*) (*2*) a group of specimens used, or of observations made, which provide information that can be used for making statistical inferences about the population(s) from which the specimens are drawn.

3.1.16 *significance level, (α), n*—the stated upper limit for the probability of a decision being made that an hypothesis about the value of a parameter is false when in fact it is true.

3.1.17 *specimen, n*—a specific portion of a material or a laboratory sample upon which a test is performed or which is selected for that purpose. (*Syn.* test specimen.)

3.1.18 *test result, n*—a value obtained by applying a given test method, expressed as a single determination or a specified combination of a number of determinations.

3.1.19 For definitions of other textile terms used in this practice, refer to Terminology D 123. For definitions of other statistical terms used in this practice, refer to Terminology D 123 or Terminology E 456.

4. Statements on Acceptance Testing

4.1 In the section on *Significance and Use*, include a statement on the use of the method for acceptance testing. If the determined precision supports such use, the test method should be recommended for use. If not, the test method should not be recommended for use. Other circumstances may cause a test method to be used for acceptance testing when precision is poor, or precision is not known. In an event such may occur, advice may be given on the consequences of such usage. In most cases, one of the recommended texts in 4.1.1, 4.1.3, or 4.1.5 will be adequate. In all cases, the recommended text in 4.2 may be part of the statement.

NOTE 2—The final decision to use a specific method for acceptance testing of commercial shipments must be made by the purchaser and the supplier and will depend on considerations other than the precision of the method, including the cost of sampling and testing and the value of the lot of material being tested.

4.1.1 If serious disagreements between laboratories is relatively unlikely, consider the following statement (Note 3).

NOTE 3—In these recommended texts, the numbers of sections, notes, footnotes, equations, and tables are for illustrative purposes and are not intended to conform to the numbers of other parts of this practice. In correspondence they can be best referenced by such phrases as: "the illustrative text in 4.1.1 numbered as 4.1.2."

4.1.2 Method D 0000 for the determination of (insert here the name of the property) is considered satisfactory for acceptance testing of commercial shipments of (insert here the name of the material) since (insert here the specific reason or reasons, such as: (*1*) current estimates of between-laboratory precision are acceptable, (*2*) the method has been used extensively in the trade for acceptance testing, or (*3*) both of the preceding reasons.)

4.1.3 If it is relatively likely that serious disagreements between laboratories may occur but the method is the best available, consider the following statement (Note 3).

4.1.4 Method D 0000 for the determination of (insert here the name of the property) may be used for the acceptance testing of commercial shipments of (insert here the name of the material) but caution is advised since (insert here the specific reason or reasons, such as: (*1*) information on between-laboratory precision is lacking or incomplete or (*2*) between-laboratory precision is known to be poor.) Comparative tests as directed in 4.2.1 may be desirable.

4.1.5 If a method is not recommended for acceptance testing because a more appropriate method is available, because the test is intended for development work only, or because experience has shown that results in one laboratory cannot usually be verified in another laboratory, consider the following statement.

4.1.6 Method D 0000 for the determination of (insert here the name of the property) is not recommended for the acceptance testing of commercial shipments of (insert here the name of the material) since (insert here the specific reason or reasons, such as: (*1*) an alternative, Method D 0000 is recommended for this purpose because (insert here reasons such as those in the illustrative text following 4.1.1), (*2*) experience has shown that results in one laboratory cannot usually be verified in another laboratory, or (*3*) the scope of the method states that the method is recommended only for development work within a single laboratory). Although Method D 0000 is not recommended for use in acceptance testing, it is useful because (insert here the reason or reasons the subcommittee thinks the method should be included in the *Annual Book of ASTM Standards*).

4.2 Include the following statement on conducting comparative tests between laboratories as part of all statements on the use of a method for acceptance testing.

4.2.1 In case of a dispute arising from differences in reported test results when using Method D 0000 for acceptance testing of commercial shipments, the purchaser and the supplier should conduct comparative tests to determine if there is a statistical bias between their laboratories. Competent statistical assistance is recommended for the investigation of bias. As a minimum, the two parties should take a group of test specimens that are as homogeneous as possible and that are from a lot of material of the type in question. The test specimens should then be randomly assigned in equal numbers to each laboratory for testing. The average results from the two laboratories should be compared using Student's *t*-test for unpaired data and an acceptable probability level chosen by the two parties before the testing is begun. If a bias is found, either its cause must be found and corrected or the purchaser and the supplier must agree to interpret future test results in the light of the known bias.

NOTE 4—The test of significance specified in the illustrative text for 4.2.1 is appropriate only for unpaired data from normal distributions. If the type of distribution is not known or is known not to be normal, substitute "a nonparametric test for unpaired data" for "Student's *t*-test for unpaired data" in the next to last sentence in 4.2.1. If the same specimens are evaluated in both laboratories, the description of the preparation of specimens will need to be altered and either "Student's *t*-test for paired data" or "a nonparametric test for paired data" used to describe the test of significance in the next to last sentence of the proposed text for 4.2.1.

5. Statistical Data in Two Sections of Methods

5.1 Many methods approved by Committee D-13 include a section on "*Number of Specimens*" which normally does not describe any interlaboratory testing done during the development of the method or include any estimates of the components of variance obtained from such a study. When that section is written as directed in Practice D 2905, the text consists of three parts. The first part specifies the allowable variation, the probability level, and whether one-sided or two sided limits are required. The second part specifies how the number of observations required for the desired precision can be calculated from an estimate of the single-operator component of variance based on records of the specific laboratory involved. The third part specifies a definite number of observations to be made in the absence of adequate information about the single-operator precision. In the last case, the recommended number of observations is based on a value of the single-operator component of variance somewhat larger than is usually found in practice. Thus, the inexperienced user has the protection of a relatively large number of observations. However, the section on *Number of Specimens* does not allow the inexperienced user of the method to visualize the single-operator precision of the method to be expected for averages based on different numbers of specimens tested by experienced operators. The desirability of supplying such information is the primary reason for requiring information about single-operator precision even when the section on *Number of Specimens* is based on Practice D 2905.

6. Sources of Data

6.1 Plan and conduct an interlaboratory study as directed in Practice D 2904 or in Practice E 691 to secure the information needed to estimate the necessary components of variance. For new test methods for which an interlaboratory test has not been run, see Section 16.

7. Categories of Data

7.1 *General*—Individual observations obtained as directed in a test method fall into a number of categories that require different treatments of the data. The more important of such categories are discussed in standard statistical texts **(1), (2), (3), (4), (5), (6),** and **(7)** and are briefly described in the following sections:

7.2 *Normal Distribution*—If the frequency distribution of individual observations approximates the normal curve and the size of the standard deviation is independent of the average level of the observations, the data can be assumed to be normally distributed and the standard deviation should be used as the measure of variability. Generally, frequency distributions having a hump somewhere near the middle of the distribution and tailing off on either side of the hump approach the normal curve closely enough to permit using data handling techniques based on the normal curve without seriously distorting the conclusions.

