INTERNATIONAL STANDARD

**ISO 19983**

Second edition
2022-06

# Rubber — Determination of precision of test methods

*Caoutchouc — Détermination de la fidélité des méthodes d'essai*



Reference number
ISO 19983:2022(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

# Contents

Page





# Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 45, *Rubber and rubber products*, Subcommittee SC 2, *Testing and analysis*.

This second edition cancels and replaces the first edition (ISO 19983:2017), which has been technically revised.

The main changes compared to the previous edition are as follows:

— detection and treatment of outliers have been explained in more detail in 6.8 and 6.9;

— a new Annex F has been added to provide an example of outlier treatment for method B.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.



# Introduction

The procedures used for several years by ISO/TC 45/SC 2 for estimating precision of test methods by means of interlaboratory tests (ISO/TR 9272) were closely related to ASTM D4483. ISO/TR 9272 was found to have serious flaws which users were using work-arounds to counteract. It was the desire of the ISO TC 45/SC2/WG4 members that ISO/TR 9272 be replaced with a new standard that included using ISO 5725 (all parts) with specific choices and variations of procedures to suit the particular requirements of rubbers.

This document provides two methods for determining the precision values of a test method:

— Method A based on ISO 5725 (all parts) to calculate repeatability, day-to-day repeatability, and reproducibility;

— Method B based on ASTM D4483 to calculate day-to-day repeatability and reproducibility.

# Rubber — Determination of precision of test methods

## 1 Scope

This document provides guidelines and specifies requirements for estimating the precision of rubber test methods by means of interlaboratory test programmes based on the procedures given in:

— Method A using ISO 5725 (all parts);

— Method B using ASTM D4483

## 2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 3534-1, *Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability*

ISO 3534-2, *Statistics — Vocabulary and symbols — Part 2: Applied statistics*

ISO 5725-1:1994, *Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions*

ISO 5725-2, *Accuracy (trueness and precision) of measurement methods and results — Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method*

ISO 5725-3, *Accuracy (trueness and precision) of measurement methods and results — Part 3: Intermediate measures of the precision of a standard measurement method*

ISO 5725-4, *Accuracy (trueness and precision) of measurement methods and results — Part 4: Basic methods for the determination of the trueness of a standard measurement method*

ISO 5725-5, *Accuracy (trueness and precision) of measurement methods and results — Part 5: Alternative methods for the determination of the precision of a standard measurement method*

ISO 5725-6, *Accuracy (trueness and precision) of measurement methods and results — Part 6: Use in practice of accuracy values*

ASTM D4483, *Standard Practice for Determining Precision for Test Method Standards in the Rubber and Carbon Black Industries*

## 3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 3534-1, ISO 3534-2, ISO 5725 (all parts), ASTM D4483, and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

**3.1**
**day-to-day repeatability**
precision under the conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment

Note 1 to entry: The time interval between repeated tests is normally between one and seven days.

**3.2**
**type 1 precision**
precision determined directly on a target material

Note 1 to entry: Prepared test pieces or test portions of the target material (class of elements) drawn from a homogeneous source are tested, with no processing or other operations required prior to testing.

**3.3**
**type 2 precision**
precision determined indirectly for a target material

Note 1 to entry: The target material is usually combined with a number of homogeneous ancillary materials to form a composite material and testing is conducted on samples of this and the property response of the target material is determined.

**3.4**
**pooled standard deviation**
square root of the average variance of a set of selected individual variances

Note 1 to entry: The pooled standard deviation, as well as the average variance, is intended as an overall or general descriptor of some set of variances and their standard deviations.



## 4 Symbols

| | |
|---|---|
| $D_{ij}$ | day-to-day effect, the day-to-day variance component of which is $\sigma_D^2$ |
| $h$ values | Mandel's between-laboratory consistency test statistic |
| $k$ values | Mandel's within-laboratory consistency test statistic |
| $L_i$ | between-laboratory effect, the between-laboratory variance component of which is $\sigma_L^2$ |
| $M_{ijk}$ | repeatability effect, the repeatability variance component of which is $\sigma_M^2$ |
| $n$ | number of measurements |
| $p$ | number of laboratories |
| $q$ | number of days |
| $r$ | repeatability |
| $r_{DA}$ | day-to-day repeatability as determined from method A calculations |
| $r_{DB}$ | day-to-day repeatability as determined from method B calculations |
| $R$ | reproducibility |
| $(r)$ | relative repeatability |
| $(r_{DA})$ | relative day-to-day repeatability as determined from method A calculations |
| $(r_{DB})$ | relative day-to-day repeatability as determined from method B calculations |
| $(R)$ | relative reproducibility |
| $s_M^2$ | repeatability variance |
| $s_{rD}^2$ | day-to-day repeatability variance as determined from method A calculations |
| $s_R^2$ | reproducibility variance |
| $s_D^2$ | day-to-day variance as determined from method B calculations |
| $s_L^2$ | between-laboratory variance |

| | |
|---|---|
| $s$ | standard deviation of data |
| $s_r$ | repeatability standard deviation |
| $s_{r\mathrm{D}}$ | day-to-day repeatability standard deviation as determined from method A calculations |
| $s_R$ | reproducibility standard deviation |
| $S_\mathrm{D}$ | day-to-day repeatability standard deviation as determined from method B calculations |
| $SS_\mathrm{T}$ | total sum of squares |
| $SS_\mathrm{L}$ | between-laboratory sum of squares |
| $SS_\mathrm{D}$ | day-to-day sum of squares |
| $SS_\mathrm{M}$ | repeatability sum of squares |
| $T$ | total sum of data |
| $V_\mathrm{L}$ | between-laboratory mean square |
| $V_\mathrm{D}$ | day-to-day mean square |
| $V_\mathrm{M}$ | repeatability mean square |
| $y_{ijk}$ | data<br>$i$, $j$, $k$: each data of laboratory, day, repeat |
| $\bar{y}$ | mean value of data |
| $\bar{\bar{y}}$ | mean value of $\bar{y}$ |
| $\phi_\mathrm{T}$ | total degree of freedom |
| $\phi_\mathrm{L}$ | between-laboratory degree of freedom |
| $\phi_\mathrm{D}$ | day-to-day degree of freedom |
| $\phi_\mathrm{M}$ | repeatability degree of freedom |
| $\mu$ | population mean |
| $\sigma_\mathrm{M}{}^2$ | repeatability variance component |
| $\sigma_\mathrm{D}{}^2$ | day-to-day variance component |
| $\sigma_\mathrm{L}{}^2$ | between-laboratory variance component |

NOTE The symbols $r_\mathrm{DB}$ and $(r_\mathrm{DB})$ in this document are the same as $r$ and $(r)$, respectively, in ASTM D4483.

# 5 Interlaboratory test programme

To evaluate precision for test method standards by means of interlaboratory test programmes (ITPs), use either one of the two methods:

a) Method A, where three precisions, namely the repeatability, the day-to-day repeatability and the reproducibility, are calculated in accordance with ISO 5725-3;

b) Method B, where two precisions, namely the day-to-day repeatability and the reproducibility, are calculated in accordance with ASTM D4483.

NOTE If two or more results are available from within-a-day repeated tests, method A is applicable to evaluate the variance of measurement errors.

# 6 Procedures

## 6.1 Application

A standard measurement method is taken to mean an established international test method for rubber.

A determination of the precision of a test method is normally conducted with a selected group of materials typical of those used with that method, and by a group of volunteer laboratories that have experience of the method.

