Designation: F 1082 – 00 (Reapproved 2000)

Standard Practice for Tires—Determining Precision for Test Method Standards[1]

This standard is issued under the fixed designation F 1082; 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

Knowing the precision (and where possible the bias and accuracy, or both) of test measurements is vital for efficient technical decision making in any area of technology. For many years the chemical and allied material industries have addressed the issue of test precision, especially as it applies to inter-laboratory testing. Test method precision is important in the tire industry as well.

Some of the specific details that are important in laboratory testing frequently do not apply when objects such as tires are tested, especially when tested for various performance features at proving grounds or with other outdoor test methods. However, the basic methodology of "within" and "between" laboratory test precision assessment can be applied to tire testing provided the unique characteristics of some tire tests are kept in mind. This practice gives broad guidelines for tire test precision assessment.

When special test requirements arise that differ from the more orthodox precision methodology, they will have to be addressed in a special "ad hoc" manner. As experience is gained with these "special cases", the procedures for handling them can be formalized and incorporated into this practice.

1. Scope

1.1 This practice presents guidelines for preparing clear and meaningful precision statements for test method standards on tires and related objects pertinent to the tire industry and within the scope of ASTM Committee F-9. It gives definitions, explains the potential use of precision for standard test methods and gives the requirements for interlaboratory or inter-test-site programs. The calculation algorithms for determining precision and the format for expressing precision are also given.

2. Referenced Documents

2.1 *ASTM Standards:*

E 691 Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method[2]

F 538 Terminology Relating to the Characteristics and Performance of Tires[3]

3. Terminology

3.1 *Definitions of Terms*

3.1.1 *accuracy*, *n*—a measurement concept that describes the degree of correspondence between an average measured value and an accepted reference or standard value for the object, material or phenomenon under test.

3.1.1.1 *Discussion*—The reference value may be established by theory, by reference to an *accepted* standard, to another test method, or in some cases the average that could be obtained by applying the test method to all of the sampling units comprising a lot.

3.1.2 *bias*, *n*—the difference between the average measured test result and the accepted reference value; it measures in an inverse manner the accuracy of a test.

3.1.2.1 *Discussion*—A large bias implies poor accuracy, and a small or negligible bias denotes a high accuracy; when bias exists, increased testing does not increase accuracy, but merely gives an increased confidence in the bias estimate.

3.1.3 *determination*, *n*—the application of the complete measurement procedure to one piece, specimen or object to produce *one* numerical measured value to be used to form an average or median.

3.1.4 *precision*, *n*—a measurement concept that expresses the ability to generate test results that *agree with each other* in absolute magnitude.

3.1.4.1 *Discussion*—The degree of agreement is normally measured inversely by the standard deviation, high precision corresponds to a low (small) standard deviation. High precision may exist simultaneously with a large bias or poor accuracy.

3.1.5 *repeatability,r*, *n*—an established value, below which

[1] This practice is under the jurisdiction of ASTM Committee F09 on Tires and is the direct responsibility of Subcommittee F09.10 on Equipment, Facilities, and Calibration.

Current edition approved May 10, 2000. Published July 2000. Originally published as F 1082 – 87. Last previous edition F 1082 – 93.

[2] *Annual Book of ASTM Standards*, Vol 14.02.

[3] *Annual Book of ASTM Standards*, Vol 09.02.

the absolute difference between two "within-laboratory" or "within test-site" test results may be expected to lie, with a specified probability.

3.1.5.1 *Discussion*—The two test results are obtained with the *same* method on nominally identical test materials under the *same* conditions (same operator, apparatus, laboratory, location, and specified time period), and in the absence of other indications, the specified probability is 0.95 (sometimes written as 95 %). The "established value" also may be called a "critical difference."

3.1.6 *repeatability, relative (r), n*—a repeatability estimate expressed as a percentage of the average of the property for which the estimate was obtained.

3.1.6.1 *Discussion*—It is often appropriate to express repeatability on a relative basis, as a percent of a mean value. This form is similar to a coefficient of variation. Such expression is useful when *r* varies with the average level of the property being measured. Relative values for *r* cannot be unambiguously expressed as percentage (%) alongside the actual measured values in usual test result units because some test methods have "percent" as their unit. To avoid this ambiguity, the symbol (*r*) is used.

3.1.7 *reproducibility, R, n*—an established value, below which the absolute difference between two" between-laboratory" or "between test-site" test results may be expected to lie, with a specified probability.

3.1.7.1 *Discussion*—The two test results are obtained with the *same* method on nominally identical test materials under *different* conditions (different laboratories, locations, operators, apparatus and in a specified time period), and in the absence of other indications, the specified probability is 0.95. The essential characteristic of reproducibility is the variability of the *different* laboratories or test sites in which the testing is conducted.

3.1.8 *reproducibility, relative (R), n*— a reproducibility estimate expressed as percentage of the average of the property for which the estimate was obtained.

3.1.8.1 *Discussion*—It is often appropriate to express reproducibility on a relative basis, as a percent of a mean value. This form is similar to a coefficient of variation. Such expression is useful when *R* varies with the average level of the property being measured. Relative values for *R* cannot be unambiguously expressed as percentages (%) alongside the actual measured values in usual test result units because some test methods have "percent" as their units. To avoid this ambiguity, the symbol (*R*) is used.

3.1.9 *test result, n*—the average or median of a specified number of determinations; it is the reported value for a test.

4. Significance and Use

4.1 This practice applies to the following test method standards:

4.1.1 Those that have test results expressed in terms of a quantitative continuous variable.

4.1.2 Those that are fully developed and are in routine use as Committee F-9 test method standards.

4.2 General theory is included to better understand the basis of precision calculations (See Section 7 and Annex A1, Annex A2, and Annex A4). For those who have a familiarity with this theoretical basis and for those engaged in frequent precision calculations, the computational formulas in Annex A4 will prove helpful.