NOTE 5—It is recommended that qualified assistance be sought when data do not conform to the normal distribution, when the response is not the arithmetic average of the observations, or both. Within ASTM Committee D-13 such help is available through Subcommittee D13.93 on Statistics.

7.3 *Transformed Data*—If the individual observations have a frequency distribution that is markedly skewed, if the standard deviation seems to be correlated with the average of the observations, or if both these conditions exist; consider transforming the original data to obtain values that are approximately normally distributed with a standard deviation that is independent of the average. Arbitrary grades or classifications and scores of ranked data are among the types of data that usually require transformation before they can be treated as being normally distributed variables.

NOTE 6—In the case of arbitrary grades or classifications and of scores for ranked data, the observations may have such a complex nonlinear relationship that meaningful transformations may not be practicable. If this is so, precision statements must be based on subjective judgement rather than on the analysis of observed data.

NOTE 7—An empirically chosen transformation is often considered satisfactory if a cumulative frequency distribution of the transformed data gives a reasonably straight line when plotted on normal probability graph paper.[8] A number of articles and standard statistical texts discuss the choice of suitable transformations **(2), (3), (8), (9),** and **(10).** (See also Practice D 2904.)

7.3.1 When the shape of the distribution of individual observations is reasonably symmetrical but the standard deviation is proportional to the average of the observations, consider converting the standard deviation to the coefficient of variation, *CV* %, using Eq 1 :

$$CV\,\% = 100\,s / \bar{X} \tag{1}$$

[8] Normal probability graph paper may be bought from most suppliers. The equivalent of Keuffel and Esser Co. Style 46-8000 or of Codex Book Co., Inc., Norwood, MA 02062, Style 3127, is acceptable.

where:

$CV\,\%$ = coefficient of variation as a percent of the average,
s = standard deviation in units of measure, and
$\bar{X}$ = average of all the observations for a specific material.

NOTE 8—Because the transformation is made on the standard deviation and not on the individual observations, the coefficient of variation is not always recognized as a transformation. The same results can be obtained, however, by transforming the individual observations.

NOTE 9—Use of the coefficient of variation when the standard deviation is the more appropriate measure of variability can cause serious errors. Likewise, the use of the standard deviation when the coefficient of variation is the more appropriate measure can result in serious errors.

7.4 *Binomial Distribution*—The binomial distribution applies to test results that are discrete variables reporting the number of successes or failures in a specified number of observations. Each observation in such a test result is an attribute; that is, a nonnumerical report of success or failure based on criteria specified in the procedure (see 7.6).

7.5 *Poisson Distribution*—The Poisson distribution applies to test results that report a count of the number of incidents, such as a specified type of defect, observed over a specified period of time or in one or more specimens of a specified size. The observed count in a Poisson distribution must be small in comparison to the potential count. Examples of data having Poisson distributions are the number of defects of a specified type counted in a specified area of a textile material and the number of stops or other incidents reported for a specified block of equipment over a specified time span.

7.6 *Attributes*—No justifiable statement can be made about the precision or the bias of a test result that is an attribute; that is, a nonnumerical report of success or failure based on criteria specified in the procedure. Test results that are a number summarizing the attributes of individual observations are discrete variables about which justifiable statements can be made on precision and bias (see 7.4 and 7.5).

8. Calculations for Normal Distributions and Transformed Data

8.1 *General*—The same calculations are required for normal distributions having variability measured by standard deviations and for all distributions for which the data have been transformed in order to approach a normal distribution, to make the measure of variability independent of the average, or to obtain both of these objectives. The use of the coefficient of variation as a substitute for the standard deviation is also covered.

8.2 *Calculating Critical Differences*— Calculate the critical differences for averages of observations using Eq 2 or Eq 3:

$$\text{Critical difference between averages, units of measure} = 1.414\, z\, s_T \quad (2)$$

$$\text{Critical difference between averages, percent of average} = 1.414\, z\, v_T \quad (3)$$

where:

1.414 = square root of 2, a constant that converts the standard error of an average to the standard error for the difference between two such averages,
z = standard normal deviate for two-sided limits and the specified probability level (z = 1.960 for the 95 % probability level),
s_T = standard error for the specific size and type of averages being compared (see 8.4), and
v_T = coefficient of variation for the specific size and type of averages being compared (see 8.4).

NOTE 10—Generally, infinite degrees of freedom are assumed when calculating critical differences and confidence limits based on the best information obtainable from existing interlaboratory tests. There are reasonable statistical arguments for this usage. Even if the degrees of freedom associated with each component of variance has been calculated by Satterthwaite's approximation **(1)** or a comparable procedure, there are no generally accepted methods known for assigning degrees of freedom to a standard error which combines two or more components of variance, each having a different number of degrees of freedom, as is done in Eq 7 and 8 and in Eq 10 and 11.

8.3 *Calculating Confidence Limits*— Calculate the width of the confidence limits for averages of observations using Eq 4 or Eq 5:

$$\text{Width of confidence limits for averages, units of measure} = \pm z\, s_T \quad (4)$$

$$\text{Width of confidence limits for averages, percent of average} = \pm z\, v_T \quad (5)$$

where the terms in the equations are defined in 8.2.

NOTE 11—The property being evaluated may need to be controlled in only one direction; for example, excessive fabric shrinkage is important but too little shrinkage is not likely to be undesirable. Nevertheless, "plus and minus" confidence limits are suggested even in these cases since confidence limits are normally used to express the variability in a single average. Critical differences should be used to compare pairs of averages.

8.4 *Combining Components of Variance*— Calculate the standard error of the specific size and type of averages that are to be compared using Eq 6, Eq 7, or Eq 8:

$$s_T\ (\text{single}-\text{operator}) = (s_s^2/n)^{1/2} \quad (6)$$

$$s_T\ (\text{within}-\text{laboratory}) = [s_w^2 + (s_s^2/n)]^{1/2} \quad (7)$$

$$s_T\ (\text{between}-\text{laboratory}) = [s_B^2 + s_w^2 + (s_s^2/n)]^{1/2} \quad (8)$$

where the equations respectively calculate the standard error of averages of observations under the conditions of single-operator, within-laboratory, and between-laboratory precision, and

where:

s_S^2 = single-operator component of variance or the residual error component of variance,
$s_{W_2}^2$ = within-laboratory component of variance (Note 12),
s_B^2 = between-laboratory component of variance, and
n = number of observations by a single operator in each average.

NOTE 12—The within-laboratory component of variance is the sum of all the individual components of variance, except the single-operator component of variance, that contribute to the variability of observations within a single laboratory. Included are such components of variance as those for days of testing, units of apparatus, and different operators within a single laboratory. If the within-laboratory component of variance is not calculated separately, all sources of variability except the single-operator component are included in the between-laboratory component. Under

these conditions, calculate the standard error (between-laboratory) using zero for the within-laboratory component.