Caution is necessary in applying precision results for a particular test method to product testing for commercial product accepted procedures. For this purpose, the precision estimates should be obtained from special programmes that are specific to the product in question and carried out by the interested laboratories.

## 6.2 Repeatability conditions

Repeatability conditions are where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time.

NOTE “Short interval of time” indicates that tests are repeated within a day, when the time needed to complete a test allows repeating the test within the same day.

"Identical test items" is interpreted as nominally identical, i.e. no intentional differences.

For rubbers, repeatability can be dependent on the magnitude or level of the measured property and is usually reported for each of several materials having particular property levels.

## 6.3 Day-to-day repeatability conditions

Day-to-day repeatability conditions are where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment.

The “intervals of time” between repeated measurements of test results may be selected by the consensus of a particular testing community. For ISO/TC 45 and the international rubber manufacturing industry, the time interval between repeat tests is of the order of one to seven days, but most commonly seven days. For special tests (long ageing periods), however, replicate tests can require a longer time span.

NOTE The “repeatability” traditionally used by ISO/TC 45/SC 2 is equivalent to the day-to-day repeatability defined in this document.

## 6.4 Reproducibility conditions

Reproducibility conditions are where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment.

"Identical test items" is interpreted as nominally identical, i.e. no intentional differences.

For ISO/TC 45, different equipment means apparatus that might have different manufacturers but complies with the requirements of the test standard in question, including calibration.

For rubbers, reproducibility might be dependent on the magnitude or level of the measured property and is usually reported for each of several materials having particular property levels.

## 6.5 Testing elements

The element that is tested is either a test piece or a test sample as defined in the test method standard. The test method standard will also define the number of test elements to be tested to obtain a result for the property.

## 6.6 Planning

Select either type 1 precision or type 2 precision as defined in 3.2 and 3.3.



It is possible that a type 1 precision programme can be conducted on test pieces or portions that require some minimum processing or other simple operations prior to actual testing.

Unless circumstances dictate otherwise, using type 1 precision is preferred.

For type 1 precision, the test pieces or test samples need to be produced from the same lot of material by the same procedures and then stored and conditioned in the same manner, in order to be nominally identical. This is best achieved by test pieces being prepared in one laboratory and distributed to the others with instructions for conditioning.

For type 2 precision, the properties of the composite material are directly related to the quality of properties of the target material. As an example, to determine the quality of a grade of SBR, a sample of the rubber plus curatives, fillers, antioxidants, etc. are mixed and cured. The precision of the resulting test pieces is determined and reflects sample preparation and the properties response of the target SBR.

The estimation of precision for rubber test methods is normally conducted using a balanced uniform level design with three or more materials sent to each participating laboratory with tests conducted to yield an independent test result by the same technician on each of two test days.

NOTE A balanced uniform level design is a plan for an interlaboratory test programme for precision, where all laboratories test all the materials selected for the programme and each laboratory conducts the same number of repeated tests, $n$, on each material.

The test method, materials, participating laboratories, test equipment and time interval for test in a laboratory are addressed in 6.1 to 6.6. Other aspects of planning shall be addressed in accordance with ISO 5725-1:1994, Clause 6.

## 6.7 Methodology

### 6.7.1 Method A

Method A determines the repeatability variance component (measurement error component) $\sigma_M{}^2$, the day-to-day variance component $\sigma_D{}^2$ and the between-laboratory variance component $\sigma_L{}^2$, by calculating the expected mean square in accordance with a suitable ANOVA table in ISO 5725-3, fully-nested experiments.

Then, the day-to-day repeatability variance $s_{rD}{}^2$ and the reproducibility variance $s_R{}^2$ are given by Formulae (1) and (2):

$$s_{rD}{}^2 = \sigma_M{}^2 + \sigma_D{}^2 \quad (1)$$

$$s_R{}^2 = \sigma_M{}^2 + \sigma_D{}^2 + \sigma_L{}^2 \quad (2)$$

The repeatability, $r$, the day-to-day repeatability, $r_{DA}$, and the reproducibility, $R$, are given by Formulae (3), (4), and (5), respectively:

$$r = 2{,}83\left(s_M{}^2\right)^{\frac{1}{2}} = 2{,}83\left(\sigma_M{}^2\right)^{\frac{1}{2}} = 2{,}83 s_M = 2{,}83\sigma_M \quad (3)$$

$$r_{DA} = 2{,}83\left(s_{rD}{}^2\right)^{\frac{1}{2}} = 2{,}83 s_{rD} \quad (4)$$

$$R = 2{,}83\left(s_R{}^2\right)^{\frac{1}{2}} = 2{,}83 s_R \quad (5)$$

Calculations for method A shall be in accordance with Annex A. An example is given in D.3.

For rubber tests, it is usually possible to have two or more repeated tests within one day.

### 6.7.2 Method B

Method B determines the day-to-day variance (between-day variance), $s_D^2$, the between-laboratory variance $s_L^2$ and the reproducibility variance $s_R^2$ (which is equal to $s_L^2 + s_D^2$), according to the calculation procedures in ASTM D4483.

The day-to-day repeatability, $r_D$, and the reproducibility, $R$, are given by Formulae (6) and (7):

$$r_{DB} = 2{,}83\left(s_D^2\right)^{\frac{1}{2}} = 2{,}83 s_D \tag{6}$$

$$R = 2{,}83\left(s_R^2\right)^{\frac{1}{2}} = 2{,}83 s_R \tag{7}$$

Calculations for method B shall be in accordance with Annex B for 2 test results or ASTM D4483 for more than 2 test results. An example is given in D.4.

When there are two or more data from repeated tests (individual determinations) within the same day, estimate the median values or the mean values, as appropriate, and apply them for the method B procedures.

### 6.7.3 Method A vs Method B — Day-to-day repeatability value

$r_{DA}$ is calculated using the day-to-day repeatability standard deviation from method A, $s_{rD}$, value. $r_{DB}$ is calculated using the day-to-day repeatability standard deviation from method B, $s_D$, value. $s_{rD} \neq s_D$. $s_{rD}$ will always be greater than $s_D$. The relationship between $s_{rD}$ and $s_D$ is given by Formula (8):

$$s_{rD} = s_D\,(n)^{1/2} \tag{8}$$

Where $n$ is the number of individual determinations used to get a test result. Therefore, $r_{DA}$ will always be greater than $r_{DB}$.

### 6.7.4 Method A vs Method B — Number of replicates

Method A has critical $h$ and $k$ values for just two replicates at 5 % significance. (See Table C.2) Two replicates are common in rubber testing ITPs.

Method B has critical $h$ and $k$ values for two, three, or four replicates at 2 % and 5 % significance. Two, three, or four replicates can be included in the ITP design. In cases of more than two replicates, method B is preferred.

## 6.8 Detection of outliers

For detecting outliers, this document adopts two measures called Mandel's $h$ and $k$ statistics. The $h$ statistic is a parameter used to review the difference between averages, while the $k$ statistics is a parameter used to review the difference between variances. This treatment is applied separately for $h$ and $k$ for each material. It may be noted that, as well as describing the variability of the measurement method, these help in laboratory evaluation. The calculation of $h$ and $k$ statistic values and the determination of their critical values at 5 % significance level for two replicates ($n$ = 2) shall be in accordance with Annex C.