5. General Principles

5.1 Although detailed definitions for repeatability and reproducibility are given later in this practice, a few words of general discussion are merited at this point.

5.1.1 Repeatability refers to the ability of the *same* laboratory or testing apparatus to obtain similar (test) results under certain specified conditions. Reproducibility refers to the ability of *different* laboratories or testing apparatus in different locations to obtain similar test results under certain specified conditions. If test results closely agree, then good repeatability or good reproducibility exists.

5.2 The precision of a test method does not of necessity characterize a test with regard to how sensitive it is in measuring the basic property it is intended to measure. Precision may be good simply because the test method is insensitive to the basic property it measures. A concept called" test sensitivity" has been defined in the statistical literature as the ratio of the responsiveness of the test measurement to finite variations in the basic property in question to the precision of the measurement. This practice does not address this issue.

5.3 Both repeatability and reproducibility should be determined under realistic or typical laboratory or test site conditions. If extraordinary care is exercised in the laboratory, the precision statement may be overly optimistic.

5.3.1 The reported value of repeatability normally quoted will include the sum of the two components of variability. As ordinarily determined, repeatability has both a test apparatus variability and any test object variability that cannot be physically removed. Object variability that is not inherent in the overall operation of the test may be removed if an appropriate test program is conducted and a statement is included with the reported value of precision.

5.4 *Discussion of Repeatability (Very Short, Short, Long Term)*:

5.4.1 There are at least three different viewpoints that have been expressed with regard to repeatability.

5.4.1.1 *View 1*—The smallest possible or "very short" time period is used to estimate the variation. The same material, apparatus and operator is used and repeat determinations are made within a period measured in minutes or at most within a period measured in hours.

5.4.1.2 *View 2*—A "short" time period is used for the repeated operations that produce test results. The same material and same operator (or set of operators) is employed but the time period for the repeat operations is most frequently measured in days.

5.4.1.3 *View 3*—A "long term" time period is used for the repeated operations that produce test results within a laboratory. This may be weeks or months. Although it may be possible to use the same material, different operators are often employed and due to the long-term nature certain other changes such as recalibration of the test apparatus may have taken place. These changed conditions produce increased variability.

5.4.2 The time period *must be specified* as each particular

test method standard is taken up for consideration.

6. Organizing a Precision Estimation Program

6.1 *Task Group*—A task group of qualified people should be organized to conduct the program; a chairman, a statistical expert and members well-experienced with the standard in question. The panel chairman should ensure that all instructions of the program are clearly communicated to all laboratories or test locations in the program.

6.2 *Laboratories, Test Sites, and Materials or Objects*:

6.2.1 The number of laboratories, test locations or sites should be determined. The number of test objects, each comprising a different level of the measured property, should be selected.

6.2.1.1 At least ten participating laboratories or test sites are recommended. A program that involves fewer than six may not lead to reliable estimates of the reproducibility of the test method.

6.2.2 The number and type of objects (materials) to be included will depend on the following:

6.2.2.1 The range of the property and how precision varies over that range,

6.2.2.2 The different types of objects to which the test method is applied,

6.2.2.3 The difficulty (expense) in performing the tests, and

6.2.2.4 The commercial or legal need for obtaining a reliable estimate of precision.

6.2.3 For each level or class of object an adequate quantity (sample) of homogeneous objects should be available for subdivision and distribution by random allocation to the participating laboratories. The term "objects" is used in a broad, generic sense. When the objects to be tested are not homogeneous, it is important to obtain or prepare the samples in a well-documented manner.

6.2.3.1 Since object or material variability is included when measuring test variability, objects with high inherent variation will cause the test to appear insensitive. High precision with large bias seems to frequently occur in destructive tire tests. It is desirable to start with a large batch of similar objects for each level, and then use techniques of sample preparation prescribed in the method being evaluated.

6.2.3.2 Extraordinary means may be necessary to obtain an adequate quantity (sample) of homogeneous objects or material. However, if extraordinary care is exercised in sample selection, the precision statement may be overly optimistic for (everyday) routine test method utilization.

6.2.4 An interlaboratory or test-site study should include at least three types of objects, each type having a different average value for the measured test parameter. For development of broadly applicable precision statements, five or more should be included. The supply of objects should include a reserve of 50 % beyond immediate requirements for possible later use in retesting in one or more laboratories. Some modifications in sample selection or preparation may be necessary to ensure that the supply of objects available is sufficient to cover the experiment and keep a stock in reserve.

6.2.5 At each level, *p* separate containers (the number of laboratories or sites) should be used where there is any danger of the objects changing or material deteriorating when the container has once been opened. Special instructions on storage and treatment should be prescribed.

6.3 *Actual Organization of the Tests*:

6.3.1 The interlaboratory test plan is shown in Fig. 1, a table that indicates the laboratories or locations, materials or objects and replicates. With q levels and n replicates, each participating laboratory or test site among the *p* total has to carry out qn tests. A decision is necessary (for each test standard) as to

Test Site or Laboratory	Class of Object or Level of Material			
	1	2	. . .	q
1				
2				
3				
⋮			y_1 ... yn	
p				

p laboratories, sites ($i \rightarrow p$)
q levels, class of objects ($j \rightarrow q$)
n replicates per cell ($k \rightarrow n$)
Y = test result (or determination)
Cell *(ij)* contains n_{ij} results Y_{ijk} ($k = 1, 2 \ldots n_{ij}$)
$\bar{y}_{ij}$ = average of n_{ij} replicates in cell (*ij*)

FIG. 1 Layout of Precision Program

whether a "replicate" is to be a "determination" or a "test result" as defined in this document. The performance of these tests should be organized and the operators instructed as follows:

6.3.2 All $q \cdot n$ tests should be performed by one and the same operator or operator set, using the same equipment throughout.