When an interlaboratory test program run as directed in Practice D 2904 results in a significant material by laboratory interaction or a material by operator (within laboratories) interaction, see Annex A1.14.2 of Practice D 2904 for instructions on estimating the components of variance for multi-material comparisons.

If components of variance are to be expressed as coefficients of variation, calculate them using Eq 1 and calculate the coefficient of variation for the specific size and type of averages that are to be compared using Eq 9, Eq 10, or Eq 11:

$$v_T \text{ (single−operator)} = (v_s^2/n)^{1/2} \quad (9)$$

$$v_T \text{ (within−laboratory)} = [v_w^2 + (v_s^2/n)]^{1/2} \quad (10)$$

$$v_T \text{ (between laboratory)} = [v_B^2 + v_w^2 + (v_s^2/n)]^{1/2} \quad (11)$$

where the meanings of the subscripts for the individual components of variance expressed as coefficients of variation are the same as in the legend for Eq 6, Eq 7, and Eq 8.

8.5 *Sample Calculations—Components of Variance as Standard Deviations*:

8.5.1 *Example 1: Single-Operator Precision*—At the 95% probability level, calculate the critical difference and confidence limits for averages of ten observations when the single-operator component of variance expressed as a standard deviation is 1.8 percentage points. Using Eq 6, $s_T = [(1.8)^2/10]^{1/2} = 0.57$ percentage points. Using Eq 2, the critical difference = $1.414 \times 1.960 \times 0.57 = 1.58$ percentage points. Using Eq 4, the width of the confidence limits = $\pm(1.960 \times 0.57) = \pm 1.12$ percentage points (Note 12).

8.5.2 *Example 2: Within-Laboratory Precision (Multi-Operator)*—At the 95 % probability level, calculate the critical difference and confidence limits for averages of ten when the single-operator and within-laboratory components of variance expressed as standard deviations are respectively 1.8 and 0.3 percentage points. Using Eq 7, $s_T = [(0.3)^2 + ((1.8)^2/10)]^{1/2} = 0.64$ percentage points. Using Eq 2, the critical difference = $1.414 \times 1.960 \times 0.64 = 1.77$ percentage points. Using Eq 4, the width of the confidence limits = $\pm(1.960 \times 0.64) = \pm 1.25$ percentage points (Note 12).

8.5.3 *Example 3: Between-Laboratory Precision*—At the 95 % probability level, calculate the critical difference and the confidence limits for averages of ten when the single-operator, within-laboratory, and between-laboratory components of variance expressed as standard deviations are respectively 1.8, 0.3, and 0.5 percentage points. Using Eq 8, $s_T = [(0.5)^2 + (0.3)^2 + ((1.8)^2/10)]^{1/2} = 0.81$ percentage points. Using Eq 2, the critical difference = $1.414 \times 1.960 \times 0.81 = 2.24$ percentage points. Using Eq 4, the width of the confidence limits = $\pm(1.960 \times 0.81) = \pm 1.59$ percentage points (Note 12).

TABLE 1 Components of Variance as Standard Deviations, Percentage Points

Names of the Properties	Single-Operator Component	Within-Laboratory Component	Between-Laboratory Component
(Insert here name of Property 1)	1.8	0.3	0.5
(Insert here name of Property 2)	1.2	0.4	0.0

8.6 *Sample Calculations—Components of Variance as Coefficients of Variation*:

8.6.1 *Example 4: Within-Laboratory Precision*—At the 95 % probability level, calculate the critical difference and the confidence limits for averages of five observations when the single-operator component of variance expressed as a coefficient of variation is 5.3 % of the average. Using Eq 9, $v_T = [(5.3)^2/5]^{1/2} = 2.37$ % of the average. Using Eq 3, the critical difference = $1.414 \times 1.960 \times 2.37 = 6.57$ % of the average. Using Eq 5, the width of the confidence limits = $\pm(1.960 \times 2.37) = \pm 4.65$ % of the average. (Note 12).

8.6.2 *Example 5: Within-Laboratory Precision (Multi-Operator)*—At the 95 % probability level, calculate the critical difference and the confidence limits for averages of five observations when the single-operator and within-laboratory components of variance expressed as coefficients of variation are respectively 5.3 and 1.0 % of the average. Using Eq 10, $v_T = [(1.0)^2 + ((5.3)^2/5)]^{1/2} = 2.57$ % of the average. Using Eq 3, the critical difference = $1.414 \times 1.960 \times 2.57 = 7.12$ % of the average. Using Eq 5, the width of the confidence limits = $\pm(1.960 \times 2.57) = \pm 5.04$ % of the average (Note 12).

8.6.3 *Example 6: Between-Laboratory Precision*—At the 95 % probability level, calculate the critical difference and the confidence limits for averages of five observations when the single-operator, within-laboratory, and between-laboratory components of variance are expressed as coefficients of variation and are respectively 5.3, 1.0, and 2.0 % of the average. Using Eq 11, $v_T = [(2.0)^2 + (1.0)^2 + ((5.3)^2/5)]^{1/2} = 3.26$ % of the average. Using Eq 3, the critical difference = $1.414 \times 1.960 \times 3.26 = 9.03$ % of the average. Using Eq 5, the width of the confidence limits = $\pm(1.960 \times 3.26) = \pm 6.39$ % of the average.

9. Calculations for Binomial Distributions

9.1 *Critical Differences for Binomial Distributions*—Determine critical differences between two test results from a binomial distribution using an exact test of significance for 2 by 2 contingency tables containing small frequencies. Prepare a table of critical differences using published tables (Table 28, Ref **7**), the methods shown in 9.1.1 and 9.1.2, or an algorithm for use with a computer.[9] See Table 3 for an example of such a table.

NOTE 13—For data from both the binomial and Poisson distributions, the tables of critical differences and of confidence limits are based on the mathematical characteristics of the applicable frequency distribution. Bias in actual test results due to systematic errors in testing for some or all of the observations will normally have the effect of reducing the actual probability level to some unknown value which is less than the value shown in the tables.

9.1.1 Calculate the probability of reporting exactly *a* successes in one of the test results using Eq 12:

$$f(a \mid r,A,B) = A!\,B!\,r!(N-r)!/a!\,b!\,(A-a)!(B-b)!N! \quad (12)$$

[9] Printouts and punched cards describing all of the algorithms mentioned in this recommended practice are available from ASTM Headquarters, 100 Barr Harbor Drive, West Conshohocken, PA 19428, at a nominal cost. Request ADJD2906.

TABLE 2 Critical Differences for the Conditions Noted, 95 % Probability Level, Percentage Points[A]

Names of the Properties	Number of Observations in Each Average	Single-Operator Precision	Within-Laboratory Precision	Between-Laboratory Precision
(Name of Property 1)	1	5.0	5.1	5.2
	4	2.5	2.6	3.0
	8	1.0	1.9	2.4
(Name of Property 2)	1	3.3	3.5	3.5
	4	1.7	2.0	2.0
	8	1.2	1.6	1.6

[A] The critical differences were calculated using $z = 1.960$.

