INTERNATIONAL STANDARD

ISO 5725-1

First edition
1994-12-15

Accuracy (trueness and precision) of measurement methods and results —

Part 1:
General principles and definitions

Exactitude (justesse et fidélité) des résultats et méthodes de mesure —
Partie 1: Principes généraux et définitions

Reference number
ISO 5725-1:1994(E)

Contents

International Organization for Standardization
Case Postale 56 • CH-1211 Genève 20 • Switzerland
Printed in Switzerland

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.

Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

International Standard ISO 5725-1 was prepared by Technical Committee ISO/TC 69, *Applications of statistical methods*, Subcommittee SC 6, *Measurement methods and results*.

ISO 5725 consists of the following parts, under the general title *Accuracy (trueness and precision) of measurement methods and results*:

- *Part 1: General principles and definitions*
- *Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method*
- *Part 3: Intermediate measures of the precision of a standard measurement method*
- *Part 4: Basic methods for the determination of the trueness of a standard measurement method*
- *Part 5: Alternative methods for the determination of the precision of a standard measurement method*
- *Part 6: Use in practice of accuracy values*

Parts 1 to 6 of ISO 5725 together cancel and replace ISO 5725:1986, which has been extended to cover trueness (in addition to precision) and intermediate precision conditions (in addition to repeatability and reproducibility conditions).

Annexes A and B form an integral part of this part of ISO 5725. Annex C is for information only.

Introduction

0.1 ISO 5725 uses two terms "trueness" and "precision" to describe the accuracy of a measurement method. "Trueness" refers to the closeness of agreement between the arithmetic mean of a large number of test results and the true or accepted reference value. "Precision" refers to the closeness of agreement between test results.

0.2 The need to consider "precision" arises because tests performed on presumably identical materials in presumably identical circumstances do not, in general, yield identical results. This is attributed to unavoidable random errors inherent in every measurement procedure; the factors that influence the outcome of a measurement cannot all be completely controlled. In the practical interpretation of measurement data, this variability has to be taken into account. For instance, the difference between a test result and some specified value may be within the scope of unavoidable random errors, in which case a real deviation from such a specified value has not been established. Similarly, comparing test results from two batches of material will not indicate a fundamental quality difference if the difference between them can be attributed to the inherent variation in the measurement procedure.

0.3 Many different factors (apart from variations between supposedly identical specimens) may contribute to the variability of results from a measurement method, including:

a) the operator;

b) the equipment used;

c) the calibration of the equipment;

d) the environment (temperature, humidity, air pollution, etc.);

e) the time elapsed between measurements.

The variability between measurements performed by different operators and/or with different equipment will usually be greater than the variability between measurements carried out within a short interval of time by a single operator using the same equipment.

0.4 The general term for variability between repeated measurements is precision. Two conditions of precision, termed repeatability and reproducibility conditions, have been found necessary and, for many practical cases, useful for describing the variability of a measurement method. Under repeatability conditions, factors a) to e) listed above are considered constants and do not contribute to the variability, while under reproducibility conditions they vary and do contribute to the variability of the test results. Thus repeatability and reproducibility are the two extremes of precision, the first describing the minimum and the second the maximum variability in results. Other intermediate conditions between these two extreme conditions of precision are also conceivable, when one or more of factors a) to e) are allowed to vary, and are used in certain specified circumstances. Precision is normally expressed in terms of standard deviations.

0.5 The "trueness" of a measurement method is of interest when it is possible to conceive of a true value for the property being measured. Although, for some measurement methods, the true value cannot be known exactly, it may be possible to have an accepted reference value for the property being measured; for example, if suitable reference materials are available, or if the accepted reference value can be established by reference to another measurement method or by preparation of a known sample. The trueness of the measurement method can be investigated by comparing the accepted reference value with the level of the results given by the measurement method. Trueness is normally expressed in terms of bias. Bias can arise, for example, in chemical analysis if the measurement method fails to extract all of an element, or if the presence of one element interferes with the determination of another.

0.6 The general term accuracy is used in ISO 5725 to refer to both trueness and precision.

The term accuracy was at one time used to cover only the one component now named trueness, but it became clear that to many persons it should imply the total displacement of a result from a reference value, due to random as well as systematic effects.

The term bias has been in use for statistical matters for a very long time, but because it caused certain philosophical objections among members of some professions (such as medical and legal practitioners), the positive aspect has been emphasized by the invention of the term trueness.