6.3.3 Each group of n tests belonging to one level must be performed under repeatability conditions, in a specified interval of time.

6.3.4 It is essential that a group of n tests under repeatability conditions be performed independently as if they were n tests on different materials.

6.3.5 The number of replicates n, must be specified. Each replicate may be *one* test result or *one* determination according to the requirements of the test method standard. Normally, n is two, but it may be larger.

6.4 *Instructions to Operators*—The operators should receive no instructions other than those contained in the standard should be asked to comment on the standard and state whether the instructions contained in it are sufficiently clear. All participating laboratories or test sites should report their test results to one more significant figure than is customary or prescribed in the Standard.

6.5 *Reporting the Test Results*—Each laboratory or test site supervisor should write a full report containing the following particulars:

6.5.1 The final test results, (avoid transcription and typing errors).

6.5.2 The original individual observations or determination values from which the final results were derived.

6.5.3 The date(s) on which the samples or objects were received and the date(s) and time(s) on which they were tested.

6.5.4 Comments and information about irregularities or disturbances that may have occurred during the test.

6.5.5 Information about the equipment used, and other relevant information.

7. Analysis of Interlaboratory Program Test Data

7.1 *General Comments*— Two tasks are required for interlaboratory precision data analysis. The data should be put into table form as shown in Fig. 1. The second task is the formal analysis.

7.2 *Statistical Model for Precision Analysis*—The statistical model is given in Annex A1. Consult this for the necessary background concepts.

7.3 *Analysis of Data*— There are three successive stages:

7.3.1 A critical examination of the data in order to identify and treat outliers or other irregularities;

7.3.2 Computation of preliminary values of r and R for each level separately; and

7.3.3 Establishment of final values of r and R including the establishment of a relation between r, R, and M (if one exists) when the analysis indicates that they depend on the level average value, M. If r or R, or both, are judged to be independent of M, the final values taken are the simple average over all the levels.

7.4 *Cells*—Each combination of a laboratory or a test site or location, and a level, is called a cell. The test results of a program with p laboratories and q levels will consist of a table with pq cells each containing n replicate results.

7.5 *Redundant and Missing Data*:

7.5.1 Sometimes more than the n replicates will be measured. In that case report all results, why this was done and which are the correct test results. If the answer is that they are all equally valid, they can all be taken into account by using the computational procedure of Annex A4.

7.5.2 Some of the test results may be missing. The analysis recommended is such that completely empty cells simply can be ignored, while partly empty cells can be taken into account by the computational procedure of A4.3. The reasons for mixing test results should be given in the supervisor's report.

7.6 *Outliers*:

7.6.1 Outliers are entries among the original test results, that deviate so much from comparable entries in the same table that they are considered as irreconcilable with the other data. Outliers cannot always be avoided and have to be taken into consideration, but great care must be exercised in investigating them.

7.6.2 There are two types of outliers, "cell variance" outliers or "cell average" outliers. For the examination of variance outliers, Cochran's maximum variance test is used as described in Annex A2. For cell average outliers, Dixon's outlier test is used. The procedure for using Dixon's Test is given in Annex A3. The following probability levels (risk level) for statistical significance of outliers are used.

7.6.2.1 $P > 5\ \%$—That is, Cochran's or Dixon's test statistic is less than its 5 % critical value. The value is accepted as correct; the test is said to be statistically insignificant.

7.6.2.2 $5\ \% > P > 1\ \%$— That is, the test statistic lies between its 5 % and 1 % critical values. The item tested or value is called a straggler and is marked with a single asterisk; the test is said to be statistically significant.

7.6.2.3 $P < 1\ \%$—That is, the test statistic is greater than its 1 % critical value. The item or value is called a statistical outlier and is marked with a double asterisk; the test is said to be statistically highly significant.

7.6.2.4 P— P is the probability of the observed value of the test statistic.

7.6.2.5 The 5 % and 1 % critical values for Cochran's and Dixon's tests are given in Annex A2 and Annex A3.

7.6.3 If the outliers can be explained by some technical, computational, or clerical error the item or value is discarded. If discordant data entries are found to be outliers solely on the basis of significance in Cochran's, or Dixon's tests, they are to be given serious consideration for elimination from the data base.

7.6.4 When several unexplained stragglers or statistical outliers occur at different levels within the same laboratory or site, that laboratory or site may be considered as an outlier, having too high a within-laboratory or site variance, or too large a systematic error in the level of its test results, or both. It may then be reasonable to discard some or all the data from such an outlying laboratory or site.

7.7 *Computation of M, r, and R*:

7.7.1 The method of analysis carrying out the computation of M, r, and R for each level separately. Subsequently, it is investigated whether r or R depend on M, and if so, what is the

functional relationship.

7.7.2 The basic data needed for the computations are presented in three tables:

7.7.2.1 Table 1—Table 1 contains the original test results,

7.7.2.2 Table 2—Table 2 contains the measures of within-cell variation, and

7.7.2.3 Table 3—Table 3 contains the cell averages.

7.7.3 Table 1, the original test results, is constructed according to the format given in 7.7.4. Table 1 is equivalent to Fig. 1.