TABLE 3 Confidence Limits for the Conditions Noted, 95 % Probability Level, Percentage Points[A]

Names of the Properties	Number of Observations in Each Average	Single-Operator Precision	Within-Laboratory Precision	Between-Laboratory Precision
(Name of Property 1)	1	±3.5	±3.6	±3.7
	4	±1.8	±1.9	±2.1
	8	±1.2	±1.4	±1.7
(Name of Property 2)	1	±2.4	±2.5	±2.5
	4	±1.2	±1.4	±1.4
	8	±0.8	±1.1	±1.1

[A] The confidence limits were calculated using $z = 1.960$.

TABLE 4 95.0 % Probability Level, Significantly Different Numbers of Successes (or of Failures) in a Specified Number of Specimens[A]

Successes in One Test Result 8 Specimens	Successes in Another Test Result 8 Specimens
0	5 or more
1	6 or more
2	7 or more
3	8
4	
5	0
6	1 or less
7	2 or less
8	3 or less

[A] This table was prepared as directed in 9.1 of Practice D 2906. The probability level is for two-sided tests. Successes in one test result are compared to successes in the other. Failures are also compared only to failures.

TABLE 5 95 % Confidence Limits for Test Results of Eight Observations[A]

Observed Number of Successes	Percent of Successes	
	Lower Limit	Upper Limit
0	0.0 %	36.9 %
1	0.3 %	52.6 %
2	3.2 %	65.1 %
3	8.5 %	75.5 %
4	15.7 %	84.3 %
5	24.5 %	91.5 %
6	34.9 %	96.8 %
7	47.4 %	99.7 %
8	63.1 %	100.0 %

[A] This table was prepared as directed in 9.2 of Practice D 2906. Limits are stated as percent successes in population sampled.

where:

A = number of observations in one test result,
B = number of observations in the other test result with B equal to or less than A,
N = $A + B$,
a = number of observations in test result A which are reported as successes,
b = number of observations in test result B which are reported as successes,
r = $a + b$, and
f = probability as a fraction of observing exactly a successes for specified values of r, A, and B.

9.1.2 For every value of r from $r = 1$ to $r = N - 1$, calculate a lower limit for a which conforms to both Eq 13 and Eq 14 and an upper limit of a which conforms to both Eq 15 and Eq 16:

$$\sum_{k=b}^{j} f(a|r,A,B) \leq \alpha / 2 \quad (13)$$

$$\sum_{k=b=1}^{j} f(a|r,A,B) > \alpha / 2 \quad (14)$$

$$\sum_{k=o}^{b} f(a|r,A,B) \leq \alpha / 2 \quad (15)$$

$$\sum_{k=o}^{b+1} f(a|r,A,B) > \alpha / 2 \quad (16)$$

where:

j = smaller of the quantities B and r, and
a = alpha, the probability as a fraction that, when both test results are drawn from the same population of observations, an observed value of a either will equal or be outside the calculated limits for a (Note 14) and where the other terms are defined in 9.1.1.

NOTE 14—Alpha equals one hundredth of the quantity, 100 minus the probability level as a percent; for example, $\alpha = 0.05$ when the probability level is 95 %.

9.2 *Confidence Limits for Binomial Distributions*—Prepare a table showing the confidence limits for the fraction of success

TABLE 6 Values of *b* for Critical Differences in Defect Counts, *a* and *b*, for Two Test Results Two-Sided Tests at the 95 % Probability Level [A,B]

Probability Level		Probability Level		Probability Level		Probability Level	
r = a + b	*b*[C]	*r = a + b*	*b*	*r = a + b*	*b*	*r = a + b*	*b*
1		26	7	51	18	76	28
2		27	7	52	18	77	29
3		28	8	53	18	78	29
4		29	8	54	19	79	30
5		30	9	55	19	80	30
6	0	31	9	56	20	81	31
7	0	32	9	57	20	82	31
8	0	33	10	58	21	83	32
9	1	34	10	59	21	84	32
10	1	35	11	60	21	85	32
11	1	36	11	61	22	86	33
12	2	37	12	62	22	87	33
13	2	38	12	63	23	88	34
14	2	39	12	64	23	89	34
15	3	40	13	65	24	90	35
16	3	41	13	66	24	91	35
17	4	42	14	67	25	92	36
18	4	43	14	68	25	93	36
19	4	44	15	69	25	94	37
20	5	45	15	70	26	95	37
21	5	46	15	71	26	96	37
22	5	47	16	72	27	97	38
23	6	48	16	73	27	98	38
24	6	49	17	74	28	99	39
25	7	50	17	75	28	100	39

[A] This table was prepared as directed in 10.1 of Practice D 2906.

[B] Additional probability levels for one-sided tests are given in Table 36A of Ref 7.

[C] If the observed value of $\overline{b<}$ the tabulated value, the two results should be considered significantly different at the 95 % probability level.

a = the larger of two defect counts, each of which is the total count for all specimens in a test result and each of which is based on the same number of specimens,
b = the smaller of the two defect counts taken as specified for *a*, and
r = a + b.

When *r* > 100, use the following approximation:

$$b = c - 1 - 1.386\sqrt{c}$$
$$k = 0.707\,z \qquad (27)$$

where:
b = calculated value of *b*, rounded to the nearest whole number,
c = *r* /2.

in the population of observations being sampled when test results have specific numbers of successes out of the specified number of observations in each test result. Use existing tables or charts **(4),** (Table 41, Ref **7**), the methods specified in 9.2.1 (Note 14), or an algorithm for use with a computer.[7] See Table 4 for an example of such a table.

9.2.1 Calculate upper and lower confidence limits for *p*, the fraction of successes in the population of observations being sampled, using Eq 17 for the upper limits and Eq 18 for the lower limits:

$$\sum_{j=0}^{k} \binom{n}{j} p^i q^{n-j} = \alpha/2 \qquad (17)$$

$$\sum_{j=0}^{n} \binom{n}{j} p^i q^{n-j} = \alpha/2 \qquad (18)$$

where:
q = $1 - p$ or the fraction of failures in the population of observations being sampled,
n = specified number of observations per test result,
k = observed number of successes in a test result,
j = range of values of k to be considered in calculating the confidence limits, and
α = alpha, the probability as a fraction that, when k successes are observed in a test result of n observations, the true value of p equals or falls outside the calculated confidence limits (Note 14).

NOTE 15—Since Eq 17 and 18 cannot be directly solved for *p*, the value of *p* is normally obtained by successive estimations which are terminated when the calculated value of $\alpha/2$ agrees, within an acceptable limit of calculational error, with the desired value of $\alpha/2$.