Accuracy (trueness and precision) of measurement methods and results —

Part 1:
General principles and definitions

1 Scope

1.1 The purpose of ISO 5725 is as follows:

a) to outline the general principles to be understood when assessing accuracy (trueness and precision) of measurement methods and results, and in applications, and to establish practical estimations of the various measures by experiment (ISO 5725-1);

b) to provide a basic method for estimating the two extreme measures of the precision of measurement methods by experiment (ISO 5725-2);

c) to provide a procedure for obtaining intermediate measures of precision, giving the circumstances in which they apply and methods for estimating them (ISO 5725-3);

d) to provide basic methods for the determination of the trueness of a measurement method (ISO 5725-4);

e) to provide some alternatives to the basic methods, given in ISO 5725-2 and ISO 5725-4, for determining the precision and trueness of measurement methods for use under certain circumstances (ISO 5725-5);

f) to present some practical applications of these measures of trueness and precision (ISO 5725-6).

1.2 This part of ISO 5725 is concerned exclusively with measurement methods which yield measurements on a continuous scale and give a single value as the test result, although this single value may be the outcome of a calculation from a set of observations.

It defines values which describe, in quantitative terms, the ability of a measurement method to give a correct result (trueness) or to replicate a given result (precision). Thus there is an implication that exactly the same thing is being measured, in exactly the same way, and that the measurement process is under control.

This part of ISO 5725 may be applied to a very wide range of materials, including liquids, powders and solid objects, manufactured or naturally occurring, provided that due consideration is given to any heterogeneity of the material.

2 Normative references

The following standards contain provisions which, through reference in this text, constitute provisions of this part of ISO 5725. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this part of ISO 5725 are encouraged to investigate the possibility of applying the most recent editions of the standards indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO 3534-1:1993, *Statistics — Vocabulary and symbols — Part 1: Probability and general statistical terms.*

ISO 5725-2:1994, *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 5725-3:1994, *Accuracy (trueness and precision) of measurement methods and results — Part 3: Intermediate measures of the precision of a standard measurement method.*

ISO 5725-4:1994, *Accuracy (trueness and precision) of measurement methods and results — Part 4: Basic methods for the determination of the trueness of a standard measurement method.*

3 Definitions

For the purposes of ISO 5725, the following definitions apply.

Some definitions are taken from ISO 3534-1.

The symbols used in ISO 5725 are given in annex A.

3.1 observed value: The value of a characteristic obtained as the result of a single observation.

[ISO 3534-1]

3.2 test result: The value of a characteristic obtained by carrying out a specified test method.

NOTE 1 The test method should specify that one or a number of individual observations be made, and their average or another appropriate function (such as the median or the standard deviation) be reported as the test result. It may also require standard corrections to be applied, such as correction of gas volumes to standard temperature and pressure. Thus a test result can be a result calculated from several observed values. In the simple case, the test result is the observed value itself.

[ISO 3534-1]

3.3 level of the test in a precision experiment: The general average of the test results from all laboratories for one particular material or specimen tested.

3.4 cell in a precision experiment: The test results at a single level obtained by one laboratory.

3.5 accepted reference value: A value that serves as an agreed-upon reference for comparison, and which is derived as:

a) a theoretical or established value, based on scientific principles;

b) an assigned or certified value, based on experimental work of some national or international organization;

c) a consensus or certified value, based on collaborative experimental work under the auspices of a scientific or engineering group;

d) when a), b) and c) are not available, the expectation of the (measurable) quantity, i.e. the mean of a specified population of measurements.

[ISO 3534-1]

3.6 accuracy: The closeness of agreement between a test result and the accepted reference value.

NOTE 2 The term accuracy, when applied to a set of test results, involves a combination of random components and a common systematic error or bias component.

[ISO 3534-1]

3.7 trueness: The closeness of agreement between the average value obtained from a large series of test results and an accepted reference value.

NOTES

3 The measure of trueness is usually expressed in terms of bias.

4 Trueness has been referred to as "accuracy of the mean". This usage is not recommended.

[ISO 3534-1]

3.8 bias: The difference between the expectation of the test results and an accepted reference value.

NOTE 5 Bias is the total systematic error as contrasted to random error. There may be one or more systematic error components contributing to the bias. A larger systematic difference from the accepted reference value is reflected by a larger bias value.

[ISO 3534-1]

3.9 laboratory bias: The difference between the expectation of the test results from a particular laboratory and an accepted reference value.

3.10 bias of the measurement method: The difference between the expectation of test results obtained from all laboratories using that method and an accepted reference value.