7.7.4 Table 2, the within-cell variance or standard deviation is constructed with the entries in Table 2 derived from Table 1 as follows:

7.7.4.1 For the general case, use the intra-cell standard deviation s_{ij}, given by Eq 1:

$$s_{ij} = \sqrt{\frac{1}{(n_y - 1)} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij})^2} \tag{1}$$

$$= \sqrt{\frac{1}{(n_{ij} - 1)} \left\{ \sum_{k=1}^{n_{ij}} (y^2_{ijk}) - \frac{1}{n_{ij}} \left[\sum_{k=1}^{n_{ij}} y_{ijk} \right]^2 \right\}}$$

The standard deviation should be expressed with one more significant figure than the results in Table 1.

7.7.4.2 For the particular case where all $n_{ij} = n = 2$, the cell range W_{ij}, may be used:

$$w_{ij} = |y_{ij}^{\ 1} - y_{ij}^{\ 2}| = s\sqrt{2} \tag{2}$$

7.7.4.3 The cell averages are shown in Table 3.

7.8 *Repeatability Variance* s^2_r *and Between-Laboratory Variance* s^2_L:

7.8.1 For a given level *j*, the values of s^2_r and s^2_L are given by the following universally applicable equations where, for convenience, the index *j* has been dropped.

$$s_r^2 = \frac{\sum_{i=1}^{p} (n_i - 1)\, s_i^2}{\left[\sum_{i=1}^{p} n_i\right] - p} \tag{3}$$

$$s_L^2 = \frac{\left(\frac{1}{p-1}\right) \sum_{i=1}^{p} n_i (\bar{y}_i - \bar{\bar{y}})^2 - s_i^2}{\bar{n}} \tag{4}$$

with

TABLE 1 Original Test Results[A]
Uniform-Level Experiment

Laboratory \ Level	1	2		*j*		*q*
1						
2						
i				y_{ij1} ... y_{ijk}		
p						

[A]The following notation is used:

(*a*) Laboratories or sites, there are *p* as a total
$L_i (i = 1,2 \ldots p)$

(*b*) Materials or levels or objects, there are *q* as a total
$M_j (j = 1,2 \ldots q)$

(*c*) Replicates, there are *n* as a total in each case of L_iM_j combination. There are normally an equal number of *n* values (usually 2) in each cell.

(*d*) y_{ijk} is a single test result value.

Example—Cell (*i, j*) contains n_{ij} results $y_{ijk} (K = 1,2, \ldots n_{ij})$.

TABLE 2 Cell Variance or Standard Deviation[A]
Uniform-Level Experiment

Laboratory \ Level	1	2	*j*		*q*
1					
2					
i			s_{ij} or w_{ij}		
p					

[A]Symbols are defined as follows:
s_{ij} = cell standard deviation, or if $n = 2$ for all the cells.
w_{ij} = cell range.

TABLE 3 Cell Averages[A]

Laboratory \ Level	1	2		*j*		*q*
1						
2						
i				$\bar{y}_{ij}$		
p						

[A] $\bar{y}_{ij}$ = cell average.

$$\bar{\bar{y}} = \frac{\sum_{i=1}^{p} n_i \bar{y}_i}{\sum_{i=1}^{p} n_i} \tag{5}$$

and:

$$\bar{\bar{n}} = \frac{1}{(p-1)} \left[\sum_{i=1}^{p} n_i - \frac{\sum_{i=1}^{p} n_i^2}{\sum_{i=1}^{p} n_i} \right] \tag{6}$$

7.8.2 For the case where all $n_i = n$, the previous formulae simplify to the following:

$$s_r^2 = \frac{1}{p} \sum_{i=1}^{p} s_i^2 \tag{7}$$

$$s_L^2 = \frac{1}{(p-1)} \sum_{i=1}^{p} (\bar{y}_i = \bar{\bar{y}})^2 - \frac{s_r^2}{n} \tag{8}$$

with:

$$\bar{\bar{y}} = \frac{1}{p} \sum_{i=1}^{p} \bar{y}_i \tag{9}$$

For the particular case where all $n_i = n = 2$, the cell range $w_i = \sqrt{2}\, s_i$ may be used, giving:

$$s_r^2 = \frac{1}{2p} \sum_{i=1}^{p} w_i^2 \tag{10}$$

$$s_L^2 = \frac{1}{(p-1)} \sum_{i=1}^{p} (\bar{y}_i - \bar{\bar{y}})^2 - \frac{s_r^2}{2} \tag{11}$$

7.8.3 To simplify calculations, formulas are given in Annex A4 in terms of sums or totals developed from Table 1, Table 2, and Table 3 for each level *M*, in the tables. The results of the calculations should be recorded in a table as described in Section 8.

8. Format for Precision Section (Clause) of Standards

8.1 *General*—The results of the formal analysis shall be contained in a specific section or clause of the test method

standard with the heading "Precision and Bias."

8.2 *Introductory Subclause*—This shall consist of one or more paragraphs that give the pertinent details of the precision program. Following this one or more tables of results that give the actual precision parameters are presented. These introductory paragraphs should answer the following questions:

8.2.1 What is the time period for repeatability, reproducibility: short term (define), long term (define)?

8.2.2 What is a test result? How many determinations? Average or median?

8.2.3 How many laboratories or test sites participated (*p*)?

8.2.4 How many materials, levels or object types (*q*)?

8.2.5 How many replicates (*n*)? What is a replicate?

8.2.6 At what time was the precision program conducted (month, year)?

8.2.7 Are there any unusual results that the reader should be aware of?

8.2.8 How do *r* and *R* vary as the mean level of the measured property varies?

8.2.9 Can these variations be described by a simple mathematical relationship?

8.3 *Table of Precision Parameters*—A table with the general format of Table 4 should be prepared. This should include the following information:

8.3.1 ASTM test method designation;

8.3.2 Measured property, time period used for *r* and *R*;

8.3.3 Materials, with mean level and units of measurement; *r*, (*r*), *R*, (*R*) and for completeness of record the within and between laboratory or test site standard deviation, s_r, and S_R. An example is given in 8.4.

