

First edition
2011-08-15

Textiles and textile products — Guidelines on the determination of the precision of a standard test method by interlaboratory trials

*Textiles et produits textiles — Lignes directrices pour la détermination
de la fidélité d'une méthode d'essai normalisée au moyen d'essais
d'interlaboratoires*



Reference number
ISO/TR 24697:2011(E)

© ISO 2011

STANDARDSISO.COM : Click to view the full PDF of ISO/TR 24697:2011



COPYRIGHT PROTECTED DOCUMENT

© ISO 2011

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword	iv
Introduction.....	v
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 Requirements for an interlaboratory precision trial	2
4.1 General	2
4.2 Personnel requirements	3
4.3 Laboratory requirements	3
4.4 Sample requirements	3
4.5 Organization of the interlaboratory trial.....	3
4.6 Conducting the interlaboratory trial	4
4.7 Analysis of the results	4
Annex A (informative) Form examples	5
Annex B (informative) Statistical assessment	7

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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. 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.

In exceptional circumstances, when a technical committee has collected data of a different kind from that which is normally published as an International Standard ("state of the art", for example), it may decide by a simple majority vote of its participating members to publish a Technical Report. A Technical Report is entirely informative in nature and does not have to be reviewed until the data it provides are considered to be no longer valid or useful.

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.

ISO/TR 24697 was prepared by Technical Committee ISO/TC 38, *Textiles*, Subcommittee SC 24, *Conditioning atmospheres and physical tests for textile fabrics*.

Introduction

It is well known that developing a standardized test method is not always an easy task. Most of the effort involves going through lots of details and trying to reach agreement between all the parties involved. As a consequence, it is wise to also dedicate part of the job to define what level of reliability the result of the standardized test method will have once it is applied.

The participation of interested laboratories is welcome, possibly those having a delegate in the commission in charge of developing the standardized test method.

Following this consideration, the aim of this Technical Report is to supply guidelines in case there is an intention to evaluate the uncertainty of that standardized test method by carrying out interlaboratory tests.

STANDARDSISO.COM : Click to view the full PDF of ISO/TR 24697:2011

Textiles and textile products — Guidelines on the determination of the precision of a standard test method by interlaboratory trials

1 Scope

This Technical Report can be applied to textiles and textile products and is concerned only with test methods which operate in a continuous scale to yield a single numerical figure as the test result. However, this single figure can be the outcome of a calculation from a set of measurements.

The distribution of test results is required to be unimodal and is assumed to be normal. With non-Gaussian distributions, other evaluation procedures will be necessary.

It does not cover methods which yield discrete values, 'pass/fail' (go/no go) type results, (accept/reject) tests or where a ranking scheme is in operation.

2 Normative references

The following referenced documents are indispensable for the application 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 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 5725-6, *Accuracy (trueness and precision) of measurement methods and results — Part 6: Use in practice of accuracy values*

3 Terms and definitions

For the purposes of this document, the terms and definitions in ISO 3534-1, ISO 5725-2 and ISO 5725-6 and the following apply.

3.1

observed value

value of a characteristic obtained as a result of a single observation

3.2

test results

value of a characteristic obtained by carrying out a specified test method

NOTE The test method should specify that a number of individual observations to be made and their average and other appropriate function (such as the median and the indication of the dispersion measured by a standard deviation) be reported as the test result.

3.3

level of the test in a precision experiment

general average of the test results from all laboratories for one particular material or specimen tested

3.4

cell in precision experiment

test results at a single level obtained by one laboratory

3.5

precision

closeness of agreement between independent test results obtained under stipulated conditions such that they are not influenced by any previous result on the same or similar material

NOTE The measure of precision is usually expressed as, or derived from, a standard deviation, which is a measure of imprecision computed from the test data. Less precision is reflected by a larger standard deviation.

3.6

accuracy

closeness of agreement between a test result and the accepted reference value

3.7

trueness

closeness of agreement between the average value from a large series of test results and an accepted reference value

3.8

repeatability

measure of the dispersion of test results under conditions where test results are obtained with the same method on identical test material in the same laboratory by the same operator using the same equipment within short intervals of time

3.9

reproducibility

measure of the dispersion of test results under conditions where test results are obtained with the same method on the same test material in different laboratories with different operators using different equipment

3.10

outlier

member of a set of values which is inconsistent with the other members of that set

3.11

degree of freedom

number of independent observations

NOTE In the evaluation of a test method, an absolute minimum of five laboratories should be used from at least three different countries

4 Requirements for an interlaboratory precision trial

4.1 General

For a successful trial it is required that:

- The participating laboratories and personnel are given all the details before the start of the exercise;
- All participating laboratories keep to the instructions for carrying out the experiment;
- All operators are familiar with the test method;

- All measurements taken shall be reported;
- It is not acceptable to carry out more than the number of replicates specified;
- It is not acceptable to report the mean of a series of replicates as a single observed value.