NOTE 6 One example of this in operation would be where a method purporting to measure the sulfur content of a compound consistently fails to extract all the sulfur, giving a negative bias to the measurement method. The bias of the measurement method is measured by the displacement of the average of results from a large number of different laboratories all using the same method. The bias of a measurement method may be different at different levels.

3.11 laboratory component of bias: The difference between the laboratory bias and the bias of the measurement method.

NOTES

7 The laboratory component of bias is specific to a given laboratory and the conditions of measurement within the laboratory, and also it may be different at different levels of the test.

8 The laboratory component of bias is relative to the overall average result, not the true or reference value.

3.12 precision: The closeness of agreement between independent test results obtained under stipulated conditions.

NOTES

9 Precision depends only on the distribution of random errors and does not relate to the true value or the specified value.

10 The measure of precision is usually expressed in terms of imprecision and computed as a standard deviation of the test results. Less precision is reflected by a larger standard deviation.

11 "Independent test results" means results obtained in a manner not influenced by any previous result on the same or similar test object. Quantitative measures of precision depend critically on the stipulated conditions. Repeatability and reproducibility conditions are particular sets of extreme conditions.

[ISO 3534-1]

3.13 repeatability: Precision under repeatability conditions.

[ISO 3534-1]

3.14 repeatability conditions: 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 within short intervals of time.

[ISO 3534-1]

3.15 repeatability standard deviation: The standard deviation of test results obtained under repeatability conditions.

NOTES

12 It is a measure of dispersion of the distribution of test results under repeatability conditions.

13 Similarly "repeatability variance" and "repeatability coefficient of variation" could be defined and used as measures of the dispersion of test results under repeatability conditions.

[ISO 3534-1]

3.16 repeatability limit: The value less than or equal to which the absolute difference between two test results obtained under repeatability conditions may be expected to be with a probability of 95 %.

NOTE 14 The symbol used is r.

[ISO 3534-1]

3.17 reproducibility: Precision under reproducibility conditions.

[ISO 3534-1]

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

[ISO 3534-1]

3.19 reproducibility standard deviation: The standard deviation of test results obtained under reproducibility conditions.

NOTES

15 It is a measure of the dispersion of the distribution of test results under reproducibility conditions.

16 Similarly "reproducibility variance" and "reproducibility coefficient of variation" could be defined and used as measures of the dispersion of test results under reproducibility conditions.

[ISO 3534-1]

3.20 reproducibility limit: The value less than or equal to which the absolute difference between two test results obtained under reproducibility conditions may be expected to be with a probability of 95 %.

NOTE 17 The symbol used is R.

[ISO 3534-1]

3.21 outlier: A member of a set of values which is inconsistent with the other members of that set.

NOTE 18 ISO 5725-2 specifies the statistical tests and the significance level to be used to identify outliers in trueness and precision experiments.

3.22 collaborative assessment experiment: An interlaboratory experiment in which the performance of each laboratory is assessed using the same standard measurement method on identical material.

NOTES

19 The definitions given in 3.16 and 3.20 apply to results that vary on a continuous scale. If the test result is discrete or rounded off, the repeatability limit and the reproducibility limit as defined above are each the minimum value equal to or below which the absolute difference between two single test results is expected to lie with a probability of not less than 95 %.

20 The definitions given in 3.8 to 3.11, 3.15, 3.16, 3.19 and 3.20 refer to theoretical values which in reality remain unknown. The values for reproducibility and repeatability standard deviations and bias actually determined by experiment (as described in ISO 5725-2 and ISO 5725-4) are, in statistical terms, estimates of these values, and as such are subject to errors. Consequently, for example, the probability levels associated with the limits r and R will not be exactly 95 %. They will approximate to 95 % when many laboratories have taken part in the precision experiment, but may be considerably different from 95 % when fewer than 30 laboratories have participated. This is unavoidable but does not seriously detract from their practical utility as they are primarily designed to serve as tools for judging whether the difference between results could be ascribed to random uncertainties inherent in the measurement method or not. Differences larger than the repeatability limit r or the reproducibility limit R are suspect.

21 The symbols r and R are already in general use for other purposes; in ISO 3534-1 r is recommended for the correlation coefficient and R (or W) for the range of a single series of observations. However, there should be no confusion if the full wordings repeatability limit r and reproducibility limit R are used whenever there is a possibility of misunderstanding, particularly when they are quoted in standards.

4 Practical implications of the definitions for accuracy experiments

4.1 Standard measurement method

4.1.1 In order that the measurements are made in the same way, the measurement method shall have been standardized. All measurements shall be carried out according to that standard method. This means that there has to be a written document that lays down in full detail how the measurement shall be carried out, preferably including a description as to how the measurement specimen should be obtained and prepared.