TABLE 7 95 % Confidence Limits for Number of Counts per Test Result[A]

Observed Count	Lower Limit[B]	Upper Limit
0	0.0	3.7
5	1.6	11.7
10	4.8	18.4
15	8.4	24.7
20	12.2	30.9
25	16.2	36.9
30	20.2	42.8
35	24.4	48.7
40	28.6	54.5
45	32.8	60.2
50	37.1	65.9
60	45.8	77.2
70	54.6	88.4
80	63.4	99.6
90	72.4	110.6
100	81.4	121.6
120	99.5	143.5
140	117.8	165.2
160	136.2	186.8
180	154.7	208.3
200	173.2	229.7

[A] This table was prepared as directed in 10.2 of Practice D 2906.
[B] Lower confidence limit for counts = $c\,[1 - (1/9c) - z\,(1/9c)^{1/2}]^3$
[C] Upper confidence limit for count = $d\,[1 - (1/9d) + z\,(1/9d)^{1/2}]^3$

where:
c = observed number of counts,
d = $c + 1$, and
z = 1.960, the standard normal deviate for two-sided limits at the 95 % probability level.

10. Calculations for Poisson Distributions

10.1 *Critical Differences for Poisson Distributions*—See Table 6 for critical differences in defect counts at the 95 % probability level. For other probability levels, prepare a table of critical differences for observed counts from test results based on the Poisson distribution using existing tables (Table 36A, Ref **7**), the methods specified in 10.1.1 and 10.1.2 (Note 13), or an algorithm for use with a computer.[8]

10.1.1 Calculate the value of b, the smaller of two observed counts from data having a Poisson distribution, which conforms to both of the binomial expressions shown as Eq 19 and 20:

$$\sum_{j=0}^{b} \binom{r}{j} 0.5^r \leq \alpha/2 \tag{19}$$

$$\sum_{j=0}^{b+1} \binom{r}{j} 0.5^r > \alpha/2 \tag{20}$$

where:
r = $a + b$
a = larger of two observed counts from data having a Poisson distribution, and
α = alpha, the probability as a fraction that, when two counts which total exactly r are made on data from the same Poisson distribution, the observed value of b will be equal to or less than the calculated value of b (Note 14).

10.1.2 When $r > 100$ or when approximating the value of b for $r \leq 100$, use the empirical relationship shown as Eq 21:

$$b = c - 1 - k\sqrt{c} \tag{21}$$

where:
b = calculated value of b, rounded to the nearest whole number,
c = $r/2$, and
k = 1.386 for the 95 % probability level.

NOTE 16—The value of the constant k in Eq 21 is determined using Eq 27:

where:
z = 1.960, the standard normal deviate for two-sided limits at the 95 % probability level, and
0.707 = a constant equal to 1.414/2 with 1.414 being the square root of 2 and serving the purpose of converting the standard error for a count to the standard error for the difference between two counts with division by 2 required because $c - b$ is one-half of the difference between a and b.

10.2 *Confidence Limits for Poisson Distributions*—See Table 7 for confidence limits for defect counts at the 95 % probability level. For other probability levels, prepare a table of confidence limits for observed counts from test results based on the Poisson distribution using existing tables **(6)**, (Table 40, Ref **7**), or one of the methods specified in 10.2.1 or 10.2.2 (Note 13).

10.2.1 For an observed count, c, from a test result obtained as directed in the method, calculate confidence limits using Eq 22 and 23:

$$\text{Lower confidence limit for counts} = \tfrac{1}{2}\,C \tag{22}$$

$$\text{Upper confidence limit for counts} = \tfrac{1}{2}\,D \tag{23}$$

where:
C = value of chi-square taken from tables of the chi-square distribution with degrees of freedom equal to $2c$ and a significance level of $(1 - \alpha/2)$ (Note 14), and,
D = value of chi-square taken from tables of the chi-square distribution with degrees of freedom equal to $2(c + 1)$ and a significance level of $\alpha/2$ (Note 14).

10.2.2 For observed counts, c, which exceed 50 or for approximations for any value of c that are correct to within two digits in the second decimal place, calculate confidence limits using the normal approximations that are shown as Eq 24 and 25:

Lower confidence limit for counts

$$= c[1 - (1/9c) - z(1/9c)^{1/2}]^3 \tag{24}$$

Upper confidence limit for counts

$$= d[1 - (1/9d) + z(1/9d)^{1/2}]^3 \tag{25}$$

where:
d = $c + 1$ or one more than the observed number of counts, and
z = 1.960, the standard normal deviate for two-sided limits at the 95 % probability level.

RECOMMENDED TEXTS 1 AND 2—VARIABLES

11. Statements Based on Normal Distributions and Transformed Data

11.1 *General*—Include a summary, description of interlaboratory test and the components of variance obtained in the test, a statement on precision giving typical critical differences or confidence limits, or both, and a statement on bias. For a method which specifies two or more procedures for a single property or for a method specifying procedures for testing two or more properties, decide whether to write a single statement on precision and bias or two or more statements on precision and bias.

11.1.1 State as a footnote where the data for the interlaboratory tests are filed. Preferably, file the data at ASTM Headquarters. To do this, get copies of the combined cover and title page from ASTM Headquarters. Send two copies of the report to Headquarters for filing with the cover page properly titled and signed by the officers of the subcommittee. The Headquarters staff will assign a number to the report and notify the subcommittee officers. A typical footnote wording is "ASTM Research Report No. D-13-XXXX. A copy is available on loan from ASTM Headquarters, 100 Barr Harbor Drive, West Conshohocken, PA 19428."

11.1.2 If there are fewer than five laboratories in the interlaboratory test and if data are listed for between-laboratory precision, consider including the following note just ahead of the general note on between-laboratory precision which is illustrated as Note 20 in Recommended Text 1 and as Note 22 in Recommended Text 2.

NOTE 17—Since the interlaboratory test included only (insert here the number) laboratories, estimates of between-laboratory precision may be either underestimated or overestimated to a considerable extent and should be used with special caution.

11.1.3 In preparing the statement on bias, consider whether (*1*) the true value of the property, such as the elongation of a yarn, can only be defined in terms of a specific test or (*2*) the true value of the property, such as the moisture content of a yarn, can be defined independently of the method of testing. Prepare a statement based on which of these alternatives exists, on judgment about the existence of either actual or suspected statistical bias, and on the reputation of the test method in the trade. Consider using one of the texts illustrated as 11.1.4, 11.1.5, 11.1.6, and11.1.7:

NOTE 18—In recommended texts, the numbers of sections, notes, footnotes, equations, and tables are for illustrative purposes and are not intended to conform to the numbers assigned to the other parts of this practice. In correspondence, they can be best referenced by such phrases as "the illustrative text numbered as 11.2.1."

11.1.4 *Bias*—The procedure in Method D 0000 for measuring (insert here the name of the property) has no known bias because the value of (insert here the name of the property) can be defined only in terms of a test method. (If applicable, the words "Method D 0000 is generally accepted as a referee method." may be added to the previous text. See 11.3.1.5 of Recommended Text 2 for an alternative wording to be used when referring to the bias of two or more properties.)