For some test methods, a test result is defined as some statistical parameter, such as mean or median, calculated from the individual measurements. For method A, Mandel's $h$ and $k$ statistics are calculated using average and standard deviation values of repeated individual test measurement data usually within the same day. These individual measurements are used to calculate a test result. For method

B, Mandel's *h* and *k* statistics are calculated using average and standard deviation values of test result data from performing the test method on two different days.

## 6.9 Treatment of outliers

As with outlier detection, the treatment of outliers is performed separately for the *h* and *k* statistics for each material. There are several techniques that can be used for outlier treatment, such as data deletion, data replacement, parameter replacement, parameter deletion and retesting. It is also possible to not treat the outliers and keep them in the precision calculations. There are several methods for deriving the replacement values. Once an outlier treatment method is selected, that method is applied to all outliers at a given significance level.

Each treatment option has advantages and disadvantages. Each laboratory/material is evaluated by the *h* and *k* statistics. A given laboratory/material may be an outlier for *h*, *k*, or both. Most laboratory/materials will not be outliers. A small percentage of laboratory/materials will be outliers for *h* or *k* but not for both. A rare few laboratories will be outliers for all, or nearly all, materials. When a laboratory is an outlier for all materials, all of that laboratory's data should be deleted, the critical *h* and *k* statistics recalculated for a lower number of laboratories, and the *h* and *k* detection of outliers repeated at the new critical *h* and *k* values. When a laboratory is an outlier for nearly all the materials, the analyst should consider removing all that laboratory's data and proceeding as above.

a) Data deletion. Most laboratories that have an outlier value only have it for either *h* or *k* but not both. Deleting a laboratory's data that has an outlier for *h* or *k* also removes the non-outlier data from the precision calculations. For example, a laboratory has an outlier for *h* but not *k*. Deleting the laboratory's data removes both the mean and standard deviation values, so no *h* or *k* values are calculated. However, only the mean was an outlier so the standard deviation value could have been used to calculate the precision. Parameter replacement or parameter deletion for the mean outlier would have kept the non-outlier standard deviation value in the precision calculations. A similar condition occurs with *k* outliers.

b) Parameter replacement. There are several techniques that can be used to calculate a replacement value for the average or standard deviation parameters for outlier laboratories. Parameter replacement may alter the data distribution and impact the precision values. Some examples of techniques to use to obtain parameter replacement values are:

    1) Mean:

        i) average of all non-outlier mean values;

        ii) ascending order trend (AOT) of mean values;

        iii) mean ± 2 standard deviations;

    2) Standard Deviation:

        i) average of all non-outlier variances to calculate the replacement standard deviation value;

        ii) ascending order trend (AOT) of standard deviation values;

        iii) mean ± 2 standard deviations is NOT an option for standard deviation replacement values because standard deviations and variances do not have a normal distribution.

c) Data replacement. When there are only two replicates, the parameter replacement value and the data range can be used to calculate replacement data values for any outlier. This requires more effort than parameter replacement and has the same possibility of altering the data distribution and precision values.

d) Parameter deletion. Parameter deletion requires the least effort. However, this may not be an option for small data sets because it reduces the number of laboratories data that is used in the precision calculations.

e) Retesting. Retesting introduces additional elements of variation into the precision parameters. During the elapsed time between the initial testing and the retest, laboratory conditions, such as environment, equipment, personnel, might have changed. This can be especially critical for small data sets. Waiting for the retest data will delay the final calculation of the precision values.

f) No treatment. There are occasions, particularly with small data sets or data collected for the initial precision calculations for a new test method, when it is best to keep the outliers in the precision calculations.

If outliers are found at the 5 % significance level, any of the options listed above can be utilized at the discretion of the analyst. The ascending order trend (AOT) replacement value technique was developed to salvage as much information as possible from an ITP with 6 or fewer participants and is not the best option for ITP data sets with more than 6 participants. For ITP data sets with 7 or 8 participants, the parameter replacement option gives the best precision values. For ITP data sets with 9 or more participants, the parameter deletion option gives the best precision values.

An example of the outlier identification process for method A is given in D.2. The data set used in this example does not contain any outliers, so it is not possible to present an example of outlier treatment for method A using this data set. An example for Method A using a data set that does contain outliers is being developed and will be included in a future revision. An example of the outlier identification process and the AOT outlier treatment process for method B is given in Annex F.

## 7 Report

Each summary precision-data table should have a heading to indicate:

— the type of precision, type 1 or type 2, used;

— the property measured and its measurement units, and

— which method (A or B) was used to perform the precision calculations.

For each material tested, the following shall be recorded:

a) the material identification;

b) the mean level of measured property;

c) the repeatability standard deviation, $s_r$;

d) the repeatability, $r$, in measurement units;

e) the relative repeatability, ($r$), in per cent of the mean level;

f) the day-to-day repeatability standard deviation, $s_{r\mathrm{D}}$ or $S_\mathrm{D}$;

g) the day-to-day repeatability, $r_\mathrm{DA}$ or $\mathrm{r}_\mathrm{DB}$, in measurement units;

h) the relative day-to-day repeatability, ($r_\mathrm{DA}$) or ($\mathrm{r}_\mathrm{DB}$), in per cent of the mean level;

i) the reproducibility standard deviation, $s_R$;

j) the reproducibility, $R$, in measurement units;

k) the relative reproducibility, ($R$), in per cent of the mean level;

l) the number of laboratories in the final database used to determine the precision.

NOTE Guidance on how to use precision results is provided in Annex E.

# Annex A
(normative)

# Calculations for method A

## A.1 General

When ITP data contain the results from repeated tests within a day, and especially when the dispersion of single-day repeated test results is in question, adopt method A for evaluation without using the mean values or the median values.

To evaluate ITP results with method A, a set of data of $r$-time results [indicated by $k$ ($k$ = 1, 2, ..., $n$)] from two-day tests [indicated by $j$ ($j$ = 1, 2, ..., $q$); in this document $q$ = 2, so ($j$ = 1, 2)] in $p$ laboratories [indicated by $i$ ($i$ = 1, 2, ..., $p$)] is required to generate a precision table (see Table A.1).

**Table A.1 — Basic data for method A**

| | | Measurement | | |
|---|---|---|---|---|
| **Laboratory** | **Day** | 1 | ... | $n$ |
| 1 | Day 1 | | | |
| | Day 2 | | | |
| 2 | Day 1 | | | |
| | Day 2 | | | |
| ... | Day 1 | | | |
| | Day 2 | | | |
| $i$ | Day 1 | | $y_{i1k}$ | |
| | Day 2 | | $y_{i2k}$ | |
| ... | Day 1 | | | |
| | Day 2 | | | |
| $p$ | Day 1 | | | |
| | Day 2 | | | |

There is a total of $p$ laboratories: $i$ = 1, 2, ..., $p$.

There is a total of two-day results for each laboratory: $j$ = 1, 2.

There is a total of $r$ measurements for each day: $k$ = 1, 2, ..., $n$.



## A.2 Generation of an ANOVA table

The following is the data structure for method A:

$$y_{ijk} = \mu + L_i + D_{ij} + M_{ijk} \quad \text{(A.1)}$$

where

$\mu$ is the population mean;

$L_i$ is the between-laboratory variance, whose day-to-day variance component is $\sigma_L{}^2$, i.e. $V(L_i) = \sigma_L{}^2$;

$D_{ij}$ is the day-to-day variance, whose day-to-day variance component is $\sigma_D^2$, i.e. $V(D_{ij}) = \sigma_D^2$;

$M_{ijk}$ is the repeatability variance (measurement errors), whose repeatability variance component is $\sigma_M^2$, i.e. $V(M_{ijk}) = \sigma_M^2$.