8.4 *Statements for Precision*:

8.4.1 Statements or recommendations for use and interpretation of statistical parameters *r* and *R* are as follows.

8.4.1.1 The *difference* between two single test results (or determinations) found on identical test material or test objects under the repeatability conditions prescribed for a particular test, will exceed the *repeatability* on average not more than once in 20 cases in the normal and correct operation of method.

8.4.1.2 The *difference* between two single and independent test results found by two operators working under the prescribed reproducibility conditions in different laboratories or at different test sites on identical test material or objects will exceed the *reproducibility* on average not more than once in 20 cases in the normal and correct operation of method.

8.4.1.3 These two statements apply to a particular mean level in a precision table (as per Table 4).

8.4.2 Alternatively, statements of the following form may be prepared for use in the Precision clause of any test method standard.

8.4.2.1 *Repeatability*— The repeatability of test *xxxx* has been established as *xxxx*. Two single test results that differ by more than *xxxx* (expressed in appropriate terms), must be considered suspect, that is, to have come from different sample populations. Such a decision dictates that some appropriate action be taken.

NOTE 1—Appropriate action may be an investigation of the test method procedure or apparatus for faulty operation or the declaration of a significant difference in the two materials, samples, etc., that generated the two test results.

8.4.2.2 *Reproducibility*— The reproducibility of test *xxxx* has been established as *xxxx*. Two single test results produced in separate laboratories or test sites, that differ by more than *xxxx*, (expressed in appropriate terms) must be considered as suspect, that is, that they represent sample populations. Such a decision dictates that appropriate investigative or technical/commercial actions be taken.

8.4.2.3 These two statements apply to particular mean levels as they appear in a precision table unless precision does not vary with mean level in which case they apply across the entire range of mean level values *or* unless the repeatability or reproducibility are expressed on a relative basis that is, (*r*) or (*R*) and the relative precision does not vary with mean level.

8.4.3 *Bias Statement*— For most test methods bias cannot be determined. The following statement is recommended.

8.4.3.1 *Bias*—In test method terminology, bias is the difference between an average test value and the reference (true) test property value. Reference values do not exist for this test method since the value or level of the test property is exclusively defined by the test method. Bias therefore cannot be determined.

8.4.3.2 For those test methods where bias can be determined a statement as to its magnitude should be included.

TABLE 4 Example—ASTM XXXX—Precision (Measured Property = XXXX)[A]

NOTE 1—p = xx, q = 4, n = 2.

NOTE 2—Pooled or average values for all tabulated parameters may be given if appropriate.

Class of Object or Material	Mean Level	Within Labs or Sites			Between Labs or Sites		
		s_r	r	(r)	S_R	R	(R)
A	XX	X	X	X	X	X	X
B	XX	X	X	X	X	X	X
C	XX	X	X	X	X	X	X
D	XX	X	X	X	X	X	X
Pooled or Average Values	XX	X	X	X	X	X	X

[A]Symbols are defined as follows:

s_r = within lab/site standard deviation.
S_R = standard deviation for total between lab/site variation.
r = repeatability (in measurement units).
(r) = repeatability (in percent).
R = reproducibility (in measurement units).
(R) = reproducibility (in percent).

9. Keywords

9.1 accuracy; bias; deviation; outliers; precision; repeatability; reproducibility; tires; variance

ANNEXES

(Mandatory Information)

A1. STATISTICAL MODEL FOR PRECISION

A1.1 *Model*—To estimate precision each individual test result is assumed to be the sum of three components.

$$y = M + B + e \quad \text{(A1.1)}$$

where y is the test result, M is the average of all measurements (defined in a specific way), B is a term representing the differences between the participating laboratories or test sites, and e is a random error occurring in every test.

A1.2 *Average, M*—The average M is called the level of the property, different materials or objects in a series will have different levels or averages, M.

A1.2.1 The concept of a reference value μ is important. This has frequently been referred to in statistical applications as the *true* value. However modern terminology uses the word *reference* in place of true. The bias is the difference $(\mu - M)$, where M is an average well established by repeated testing.

A1.2.2 Reference values may not exist or not be determinable. In such cases the level of the test property is exclusively defined by the test method and the concept of an independent reference value does not apply. Bias therefore cannot be determined.

A1.3 *Term, B*:

A1.3.1 This term is considered to be constant during any series of tests performed under repeatability conditions, such as, in a given laboratory, but to behave as a random variable in a series of tests performed, in many laboratories under reproducibility conditions, or a non-random variable associated with a finite population of N members, test sites or laboratories, all N being included in the test program. The distribution of this variable is assumed to be approximately normal but in practice, it is sufficient that it is unimodal. Its variance will be denoted as follows:

$$\text{var(B)} = \sigma_L^2 = \text{the between-laboratory or between-test site variance} \quad \text{(A1.2)}$$

A1.3.2 The variance $\sigma^2{}_L$ includes the between-operator and between-equipment variabilities but does not contain the day-to-day or within laboratory or test site replication source of variation.

A1.3.3 In general, B can be considered as the sum of a random component B_o and a systematic component B_s, as follows:

$$B = B_o + B_s \quad \text{(A1.3)}$$

A1.3.4 In a single laboratory or at a single test site, conditions cannot be kept completely constant in the long run. Hence, long-term variabilities will exist larger than those accounted for by the repeatability. These long term variations will contribute to a random component B_o.