4.2 Personnel requirements

4.2.1 The project manager

The working group or committee shall appoint a Project Manager by one of its members who will take full responsibility for the organization of the experiment, supervise its execution, collation the results and determination the precision of the test method.

The project manager should be fully familiar with the test method, and should have knowledge of statistical design and analysis. If necessary he may appoint a statistician to assist with the analysis of the results.

4.2.2 Laboratory contact person

A suitable contact person – the laboratory contact – shall be identified within each participating laboratory, to which the samples and information about the trial should be sent. This person is responsible for supervision of the testing by the operator(s) and for the reporting of results to the project manager.

4.2.3 The operator(s)

In each of the participating laboratories the trial must be performed by an operator who is competent in carrying out this sort of measurement.

4.3 Laboratory requirements

In the evaluation of a test method, an absolute minimum of five laboratories should be used from at least three different countries.

The participation of interested laboratories is welcome; possibly those having a delegate in the commission in charge of developing the standardized test method.

4.4 Sample requirements

4.4.1 The number of types of material (levels) tested in each laboratory should be selected such that the total number of samples tested across all laboratories is not less than 30, preferably nearer to 60. Thus, if there are 5 participating laboratories, a minimum of 6 materials (levels) are needed.

4.4.2 The working group should agree on the types of materials required to cover the whole field of application of the test (different levels).

4.4.3 The quantity of material prepared shall be sufficient to cover the trial, and to allow a reserve.

4.5 Organization of the interlaboratory trial

The project manager is responsible for the organization of the trial as follows:

4.5.1 The design of the trial, based on ISO 5725-2, to include the number of levels required (see 4.4.1), a number of times that the test should be carried out, and the order in which samples should be tested.

4.5.2 The preparation of sufficient samples and their randomization to ensure that each laboratory receives as nearly as possible homogeneous samples. Additional samples shall be prepared for the replacement of any lost or damaged samples if necessary.

4.5.3 Labelling of samples

Each sample should be labelled preferably with a three or five digit random number. The allocation of random numbers to the samples should be known only to the Project Manager.

Preparation of an instruction sheet for the participating laboratories to include, at least, the following:

- the test method to be used;
- the number of repeat measurements to be made;
- the number of operators to be used;
- to specify how the samples are to be conditioned prior to the test;
- the order in which the samples should be tested;
- the deadline for completion of tests;
- the questionnaire for feed-back;
- the standard sheet for the reporting of the results (see Annex A, for an example).

4.5.4 Distribution of the samples and instructions to the laboratories.

4.6 Conducting the interlaboratory trial

4.6.1 Testing should be carried out by the participating laboratories according to the instructions provided by the Project Manager.

4.6.2 Results should be sent back to the Project Manager within the required time-scale. Any deviations from the required procedure or any problems experienced should be reported.

4.7 Analysis of the results

4.7.1 Data correction

4.7.1.1 Missing data

Unless these are so excessive as to hazard the validity of the study, they should be ignored in the analysis apart from necessary procedural adjustments.

4.7.1.2 Outliers

Experience has taught that outliers cannot be avoided and have to be taken into consideration. As a general rule no readings should be rejected unless either, there is evidence for a definite source of error or, they fail some statistical criteria. It should be noted that not only individual results but data from a source (i.e. a laboratory) may be subject to this procedure. Under no circumstances, after rejection of outliers, may be a further analysis be undertaken to detect further outliers inconsistent with the adjusted data set. For an extensive treatment on the subject see ISO 5725-2.

4.7.2 When the standards deviation for both repeatability and reproducibility do not show any dependence on the level of tests it is permissible to average the values before calculation of the precision. Otherwise, following suitable statistical test to check for homogeneity (see ISO 5725.2:1994, Clause 7.3.3 - Cochran's test) separate precision values may be assigned to each level.

4.7.3 Calculation of precision

See Annex B.

Annex A (informative)

Form examples

Form A - Recommended form for the collation of the original data						
Laboratory	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
1	n_1					
	n_k					
2						
3						
4						
5						
p						

n_k : total number of replicates per cell.

Form B - Recommended form for collation of calculated averages						
Laboratory	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
1						
2						
3						
4						
5						
p						

Form C - Recommended form for collation of measured spread within cells						
Laboratory	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
1						
2						
3						
4						
5						
p						

Symbol used for laboratory i

Symbol used for level j

Annex B (informative)

Statistical assessment

B.1 Foreword

It is assumed that this particular exercise of interlaboratory test is needed to give more reliability and better confidence on the action that the results of any type of test may suggest to undertake.