$T_{ij}$, $T_i$ and $T$ are defined as shown in Formulae (A.2) to (A.4):

$$T_{ij} = \sum_k y_{ijk} \tag{A.2}$$

$$T_i = \sum_j \sum_k y_{ijk} \tag{A.3}$$

$$T = \sum_i \sum_j \sum_k y_{ijk} \tag{A.4}$$

The total sum of squares, $SS_T$, the total between-laboratory sum of squares, $SS_L$, the total day-to-day (between days) sum of squares, $SS_D$, and the total repeatability (measurement errors) sum of squares, $SS_M$, are determined by Formulae (A.5) to (A.8):

$$SS_T = \sum_i \sum_j \sum_k y_{ijk}^2 - \frac{T^2}{pqn} \tag{A.5}$$

$$SS_L = \sum_i \frac{T_i^2}{qn} - \frac{T^2}{pqn} \tag{A.6}$$

$$SS_D = \sum_i \sum_j \frac{T_{ij}^2}{n} - \sum_i \frac{T_i^2}{qn} \tag{A.7}$$

$$SS_M = \sum_i \sum_j \sum_k y_{ijk}^2 - \sum_i \sum_j \frac{T_{ij}^2}{n} \tag{A.8}$$

Each degree of freedom $\phi_T$, $\phi_L$, $\phi_D$ and $\phi_M$ are determined by Formulae (A.9) to (A.12):

$$\phi_T = pqn - 1 = 2pn - 1 \tag{A.9}$$

$$\phi_L = p - 1 \tag{A.10}$$

$$\phi_D = p(q-1) = p \tag{A.11}$$

$$\phi_M = pq(n-1) = 2p(n-1) \tag{A.12}$$



Using these values, generate an ANOVA table as shown in Table A.2.

**Table A.2 — ANOVA table for a three-factor fully-nested experiment**

| Source | Sum of squares | Degree of freedom | Mean square | Expected mean square |
|---|---|---|---|---|
| Laboratory | $SS_L$ | $p - 1$ | $V_L = SS_L/(p - 1)$ | $\sigma_M^2 + n\sigma_D^2 + 2n\sigma_L^2$ |
| Day | $SS_D$ | $p$ | $V_D = SS_D/p$ | $\sigma_M^2 + n\sigma_D^2$ |
| Measurement | $SS_M$ | $2p(n - 1)$ | $V_M = SS_M/2p(n - 1)$ | $\sigma_M^2$ |
| Total | $SS_T$ | $2pn - 1$ | | |
| NOTE The number of days $q$ is fixed as 2. | | | | |

## A.3 Estimation of variance components

Method A determines the repeatability variance component (measurement error component), $\sigma_M{}^2$, the day-to-day variance component, $\sigma_D{}^2$, and the between-laboratory variance component, $\sigma_L{}^2$, from the mean squares $V_L$, $V_D$, and $V_M$, respectively, as Formulae (A.13), (A.14) and (A.15):

$$\sigma_L{}^2 = \frac{1}{2n}(V_L - V_D) \tag{A.13}$$

$$\sigma_D{}^2 = \frac{1}{n}(V_D - V_M) \tag{A.14}$$

$$\sigma_M{}^2 = V_M \tag{A.15}$$

Then, the day-to-day repeatability variance, $s_{rD}{}^2$, and the reproducibility variance, $s_R{}^2$, are given by Formulae (A.16) and (A.17):

$$s_{rD}{}^2 = \sigma_M{}^2 + \sigma_D{}^2 \tag{A.16}$$

$$s_R{}^2 = \sigma_M{}^2 + \sigma_D{}^2 + \sigma_L{}^2 \tag{A.17}$$

The repeatability, $r$, the day-to-day repeatability, $r_{DA}$, and the reproducibility, $R$, are then given by Formulae (A.18), (A.19) and (A.20):

$$r = 2{,}83\left(s_M{}^2\right)^{\frac{1}{2}} = 2{,}83\left(\sigma_M{}^2\right)^{\frac{1}{2}} = 2{,}83 s_M = 2{,}83\sigma_M \tag{A.18}$$

$$r_{DA} = 2{,}83\left(s_{rD}{}^2\right)^{\frac{1}{2}} = 2{,}83 s_{rD} \tag{A.19}$$

$$R = 2{,}83\left(s_R{}^2\right)^{\frac{1}{2}} = 2{,}83 s_R \tag{A.20}$$



# Annex B
(normative)

## Calculations for method B

The calculations for method B shall be in accordance with ASTM D4483, with the understanding that the symbols $r_{DB}$ and $(r_{DB})$ in this document are the same as $r$ and $(r)$, respectively, in ASTM D4483.

The day-to-day repeatability, $r_{DB}$, and the reproducibility, $R$, are given by Formulae (B.1) and (B.2):

$$r_{DB} = 2{,}83\left(s_D{}^2\right)^{\frac{1}{2}} = 2{,}83 s_D \tag{B.1}$$

$$R = 2{,}83\left(s_R{}^2\right)^{\frac{1}{2}} = 2{,}83 s_R \tag{B.2}$$

# Annex C
(normative)

# Calculating the *h* and *k* values (Mandel's statistics)

## C.1 General

When evaluating outliers, two measures called Mandel's *h* and *k* statistics are used. It may be noted that, as well as describing the variability of the measurement method, these help in laboratory evaluation.

## C.2 *h*-value, for the between-laboratory dispersion

Calculate the between-laboratory consistency statistic, *h*, for each laboratory by dividing the cell deviation (cell mean minus the grand mean for that level) by the standard deviation among cell means, shown as Formula (C.1):

$$h_i = \frac{\bar{y}_{i\bullet} - \bar{\bar{y}}}{\sqrt{\frac{1}{(p-1)} \sum_{i=1}^{p} (\bar{y}_{i\bullet} - \bar{\bar{y}})^2}} \tag{C.1}$$

in which $\bar{y}_{i\bullet}$ is defined in Table C.1.

$$\bar{\bar{y}} = \frac{\sum_{i=1}^{p} \bar{y}_{i\bullet}}{p} \tag{C.2}$$

## C.3 *k*-value, for the within-laboratory dispersion

Calculate the within-laboratory consistency statistic, *k*, by first calculating the pooled within-cell standard deviation using Formula (C.3):

$$\sqrt{\frac{\sum_{i=1}^{p} s_i^2}{p}} \tag{C.3}$$

and then calculate Formula (C.4):

$$k_i = s_i \sqrt{\frac{p}{\sum_{i=1}^{p} s_i^2}} \tag{C.4}$$

for each laboratory.

**Table C.1 — Format for statistics analyses**

| Laboratory | Day 1 | Day 2 | Means of day | Standard deviations of day |
|---|---|---|---|---|
| 1 | | | | |
| 2 | | | | |
| …<br>$i$<br>… | $y_{i1}$ | $y_{i2}$ | $\bar{y}_{i\bullet}$ | $s_i$ |
| $p$ | | | | |

## C.4 Statistical table

Indicators for Mandel's *h* and *k* statistic at the 5 % significance level are given in Table C.2.