A1.3.5 Permanent systematic differences between laboratories or sites may exist. They should be investigated and corrected. These may be due to the use of different measuring instruments and operator technique or different environmental factors. These will all contribute a systematic component B_s.

A1.3.6 If there are N laboratories or sites likely to use the method at any time, B_s will take only N discrete values and the term B in the model can only be considered as a random variable if either the systematic differences B_s are so small that they can be ignored, or else if the test results from which the reproducibility criterion is obtained were carried out by laboratories or sites that can be considered as selected at random from all those likely to use the method. Therefore, caution is needed when the test results to be compared are always performed by the same two laboratories or sites.

A1.4 *Error Term, e*:

A1.4.1 This term represents a random error occurring in every single test result. The distribution of e is assumed to be approximately normal but, in practice, it is sufficient that it be unimodal. Its variance is

$$\text{var(e)} = \sigma_w^2 = \text{the within-laboratory or within test-site variance} \quad \text{(A1.4)}$$

A1.4.2 This variance will vary between laboratories or sites due to differences in the skills of the operators, in the quality of the equipment used and other factors. However, when a test method has been properly standardized, the differences in the value of $\sigma^2{}_w$ between laboratories or sites should besmall so that it is justifiable to establish a common value for the within-laboratory or within-site variance valid for all laboratories and sites using the standard method.

A1.4.3 This common value, which is an average of the variances taken over the laboratories or sites participating in the precision experiment, will be called the repeatability variance and be designated by

$$\text{var(e)} = \sigma_r^2 = \text{repeatibility variance} \quad \text{(A1.5)}$$

A1.5 *Total Between Site or Laboratory Variability*:

A1.5.1 The (variance of) total between laboratory or site variability (for single test results) is comprised of two components, within laboratory/site variance and a between-laboratory/site variance. Thus the term $S^2{}_R$ the reproducibility variance, is defined as follows:

$$(S_R)^2 = (S_r)^2 + (S_L)^2 \quad \text{(A1.6)}$$

$$S_R = \sqrt{(S_r)^2 + (S_L)^2} \quad \text{(A1.7)}$$

A1.5.2 The difference between S_R and S_L is as follows: S_R (as defined above) represents total variability of measured test results among (between) typical laboratories or sites; S_L represents the variability of averages or means among typical laboratories or sites on the basis that there is no within laboratory/site variability. It is in a sense, an artificial statistical

parameter that is derived directly from the unique character of the analysis procedures used to calculate precision. Normally the difference between S_L and S_R is small provided that S_r is small in comparison to S_L.

A1.6 *Relation Between the Model, r and R*:

A1.6.1 On a theoretical basis repeatability, *r*, and the reproducibility, *R*, are given as follows:

$$r = f\sqrt{2}\cdot\sigma_r \quad \text{(A1.8)}$$

$$R = f\sqrt{s}\sqrt{\sigma_L^2 + \sigma_r^2} = f\sqrt{2}\cdot\sigma_R \quad \text{(A1.9)}$$

where ($\delta^2_R = \sigma^2_L + \sigma^2_r$) is called the reproducibility variance.

A1.6.2 The coefficient $\sqrt{2}$ is derived from the fact that *r* and *R* refer to the difference between two single test results, and *f* is equivalent to Student's *t* when used with equations (Eq A1.8) and (Eq A1.9), is a factor whose value depends both on the number of test results available for estimating the variances σ^2_r and σ^2_R, and on the shape of the distributions of the random components *B* and *e* in the model. If these distributions are approximately normal (in practice unimodal), the number of test results is not too small, and if the probability level is 95 %, the factor *f* (or *t*) will never differ much from the value 2 and the use of this value throughout is therefore recommended in this practice.

Therefore:

$$r = 2.83\,\sigma_r = 2.83\,S_r \quad \text{(A1.10)}$$

$$R = 2.83\,\sigma_r = 2.83\,S_R \quad \text{(A1.11)}$$

A1.6.3 As the values of repeatability variance (σ^2_r) and the reproducibility variance (σ^2_R) are not exactly known, their estimates of s^2_r and s^2_R are therefore used instead.

A2. COCHRAN'S MAXIMUM VARIANCE TEST

A2.1 This practice assumes that between laboratories, only small differences exist in the within-laboratory variance. As experience shows that this condition is not always satisfied, a test has been included to investigate the validity of this assumption.

A2.1.1 Given a set of *p* standard deviations s_i, all computed from the same number *n* of replicate test results, Cochran's criterion *C* is given by the following:

$$C = \frac{s^2\,\max}{\sum_{i=1}^{p} s_i^2} \quad \text{(A2.1)}$$

A2.1.2 In this expression, s^2_{max}stands for the highest (variance) value in the set. If the test is significant, $s^2_{max}$is classified as straggler or statistical outlier according to the procedure of 7.7. Critical values for Cochran's criterion at the 5 % and 1 % levels are given for $p = 2$ to 40 and $n = 2$ to 6 in Table A2.1.

A2.1.3 Cochran's criterion must be applied to Table A2.1 at each level separately.

A2.2 As previously stated, Cochran's criterion applies strictly only when all standard deviations are derived for the same number of test results obtained under conditions of repeatability. In actual cases, this number may vary due to redundant, missing or discarded data. In a properly organized experiment, such variations in the number of test results per cell will be limited and can be ignored, Cochran's criterion being applied using for *n* the number of results occurring in the majority of cells.

A2.3 Cochran's criterion tests only the highest value in a set of standard deviations and is therefore a one-sided outlier test. Variance heterogeneity may, of course, also manifest itself in some of the standard deviations being comparatively too low. Small values of standard deviation may be very strongly influenced by the degree of rounding of the original test results and are for that reason not very reliable. It does not seem appropriate to reject the data of some laboratory because of its higher precision. Hence, Cochran's criterion is considered adequate.