This is particularly true when the final outcome of a normative is the completion of a test.

It can be agreed upon the fact that dedicating only most effort in “assembling” a normative, with lots of details, finding agreement from all parts involved and not attaching to it some sort of reliability to the results of the test derived from the normative itself, it's not acceptable.

To comply with the spirit of this technical report, the statistical approach proposed is what is normally used in the industrial area, in term of process control – quality control and quality assurance.

It is almost normal practise to look at any test finalized to obtain a value of a certain characteristic as a way to draw conclusion only on the average of a various number of measurements.

In other words the normal behaviour is to consider the average as a sound and sufficient factor to decide to take any action.

This “*modus operandi*” is unfortunately more spread than one might think, at least outside the operators directly involved in laboratory testing.

Any effort must than be dedicated to explain that, once the results are available, a correct action is to consider not only the average obtained, but also carefully the range in which this average can “ move “on, due to the inevitable error in measuring.

This range is normally referred to as standard error and is strictly connected to the variability of the characteristic under test as well as other factors.

B.2 Some basic statistics

It may be useful at this point, to summarize the law which is called Normal or Bell Shaped Distribution, in relation to the measurement of continuous quantity value.

(i.e it can take all values from 0 to 1)

In theoretical terms we may consider a population of certain characteristic (a bulk of whole possible values $x_1, \dots, x_n$ of this) of which we know the mean μ (the central point of the population) and a variation σ (the dispersion of the individual elements from the centre).

1) The mathematical formula are: $\mu = \frac{1}{n} \sum_{i=1}^n x_i$ and $\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2}$

This last parameter is expressed in the same value of the mean and it's quite useful because it gives a numerical indication on how the individuals are close to the mean.

2) In particular

- 68 % of the individuals lies between the mean $\pm 1,00 \sigma$
- 95 % of the individuals lies between the mean $\pm 1,96 \sigma$
- 99 % of the individuals lies between the mean $\pm 2,57 \sigma$

This range is referred to as **Coverage interval**.

The factors 1,00 – 1,96 – 2,57 are related to the theoretical Normal Distribution, and they are referred to as **Coverage factor**.

In practical terms any time we make a test, which consist of a limited series of **n** measurement of the characteristic under examination (measurand) three parameters can be calculated.

3) An average $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

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

5) A standard deviation $s = \sqrt{\text{variance}} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$

With limited number of measurement, the symbol used for average is $\bar{x}$ instead of μ and for the variance **s** instead of σ .

This last formula, that in short is the average of the squared differences of the individuals' observation from their average, is similar to the one in 1), but with a denominator **n-1**, which is worthwhile to explain in practical terms: the mathematical role of that **-1** is effective with low value of **n** and is loosing importance as **n** increase. In this way it is included in the formula a certain assurance that by having limited number of observation, some extreme results may not be included due to the lower probability of being obtained.

n-1 is referred to as Degrees of freedom

It is convenient here to recall the definition of Uncertainty:

Parameter associated with the results of a measurement, which characterize the dispersion of the values that could reasonably be attributed to the measurand.

With this single test we can have some knowledge of the variability of the characteristic, but it's important to know also how close we are to the true average of the measurand.

Statistic theory helps in giving the answer.

If we proceed with a second test on the same material, same testing condition and same number of observations, we would get a different average and standard deviation.

However if the material tested is really homogeneous, it exists a relation between this subsequent results.

By going on making further tests (always of **n** measurement) and calculating each time the average, we obtain a series of averages from which in turns we can calculate a new Grand average and a Std. Dev of these averages.

In particular this new Standard deviation is related to the previous one by the following equation:

Standard deviation of averages = Standard deviation of single test / Square root of $n - 1$

$$6) \quad s_{\bar{x}} = \frac{s}{\sqrt{n-1}} \quad \text{for } n \rightarrow \infty \quad s_{\bar{x}} = 0 \text{ (zero)}$$

The standard deviation of average so calculated is also referring as Random error of the test.

It's easy to appreciate that if n gets higher and higher; the ratio gets smaller and smaller. For n to infinity, the ratio goes to 0 (zero) and then there is no more dispersion and no error in the test, because we have reached the true average (the mean) having tested theoretically the whole population of the characteristic under test.

It's then quite obvious that the more measurement n we make the more close we are to the true value, and the error is reduced (as well as the uncertainty of a measurement). Clearly there are limitations to the value of n , especially when to make the test the sample is destroyed (as for instance breakage test of a thread to measure its strength).

Other point to keep in mind is the increasing cost of the test by increasing n .