**Table C.2 — Critical *h*-values and *k*-values at 5 % significance level**

| *p* | *h* | *k* |
|---|---|---|
| 3 | 1,15 | 1,65 |
| 4 | 1,42 | 1,76 |
| 5 | 1,57 | 1,81 |
| 6 | 1,66 | 1,85 |
| 7 | 1,71 | 1,87 |
| 8 | 1,75 | 1,88 |
| 9 | 1,78 | 1,90 |
| 10 | 1,80 | 1,90 |
| 11 | 1,82 | 1,91 |
| 12 | 1,83 | 1,92 |
| 13 | 1,84 | 1,92 |
| 14 | 1,85 | 1,92 |
| 15 | 1,86 | 1,93 |
| 16 | 1,86 | 1,93 |
| 17 | 1,87 | 1,93 |
| 18 | 1,88 | 1,93 |
| 19 | 1,88 | 1,93 |
| 20 | 1,89 | 1,94 |
| NOTE The day number is fixed as 2. | | |



# Annex D
(informative)

# Example of general precision determination

## D.1 General

This annex presents a detailed example of a precision determination with method A for the ITP of ISO 37, which was performed in 2002. The number of participating laboratories was eight, the number of the set of tests was two days and the number of measurements was five. In addition, D.4 gives another example of how the same data as in Table D.1 can be treated with method B using averaged Day 1 and Day 2 results.

## D.2 Identification of outliers

Table D.1 shows the results of the ITP (for Type 1A dumbbells), as well as supplementary statistics.

**Table D.1 — Results of the ITP for ISO 37 and supplementary statistics**

| Laboratory | Day | Measurement | | | | | Mean of data $\bar{y}_{ij\bullet}$ | Mean of day $\bar{y}_{i\bullet\bullet}$ | Standard deviation of day $s_i$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Day 1 | 31,60 | 32,55 | 32,40 | 32,52 | 31,48 | 32,110 | 32,295 | 0,262 |
| | Day 2 | 32,98 | 33,33 | 30,02 | 33,00 | 33,07 | 32,480 | | |
| 2 | Day 1 | 34,66 | 32,17 | 33,02 | 33,24 | 33,55 | 33,328 | 32,839 | 0,692 |
| | Day 2 | 31,80 | 32,75 | 31,74 | 31,33 | 34,13 | 32,350 | | |
| 3 | Day 1 | 34,50 | 33,80 | 34,10 | 33,90 | 31,20 | 33,500 | 34,090 | 0,834 |
| | Day 2 | 35,20 | 35,50 | 33,40 | 35,30 | 34,00 | 34,680 | | |
| 4 | Day 1 | 34,70 | 32,70 | 34,30 | 32,50 | 33,30 | 33,500 | 33,870 | 0,523 |
| | Day 2 | 33,10 | 35,00 | 34,00 | 33,20 | 35,90 | 34,240 | | |
| 5 | Day 1 | 33,47 | 35,56 | 32,09 | 33,20 | 33,27 | 33,518 | 33,256 | 0,371 |
| | Day 2 | 32,98 | 30,97 | 32,72 | 34,04 | 34,26 | 32,994 | | |
| 6 | Day 1 | 31,29 | 30,89 | 31,43 | 30,53 | 31,98 | 31,224 | 31,385 | 0,228 |
| | Day 2 | 31,20 | 30,73 | 31,39 | 32,45 | 31,96 | 31,546 | | |
| 7 | Day 1 | 34,60 | 31,70 | 34,70 | 30,90 | 32,20 | 32,820 | 32,550 | 0,382 |
| | Day 2 | 33,00 | 33,10 | 31,90 | 32,00 | 31,40 | 32,280 | | |
| 8 | Day 1 | 34,70 | 32,70 | 34,30 | 32,50 | 33,30 | 33,500 | 33,870 | 0,523 |
| | Day 2 | 33,10 | 35,00 | 34,00 | 33,20 | 35,90 | 34,240 | | |

Mandel's *h* and *k* statistics are calculated by average values of repeated test data within the same day.

Using the data in Table D.1, the $h$ values for between-laboratory dispersion and the $k$ values for within-laboratory dispersion were determined by Formulae (D.1) and (D.2). For this programme, $p_j$ is 8:

$$h_i = \frac{\bar{y}_{i\bullet} - \bar{\bar{y}}}{\sqrt{\frac{1}{(p-1)}\sum_{i=1}^{p}\left(\bar{y}_{i\bullet} - \bar{\bar{y}}\right)^2}} \tag{D.1}$$

$$k_i = s_i \sqrt{\frac{p}{\sum_{i=1}^{p} s_i^2}} \tag{D.2}$$

The obtained $h$-values and $k$-values are shown in Table D.2 and Table D.3.

**Table D.2 — Obtained *h*-values**

| Laboratory | $h_i$ |
|---|---|
| 1 | −0,78 |
| 2 | −0,19 |
| 3 | 1,15 |
| 4 | 0,91 |
| 5 | 0,25 |
| 6 | −1,75 |
| 7 | −0,50 |
| 8 | 0,91 |

**Table D.3 — Obtained *k*-values**

| Laboratory | $k_i$ |
|---|---|
| 1 | 0,51 |
| 2 | 1,34 |
| 3 | 1,62 |
| 4 | 1,02 |
| 5 | 0,72 |
| 6 | 0,44 |
| 7 | 0,74 |
| 8 | 1,02 |

Each $h$ and $k$ value was compared with the corresponding critical values. According to Table C.2, the critical $h$ value is 1,75 and the critical $k$ value is 1,88 when $p$ is 8. Since none of the values in Tables D.2 and D.3 exceeded those critical values, it was concluded that there was no outlier in the results and that all the data were valid.

## D.3 Precision analyses with method A

Table D.4 shows the ITP results and the supplementary statistics that were required for precision evaluation with method A.

**Table D.4 — ITP results and supplementary statistics (method A)**

| Laboratory | Day | Measurement | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | M1 | M2 | M3 | M4 | M5 | $T_{ij.}$ | $T_{i..}$ |
| 1 | Day 1 | 31,60 | 32,55 | 32,40 | 32,52 | 31,48 | 160,55 | 322,95 |
| | Day 2 | 32,98 | 33,33 | 30,02 | 33,00 | 33,07 | 162,40 | |
| 2 | Day 1 | 34,66 | 32,17 | 33,02 | 33,24 | 33,55 | 166,64 | 328,39 |
| | Day 2 | 31,80 | 32,75 | 31,74 | 31,33 | 34,13 | 161,75 | |
| 3 | Day 1 | 34,50 | 33,80 | 34,10 | 33,90 | 31,20 | 167,50 | 340,90 |
| | Day 2 | 35,20 | 35,50 | 33,40 | 35,30 | 34,00 | 173,40 | |
| 4 | Day 1 | 34,70 | 32,70 | 34,30 | 32,50 | 33,30 | 167,50 | 338,70 |
| | Day 2 | 33,10 | 35,00 | 34,00 | 33,20 | 35,90 | 171,20 | |
| 5 | Day 1 | 33,47 | 35,56 | 32,09 | 33,20 | 33,27 | 167,59 | 332,56 |
| | Day 2 | 32,98 | 30,97 | 32,72 | 34,04 | 34,26 | 164,97 | |
| 6 | Day 1 | 31,29 | 30,89 | 31,43 | 30,53 | 31,98 | 156,12 | 313,85 |
| | Day 2 | 31,20 | 30,73 | 31,39 | 32,45 | 31,96 | 157,73 | |
| 7 | Day 1 | 34,60 | 31,70 | 34,70 | 30,90 | 32,20 | 164,10 | 325,50 |
| | Day 2 | 33,00 | 33,10 | 31,90 | 32,00 | 31,40 | 161,40 | |
| 8 | Day 1 | 34,70 | 32,70 | 34,30 | 32,50 | 33,30 | 167,50 | 338,70 |
| | Day 2 | 33,10 | 35,00 | 34,00 | 33,20 | 35,90 | 171,20 | |