11.1.5 *Bias*—Method D 0000 for measuring (insert here the name of the property) has no known bias and is generally used as a referee method.

11.1.6 *Bias*—The average results secured using Method D 0000 for measuring (insert name of property) were 1.7 percentage points higher than the results obtained using Method D 0000, the referee method.

11.1.7 *Bias*—The average of 20 observations made on a National Bureau of Standards standard using Method D 0000 was 23.75± 0.23 % of (insert name of property) at the 95 % probability level as compared to a stated value of 24.18 %.

11.2 *Recommended Text 1 for a Single Property*—Use the text illustrated as 11.2.1.1-11.2.1.5 for statements on a single property for which the variability is expressed as coefficients of variation (Note 16). See 11.2.2-11.2.6 for instructions on variations in the recommended text:s

11.2.1 **Precision and Bias**

11.2.1.1 *Summary*—In comparing two single observations, the difference should not exceed 14.7 % of the average of the two observations in 95 out of 100 cases when both observations are taken by the same well-trained operator using the same piece of test equipment and specimens randomly drawn from the same sample of material. Larger differences are likely to occur under all other circumstances. The true value of (insert the name of the property) can only be defined in terms of a specific test method. Within this limitation, Method D 0000 has no known bias. Sections 11.2.1.2-11.2.1.5 explain the basis for this summary and for evaluations made under other conditions.

11.2.1.2 *Interlaboratory Test Data*[4]—An interlaboratory test was run in 19XX in which randomly drawn samples of two materials were tested in each of six laboratories. One operator in each laboratory tested four specimens of each material. The components of variance for (insert here the name of the property) results expressed as coefficients of variation were calculated to be:

Single-operator component	5.3 % of the average
Between-laboratory component	3.0 % of the average

NOTE 19—The square roots of the components of variance are being reported to express the variability in the appropriate units of measure rather than as the squares of those units of measure.

11.2.1.3 *Critical Differences*—For the components of variance reported in 11.2.1.2 averages of observed values should be considered significantly different at the 95 % probability level if the difference equals or exceeds the following critical differences:

Number of Observations in Each Average	Critical Difference, Percent of Grand Average, for the Conditions Noted[A,B]	
	Single-Operator Precision	Between-Laboratory Precision
1	14.7	16.9
5	6.6	10.6
10	4.6	9.5

[A] The critical differences were calculated using $z = 1.960$.

[B] To convert the values of the confidence limits to units of measure, multiply the critical differences by the average of the two specific sets of data being compared and then divide by 100.

NOTE 20—The tabulated values of the critical differences and confidence limits should be considered to be a general statement, particularly

with respect to between-laboratory precision. Before a meaningful statement can be made about two specific laboratories, the amount of statistical bias, if any, between them must be established, with each comparison being based on recent data obtained on specimens taken from a lot of material of the type being evaluated so as to be as nearly homogeneous as possible and then randomly assigned in equal numbers to each of the laboratories.

11.2.1.4 *Confidence Limits*—For the components of variance reported in 11.2.1.2, single averages of observed values have the following 95 % confidence limits (Note 20):

Number of Observations in Each Average	Width of 95 % Confidence Limits, Percent of the Grand Average for the Conditions Noted[A,B]	
	Single-Operator Precision	Between-Laboratory Precision
1	±10.4	±11.9
5	±4.6	±7.5
10	±3.3	±6.7

[A] The confidence limits were calculated using $z = 1.960$.

[B] To convert the values of the confidence limits to units of measure, multiply the critical differences by the average of the two specific sets of data being compared and then divide by 100.

11.2.1.5 *Bias*—The procedure in Method D 0000 for measuring (insert here the name of the property) has no known bias because the value of (insert here the name of the property) can be defined only in terms of a test method.

(*Editorial Comment*—See 11.1.1 for the wording of the footnote indicated in the title of 11.2.1.2.)

11.2.2 If variability is expressed as standard deviations, substitute the words "standard deviations" for "coefficient of variations" as needed, and omit the footnotes *B* which are used in 11.2.1.3 and 11.2.1.4 of the recommended text. If a transformation other than the coefficient of variation has been used, see 11.4.

11.2.3 If test results are normally the average of more than a single observation, use a first sentence in the summary such as: "In comparing two averages of five observations, the differences should not exceed 6.6 % of the grand average of all of the observations in 95 cases out of 100 when all of the observations are taken by the same well-trained operator using the same piece of test equipment and specimens randomly drawn from the same sample of material."

11.2.4 When more than one operator per laboratory was used in the interlaboratory test, change the description of the interlaboratory test by using a sentence such as: "Each laboratory used two operators, each of whom tested four specimens of each material" and by listing the value for the within-laboratory component of variance. In addition, add a column of data with the caption "Within-laboratory Precision" in 11.2.1.3 and 11.2.1.4 of Recommended Text 1.

11.2.5 If between-laboratory precision is not evaluated, omit the note illustrated as Note 20 in Recommended Text 1.

11.2.6 If either the section on confidence limits or the section on critical differences is omitted from the recommended text, change the title of the remaining section to *Precision*.

11.3 *Recommended Text 2 for More than One Property*—Use the text illustrated as 11.3.1.1-11.3.1.5 for statements on more than one property for which variability is expressed as standard deviations (Note 18). See 11.3.2 and 11.3.3 for instructions on variations in the recommended text:

11.3.1 **Precision and Bias**

11.3.1.1 *Summary*—In comparing two averages, the differences should not exceed the following critical differences in 95 cases out of 100 when all of the observations are taken by the same well-trained operator using the same piece of test equipment and specimens randomly drawn from the same sample of material and tested on the same day.

(Name of property 1)	2.5 percentage points for averages of 4
(Name of property 2)	1.7 percentage points for averages of 8

The size of the differences is likely to be affected adversely by different circumstances. The true values of (insert name of property 1) and (insert name of property 2) can be defined only in terms of specific test methods. Within this limitation, the procedures in Method D 0000 for determining these properties have no known bias. Sections 11.3.1.2-11.3.1.5 explain the basis for this summary and for evalutions made under other conditions.

11.3.1.2 *Interlaboratory Test Data*[4]—An interlaboratory test was run in 19XX in which randomly drawn samples of three materials were tested in eight laboratories. Each laboratory used two operators, each of whom tested six specimens of each material. The components of variance expressed as standard deviations were calculated to be the values listed in Table 1.

NOTE 21—The square roots of the components of variance are being reported to express the variability in the appropriate units of measure rather than as the squares of those units of measure.

11.3.1.3 *Critical Differences*—or the components of variance listed in Table 1, two averages of observed values should be considered significantly different at the 95 % probability level if the difference equals or exceeds the critical differences listed in Table 2.

NOTE 22—The tabulated values of the critical differences and confidence limits should be considered to be a general statement, particularly with respect to between-laboratory precision. Before a meaningful statement can be made about two specific laboratories, the amount of statistical bias, if any, between them must be established, with each comparison being based on recent data obtained on specimens taken from a lot of material of the type being evaluated so as to be as nearly homogeneous as possible and then randomly assigned in equal numbers to each of the laboratories.