TABLE A2.1 Critical values for Cochran's maximum variance test

p^A	$n^B = 2$		$n^B = 3$		$n^B = 4$		$n^B = 5$		$n^B = 6$	
	1 %	5 %	1 %	5 %	1 %	5 %	1 %	5 %	1 %	5 %
2	...	...	0.995	0.975	0.979	0.939	0.959	0.906	0.937	0.877
3	0.993	0.967	0.942	0.871	0.883	0.798	0.834	0.746	0.793	0.707
4	0.968	0.906	0.864	0.768	0.781	0.684	0.721	0.629	0.676	0.590
5	0.928	0.841	0.788	0.684	0.696	0.598	0.633	0.544	0.588	0.506
6	0.883	0.781	0.722	0.616	0.626	0.532	0.564	0.480	0.520	0.445
7	0.838	0.727	0.664	0.561	0.568	0.480	0.508	0.431	0.466	0.397
8	0.794	0.680	0.615	0.516	0.521	0.438	0.463	0.391	0.423	0.360
9	0.754	0.638	0.573	0.478	0.481	0.403	0.425	0.358	0.387	0.329
10	0.718	0.602	0.536	0.445	0.447	0.373	0.393	0.331	0.357	0.303
11	0.684	0.570	0.504	0.417	0.418	0.348	0.366	0.308	0.332	0.281
12	0.653	0.541	0.475	0.392	0.392	0.326	0.343	0.288	0.310	0.262
13	0.624	0.515	0.450	0.371	0.369	0.307	0.322	0.271	0.291	0.246
14	0.599	0.492	0.427	0.352	0.349	0.291	0.304	0.255	0.274	0.232
15	0.575	0.471	0.407	0.335	0.332	0.276	0.288	0.242	0.259	0.220
16	0.553	0.452	0.388	0.319	0.316	0.262	0.274	0.230	0.246	0.208
17	0.532	0.434	0.372	0.305	0.301	0.250	0.261	0.219	0.234	0.198
18	0.514	0.418	0.356	0.293	0.288	0.240	0.249	0.209	0.223	0.189
19	0.496	0.403	0.343	0.281	0.276	0.230	0.238	0.200	0.214	0.181
20	0.480	0.389	0.330	0.270	0.265	0.220	0.229	0.192	0.205	0.174
21	0.465	0.377	0.318	0.261	0.255	0.212	0.220	0.185	0.197	0.167
22	0.450	0.365	0.307	0.252	0.246	0.204	0.212	0.178	0.189	0.160
23	0.437	0.354	0.297	0.243	0.238	0.197	0.204	0.172	0.182	0.155
24	0.425	0.343	0.287	0.235	0.230	0.191	0.197	0.166	0.176	0.149
25	0.413	0.334	0.278	0.228	0.222	0.185	0.190	0.160	0.170	0.144
26	0.402	0.325	0.270	0.221	0.215	0.179	0.184	0.155	0.164	0.140
27	0.391	0.316	0.262	0.215	0.209	0.173	0.179	0.150	0.159	0.135
28	0.382	0.308	0.255	0.209	0.202	0.168	0.173	0.146	0.154	0.131
29	0.372	0.300	0.248	0.203	0.196	0.164	0.168	0.142	0.150	0.127
30	0.363	0.293	0.241	0.198	0.191	0.159	0.164	0.138	0.145	0.124
31	0.355	0.286	0.235	0.193	0.186	0.155	0.159	0.134	0.141	0.120
32	0.347	0.280	0.229	0.188	0.181	0.151	0.155	0.131	0.138	0.117
33	0.339	0.273	0.224	0.184	0.177	0.147	0.151	0.127	0.134	0.114
34	0.332	0.267	0.218	0.179	0.172	0.144	0.147	0.124	0.131	0.111
35	0.325	0.262	0.213	0.175	0.168	0.140	0.144	0.121	0.127	0.108
36	0.318	0.256	0.208	0.172	0.165	0.137	0.140	0.117	0.124	0.106
37	0.312	0.251	0.204	0.168	0.161	0.134	0.137	0.116	0.121	0.103
38	0.306	0.246	0.200	0.164	0.157	0.131	0.134	0.113	0.119	0.101
39	0.300	0.242	0.196	0.161	0.154	0.129	0.131	0.111	0.116	0.099
40	0.294	0.237	0.192	0.158	0.151	0.126	0.128	0.108	0.114	0.097

[A] p = the number of laboratories at a given level.
[B] n = number of results per cell.

A3. DIXON's OUTLIER TEST

A3.1 Given a set of data *Z(h)* where $h = 1,2,3, ..., H$, arranged in order of magnitude, then Dixon's test uses the test statistics given in Table A3.1. Critical values of these test statistics at the 5 % and 1 % level and for $H = 3$ to 40 are reproduced in Table A3.2.

A3.2 In analyzing a precision experiment, Dixon's test can be applied as follows:

A3.2.1 To the test results within a cell of Table 1 when $n_{ij} \geq 3$, but this procedure should only be used where Cochran's test has suggested an outlier or a straggler, in order to see whether this was due solely to one observation; in that case $h = k$, $H = n_{ij}$, and $z(h) = y_{ijk}$, i and j both being fixed.

A3.2.2 To the cell averages for a given level j in Table 3, when in that case $h = i$, $H = p_j$, and $z(h) = \bar{y}_{ij}$, j being fixed.

A3.3 If Dixon's test reveals one of the extreme values in a series (the highest or the lowest) as a straggler or statistical outlier, the test should again be applied to the remaining $H - 1$ values; and if this once more proves one of the extremes as suspect, the test should be applied afresh to the remaining set of $H - 2$ values.