Those values were applied to Formulae (D.3) to (D.7) to generate an ANOVA table (see Table D.5):

$$T=\sum_i\sum_j\sum_k y_{ijk}=2\ 641{,}55 \tag{D.3}$$

$$S\,S_T=\sum_i\sum_j\sum_k {y_{ijk}}^2-\frac{T^2}{pqn}=87\ 370{,}854\ 7-\frac{2\ 641{,}55^2}{8\times2\times5}=148{,}525 \tag{D.4}$$

$$S\,S_L=\sum_i\frac{T_i^2}{qn}-\frac{T^2}{pqn}=\frac{872\ 833{,}110\ 7}{2\times5}-\frac{2\ 641{,}55^2}{8\times2\times5}=60{,}981 \tag{D.5}$$

$$S\,S_D=\sum_i\sum_j\frac{{T_{ij}}^2}{n}-\sum_i\frac{{T_i}^2}{qn}=\frac{436\ 469{,}690\ 9}{5}-\frac{872\ 833{,}110\ 7}{2\times5}=10{,}627 \tag{D.6}$$

$$S\,S_M=\sum_i\sum_j\sum_k {y_{ijk}}^2-\sum_i\sum_i\frac{{T_{ij}}^2}{n}=87\ 370{,}854\ 7-\frac{436\ 469{,}690\ 9}{5}=76{,}917 \tag{D.7}$$

**Table D.5 — Obtained ANOVA table**

| Source | Sum of squares | Degrees of freedom | Mean square | Expected mean square |
|---|---|---|---|---|
| Laboratory | 60,981 | 7 | 8,712 | ${\sigma_M}^2+5{\sigma_D}^2+10{\sigma_L}^2$ |
| Day | 10,627 | 8 | 1,328 | ${\sigma_M}^2+5{\sigma_D}^2$ |
| Measurement | 76,917 | 64 | 1,202 | ${\sigma_M}^2$ |
| Total | 148,525 | 79 | | |

From the ANOVA table, variance components were estimated as shown in Formulae (D.8) to (D.10):

$$\sigma_L{}^2 = \frac{1}{2n}(V_L - V_D) = \frac{1}{2 \times 5}(8{,}712 - 1{,}328) = 0{,}738\,4 \quad \text{(D.8)}$$

$$\sigma_D{}^2 = \frac{1}{n}(V_D - V_M) = \frac{1}{5}(1{,}328 - 1{,}202) = 0{,}025\,2 \quad \text{(D.9)}$$

$$\sigma_M{}^2 = V_M = 1{,}201\,8 \quad \text{(D.10)}$$

Then, the day-to-day repeatability variance, $s_{rD}{}^2$, and the reproducibility variance, $s_R{}^2$, were determined as shown in Formulae (D.11) and (D.12):

$$s_{rD}{}^2 = \sigma_M{}^2 + \sigma_D{}^2 = 1{,}201\,8 + 0{,}025\,2 = 1{,}227\,0 \quad \text{(D.11)}$$

$$s_R{}^2 = \sigma_M{}^2 + \sigma_D{}^2 + \sigma_L{}^2 = 1{,}201\,8 + 0{,}025\,2 + 0{,}738\,4 = 1{,}965\,4 \quad \text{(D.12)}$$

Finally, the repeatability, $r$, the day-to-day repeatability, $r_D$, and the reproducibility, $R$, are determined by Formulae (D.13) to (D.15):

$$r = 2{,}83\left(s_M{}^2\right)^{\frac{1}{2}} = 2{,}83\left(\sigma_M{}^2\right)^{\frac{1}{2}} = 3{,}102 \quad \text{(D.13)}$$

$$r_{DA} = 2{,}83\left(s_{rD}{}^2\right)^{\frac{1}{2}} = 3{,}134 \quad \text{(D.14)}$$

$$R = 2{,}83\left(s_R{}^2\right)^{\frac{1}{2}} = 3{,}967 \quad \text{(D.15)}$$

## D.4 Precision analyses with method B

For reference, the following shows the precision analysis procedure with method B for the same data.

Calculate the mean values for the within-a-day repeated results. Table D.6 shows the obtained data, as well as the supplementary statistic values for method B.

**Table D.6 — Obtained data and supplementary statistics (method B)**

| Laboratory | Day 1 | Day 2 | Means of day | Day 1 - Day 2 |
|---|---|---|---|---|
| 1 | 32,11 | 32,48 | 32,295 | −0,37 |
| 2 | 33,33 | 32,35 | 32,839 | 0,98 |
| 3 | 33,50 | 34,68 | 34,090 | −1,18 |
| 4 | 33,50 | 34,24 | 33,870 | −0,74 |
| 5 | 33,52 | 32,99 | 33,256 | 0,53 |
| 6 | 31,22 | 31,55 | 31,385 | −0,33 |
| 7 | 32,82 | 32,28 | 32,550 | 0,54 |
| 8 | 33,50 | 34,24 | 33,870 | −0,74 |

From Table D.6, the day-to-day variance, $s_D{}^2$, and the reproducibility variance, $s_L{}^2$, are obtained as follows. See Formulae (D.16) to (D.19):

$$\bar{\bar{y}}_j = 33{,}019\,4 \quad \text{(D.16)}$$



$$s_D{}^2 = \frac{1}{2p}\sum_{i=1}^{p}(y_{i1} - y_{i2})^2 = \frac{1}{2\times 8}\times 4{,}266 = 0{,}266 \quad \text{(D.17)}$$

$$s_L{}^2 = \frac{1}{p-1}\sum_{i=1}^{p}(\bar{y}_{ij} - \bar{\bar{y}}_j)^2 - \frac{s_D{}^2}{2} = \frac{1}{8-1}\times 6{,}098\ 1 - \frac{0{,}265\ 7}{2} = 0{,}738\ 3 \quad \text{(D.18)}$$

From these data,

$$s_R{}^2 = s_L{}^2 + s_D{}^2 = 0{,}738\ 3 + 0{,}265\ 7 = 1{,}00\ 4 \quad \text{(D.19)}$$

NOTE Method B does not evaluate within-day variance (measurement errors) since it employs the mean values of five measurements.

The expected $s_D{}^2$ for method B can be described as $\sigma_D{}^2 + \frac{\sigma_M{}^2}{5}$ if converted using the method A structure. $s_D{}^2$ here is determined as shown in Formula (D.20):

$$s_D{}^2 = \frac{\sigma_M{}^2}{5} + \sigma_D{}^2 = \frac{1{,}201\ 8}{5} + 0{,}025\ 2 = 0{,}265\ 6 \quad \text{(D.20)}$$

The $s_{rD}{}^2$ for method A, on the other hand, evaluates within-a-day variance (measurement errors) and is determined as shown in Formula (D.21):

$$s_{rD}{}^2 = \sigma_M{}^2 + \sigma_D{}^2 = 1{,}201\ 8 + 0{,}025\ 2 = 1{,}227\ 0 \quad \text{(D.21)}$$

The obtained $s_L{}^2$ is the same with the $\sigma_L{}^2$ value as in method A.