11.3.1.4 *Confidence Limits*—For the components of variance listed in Table 1, single averages of observed values have the 95 % confidence limits listed in Table 3 (Note 22).

11.3.1.5 *Bias*—The procedures in Method D 0000 for measuring the properties listed in Tables 1-3 have no known bias because the value of these properties can be defined only in terms of a test method.

(*Editorial Comment*—See 11.1.1 for the wording of the footnote indicated in the title of 11.3.1.2. Tables 1-3 are at the end of this practice.)

11.3.2 If variability is expressed as coefficients of variation, (*1*) substitute the words "coefficients of variation" for "standard deviation" as needed, (*2*) use "% of the average" as the unit of measure, and (*3*) add a footnote *B* to Table 2 and to Table 3 respectively, equivalent to the footnotes *B* in 11.2.1.1 and 11.2.1.2 of Recommended Text 1. If a transformation other than the coefficient of variation has been used, see 11.4.

11.3.3 If all the properties listed in the tables are not reported in the same units of measure, use "Units as Indicated" in the table headings and show the units of measure after each property name in the leftmost column of each table.

11.4 *Transformations*—When a transformation other than the coefficient of variation has been used, make the following changes in the appropriate recommended text.

11.4.1 Use a text comparable to the one illustrated as 11.4.1.1 and 11.4.1.2 for a single property or modify those sections as required for statements on two or more properties:

11.4.1.1 *Summary*—In 95 cases out of 100 when comparing two averages of two observations each, the difference should not exceed that one of the following critical differences which is appropriate for the level of (insert here the name of the property) when all of the observations are taken by the same well-trained operator using the same piece of test equipment and specimens randomly drawn from the same sample of material and tested on the same day.

Smaller Average, %	Larger Average, %	Critical Difference, percentage points
1.00	1.45	0.45
3.00	3.63	0.63
5.00	5.76	0.76
7.00	7.87	0.87

The size of the difference is likely to be adversely affected when testing is done under other circumstances. The true but unknown value of (insert the name of the property) can be defined only in terms of a specific test method. Within this limitation, Method D 0000 has no known bias. Sections 11.4.1.2 through 11.4.1.5 explain the basis for this summary and for evaluations made under other conditions.

11.4.1.2 *Interlaboratory Test Data*[4]—An interlaboratory test was run in 1970 in which randomly drawn samples of three materials were tested in each of five laboratories. Each laboratory used two operators, each of whom tested two specimens of each material. The observed data did not conform to the assumptions underlying the analysis of variance. Before analysis, the data were transformed using Eq 26:

$$C = (D + 1)^{1/2} \quad (26)$$

where:
C = transformed datum, and
D = observed datum.

The components of variance for (insert here the name of the property) expressed as standard deviations were calculated to be:

Single-operator component	0.077 transformed units
Within-laboratory component	0.000 transformed units
Between-laboratory component	0.068 transformed units

NOTE 23—The square roots of the components of variance are being reported to express the variability in the appropriate units of measure rather than as the squares of those units of measure.

(*Editorial Comment*—The equation for transforming the observed data will vary with the method. See 7.3. The data illustrated in 11.4.1.1 were obtained by (*1*) selecting a series of smaller averages spanning the range of interest, (*2*) converting each to transformed units, (*3*) adding the appropriate critical difference in transformed units, (*4*) obtaining the corresponding larger average by converting the sum back to observed units, and (*5*) obtaining the critical difference in observed units by subtraction.)

11.4.2 Use "transformed units" as the unit of measure throughout the statements.

NOTE 24—When both the statement on number of specimens and the statement on precision and bias refer to a measure of variability expressed in transformed units, information about the reason for and the basis of the transformation may be moved from the section on *Precision and Bias* to a new paragraph in the section on *Significance and Use* with a reference to that new paragraph in the format, "transformed units (see X.X)" being used instead of "transformed units" in both the text and tables of the method.

11.4.3 Use the following footnote as a footnote to the data on critical differences:

11.4.4 Use the following footnote as a footnote to the data on confidence limits:

[B] Before evaluating the precision of a single average expressed in observed units of measure; transform the average using Eq 26, calculate in transformed units the average plus and minus the tabulated values, and convert the resulting values into observed units of measure using Eq 26. The values of the upper and lower confidence limits in observed units of measure will not be symmetrical about the average.

RECOMMENDED TEXT 3—BINOMIAL DISTRIBUTIONS

12. Statements Based on Binomial Distributions

12.1 *General*—Include an introductory statement, a statement on precision giving typical critical differences and confidence limits, and a statement on bias.

12.1.1 In preparing the statement on bias, consider using one of the texts illustrated in 11.1.3 as 11.1.4 or 11.1.5.

12.2 *Recommended Text 3 for Binomial Distributions*—Use the text illustrated in 12.2.1.1-12.2.1.4 for statements on test results having a binomial distribution (Note 18):

12.2.1 **Presicion and Bias**

12.2.1.1 *Introduction*—Individual observations are attributes; that is, each observation is reported as a success or a failure based on criteria specified in the method. Test results report the number of successes in a specified number of observations on a specific material tested in a single laboratory under comparable conditions and have a binomial distribution. If actual test results include bias due to systematic sampling or testing errors in some or all of the observations, the critical differences in Table 4 will be overly optimistic and the confidence limits in Table 5 will be widened by the existence of such bias.

12.2.1.2 *Critical Differences*—Table 4 contains criteria for comparing two test results at the 95 % probability level, each comparing eight observations obtained under comparable conditions in the same laboratory.

NOTE 25—No justifiable statement can be made about the critical differences or confidence limits for Method D 0000 for measuring (insert here the name of the property) under conditions of between-laboratory precision.

12.2.1.3 *Confidence Limits*—Table 5 lists the confidence limits at the 95 % probability level for single test results based on eight observations (Note 25).

12.2.1.4 *Bias*—The true value of (insert name of property) can only be defined in terms of a test method. Within this limitation, Method D 0000 has no known bias.

(*Editorial Comment*—Tables 4 and 5 are at the end of this practice.)

RECOMMENDED TEXT 4—POISSON DISTRIBUTIONS

13. Statements Based on Poisson Distributions

13.1 *General*—Include an introductory statement, a statement on precision giving typical critical differences and confidence limits, and a statement on bias.

13.1.1 In preparing the statement on bias, consider using one of the texts illustrated in 11.1.3 as 11.1.4 and 11.1.5.

13.2 *Recommended Text 4 for Poisson Distributions*—Use the text illustrated in 13.2.1.1-13.2.1.4 for statements on test results having a Poisson distribution (Note 18):

13.2.1 **Precision and Bias**

13.2.1.1 *Introduction*—Test result are reported as the average defect count per specimen for a specific material. The precision of test results is evaluated in terms of the total defect count for all specimens included in each test result since such total counts have a Poisson distribution while the average defect counts do not have such a distribution. If the total counts for actual test results include bias due to systematic sampling or testing errors, the critical differences in Table 6 will be overly optimistic and the confidence limits in Table 7 will be widened by the existence of such bias.