We then have to accept a compromise between the reliability of the requested results from the test and its cost, or in other words what numerical level of confidence can we attached to the results.

The way to obtain a reasonable answer connected to a certain degree of confidence is by following the concept explained in 6).

The average obtained from a test, it's not, as we have seen, the true average, but we can reasonably expect this to lie in a range of values with a level of probability connected to a **factor t** as follows:

$$7) \quad \text{Field of possible values: } = \bar{x} \pm t \cdot \frac{s}{\sqrt{n-1}}$$

Where $\bar{x}$ is the average and s is the standard deviation from the test.

The coverage factor t must be used instead of that in the theoretical normal distribution, again due to the fact that we are dealing with a limited n (see the attached table for the value of t in relation to $n-1$ (degree of freedom)).

To be clearer any average value included in this range is an estimate of the accepted average, and the difference between them is not significant simply because they are averages of limited different individuals belonging to the same population.

This final consideration is the basis to find out instead if there is a real difference between the results of a series of test on the same material in the same or different conditions, once we have determined the relative standard deviation.

The statistical approach to accept one or the other possibility is the so called "Null Hypothesis".

It is accepted in principle that the difference between the results of samples means values (i.e. two test result with average value $\bar{x}$ and $\bar{y}$) are not significant at the level of confidence stated unless the calculation shows the contrary, according to the formula:

$$8) \quad \frac{\bar{x} - \bar{y}}{\frac{s}{\sqrt{n-1}}} \leq t \text{ not significant difference: the Hypothesis is accepted}$$

If $t > t$ the difference is significant and the Null Hypothesis rejected.

In both cases the conclusion is connected at the % of confidence given by the value of t in relation to n .

The levels of confidence (coverage probability) applied in normal practice are at 95 % and 99 %.

This is to say that if we accept the difference in 8) to be not significant, we might be wrong only in one case out of twenty or even one out of forty if we are interest only in one side of the distribution around the mean.

We have now to define the value of s which the objective of this study, that is to define the variability of the test proposed by any normative, i.e. its precision.

The importance to attach a precision figure to a test can be best evaluated if one considers the dispute that can take place between laboratories, in case of different acceptance / rejection limits (tolerance) against a standard quantity value, not having as a reference the accepted variability of that test.

B.3 Sequence of calculation

It is proposed, using the suggested tables in Annex A, the following:

1 – Enter the individual test results in the cell (level / laboratory)

2 – Calculate the average in each cell: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

Where $x_1, \dots, x_n$ are individual results of n replicates

3 – Calculate the variance for each cell: (simplified method) $= \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right]$

i.e. sum of the squared n replicates minus the squared sum of n replicates

4 – Proceed searching for possible outlier in each cell

We could follow the table of the F distribution to analyse significant difference

between variance, but probably the more known Cochran's test can be used.

$$C = \frac{s^2_{\max}}{\sum_{i=1}^p s_i^2}$$

this ratio of the higher variance and sum of the variances from all laboratories, to be compared against critical values in the relative table attached of Cochran's test

5 – After decision on the outcome of 4) proceed with the estimate of repeatability

variance $s_{r,j}^2$ at each level

$$S_{r,j}^2 = \frac{\sum_{i=1}^p (n_i - 1) \cdot s_i^2}{\sum_{i=1}^p (n_i - 1)} \quad \frac{\text{Sum of the cell variances}}{\text{Degrees of freedom}} \quad j = \text{each level}$$

6 – Calculate the between laboratory variance $S_{L,j}^2$

$$S_{L,j}^2 = \frac{s_{dj}^2 - s_{rj}^2}{n_j} \quad \text{where} \quad s_{d,j}^2 = \frac{1}{p-1} \sum_{i=1}^p n_i \left(\bar{x}_i - \bar{X}_j \right)^2 \quad \text{variance per level all laboratories}$$

and

$$\bar{n} = \frac{1}{p-1} \left(\sum_{i=1}^p n_i - \frac{\sum_{i=1}^p n_i^2}{\sum_{i=1}^p n_i} \right) \quad n_j \text{ is number of replicates per cell}$$

7 – Determine the reproducibility variance $S_{R,j}^2$

$$S_{R,j}^2 = S_{L,j}^2 + S_{r,j}^2$$

8 – The standard deviations will be as follows.

$$S_{r,j} = \sqrt{\frac{\sum_{i=1}^p (n_i - 1) \cdot s_i^2}{\sum_{i=1}^p (n_i - 1)}} \quad \text{Repeatability}$$

$$S_{R,j} = \sqrt{S_{L,j}^2 + S_{r,j}^2} \quad \text{Reproducibility}$$