A3.4 However, caution should be exercised in drawing conclusions from the results of repeated applications of Dixon's test. If several stragglers or statistical outliers are found at only a single level, this may not be really significant, but if they occur at differing levels within a single laboratory, this may be considered as indicating that that is an outlying laboratory.

A3.4.1 The strategy in dealing with stragglers or statistical outliers in 7.7 should be followed in the case of Dixon's test.

TABLE A3.1 Dixon's Outlier Test Statistics

H	Test Statistic
3 to 7	Q_{10} = the larger of $\frac{z(2) - z(1)}{z(H) - z(1)}$ and $\frac{z(H) - z(H-1)}{z(H) - z(1)}$
8 to 12	Q_{11} = the larger of $\frac{z(2) - z(1)}{z(H-1) - z(1)}$ and $\frac{z(H) - z(H-1)}{z(H) - z(2)}$
13 or more	Q_{22} = the larger of $\frac{z(3) - z(1)}{z(H-2) - z(1)}$ and $\frac{z(H) - z(H-2)}{z(H) - z(3)}$

TABLE A3.2 Critical Values for Dixon's Outlier Test[A]

Test Criterion	*H*	Critical values	
		5 %	1 %
$Q_{10} = \frac{z(2) - z(1)}{z(H) - z(1)}$ or $\frac{z(H) - z(H-1)}{z(H) - z(1)}$ whichever is the greater	3	0.970	0.994
	4	0.829	0.926
	5	0.710	0.821
	6	0.628	0.740
	7	0.569	0.680
$Q_{11} = \frac{z(2) - z(1)}{z(H-1) - z(1)}$ or $\frac{z(H) - z(H-1)}{z(H) - z(2)}$ *whichever is the greater*	8	0.608	0.717
	9	0.504	0.672
	10	0.530	0.635
	11	0.502	0.605
	12	0.479	0.579
$Q_{22} = \frac{z(3) - z(1)}{z(H-2) - z(1)}$ or $\frac{z(H) - z(H-2)}{z(H) - z(3)}$ *whichever is the greater*	13	0.611	0.697
	14	0.586	0.670
	15	0.565	0.647
	16	0.546	0.627
	17	0.529	0.610
	18	0.514	0.594
	19	0.501	0.580
	20	0.489	0.567
	21	0.478	0.555
	22	0.468	0.544
	23	0.459	0.535
	24	0.451	0.526
	25	0.443	0.517
	26	0.436	0.510
	27	0.429	0.502
	28	0.423	0.495
	29	0.417	0.489
	30	0.412	0.483
	31	0.407	0.477
	32	0.402	0.472
	33	0.397	0.467
	34	0.393	0.462
	35	0.388	0.458
	36	0.384	0.454
	37	0.381	0.450
	38	0.377	0.446
	39	0.374	0.442
	40	0.371	0.438

[A] This is R. S. Gardner's version of Dixon's test. This version applies when it is not known at which end of a series of data an outlier may occur. $z(h)$, $h = 1, 2, ...$ H, is the series of data to be tested arranged in order of magnitude.

A4. COMPUTATIONAL FORMULA FOR *r* AND *R*

A4.1 *With n = 2 replicates per cell*:

$$T_1 = \Sigma\, \bar{y}_i \tag{A4.1}$$

$$T_2 = \Sigma\, (\bar{y})^2 \tag{A4.2}$$

$$T_3 = \Sigma\, (w_i)^2 \tag{A4.3}$$

$$T_4 = \Sigma\, (s_i)^2 \tag{A4.4}$$

NOTE A4.1—Use either T_3 or T_4

$$s_r^2 = \frac{T_3}{2p} = \frac{T_4}{p} \tag{A4.5}$$

$$s_L^2 = \left[\frac{pT_2 - T_i^2}{p(p-1)}\right] - \frac{s_r^2}{2} \tag{A4.6}$$

$$s_R^2 = s^2{}_L + s_r^2 \tag{A4.7}$$

$$m = T_1 / p \tag{A4.8}$$

$$r = 2.83 \sqrt{s_r^2} \tag{A4.9}$$

$$R = 2.83 \sqrt{s_R^2} \tag{A4.10}$$

NOTE A4.2—If $s^2{}_L$ is negative, substitute $s^2{}_L = 0$ in Eq A4.7.

A4.2 *With n > 2 (a constant value over all cells)*—The computational formulas are identical to A4.1 except that the value of *n* is used in place of 2 in the denominator of the second term of Eq A4.6. The value of $s^2{}_r$ is obtained by means of the T_4/p expression of Eq A4.5.

A4.3 *With unequal numbers of n replicates per cell*:

$$T_5 = \Sigma\, n_i \bar{y}_i \tag{A4.11}$$

$$T_6 = \Sigma\, n_i\, (\bar{y}_i)^2 \tag{A4.12}$$

$$T_7 = \Sigma\, n_i \tag{A4.13}$$

$$T_8 = \Sigma\, (n_i)^2 \tag{A4.14}$$

$$T_9 = \Sigma\, (n_i - 1)(s_i)^2 \tag{A4.15}$$

$$s_r^2 = \frac{T_9}{(T_7 - p)} \tag{A4.16}$$

$$S_L^2 = \left(\frac{T_6 T_7 - T_5^2}{T_7\,(p-1)} - S_r^2\right)\left(\frac{T_7\,(p-1)}{T_7^2 - T_8}\right) \tag{A4.17}$$

$$s_R^2 = s_L^2 + s_r^2 \tag{A4.18}$$

Calculate *M*, *r*, and *R* as per A4.1 or $M = T_5/T_7$.