13.2.1.2 *Critical Differences*—Table 6 contains criteria for determining if the total defect counts for two test results, each based on the same number of specimens of a stated size, should be considered significantly different at the indicated probability levels.

NOTE 26—No justifiable statement can be made about the critical differences or confidence limits for Method D 0000 for measuring (insert here the name of the property) under conditions of between-laboratory precision.

NOTE 27—Although the preparation of specimens invariably increases the number of defects in the original material, specimens prepared in the same way and under the same conditions provide useful comparisons of the actual or potential defect count in different materials or in different lots of the same material.

13.2.1.3 *Confidence Limits*—Table 7 shows the 95 % confidence limits for the total defect count in a single test result obtained as directed in the method (Note 27).

13.2.1.4 *Bias*—The true value of (insert name of property) can only be defined in terms of a test method. Within this limitation, Method D 0000 has no known bias.

(*Editorial Comment*—The note illustrated in the text as Note 26 may need to be omitted or altered based on the judgement of the task group writing the method.)

RECOMMENDED TEXT 5—ATTRIBUTES

14. Statements Based on Attributes

14.1 When a method specifies that the test result is an attribute, that is, a nonnumerical report of success or failure based on criteria specified in the procedure, use the text illustrated as 14.1.1.1 (Note 18):

14.1.1 **Precision and Bias**

14.1.1.1 No justifiable statements can be made either on the precision or on the bias of Method D 0000 for measuring (insert here the name of the property) since the test result merely states whether there is conformance to the criteria for success specified in the procedure. (This statement should only be used after determining a statement based on counts of success or failure over a material range performed in different laboratories is not a viable means of arriving at test method precision.)

RECOMMENDED TEXT 6—ANOTHER METHOD

15. Precision and Bias Based on Another Method

15.1 When a method specifies that the procedures in another ASTM method are to be used without modification, do not write statements on precision and bias since those in the other method are applicable. When a method specifies that the procedures in another ASTM method are to be used with significant revisions, write statements on precision and bias as directed in this practice. When a method specifies that the procedures in another ASTM method are to be used with only insignificant modification(s), use the text illustrated in 15.1.1.1 to assure the user that precision and bias are not affected by the modification(s) in the procedure (Note 18).

15.1.1 **Precision and Bias**

15.1.1.1 The precision and bias of Method D 0000 for measuring (insert here the name of the property) are as specified in Method (insert here the designation of the other method).

RECOMMENDED TEXT 7—PRECISION NOT ESTABLISHED

16. Statements When Precision Not Established

16.1 For a properly developed new test method, a statement on within-laboratory precision should be possible based on evaluations made in at least one or two laboratories, including a ruggedness test. If the responsible subcommittee decides an interlaboratory study should be delayed until after the method has been published, a temporary statement addressing only within-laboratory precision is permitted. In the event that no information is available on within-laboratory precision, the responsible subcommittee may use a temporary statement like that illustrated as 16.1.1.1 and an appropriate statement on bias as directed in 11.1.3 (Note 18):

16.1.1 **Precision and Bias**

16.1.1.1 *Precision*—The precision of the procedure in Method D 0000 for measuring (insert here the name of the property) is being established.

16.1.1.2 *Bias*—(Insert a statement as directed in 11.1.3.).

16.2 Such temporary statements imply a promise to obtain more complete information about precision and shall not appear in a method for more than five years. (This statement should only be used when the responsible subcommittee has found it not possible to determine single-operator, single-laboratory precision, for a standard reason.)

RECOMMENDED TEXT 8—SPECIAL CASES OF RATINGS

17. Statements Based on Special Cases of Ratings

17.1 In the case of arbitrary grades or classifications and of scores for ranked data, the observations may have such a complex nonlinear relationship that meaningful transformations may not be practicable. If this is so, use the text illustrated as 17.1.1.1 and 17.1.1.2 as a guide in giving a subjective basis for evaluating the precision of test results (Note 18):

17.1.1 **Precision and Bias**

17.1.1.1 *Interlaboratory Test Data*[4]—An interlaboratory test was run in 19XX in which randomly drawn samples of two materials were tested in each of five laboratories. Each laboratory used two operators, each of whom tested four specimens of each material. Calculation of components of variance was thought to be inappropriate due to the restricted and discontinuous rating scales, the non-linear relationships between the rating scales and color difference units, and the increased variability in color difference units as the true value of the ratings decrease.

17.1.1.2 *Precision*—Based on the observations, described in 17.1.1.1 and on general practice in the trade, a lot or consignments is generally considered as having a rating that is significantly worse than a specified value when a specimen from the lot or consignment has a rating for (insert here the name of the property) that is more than one-half step below the specified rating on the AATCC Gray Scale for Color Change.

17.1.1.3 *Bias*—The true value of (insert name of property) can only be defined in terms of a test method. Within this limitation, Method D 0000 has no known bias.

(*Editorial Comment*—Where applicable, AATCC Evaluation Procedure 1, Gray Scale For Color Change, should be included in the list of applicable documents. (See also 11.1.1 and 11.1.3.))

18. Keywords

18.1 bias; precision; statistics; writing statements

REFERENCES

(**1**) Anderson, R. L. and Bancroft, T. A., *Statistical Theory in Research*, McGraw-Hill Book Co., Inc., New York, NY, 1952.

(**2**) Brownlee, K. A., *Industrial Experimentation*, Chemical Publishing Co., Brooklyn, NY, 1953.

(**3**) Davies, O. L., *The Design and Analysis of Industrial Experiments*, Hafner Publishing Co., New York, NY, 1956.

(**4**) Dixon, W. J. and Massey, F. J., Jr., *Introduction to Statistical Analysis*, McGraw-Hill Book Co., Inc., New York, NY, 1957.

(**5**) Ostle, Bernard, *Statistical Research*, Iowa State University Press, Ames, Iowa, 1963.

(**6**) Johnson, N. L. and Kotz, Samuel, *Distributions in Statistics, Discrete Distributions*, Houghton Mifflin Publishing Co., New York, NY, 1969, available through John Wiley& Sons, New York, NY.

(**7**) Pearson, E. S. and Hartley, H. O., *Biometrika Tables for Statisticians, Vol 1*, Cambridge University Press, Cambridge, England, 1954.

(**8**) Bartlett, M. S., "The Use of Transformations," *Biometrics*, March 1947, Vol 3, pp. 39–52.

(**9**) Hald, A., *Statistical Theory with Engineering Applications*, John Wiley & Sons, Inc., New York, NY; Chapman & Hall, Ltd., London, 1952.

(**10**) Snedecor, G. W., *Statistical Methods*, Iowa State College Press, Ames, Iowa, 1946.