Finally, the day-to-day repeatability, $r_{DB}$, and the reproducibility, $R$, are determined as shown in Formulae (D.22) and (D.23):

$$r_{DB} = 2{,}83(s_D{}^2)^{\frac{1}{2}} = 1{,}459 \quad \text{(D.22)}$$

$$R = 2{,}83(s_R{}^2)^{\frac{1}{2}} = 2{,}836 \quad \text{(D.23)}$$

While it is reasonable to expect that the day-to-day repeatability from the two methods, $r_{DA}$ and $r_{DB}$, should have the same value when calculating them using the same data set and outlier handling procedures, in this example they do not have the same value.



# Annex E
(informative)

# Guidance for using precision results

**E.1** The general procedure for using precision results is as described in E.2 to E.5, with the symbol $|x_1 - x_2|$ designating a positive difference in any two measurement values (i.e. without regard to sign).

**E.2** Enter the appropriate precision table (for whatever test parameter is being considered) at an average value (of the measured parameter) nearest to the "test" data average under consideration. This line will give the applicable $r$, $(r)$, $r_{DA}$, $r_{DB}$, $(r_{DA})$, $(r_{DB})$, $R$ or $(R)$ for use in the decision process.

**E.3** With the $r$ and $(r)$ values, the following general repeatability statements may be used to make decisions.

a) For an absolute difference: the difference $|x_1 - x_2|$ between two test (value) averages, found on nominally identical material samples under normal and correct operation of the test procedure, will exceed the tabulated repeatability, $r$, on average of not more than once in 20 instances of testing under the given repeatability conditions.

b) For a percentage difference between two test (value) averages, the percentage difference is shown as Formula (E.1):

$$\frac{|x_1 - x_2|}{\frac{1}{2}(x_1 + x_2)} \times 100 \qquad \text{(E.1)}$$

between two test values, found on nominally identical material samples under normal and correct operation of the test procedure, will exceed the tabulated repeatability, $(r)$, on average of not more than once in 20 cases.

**E.4** With the $r_{DA}$ or $r_{DB}$ and $(r_{DA})$ or $(r_{DB})$ values, the following general day-to-day repeatability statements may be used to make decisions.

a) For an absolute difference: the difference $|x_1 - x_2|$ between two day test (value) averages, found on nominally identical material samples under normal and correct operation of the test procedure, will exceed the tabulated day-to-day repeatability, $r_D$, on average of not more than once in 20 instances of testing under the given repeatability conditions.

b) For a percentage difference between two-day test (value) averages, the percentage difference is shown as Formula (E.2):

$$\frac{|x_1 - x_2|}{\frac{1}{2}(x_1 + x_2)} \times 100 \qquad \text{(E.2)}$$

between two test values, found on nominally identical material samples under normal and correct operation of the test procedure, will exceed the tabulated repeatability $(r_D)$ on average of not more than once in 20 instances of testing under the given repeatability conditions.

**E.5** With the $R$ and ($R$) values, the following general reproducibility statements may be used to make decisions.

a) For an absolute difference: the absolute difference $|x_1 - x_2|$ between two independently measured test (value) averages, found in two laboratories using normal and correct test procedures on nominally identical material samples, will exceed the tabulated reproducibility, $R$, not more than once in 20 instances of testing under the given reproducibility conditions.

b) For a percentage difference between two test (value) averages, the percentage difference is shown as Formula (E.3):

$$\frac{|x_1 - x_2|}{\frac{1}{2}(x_1 + x_2)} \times 100 \tag{E.3}$$

between two independently measured test (value) averages, found in two laboratories using normal and correct test procedures on nominally identical material samples, will exceed the tabulated reproducibility, ($R$), not more than once in 20 instances of testing under the given reproducibility conditions.

# Annex F
(informative)

# Example of outlier treatment for method B

## F.1 General

This annex presents a method for treating outliers when carrying out the precision analysis and calculations using method B. The method B precision analysis and calculations do not require the creation of an ANOVA table as presented in method A and Annex A, because it is not needed in this case. However, it is required to build tables to perform many of the calculations whose results are displayed in an ANOVA table.

The tables presented in this annex are only those needed for the outlier analysis, detection, and treatment in method B. The additional tables needed to perform the precision calculations are not given. The complete set of tables and their use can be found in ASTM D4483.

The data and analysis presented in Annex D does not contain any outliers at the 5 % significance level, so no information was presented on the treatment of outliers for comparison to the method B process. An example of outlier treatment using Method A will be included in a future revision.

This annex presents a detailed example of a precision determination with method B for an ITP of ISO 289, which was performed in 1982 using the version of ISO 289 that was available in 1982. The number of participating laboratories was nine, the number of materials tested was four, with each laboratory testing each material once on two different days. Each test result is from a single determination.

Table F.1 shows the results of the ITP as well as statistical summaries, the critical *h* and *k* values for the number of laboratories (9) and replicates (2) at the 5 % significance level.



**Table F.1 — Test data (raw input data)**

| | **Material 1** | | **Material 2** | | **Material 3** | | **Material 4** | |
|---|---|---|---|---|---|---|---|---|
| **Laboratory** | **Day 1** | **Day 2** | **Day 1** | **Day 2** | **Day 1** | **Day 2** | **Day 1** | **Day 2** |
| 1 | 50,8 | 51,9 | 68,0 | 68,3 | 73,3 | 75,2 | 99,0 | 98,5 |
| 2 | 53,0 | 53,0 | 66,0 | 66,5 | 70,0 | 71,0 | 96,5 | 97,0 |
| 3 | 52,4 | 51,9 | 66,1 | 66,6 | 73,6 | 74,6 | 97,7 | 98,6 |
| 4 | 53,0 | 51,5 | 66,0 | 66,0 | 80,0 | 76,5 | 97,0 | 94,0 |
| 5 | 52,3 | 52,1 | 66,5 | 66,5 | 77,0 | 78,1 | 99,2 | 99,4 |
| 6 | 54,4 | 54,3 | 67,5 | 67,0 | 81,4 | 83,3 | 98,0 | 98,1 |
| 7 | 52,8 | 52,8 | 67,5 | 67,4 | 72,8 | 73,4 | 97,9 | 98,4 |
| 8 | 53,0 | 53,0 | 67,0 | 66,5 | 77,0 | 77,0 | 103,0 | 102,0 |
| 9 | 50,1 | 50,3 | 67,0 | 66,6 | 64,6 | 62,6 | 92,0 | 90,2 |
| Day average | 52,42 | 52,31 | 66,84 | 66,82 | 74,41 | 74,63 | 97,81 | 97,36 |
| 2-Day average | | 52,37 | | 66,83 | | 74,52 | | 97,58 |
| Between laboratory Standard deviation | 1,28 | 1,13 | 0,74 | 0,67 | 5,17 | 5,66 | 2,88 | 3,41 |
| Pooled between laboratory standard deviation | | 1,21 | | 0,71 | | 5,42 | | 3,16 |
| Number of laboratories, *p* | 9 | | 9 | | 9 | | 9 | |
| *h*(crit), 5 % | 1,78 | | 1,78 | | 1,78 | | 1,78 | |

**Table F.1** *(continued)*

| Laboratory | Material 1 | | Material 2 | | Material 3 | | Material 4 | |
|---|---|---|---|---|---|---|---|---|
| | Day 1 | Day 2 | Day 1 | Day 2 | Day 1 | Day 2 | Day 1 | Day 2 |
| $k$(crit), 5 % | 1,90 | | 1,90 | | 1,90 | | 1,90 | |

The precision results based on this initial set of data without any determination of outliers or any outlier treatment, is shown in Table F.2.