4.1.2 The existence of a documented measurement method implies the existence of an organization responsible for the establishment of the measurement method under study.

NOTE 22 The standard measurement method is discussed more fully in 6.2.

4.2 Accuracy experiment

4.2.1 The accuracy (trueness and precision) measures should be determined from a series of test results reported by the participating laboratories, organized under a panel of experts established specifically for that purpose.

Such an interlaboratory experiment is called an "accuracy experiment". The accuracy experiment may also be called a "precision" or "trueness experiment" according to its limited purpose. If the purpose is to determine trueness, then a precision experiment shall either have been completed previously or shall occur simultaneously.

The estimates of accuracy derived from such an experiment should always be quoted as being valid only for tests carried out according to the standard measurement method.

4.2.2 An accuracy experiment can often be considered to be a practical test of the adequacy of the standard measurement method. One of the main purposes of standardization is to eliminate differences between users (laboratories) as far as possible, and the data provided by an accuracy experiment will reveal how effectively this purpose has been achieved. Pronounced differences in the within-laboratory variances (see clause 7) or between the laboratory means may indicate that the standard measurement method is not yet sufficiently detailed and can possibly be improved. If so, this should be reported to the standardizing body with a request for further investigation.

4.3 Identical test items

4.3.1 In an accuracy experiment, samples of a specific material or specimens of a specific product are sent from a central point to a number of laboratories in different places, different countries, or even in different continents. The definition of repeatability conditions (3.14) stating that the measurements in these laboratories shall be performed on identical test items refers to the moment when these measurements are actually carried out. To achieve this, two different conditions have to be satisfied:

a) the samples have to be identical when dispatched to the laboratories;

b) they have to remain identical during transport and during the different time intervals that may elapse before the measurements are actually performed.

In organizing accuracy experiments, both conditions shall be carefully observed.

NOTE 23 The selection of material is discussed more fully in 6.4.

4.4 Short intervals of time

4.4.1 According to the definition of repeatability conditions (3.14), measurements for the determination of repeatability have to be made under constant operating conditions; i.e. during the time covered by the measurements, factors such as those listed in 0.3 should be constant. In particular, the equipment should not be recalibrated between the measurements unless this is an essential part of every single measurement. In practice, tests under repeatability conditions should be conducted in as short a time as possible in order to minimize changes in those factors, such as environmental, which cannot always be guaranteed constant.

4.4.2 There is also a second consideration which may affect the interval elapsing between measurements, and that is that the test results are assumed to be independent. If it is feared that previous results may influence subsequent test results (and so reduce the estimate of repeatability variance), it may be necessary to provide separate specimens coded in such a way that an operator will not know which are supposedly identical. Instructions would be given as to the order in which those specimens are to be measured, and presumably that order will be randomized so that all the "identical" items are not measured together. This might mean that the time interval between repeated measurements may appear to defeat the object of a short interval of time unless the measurements are of such a nature that the whole series of measurements could all be completed within a short interval of time. Common sense must prevail.

4.5 Participating laboratories

4.5.1 A basic assumption underlying this part of ISO 5725 is that, for a standard measurement method, repeatability will be, at least approximately, the same for all laboratories applying the standard procedure, so that it is permissible to establish one common average repeatability standard deviation which will be applicable to any laboratory. However, any laboratory can, by carrying out a series of measurements under repeatability conditions, arrive at an estimate of its own repeatability standard deviation for the measurement method and check it against the common standard value. Such a procedure is dealt with in ISO 5725-6.

4.5.2 The quantities defined in 3.8 to 3.20 in theory apply to all laboratories which are likely to perform the measurement method. In practice, they are determined from a sample of this population of laboratories. Further details of the selection of this sample are given in 6.3. Provided the instructions given there regarding the number of laboratories to be included and the number of measurements that they carry out are followed, then the resulting estimates of trueness and precision should suffice. If, however, at some future date it should become evident that the laboratories participating were not, or are no longer, truly representative of all those using the standard measurement method, then the measurement shall be repeated.

4.6 Observation conditions

4.6.1 The factors which contribute to the variability of the observed values obtained within a laboratory are listed in 0.3. They may be given as time, operator and equipment when observations at different times include the effects due to the change of environmental conditions and the recalibration of equipment between observations. Under repeatability conditions, observations are carried out with all these factors constant, and under reproducibility conditions observations are carried out at different laboratories; i.e. not only with all the other factors varying but also with additional effects due to the difference between lab-