**Table F.2 — Precision results for ISO 289, Mooney viscometer (Type 1 precision)**

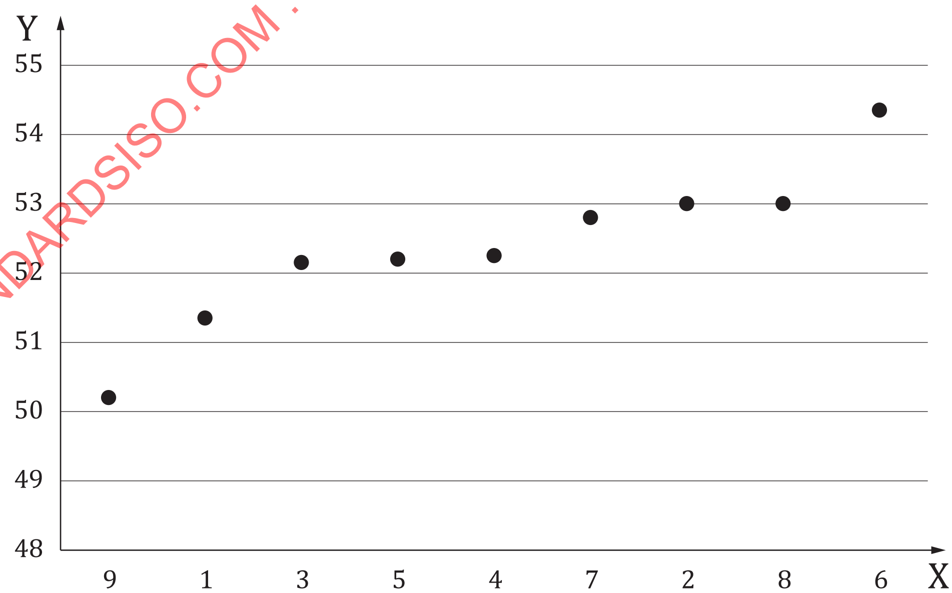
| Material | Units | Mooney torque | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Number of laboratories | Mean level | *Sr* | *r* | (*r*) | *SR* | *R* | (*R*) |
| Material 1 | 9 | 52,37 | 0,459 | 1,300 | 2,48 | 1,203 | 3,41 | 6,50 |
| Material 2 | 9 | 66,83 | 0,265 | 0,749 | 1,12 | 0,703 | 1,99 | 2,98 |
| Material 3 | 9 | 74,52 | 1,226 | 3,469 | 4,65 | 5,411 | 15,31 | 20,55 |
| Material 4 | 9 | 97,58 | 0,908 | 2,570 | 2,63 | 3,157 | 8,93 | 9,15 |
| Average | | 72,83 | | | | | | |
| Pooled values | | | 0,808 | 2,285 | 3,14 | 3,209 | 9,08 | 12,47 |

NOTE The average and pooled values represent the best estimates of the precision of a test method when used to test materials at varying levels of the test result, particularly for a material not given in the table. When testing the same material or a material at a similar level as one presented in the precision table, the given mean and precision values for that material can be used.

An activity that is not required to perform the precision calculations, but might be helpful, is to prepare an ascending order trend (AOT) graph of the initial average and standard deviation values to get some indication of the data distribution and possible outliers for each material.

Figures F.1 and F.2 show the AOT graph for the average and standard deviation values respectively for Material 1. Similar graphs can be prepared for the other materials.





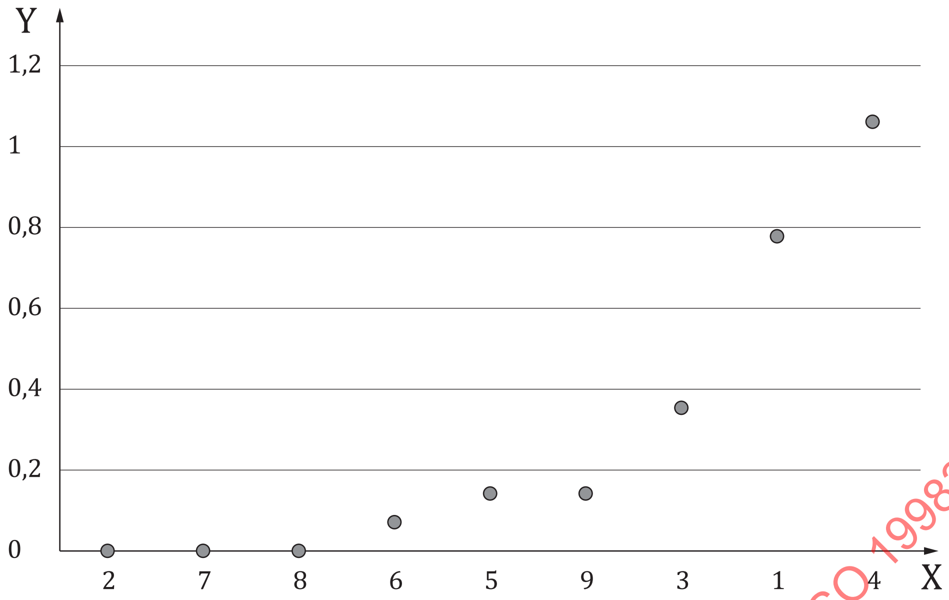

**Key**

Y Mooney average

X sorted laboratory number

**Figure F.1 — Material 1 — Averages AOT graph**




**Key**

Y Mooney standard deviation

X sorted laboratory number

**Figure F.2 — Material 1 — Standard deviations AOT graph**

For identification, the laboratory number is included with the sorted data and displayed on the x-axis. The laboratory numbers are not x-axis values.

From Figure F.1, it can be seen that laboratories 6 and 9 might be outliers and from Figure F.2, it can be seen that laboratory 4 might be an outlier.

The identification of outliers is not performed using these AOT graphs. Outliers are identified using the Mandel *h* and *k* statistics.

## F.2 Identification of outliers

The absolute value of the calculated *h* value for each laboratory and material is compared to the critical *h* value for that material. Any laboratory/material whose absolute calculated *h* value is greater than the critical value is identified as an outlier. Table F.3 shows the result of this operation.

**Table F.3 — Cell *h*-values**

| Laboratory | Material 1 | Material 2 | Material 3 | Material 4 |
|---|---|---|---|---|
| 1 | −0,88 | 1,94 * | −0,05 | 0,38 |
| 2 | 0,55 | −0,86 | −0,75 | −0,27 |
| 3 | −0,19 | −0,71 | −0,08 | 0,18 |
| 4 | −0,10 | −1,23 | 0,70 | −0,67 |
| 5 | −0,14 | −0,49 | 0,57 | 0,56 |
| 6 | 1,71 | 0,61 | 1,47 | 0,15 |
| 7 | 0,37 | 0,91 | −0,27 | 0,18 |
| 8 | 0,55 | −0,12 | 0,46 | 1,59 |
| 9 | −1,87 * | −0,05 | −2,04 * | −2,10 * |
| *h*(crit), 5 % | 1,78 | 1,78 | 1,78 | 1,78 |
| * indicates outlier value. | | | | |